SPARTA

D4.6

Visual Analytics System for Cybersecurity threat analysis

Project number	830892
Project acronym	SPARTA
Project title	Strategic programs for advanced research and technology in Europe
Start date of the project	1st February, 2019
Duration	36 months
Programme	H2020-SU-ICT-2018-2020

Deliverable type	Demonstrator
Deliverable reference number	SU-ICT-03-830892 / D4.6 / V1.0
Work package contributing to the deliverable	WP4
Due date	January 2022 – M36
Actual submission date	4th February, 2022

Responsible organisation	KTU
Editor	Šarūnas Grigaliūnas
Dissemination level	PU
Revision	V1.0

Abstract	Information visualization approaches help facilitate successful data analysis by employing visual and graphical representation in Visual Analytics. The primary objectives of the report are to present Visual Analytic Systems (VAS) research publications, to search for and evaluate currently available open source and to implement them by cross-integrating them into T-SHARK.
Keywords	Visual Analytics System, VAS, Gathering methods, Vulnerability analysis methods, Impact Assessment methods, Knowledge Gatherer Model, Cyber Mission Impact Assessment Tool, Cyber Attack Modelling, Impact Assessment Framework, Cyber Security Domain

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 830892.

Editor

Šarūnas Grigaliūnas (KTU)

Contributors (ordered according to beneficiary numbers)

Romain Ferrari (TCS)

Porreti Claudio (LEO)

Robertas Damaševičius, Algimantas Venčkauskas, Jevgenijus Toldinas (KTU)

Jocelyn Aubert (LIST)

Filipe Apolinario (INOV)

Miguel Correia (IST)

Reviewers

Herve Debar (IMT)

Claudio Porretti (LEO)

Executive Summary

Deliverable 4.6, **Visual Analytics System for Cybersecurity threat analysis**, is the deliverable of the T-SHARK research programme. T-SHARK stands for Full spectrum Cybersecurity Situational Awareness. This scientific activity of the SPARTA project addresses the issue of emerging complex cybersecurity threats for sensitive and critical industrial and governmental ecosystems. Timely and well-informed supervision of dependent systems allows their operators to build accountable trust in the face of complex cybersecurity contexts, even as these systems become more intricate and more heterogeneous.

All of the significant results produced during the project, and specifically derivable D4.6 research, was published in scientific articles and in journals with high impact factors. The partners presented their subcases during SPARTA days and other relevant events to the external partners of SPARTA. To demonstrate the comprehensive and collaborative SPARTA approach, the partners integrated their subcases - either through visual analytics or through the sharing platforms MISP and C3ISP. The subcases were presented to Arbitrage Group members organized in Task 4.6. AG assessed the subcases from multiple perspectives and provided their valuable feedback. Finally, the subcases were deployed either internally, within the partner's organization, or publicly, to improve capabilities in the EU cybersecurity community.

Table of Content

List of Figures

List of Tables

Chapter 1 Introduction

The use of sophisticated tools and procedures to evaluate datasets using visual representations of the data is known as visual analytics. Users can find patterns and produce actionable insights by visualizing data in graphs, charts, and maps. These insights help businesses make better data-driven decisions. Visual analytics is not the same as data visualization in that it is not just an issue of showing data graphically. Modern, interactive visual analytics, on the other hand, makes it simple to mix data from numerous sources and examine it in depth right within the visualization. Furthermore, AI and machine learning algorithms can make suggestions to assist the user in their exploration.

Visual analytics enables analysts to use various techniques to interactively bring critical data out of the abyss and make it instantly understandable. Visual analytics, in essence, cuts the time it takes to convert data into knowledge by an order of magnitude or more. This is conceivable due to several factors:

Computer security experts can use visual analytics to rethink how they identify hazards and defend against cyber threats. As a result, assaults can be prevented more effectively, and attacks can be isolated and mitigated more quickly.

Data collection, discovery, investigation, examination, analysis, and reporting are all aided by visual analytics in the digital forensic process. In cyber security and computer forensics, visual analytics facilitates sense making in three distinct ways.

1. First, log files from a computer network intrusion detection system can be loaded, and suspicious connections between machines are investigated. This information can be paired with other log data to provide a better picture of security breach events.
2. Second, forensic analysis of e-mails can be used to model communication patterns and summarize e-mail content. Finally, directory structures can be seen and filtered based on file modification timings to discover what activities took place on which dates. Identical files that may have been copied from one machine to another can be rapidly discovered by comparing multiple file systems.
3. Visual analytics provides information discovery, processing, and visualization capabilities that can be used in a variety of computer security and forensics applications, including:
    - After an intrusion, analyse a computer system to determine how the attacker obtained access and what the attacker did.
    - Analyse data on confiscated gear, particularly among intelligence, military, and law enforcement agencies.
    - In court situations, computer forensic techniques are used to examine the computer systems of defendants.

To achieve this functionality requires tools, models, and methods. To evaluate the Visual Analytics System (VAS) for Cybersecurity threat analysis, it is necessary to have data sets. The proposed Visual Analytics System (VAS) for cybersecurity threat analysis is based on our scientific papers, findings, and experience gained from the SPARTA T-SHARK project.

1.1 T-SHARK and Visual Analytics System introduction

SPARTA T-SHARK is concerned with hybrid cyber activities, such as disinformation campaigns and fake news combined with targeted cyber operations, vary in scope, scale, duration, intensity, complexity, sophistication, and impact, and require the deployment of the full range of intelligence tools and instruments, as well as new ways of responding. The ability to predict the future is crucial to success. It allows for situational threat intelligence, informed, and effective decision-making, and so supports the EU's priority move from a reactive to a proactive approach in reacting to current and

future cyber threats. To do so, cybersecurity threat intelligence must be expanded and enriched with related external information and information from other security domains, as well as general context information, allowing for Full Spectrum Analysis of Potential and evolving Threats across current, near-future, and far-future time dimensions. The next big thing is information sharing and communication. Within the T-SHARK program, data exchange is one of the most important parts of developing a comprehensive cybersecurity approach. Shared cyber incident data retrieved using various technological techniques. Organized contextual information from multiple sources allows for the connection of isolated occurrences, a better understanding of targets, and the development of situation awareness across the ecosystem. And to fill in the missing gaps, tools are needed to do that, evaluate the information, and make future decisions and actions.

Faced with all these new obligations, it is apparent that incorporating cybersecurity threat intelligence into governmental and institutional operations will grow more vital in the coming years, as all levels and personnel are required to respond to the cybersecurity threat. This visual analytics is on the rise, especially considering the recent recognition of the success of cyber-enhanced disinformation campaigns used for political influence, illegally obtained and profiled personal data, leaked confidential strategic information, and stolen commercial and scientific developments. It all points to an obvious need to expand cybersecurity beyond a focus on raw data to focus more on visualizing data for decision making.

1.2 The scope of the work

The objective of this deliverable is to demonstrate the Visual Analytics System (VAS) for Cybersecurity threat analysis comprehensive **D4.2 Cybersecurity threat analysis model** and the collaborative SPARTA approach. The proposed Visual Analytics System (VAS) for Cybersecurity threat analysis, and through the sharing platforms **D4.1 Cybersecurity threat intelligence common data model** and **T4.4 Information sharing and integration.**

The structure of the document is as follows. Chapter 2 - this chapter analyses and reviews existing methods, models, and tools. In Chapter 3, we provide an analysis of visual analytics system for cyber security domain. Chapter 4 - it is about developing a model and a working solution for a visual analytics system and demonstrating its capabilities. Chapter 5 - presents one of the subcase scenarios and applications together with a visual analytics system.

Chapter 6 - addresses the interconnection of the work carried out in this deliverable with the work in other SPARTA programs. Finally, Chapter 7 - summarizes the summary and conclusions of the report.

It also follows References and Annexes that can be useful in creating VAS automated information exchanges.

1.3 Research results

Research and the articles of the results are publicly available at:

1. Visual Analytics for Cyber Security Domain: State-of-the-Art and Challenges. In International Conference on Information and Software Technologies (pp. 256-270). Springer, Cham. (**2019, October**).
2. LITNET-2020: An Annotated Real-World Network Flow Dataset for Network Intrusion Detection. Electronics, 9(5), 800. (**2020, May**)
3. LITNET-2020 Datasets https://dataset.litnet.lt/index.php (**2020 May**)
4. Technical Threat Intelligence Analytics: What and How to Visualize for Analytic Process. 24th International Conference ELECTRONICS 2020 and for publication in the IEEE Conference Proceedings (ISBN 978-1-7281-5868-6)). (**2020, June**)

5. "Ensemble-Based Classification Using Neural Networks and Machine Learning Models for Windows PE Malware Detection" 2021, February "Electronics 10, no. 4: 485. (**2021, February**)
6. "A Novel Approach for Network Intrusion Detection Using Multistage Deep Learning Image Recognition" 2021 August "Electronics 10, no. 15: 1854" (**2021, August**)
7. “Impact Propagation in Airport Systems 2021. In: Abie H. et al. (eds) Cyber-Physical Security for Critical Infrastructures Protection. CPS4CIP 2020’, Lecture Notes in Computer Science, vol 12618. Springer, Cham. (**2021, September**)
8. “Toolkit to Enhance Cyber-physical Security of Critical Infrastructures in Air Transport’, Cyber-Physical Threat Intelligence for Critical Infrastructures Security by John Soldatos, Isabel Praça, Aleksandar Jovanovic “. ISBN 978-1-68083-822-0. E-ISBN 978-1-68083-823-7, DOI: 10.1561/9781680838237.ch11. (**2021, September**)

Chapter 2 Models, Methods and Tools for Visual Analytics Systems

The amount of digital information created and replicated in the world grew exponentially and today is calculated in zettabytes. This has major implications for the responsibility and IT practices of organizations, such as businesses, government agencies, health insurance, non-profit organizations, and so on. Organizations collect, retain, and use personal information that individuals generate as they go about their daily lives, including video on surveillance cameras, credit card usage, copies of hospital scans, Web search histories, financial transaction journals, mailing lists, public records in social media, and many others.

To a greater or lesser extent, all user-generated content will encounter an organization at some point along the route, whether it is in the form of a network, data center, hosting site, private branch exchange (PBX), voice over Internet protocol (VoIP), Internet switch, or backup system. Many unstructured data from unconnected sources are utilized as input/output for tools like:

- Technical Threat Intelligence (TTI)
- Open Source Intelligent (OSINT)
- Human Intelligence (HUMINT)
- Geospatial Intelligence (GEOINT)
- Financial Intelligence (FININT)
- Signal Intelligence (SIGINT)
- Cyber Intelligence (CYBINT)
- Social Intelligence (SOCINT)
- Comprehensive Cybersecurity Threat Intelligence (CSTI)
- Cybersecurity Management in the sector of public administration and defences.

Current approaches in the realm of cyber security visual analytics systems (VAS) are commonly built to help organizations pinpoint flaws in their IT infrastructure that can (or may have) contributed (d) to cyber-attacks happening on their systems. VAS key objective is to provide the necessary information for system administrators to identify: the IT systems subject to visual analysis; the vulnerabilities present on those systems; how cyber-attacks are exploiting such vulnerabilities; and the actual impact of those attacks on the services and goals offered on the IT systems analysed.

The first step in performing cyber threat analysis is to gain knowledge of the IT systems that should be protected from cyber security attacks. In this step system, administrators identify the critical IT devices they want to protect and the IT infrastructure that can be used by the attacker to compromise those assets.

The most common approach to identify critical IT devices of a system is to map the services and goals of the organization (also called the mission of the organization) with the IT devices used to perform those services. The level of criticality of the IT device can be classified according to the amount of work performed in the mission or on how dependent the mission accomplishment is on the asset (level of operability if the asset is compromised). There are several tools to help system operators classify critical assets: COTS software that allows operators to manually identify critical assets (such as Microsoft Threat Modelling Tool [5]), or semi-supervised methods to automatically identify mission processes and goals on the work performed by the IT devices [6], [7].

Regarding IT infrastructure, identifying the attack surface is crucial to pinpoint the main entry points in the IT network, and the overall topology of the network. There are several tools to aid system operators in identifying the attack surface: COTS software that allows operators to manually identify entry points and network topology (such as Microsoft Threat Modelling Tool [5]), or automated cyber-

threat reconnaissance methods (such as NMAP [8]; Moloch [9], Silk [10]) that monitor traffic and perform asset identification using network topology inference.

2.1 Overview of the current vulnerability analysis methods

The main objective of vulnerability gathering methods is to identify and classify the security flaws present in the organization. To do so, common vulnerability gathering methods are typically divided into two phases: 1) vulnerability identification, where IT devices (inner hardware peripherals, operating system, and software) and network communications (made by those devices) are subject to security tests/analysis in order to identify vulnerabilities; 2) vulnerability classification, which happens after the identification phase, and involves cataloguing vulnerabilities according to hindering effect they may on the systems based on standard vulnerability classification methodologies.

Current methods to identify and catalogue vulnerabilities in organizations can be divided into two categories: product-based vulnerabilities and organization-based vulnerability identification methods.

The first category, product-based vulnerability scanning, is typically performed to COTS products (hardware devices, operating systems, or software applications) and is normally employed by vendors of the products. In such methods, product families are subjected to security tests (employed by product-vendors or, in most cases, third parties) to identify possible security vulnerabilities related with bad implementation of products or malfunctions affecting one or more versions of the product. The result of these discovered vulnerabilities is then communicated to the product consumers by placing them in public databases (e.g., CVE [11], TVD [12]). Organizations then perform an inventory of the vulnerabilities affecting their IT infrastructure by collecting information product vendors, either by directly accessing the public databases or by resorting vulnerability scanners (e.g., NESSUS [13], Nikto [14], Zap [15], Archery [16], OpenVAS [17]) that identify vulnerabilities present in the products of an organization based on the publicly available vulnerability databases.

The second category, organization-based vulnerability identification methods, are typically performed within the organizations IT infrastructure and are centred on identifying security vulnerabilities that appear due to the operational deployment of those products, namely bad configuration of products and/or deployment without assuring the necessary conditions to provide secure execution. Typically, these methods are performed manually by system administrators of the organization following threat modelling methodologies (such as STRIDE, etc.). And managed by the organization security team either by storing it on excel sheets or by resorting to Vulnerability Assessment and Management tools such as OWASP Threat Dragon [18], QRadar [19], DefectDOJO [20], ThreadFix [21].

2.2 Overview of the current intrusion detection systems

Intrusion detection systems (IDS) are an important source of information used in the context of security risk management [22] to monitor the security level of an IT infrastructure. They help to detect potential active threats, identify the exploitable vulnerabilities, and the assets directly hit by attacks. The rationale behind these technologies is to detect incidents based on data collected from sensors that monitor IT devices (host-based: Ossec [23]); network (network-based Snort [24], BroIDS [25]); or both (hybrid IDS: Trend Micro [26] BP-IDS [27]). Several methods can be used for detection: signature-based IDS where collected data are compared to patterns of known attacks; anomaly-based IDS compare data with the normal behaviour (baseline) of the system taught beforehand during an initialization period; specification-based IDS which compare data collected with a set of specification rules stipulating the acceptable behaviour of the system.

Although widely commercialized products with several COTS solution providers available [28], intrusion detection remains an important topic in academic research. Proposals for novel schemes

that improve anomaly-based IDS mechanisms with new fingerprinting techniques to detect stealthy attacks that occur in cyber-physical systems (CPS) such as the detection of these new classes of attacks in vehicle intrusion attacks [29] – [31]; water control systems attacks [32] – [35]; and SCADA attacks [34], [36], [37]. Several of these new approaches distinguish themselves based on the type of fingerprints used for profiling the normal behaviour, which can be: (a) physical fingerprinting, by measuring the time taken for the physical component (actuator) to perform an operation [36], or characteristics of the physical sensor measurements [32], [33], [35] (e.g., water level, pre-treatment the conductivity, and oxidation reduction potential), with some of these proposals incorporating noise handling mechanisms during physical fingerprinting to optimize results [33], [35];(b) fingerprints based on metrics in the communication between master-slave components [29], [31], [36]; (c) fingerprints based on code analysis of the software used in control units to determine normal and abnormal behaviours [30], [34], [37]. An alternative method to using anomaly detection to detect stealthy physical attacks has been demonstrated in the ECOSSIAN project, the proposed system used the Business Process Intrusion Detection System (BP-IDS) [27] a specification-based IDS that collected data from multiple sources of CPS sensors installed and identified incidents by comparing the information collected with the organization's business processes and rules that stipulate normal behavior. The results of this project have shown that BP-IDS is very capable of detecting incidents caused by MiTM attacks in SCADA systems that communicate through IP; IEC 60870-5; and Modbus network protocols.

Intrusion detection is thus a great solution to know when one is being attacked and react promptly with security measurements to minimize the impact of the attack. Although knowing the time and type of attack taking place is essential for the success of the contingency plans, other information needs to be evaluated regarding the impact of the attack to properly plan the reaction, namely: business assets affected, and business goals compromised by the attack. Within this generic scope, impact assessment procedures used by organizations evaluate the damage caused by a cyber threat to its assets and goals.

2.3 Overview of the current Impact Assessment methods

Impact assessment on cyber-attacks includes vulnerability assessment [38], [39], or mission impact assessment (MIA) [40]. The main objective of these methods is to evaluate whenever an attack occurs the amount of damage caused to the organization's assets and which operational objectives and goals (that is, mission) were compromised by the compromise of those assets.

The evaluation of the impact of cyberattacks is largely influenced by approaches such as M-Correlator [41], which prioritizes and clusters incident alerts published by different IDS solutions installed in organizations, according to the impact the incident has on the assets of the organization. The M-Correlator combines asset identification using network topology inference with NMAP [8] (a tool used for the automatic context gathering technique required during vulnerability assessment) and vulnerability identification and classification made by a human analyst, to classify using Bayesian networks [42] the impact of the incident reported (in one or more alerts) according to the importance of affected assets and the vulnerability rank. VTAC [43] used the foundation of this work and extended it by automating vulnerability identification and ranking with Calderon approach [44] of using vulnerability scanners to provide that information and evaluated the overall impact of the organization after an incident (or a series of incidents) caused by cyberattacks.

Although both M-Correlator and VTAC evaluate the impact of a cyberattack on organizational assets, neither method is capable of accurately pinpointing the amount of damage caused to the objectives / goals of the mission, [45] increased VTAC to allow this estimation by mapping mission information (tasks) with organizational assets and network topology using dependency graphs [46] to calculate the operational capacity of each mission after an incident (or a series of incidents) caused by cyberattacks. Further research to improve the impact assessment of cyber-attacks [38] – [40] adapted this work to be compatible with the SCADA environment using vulnerability trees, PetriNets, or SysML instead of dependency graphs; [2], [40] proposed using BPMN to model high-level mission information and was put in motion in [47] by implementing a tool that provides conversion of BPMN

to dependency graphs and integrates with the enhanced VTAC; and [48], [49] introduce the notion of temporal dependency to improve impact calculation.

2.4 Models (methods, tools) documentation

The list of the most significant visual analytics in cyber security publications (e.g., MITRE, CEN/CENELEC, ISO, ENISA, NIST) relating to the method or models.

The current method is proposed by INOV and developed within the SPARTA project, the reason why there are two publications of this method. However, the proposed method is inspired by the following publications on visual analytics:

2.4.1 A Cyber Mission Impact Assessment Tool

One of the tools used to depict cyber processes, resources, and incident consequences was a Cyber Mission Impact Business Process Modeling tool [3]. Apply the CMIA method, which has as its primary goal the representation of both the mission and its cyber dependencies.

The use of mission-level modelling based in business process modelling notation (BPMN) which can be used to represent mission threads presents a big advantage in comparison to some proprietary languages in the simulation tool [50], [51], it allows non engineers the ability to model and simulate a system behaviour using a graphical modelling language which is intuitive to the user.

This approach uses the possible effects of a cyber-attack (i.e., interception), instead of specific cyberattacks (i.e., a man-in-the-middle attack), turning cyber incident effects into cyber mission effects, to express how the impact of detected incidents affect cyber missions. There are six possible effects of a cyber-attack considered, which are: degradation, interruption, modification, fabrication, unauthorized use, and interception (DIMFUI).

In the implementation of the presented tool, the focus is set on the IT processes; therefore, they allow the user to import the IT resources from external sources (such as network captures, scanners or asset inventory products); according to the paper, the DIMFUI values are attributed to the resources automatically, and it is not presented a clear explanation on how these values are set or even propagated through the model. The IT processes can be modulated in the created software and should be carefully configured, and the different mission dependencies should also be carefully represented.

Figure 1: Example result from two IT resources

After the model is fully created, you can select one of the IT resource assessments and the result displayed shows which incident effects cause mission impacts, as depicted in Figure 1.

2.4.2 CyGraph: Graph-Based Analytics and Visualization for Cybersecurity

CyGraph [1] is an MITRE developed tool, it is described as a 'tool for the analysis, visualization, and management of cyber warfare' by the authors. CyGraph creates an attack graph that maps all the possible attack paths, to do so it correlates information from several different sources (network

topology, firewall rules, host configurations and vulnerabilities) in order to create a normalized data module that represents the whole network, where individual pieces of information are correlated into meaningful data that can be later analysed.

When it comes to its architecture, CyGraph is composed of three REST web services; one ingests the different sources of data mentioned above into a network model, the second web service transforms the created network model into a graph model where the important relationships and dependencies are represented. The third web service is a client application that provides different visual interfaces that can be used to analyse the graph model created in the previous stage.

These queries are able to: identify key vulnerabilities, suggest optimal mitigation strategies, map host-to-host trust relationships, show downstream and upstream attack response paths, etc. The paper supports the execution of queries on the previously generated graph model to increase its analytic value. CyQL (CyGraph Query Language) was created as a domain-specific query language for expressing graph patterns of interest and visualizing query results interactively. In this specific architecture, the implementation of a domain-specific query language improves query clarity, while simultaneously reducing the learning curve for users and increasing productivity. The CyQL abstraction layer also adds an additional level of complexity. CyGraph can support multiple implementations of back-end data engines, each with its own native query language.

In the client application, users submit queries using the previously mentioned CyQL and analyse their results. In Figure 2, the CyGraph user interface can be observed. The created interface has functions for graph statistics, styling, spatial layout, and evolution over time.

It is important to take into consideration that adding some complexity to the queries will result in smaller subgraphs, as such an analytic strategy should be beginning with more general queries and then refine them as more information is acquired. This helps focus analysis, manage complexity, and improve performance.

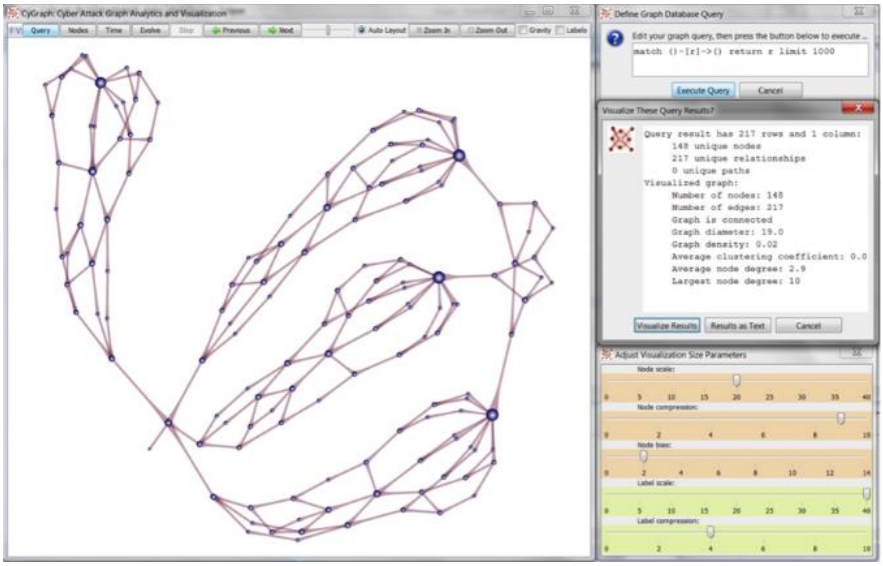

Figure 2: CyGraph client-side user interface [1]

2.4.3 A Cyber Attack Modelling and Impact Assessment Framework

Graph creation and analysis are optimized in Article [52] to enable its use in systems operating in near-real-time modes, such as Security Information and Event Management (SIEM) systems. Article [52] presents a framework for building a Cyber Attack Modelling and Impact Assessment Component (CAMIAC).

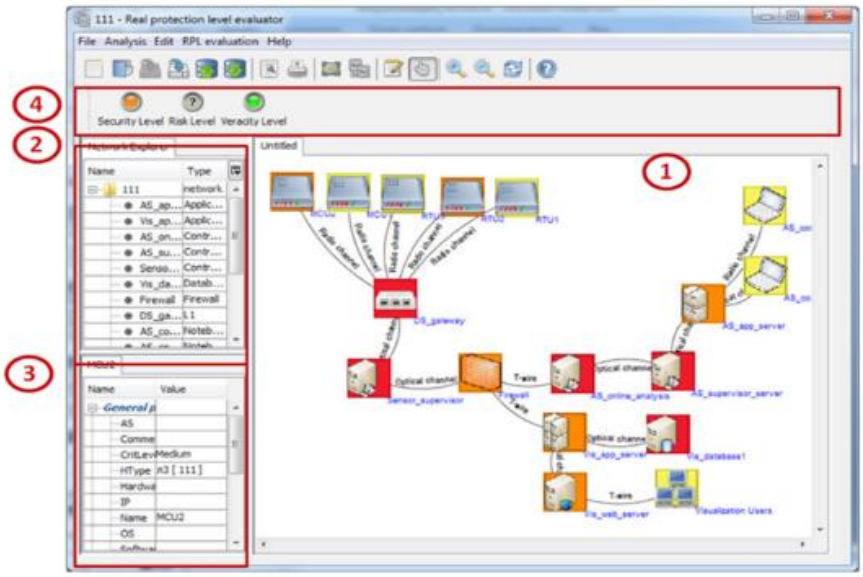

Figure 3: Example of CAMIAC dashboard [52]

CAMIAC offers a user interface (Figure 3), to set initial data, which can be divided into four subviews: (1) shows the graph-based topology of the studied network, where the colour of each network object icon reflects the values of the security metrics calculated for the given object. These metrics are chosen by the user from a predefined list Criterion, Mortality, Risk Level}; (2) depicts the hierarchical structure of the network (domains or specified network zone); (3) allows the user to configure each host and network, such as IP address, host type (web server, router, firewall, etc.), installed software and hardware, user-defined host criticality; (4) reflects the security metrics calculated for the network itself with a predefined set of values Low, Medium, Above Medium, High, Undefined.

This calculation is based on attack graphs and an anytime approach (an algorithm capable of returning a valid solution at any time of its execution, even if it is still not completed, or if is interrupted before it ends), done in two stages.

In the first stage, a graph of possible attacks is constructed for various models of potential bad actors during network deployment or as an offline procedure. To detect attack scenarios and perform real-time event analysis, a second attack graph is modified based on changes to the network being analysed. This, combined with the detection of the set of attack trees to which a security event can belong, allows the detection of attack scenarios as they occur.

2.4.4 Camus: Automatically mapping cyber assets to missions and users

Camus [4] proposes an ontology-based semantic approach to automated mapping of cyber assets to missions and users (Camus), grounded in the idea that the needed data exist in digital format, but are in different locations and formats.

In the first stage, Camus parses raw data sources into ontology instances. However, instead of populating the ontology all at once, many data sources can be used to populate small portions of a base ontology, since, on a second stage, the logical reasoning capabilities of an ontology language can infer indirect relationships by using shared concepts as alignment points. This results in a complete mapping of asset-to-mission relationships, where IP addresses, usernames, user roles, and organizational missions (as departments) are used as the main alignment points. In a third and last stage, a Knowledge Management module is deployed consisting of web APIs, to access the ontology programmatically, and visualization capabilities to display mission context information.

When a Camus user selects an alert, the cache of the data fusion engine is queried, and any raw data files required for processing are parsed on the fly. This demonstrates (Figure 4) the technique. Individuals and roles, as well as network peers linked with the specified IP address, are all included in the contextual information (on the right), which informs the operator of how the attacked assets support specific users, segments of the organization, and other cyber assets.

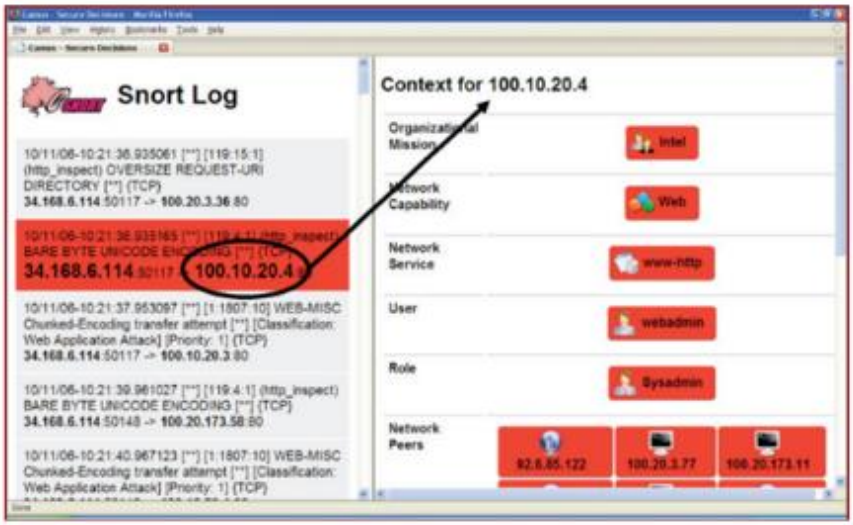

Figure 4: Camus user interface [4]

2.4.5 Cybersecurity threat analysis model (D4.2)

Cyber-attacks are becoming more sophisticated and affecting a wider range of targets. A web-based attack can evolve into a physical attack, and vice versa. For example, following a physical penetration inside a building (e.g., by social engineering), the attacker hides a network device inside the building so that the attacker may remotely access the internal network infrastructure. To achieve the intended cascading impact, complex threats integrate numerous attack vectors aimed at multiple targets in tandem. To fully analyze the materialization of these complex threats, we must bring together the skills, knowledge, and capabilities of various cybersecurity fields that have never been examined together before. In T-SHARK, A functional model of the full cyberthreat analysis process is introduced in the cybersecurity threat analysis model. The model captures critical elements that are required to combat complex threats. The goal of the model is to show and explain which functional components are involved in the analytical process. The concept is intended to assist organizations coping with complex threats in answering questions such as:

- *What is the scope for improvement in the threat analysis process in my organization?*
- *What components are not involved?*
- *What components are not adequately supported?*
- *Where are the communication gaps between the components?*
- *Is the analysis process sufficiently streamlined and automatized?*

Naturally, because the model is an abstraction, it cannot capture specific features that may vary depending on an organization's specific needs. As a result, the D4.2 Cybersecurity threat analysis model anticipates that businesses will be able to select only those pieces that are appropriate for their needs, as well as change the sequence of the components to better represent their process.

2.4.6 Measurable research indicators

The assessment of models, methods maturity based on measurable research indicators. The searching is directly focused on the topic and is based on the precise combination of keywords (Table 1).

Number of scientific papers (libraries and journals, such as ACM digital library, IEEE Xplore Digital Library, ScienceDirect journals, books and articles, Springer Link and others)	5560 articles were found by a search engine. Cited 15 articles/models: [1], [2], [4], [38]–[40], [43]–[49], [52], [53]
Number of tools for particular model or method (existing open source and commercial visualization tools – Title, vendor or abstract)	4: [1]–[4]
Number of EU Elections for building resilience to all Cyber-Enabled threats (e.g. method of fake-news investigation)	0

Table 1: The results of measurable research indicators searching

Searching details (The date of searching: 2021-10-11):

- Search by syntax: results
- Justification: This is the aim of the detailed cyber-threat analysis model proposed.

2.4.7 Results interpretation and debate

The section provides the summary of analysed models, methods, tools, and their suitability for application and use in VAS (Table 2 and Table 3, and Table 4).

Criteria Method	Use cases	Type of visualized data	Type of visualized devices	Type of visualized objects	Could be applied to cyber security visualization
A Cyber Mission Impact Assessment Tool	Detection of cyber and physical attacks on critical infrastructure across Europe	Mission, Attack graph, network infrastructure	IT devices	IT devices	Y
CyGraph: Graph-Based Analytics and Visualization for Cybersecurity	Detection of cyber and physical attacks on critical infrastructure across Europe	Mission, Attack graph, network infrastructure	IT devices	Mission business processes & activities, threats, IT devices	Y
Cyber Attack Modelling and Impact Assessment Framework	Detection of cyber and physical attacks on critical infrastructure across Europe	Attack graph, network infrastructure	IT devices	Threats, IT devices & infrastructure	Y
Camus: Automatically mapping cyber	Detection of cyber and physical attacks	Mission, network infrastructure, incident detection	IT devices	Mission business processes, IT	Y

Criteria Method	Use cases	Type of visualized data	Type of visualized devices	Type of visualized objects	Could be applied to cyber security visualization
assets to missions and users	on critical infrastructure across Europe			devices & infrastructure, incident detection	
Proposed Method (BIA)	Detection of cyber and physical attacks on critical infrastructure across Europe	Mission, Attack graph, network infrastructure	IT device & ICS devices	Mission business processes & activities, threats, IT devices & infrastructure	Y

Table 2: Comparison of visualization methods

Criteria Model	Use cases	How relevant is this model to cyber security visualization	How interpretable is this model in cyber security visualization	Does this model incorporate credible cyber security visualization	Easy to enrich this model for its adoption for cyber security visualization
A Cyber Mission Impact Assessment Tool	Detection of cyber and physical attacks on critical infrastructure across Europe	13 citations	Specifically crafted for VAS	Yes	No, closed source
CyGraph: Graph-Based Analytics and Visualization for Cybersecurity	Detection of cyber and physical attacks on critical infrastructure across Europe	36 citations	Specifically crafted for VAS	Yes	No, closed source
Cyber Attack Modelling and Impact Assessment Framework	Detection of cyber and physical attacks on critical infrastructure across Europe	112 citations	Specifically crafted for VAS	Yes	No, closed source
Camus: Automatically mapping cyber assets to missions and users	Detection of cyber and physical attacks on critical infrastructure across Europe	44 citations	Specifically crafted for VAS	Yes	No, closed source
Proposed Method (BIA)	Detection of cyber and physical attacks on critical	Specifically, being created for T-SHARK	Specifically crafted for VAS	Yes	Yes, specifically being created for T-SHARK

Criteria Model	Use cases	How relevant is this model to cyber security visualization	How interpretable is this model in cyber security visualization	Does this model incorporate credible cyber security visualization	Easy to enrich this model for its adoption for cyber security visualization
	infrastructure across Europe				
Cybersecurity threat analysis model (D4.2)	The model's goal is to show and explain which functional components are involved in the analytical process	Specifically, being created for T-SHARK	Specifically, being created for T-SHARK	Yes	Yes, specifically being created for T-SHARK

Table 3: Comparison of visualization models

Criteria Tool	Use cases	Ability for application to cyber security visualization	Ability to import / export data	Integration ability with other tools	Format of data for import	Format of data for export	Pricing
A Cyber Mission Impact Assessment Tool	Detection of cyber and physical attacks on critical infrastructure across Europe	Mission, Attack graph, network infrastructure	No	No	Not described	Not described	Not described
CyGraph: Graph-Based Analytics and Visualization for Cybersecurity	Detection of cyber and physical attacks on critical infrastructure across Europe	Mission, Attack graph, network infrastructure	No	No	Not described	Not described	Not described
Cyber Attack Modelling and Impact Assessment Framework	Detection of cyber and physical attacks on critical infrastructure across Europe	Attack graph, network infrastructure	No	No	Not described	Not described	Not described
Camus: Automatically mapping	Detection of cyber and	Mission, network infrastructure,	No	No	Not described	Not described	Not described

Criteria Tool	Use cases	Ability for application to cyber security visualization	Ability to import / export data	Integration ability with other tools	Format of data for import	Format of data for export	Pricing
cyber assets to missions and users	physical attacks on critical infrastructure across Europe	incident detection					
Proposed Method (BIA)	Detection of cyber and physical attacks on critical infrastructure across Europe	Mission, Attack graph, network infrastructure	Yes, using information exportation layer. Each layer will be crafted to allow import data.	Yes, using plugins to connect to each layer.	Json objects currently being defined. Other objects for each layer are also being considered: Infrastructure layer, pcap1; Business process layer, xes2; vulnerability layer: CVE; incident layer: STIX3; impact assessment layer, JGF4	Same considerations as data for import	Free for T-SHARK.

Table 4: Comparison of visualization tools

[1] https://www.tcpdump.org/pcap.html
[2] http://www.xes-standard.org/
[3] https://stixproject.github.io/data-model/1.2/incident/IncidentType/
[4] http://jsongraphformat.info/

Chapter 3 Visual Analytics for Cyber Security Domain

Visual analytics is a complicated subfield of data analytics that focuses on the use of information visualization approaches to facilitate successful data analysis through the use of visual and graphical representation [63]. In the realm of multi-attribute, multidimensional data processing, and cognition, visualization plays a crucial role. In complicated data, visual data analysis can help detect structures, repetitions, anomalies, patterns, and trends, as well as exceptions [90]. By providing effective tools and settings, cyber security data visualization aims to assist domain specialists in decision making to improve the safety and security of our cyber infrastructures [91]. Use data visualization techniques developed in the field of information security, particularly for visualizing network traffic and cyber-attacks [92]. Situational awareness can benefit from these visualization techniques [93]. Effective data visualization allows domain analysts to derive significant insights that help them build successful strategies and make educated decisions [94]. Identifying who (users) and what (applications) are responsible for a cyber-attack or post-attack investigation is a difficult undertaking. Stakeholders in cyber security require good visualization methods and tools that enable them to quickly explore and analyse massive amounts of log data in a visual and interactive manner.

The Five Question Method of Five W's and How (Why, Who, What, When, Where and How) approach is used as a methodological framework in this part to describe a significant number of cyber security and digital forensics visualization studies [7]. [8] used a similar method to gather a study of visual analytics in the deep learning domain. Many well-known research papers could be examined by creating the survey in this manner. Such a structure encapsulates the requirements, stakeholders, and methodologies of cyber security visualization, and places other writers' work in the context of existing literature. Finally, we will go through some of the current research challenges. This survey is intended to help scholars and practitioners understand how visualization aids research and decision making in the cyber security sector.

3.1 Methodology

Using a five-question approach [95], we present a comprehensive overview of visualization and visual analytics in cyber security research, allowing us to position each analysed research contribution in relation to the following questions: Why, Who, What, How, When, and Where, while also discussing and emphasizing multifaceted contributions known research contributions. Our survey examines the most relevant work from the confluence of cyber security, digital forensics, human-computer interface, usability engineering, and computer vision domains to emphasize and analyse the influence of visual analytics on the cyber security domain. The solution shows how visual analytics has become critical to solving some of the most difficult challenges in cyber security and digital forensics, such as interpreting network data flow.

We use the Clarivate Analytics Web of Science, Scopus, IEEE Xplore, and ACM Digital Library databases to find papers. We looked through metadata, abstracts, explicit contributions, and set problems and requirements for each study (if any). We used the Five-Question Framework to examine the existing efforts on visualization and visual analytics in cyber security using this information.

- **Why** do we want to use visualization in the cyber security domain?
- **Who** are the stakeholders who use and benefit from cyber security visualization?
- **What** data, features, and relationships can we visualize in cyber security?
- **How** can we visualize cyber security data?
- **When** is visualization used in cyber security?

- **Where** is cyber-security visualization used?

Our next step is to use J. Bertin's semiotic theory of graphics [72]. This hypothesis is based on the widely held belief that communication involves the use of meaningful signs to share information. J. Bertin listed the most important visual variables (color, size, shape, hue, texture, and orientation) and proposed how to use them to represent various forms of data, including quantitative, ordinal, selected, and associative data.

While the theoretical frameworks outlined above provided many important suggestions and directions for visualization, they lacked conceptual operationalization for use in specific areas such as cyber security [73].

3.2 Findings and results

Our findings and results are presented in the following subsections.

Why: Motivation for visualization in cybersecurity?

One of the most important reasons cybersecurity stakeholders want to visualize data is to analyze and explain massive amounts of data that would otherwise overwhelm an expert due to its quantity. Cyber security analysts have powerful analytical and visualization capabilities to deal with the magnitude and complexity of the challenges provided by these data [89]. Data visualization is frequently incorporated into cyber security platforms, providing a simple way to view activities that contribute to more effective cyber operations. Identifying an abnormality is one method to learn more about high warnings. Models and trends are fascinating, but cyber analysts frequently need to produce remarkable results. Seeing data from multiple cases or alerts at once can indicate unexpected patterns of activity that need to be investigated further.

Who: users of visualization in cybersecurity?

The effectiveness of user-based evaluation in demonstrating the measurable benefits of data visualization systems and the influence of visualization tools on achieving security goals. Graphics visualization is used by analysts to present a global and/or local picture of their related data. Its main function, according to cyber security visualization experts, is decision support [74]. Matching these aims is challenging, if not impossible, so it is important to understand the trade-offs when comparing alternative datasets or systems.

What: cybersecurity data used for visualization?

Any cyber security strategy must include graphic visualization. Analysts are attempting to extract information from complex logs, which limits their investigation capabilities and exposes systems to risk. The cyber data visualization system must aid analysts in the decision-location maker's and produce tailored network visualization results that allow analysts to analyze data at their own pace and size. Analysts must be able to communicate data insights to each other. Filtering, temporal analysis, geospatial visualization, anomaly detection, social network analysis, and node grouping are essential visualization features. A data analyst cannot avoid a cyber-attack, but visualization of the visuals can help them understand the incident and prevent it from happening again.

Visual data system objects created in conjunction with real users are frequently convincing examples of practice because they can reveal previously undiscovered information within a company. The layout, filtering, and social networking of the performers are the most significant aspects, however, they must be combined with a robust rendering engine. However, real-life data sets are rarely released, even anonymously, making meta-analysis in the following research problematic [75]. These sets of data items are also unmanageable in terms of size and complexity and cannot be compared to data sets from other organizations. As a result, the characteristics of actual displayed data item sets should specify their traits and properties to enable future research.

How: Visualization Metaphors and Techniques

Data items are visualized as pictures with icons, and data are mapped to visual elements. Shapes (such as lines, bands, or bars) and visual elements such as color, texture, and shape diameter can all be used in icons. Textures and animation can be used to enhance the portrayal [76]. Each network element is represented by an icon in VisSecAnalyzer [77], with the background color of the icon encoding values of security metrics computed for a network host, such as Risk Level, Mortality, or Criticality.

Radial icons are based on a wheel metaphor that uses a radial layout to represent geographical distribution and multidimensional properties. IDSRadar, for example, is a real-time display platform for intrusion detection system (IDS) alarms [78]. When viewing large volumes of network data, a wheel displays its advantages for compact architecture.

Word clouds are a text visualization technique based on keywords in which each visual feature symbolizes a text chunk (for example, labels) in a document. Word clouds, for example, are used in TagsNet [79] for the forensic triage of narratives from email networks.

Depending on the data distribution, a heat map (or density map) integrates discrete data points into a continuous image. Heatmap is a good tool for qualitative analysis, since it handles overlapping issues well. In SemanticPrism [80], for example, a heatmap is used to depict the geospatial distribution of a policy or activity.

The dynamics and similarities in complicated networks can be seen using graphs. Although graphs are a logical way to depict computer networks, they can have scalability issues. Additional visualization approaches, such as graph clustering or node coloring, can be utilized to solve these problems. To adequately describe actions in the computer network, for example, ENAVis [81].

Bubble charts are useful for spotting clusters and outliers in data, and their effectiveness is independent of the dataset's size. Bubble charts and cartographic maps, for example, are used in BubbleNet [82] to find and describe patterns in the geographical distribution of cybersecurity data.

Icicle plots are a method for presenting hierarchical/clustered data. For example, in LifeFlow [83], icicle plots are used to visualize the temporal distribution of sequential events.

To give interactive analytical capacity, an interactive lens (or semantic zoom) allows the user to emphasize distinct properties of a large dataset at different scales and zoom levels. For example, in Portvis [84], semantic zoom enables detailed presentation of information by allowing zooming in and out.

In temporal visualizations, the metaphor of a 'river' is used, with information flow reflecting the passage of time from left to right. The alluvial plot can be used to represent the metaphor of the river. Such a depiction, for example, is used to show temporal variation in [85].

Tree maps use a collection of tiles of varied sizes and colors to create a hierarchical display of data. It gives you a bird's-eye view of multidimensional, interconnected data. Treemap visualization is used to depict the hierarchical order of the source code in the Prefuse visualization toolkit [86], for example, while color reflects the date of last modification.

When: When is visualization used in cyber security?

The most prevalent application of visualization in cybersecurity is for historical data analysis, such as post-attack digital forensics. A growing number of papers address the challenge of real-time data processing and visualization, for example, to detect cyber-attacks or malware spread in real time [78]. Finally, a third line of inquiry looks at the difficulty of predicting future cyberattacks [87].

Where: Where is cyber-security visualization used?

Any of the attributes in the data file can be used as cyber threat intelligence objects. All of this is due to the huge, complicated, and diverse nature of the data. Intelligence is the knowledge of threats based on evidence. The data visualization system's job is to assist the user in making decisions and reacting. We must supply intelligence to them, whether it is technical knowledge about tools and

technologies used by the adversary, or contextual intelligence about actor motivation or distinct campaigns. Graphics are an excellent approach to convey such magnitude, complexity, and variation in a user-friendly way [88].

3.3 Visualization Tools for Cyber Security Domain

We have analysed eight mostly popular VAS tools and summaries analyses results in Table 5

Tool / *Availability*	Key features	Advantages	Disadvantages
BOARD *Request form for a free demo*	- Performance - Programming-free - Powerful Reporting - Multi-lingual	- All-in-One Platform - Cloud or On-premise - Real-time Modelling - Server Clustering - User Self-sufficiency	- Is extremely configurable - The solution involves a steep learning curve - Does not offer support for Linux and Mac - Does not provide detailed documentation - Setting up security profiles is a time-consuming process
Microsoft Power BI *Try free registering e-mail* **Microsoft Power BI Desktop** *Free*	- Provides powerful ad-hoc analytics - Solution is loaded with data mining features - Data source Integration - Data Modelling - Data Organization - Mobile Access - Seamless Integration	- Live 360-Degree View of Data - Securely Share Insights - Visualize Data Trends - Make Data-Driven Decisions - Collaborate Across Organization	- No direct integration available - Graphical visualization is limited - Doesn't support SQL queries - Comparatively difficult to work with huge data sets - Doesn't let users build scheduled reports
SAS *Request form for demos, pricing and free software trials. Try it free for up to 14 days.*	- Ad Hoc Reporting and Analysis - Predictive Analysis - Mobile Apps - Interactive Discovery - Location Analytics - Trend Indicators - Visualizations. - Scheduled Reporting - Customizable Dashboard - Embedded BI	- Easily Create Models and Explore Data - Discover Data Relationships and Patterns - Visualize Data in Intuitive Graphics - Glean Data-Based Insights - Make Better Business Decisions - Share and Collaborate	- SAS Visual Analytics is one of the most difficult solutions to implement - Lacks efficiency in data preparation and data management - Requires manual work in forecasting - Content creation is very difficult and time-consuming - UI isn't user-friendly
Tableau *Tableau Desktop: Start your free 14-day trial*	- Multiple Data Source Connections/Integrations - Maps and Geographic Data Analysis - Natural Language Interaction - Recommended Visualization - Smart Table Join and Source Recommendation - Time Series Analysis	- Powerful Visualization Capabilities - Ease of Use - High Performance - Multiple Data Source Connections and Impressive integration - Mobile-Friendly - Support	- Tableau Server does not support data encryption - Does not provide multi-location support - The solution is not user-friendly - Does not allow users to share datasets, and requires a separate subscription package - Requires IT support for integrations and other high-end tasks
TIBCO Spotfire *Try TIBCO Spotfire Analytics Experience the full capability of Spotfire® Analytics with a 30-day free trial.*	- Location Analytics - Real-Time Streaming Analytics - Visual Analytics - Intelligent Data Wrangling	- Scale for Large (or Small) Enterprises - Built for the Individual Future: Spotfire is equipped with advanced predictive analytics tools - Solve Tough Problems - Access to the TIBCO Community	- Calculating custom fields is fairly difficult - Does not offer robust metadata functionality - Difficult to customize visualizations - Does not let users specify custom SQL instructions when accessing a data source

Tool / *Availability*	Key features	Advantages	Disadvantages
			- Difficult to write and understand Spotfire coding
PunchPlatform *Licensed under* **Thales** *Inner Source licenses Free for demonstration only. Cannot be used in production without license. No customer support Potentially unstable and breaking compared to previous versions.*	- Pipelines - Channel: - ever-running streaming components; - batch processing - administrative tasks - Tenantse - Platform.	- Scale to hundreds of thousands of events per sec - Do not loose any data - Cheap yet resilient - Custom parser is very easy to develop	- Authorization is not fine grained - Needs to deploy a PKI - Needs to provide a company LDAP
Skydive *Free Apache License 2.0*	- Captures network topology and flows - Full history of network topology and flows - Distributed -Ability to follow a flow along a path in the topology -Supports VMs and Containers infrastructure -Unified query language for topology and flows (Gremlin) - Web and command line interfaces - REST API -Easy to deploy (standalone executable) -Connectors to OpenStack, Docker, OpenContrail, Kubernetes	- Easy to deploy - OpenSource and good community	
Maltego *Free - Community / Kali . The basic version of* **Maltego** *for getting started. Not for commercial use.* **Maltego CE** *Free version returning up to 12 Entities per Transform* **CaseFile** *Free version for commercial use to visualize connections in offline data*	- Commercial Use -Access to commercial Transform Hub -Use with Internal Transform servers Standard OSINT Transforms - Technical support - Graph Export (CSV, XLS, XLSX, PDF and Image formats) - Graph Import (CSV, XLS, XLSX) - Shared Graph Sessions (Collaboration) - Machines (Transform Macros)	- The CTAS server includes transforms for gathering OSINT from common sources on the Internet which includes querying DNS servers, search engines, social networks, various APIs and other sources. - Out-of-the-box Maltego clients use Paterva's public CTAS server to run transforms that are located on the Internet and used by other Maltego users. - CTAS is delivered as a VMware image in OVF format - The server is delivered as a virtual image - The server is (currently) built on a Ubuntu LTS 14.04.01 base - The search engine transforms (*_SE) use the Bing Web-Search API	- Maltego CE - free, non commercial use - Maltego Classic - paid, commercial use - Maltego XL - paid, commercial use - Maltego CaseFile - free, commercial, no Transform
Elastic SIEM *Register for free 14-day trial*	- Return results from ad hoc queries in seconds with the speed of Elasticsearch's schema-on-write architecture	- Better query experience - Faster historical queries and aggregations - Logs to metrics - Truth in time	- Elastic SIEM – free 14-day trial available. - Running repeated queries on top of unstructured logs that execute complex regex statements to extract fields

Tool / Availability	Key features	Advantages	Disadvantages
Free Elastic Stack Open Source - Apache 2.0 license *Basic (Everything in Open Source plus: Core security features, Solutions such as APM, SIEM, Maps, and more.). The forever-free plan.*	- Visualize and interact with your data on custom dashboards - Drill into events of interest and pursue breadcrumbs by pivoting through underlying data.	- Real time anomaly detection and alerting - Logs in observability initiatives - Data quality - Granular access control	consumes more RAM and CPU resources on an ongoing basis.

Table 5: Summary of tools analysis results

3.4 Summary of results

The results [63] of our survey are summarized in Table 6 was published: Comparison of visualization capabilities of visual analytics platforms and tools for cybersecurity according to the Bertin's [72] visual variables (Y = yes; N = no; G = good; M = marginal; P = poor; NA= not applicable) based on the characteristics of Visual Variables [72] as follows:

- **Associative (A)**: Is change in this visual variable enough to allow us to perceive them as a group?
- **Selective (S)**: Is change in this visual variable alone enough to allow us to select it from a group?
- **Ordinal (O)**: Are changes in this variable perceived as ordered?
- **Numerical (N)**: Is there a numerical reading obtainable from changes in this visual variable?

Tool	Syntactic	Size	Shape	Orientation	Color Hue	Color value	Texture	Reference
CRUMBS	Associative	**N**	**Y**	**Y**	**Y**	**N**	**Y**	[96]
	Selective	**Y**	**N**	**Y**	**Y**	**Y**	**Y**	
	Ordinal (ordered)	**P**	**P**	**M**	**G**	**G**	**M**	
	Numerical (quantitative)	**P**	**P**	**M**	**P**	**P**	**P**	
Taggle	Associative	**N**	**N**	**Y**	**Y**	**Y**	**N**	[97]
	Selective	**N**	**N**	**Y**	**Y**	**N**	**N**	
	Ordinal (ordered)	**G**	**P**	**P**	**M**	**M**	**P**	
	Numerical (quantitative)	**P**	**M**	**M**	**G**	**P**	**P**	
Bespoke	Associative	**N**	**N**	**N**	**N**	**Y**	**N**	[98]
	Selective	**Y**	**N**	**N**	**N**	**Y**	**N**	
	Ordinal (ordered)	**P**	**P**	**P**	**P**	**M**	**P**	
	Numerical (quantitative)	**P**	**P**	**P**	**P**	**P**	**P**	
FLUKES	Associative	**N**	**N**	**N**	**Y**	**Y**	**N**	[99]
	Selective	**N**	**N**	**N**	**Y**	**Y**	**N**	
	Ordinal (ordered)	**P**	**P**	**P**	**G**	**P**	**P**	
	Numerical (quantitative)	**P**	**P**	**P**	**G**	**P**	**P**	
Visualdrives	Associative	**Y**	**N**	**N**	**Y**	**N**	**N**	[100]
	Selective	**Y**	**N**	**N**	**Y**	**N**	**N**	
	Ordinal (ordered)	**G**	**P**	**P**	**G**	**M**	**P**	
	Numerical (quantitative)	**G**	**P**	**P**	**G**	**M**	**P**	

Tool	Syntactic	Size	Shape	Orientation	Color Hue	Color value	Texture	Reference
Insight	Associative	**N**	**N**	**N**	**Y**	**N**	**N**	[101]
	Selective	**N**	**N**	**N**	**Y**	**N**	**N**	
	Ordinal (ordered)	**P**	**P**	**P**	**M**	**P**	**P**	
	Numerical (quantitative)	**P**	**P**	**P**	**M**	**P**	**P**	
CyberForensic TimeLab	Associative	**N**	**N**	**Y**	**N**	**N**	**N**	[102]
	Selective	**N**	**N**	**Y**	**Y**	**N**	**N**	
	Ordinal (ordered)	**P**	**P**	**G**	**M**	**P**	**P**	
	Numerical (quantitative)	**p**	**p**	**G**	**M**	**P**	**P**	
BubbleNet	Associative	**Y**	**N**	**Y**	**Y**	**Y**	**N**	[82]
	Selective	**Y**	**N**	**Y**	**Y**	**N**	**N**	
	Ordinal (ordered)	**G**	**P**	**P**	**G**	**G**	**P**	
	Numerical (quantitative)	**G**	**P**	**P**	**G**	**G**	**P**	
Change-Link 2.0	Associative	**N**	**N**	**N**	**Y**	**Y**	**N**	[103]
	Selective	**N**	**N**	**N**	**Y**	**Y**	**N**	
	Ordinal (ordered)	**P**	**M**	**P**	**G**	**G**	**P**	
	Numerical (quantitative)	**P**	**M**	**P**	**G**	**G**	**P**	
LogAnalysis	Associative	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	[104]
	Selective	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	
	Ordinal (ordered)	**G**	**G**	**G**	**G**	**G**	**P**	
	Numerical (quantitative)	**G**	**G**	**G**	**G**	**G**	**P**	
Topic models	Associative	**N**	**N**	**N**	**Y**	**N**	**N**	[105]
	Selective	**N**	**N**	**N**	**Y**	**Y**	**N**	
	Ordinal (ordered)	**P**	**P**	**P**	**G**	**M**	**P**	
	Numerical (quantitative)	**P**	**P**	**P**	**G**	**M**	**P**	
ContraVis	Associative	**Y**	**Y**	**N**	**Y**	**Y**	**N**	[106]
	Selective	**Y**	**Y**	**N**	**Y**	**Y**	**N**	
	Ordinal (ordered)	**G**	**G**	**M**	**G**	**G**	**P**	
	Numerical (quantitative)	**G**	**G**	**M**	**G**	**G**	**P**	
XFake	Associative	**N**	**N**	**N**	**Y**	**Y**	**N**	[107]
	Selective	**N**	**N**	**N**	**Y**	**N**	**N**	
	Ordinal (ordered)	**P**	**P**	**P**	**G**	**M**	**P**	
	Numerical (quantitative)	**P**	**P**	**P**	**G**	**G**	**P**	
ExplorViz	Associative	**N**	**N**	**N**	**Y**	**Y**	**N**	[108]
	Selective	**N**	**N**	**N**	**Y**	**Y**	**N**	
	Ordinal (ordered)	**P**	**P**	**M**	**G**	**G**	**P**	
	Numerical (quantitative)	**P**	**P**	**M**	**G**	**G**	**P**	

Table 6: Comparison of visualization capabilities of visual analytics platforms and tools

Visual analytics is a complicated sub-field of data analytics that focuses on using information visualization approaches to facilitate effective data analysis through the use of visual and graphical representation. In the realm of multi-attribute, multi-dimensional data processing and cognition,

visualization plays a crucial role. Using the Four Questions approach, we give a complete overview of tools for visualization and visual analytics in the cyber security area, allowing us to place each analysed tool in relation to the following questions: Features, Key concepts, Advantages, Limitations and situations where this tool may be useful for VAS.

Chapter 4 Visual Analytics System for Cybersecurity Threat Analysis

4.1 Description of the VAS need

Every day, security experts are confronted with an increasing number of security incidents that have an impact on people's well-being, their information systems, and, in some cases, vital infrastructure. The sooner they can detect and comprehend these hazards, the better equipped they will be to minimize and examine them forensically. As a result, they must be aware of current security events and their potential consequences. However, given the enormous number of events, security analysts and researchers may find it challenging to keep up with the flow of information and answer the following questions in near-real time: What are the current security events? How long are they going to last? In this use case, we will aim to address these concerns by utilizing cyber goodies that offer a wealth of useful information on a variety of topics. However, due to the large amount of data, extracting useful information can be difficult. The proposed use case is a cyber-threat analysis system that can identify, geolocate, and categorize cybersecurity events in near-real time across a stream of cyber threats. To begin detecting events, IT is based on correlations between cybersecurity incidents and a collection of seed keywords that characterize the type of events that follow. It automatically discovers additional relevant phrases, such as virus names or compromised IP addresses, using these seed keywords to broaden the detection range while remaining in the same domain.

The data and threat analysis process provides a high-level overview of the architecture. The inner workings of this use case can be split into three sections. To retrieve a continuous stream of Reputation in Phase 1, the stream generator uses Threat intelligence, Netflows, Reputation data providers APIs, and a set of keywords. The feeds are pre-processed before being stored in a database. The event detection algorithm then recognizes events across the feed stream in phase 2. Events are made up of a collection of the same IP's reputation over a period. These occurrences are geolocated, categorised, and shown on a user-friendly interface. The system is unaware of new keywords such as new malware, spam, bot-networking (c&c) naming, or a new sort of cyberattack because we employ predetermined keywords to construct the stream in phase 1. To obtain those papers, we developed a component called the keyword finder, which automatically tracks new security keywords and adds them to the existing list of keywords. This component searches the previous compromised IPs for new relevant keywords and adds them to the stream generator's keyword list (phase 3). We were assessing potential dangers at the same time. The stream is then supplemented with VAS characteristics containing the detected keywords. Visualization has parallels to achieve more integration between the VAS system and the analyst.

4.2 Business Impact Assessment

Proposed Business Impact Assessment (BIA) [68], a tool gathers knowledge about an infrastructure, assesses the impact of a cyber-threat and performs VAS. The overall model is displayed in in Figure 5. The method collects, aggregates, and shares information to VAS about IT system infrastructure, cyber-threats, affected business critical operations, and impact assessment results; and is constructed as a multi-layered mission-aware impact assessment model.

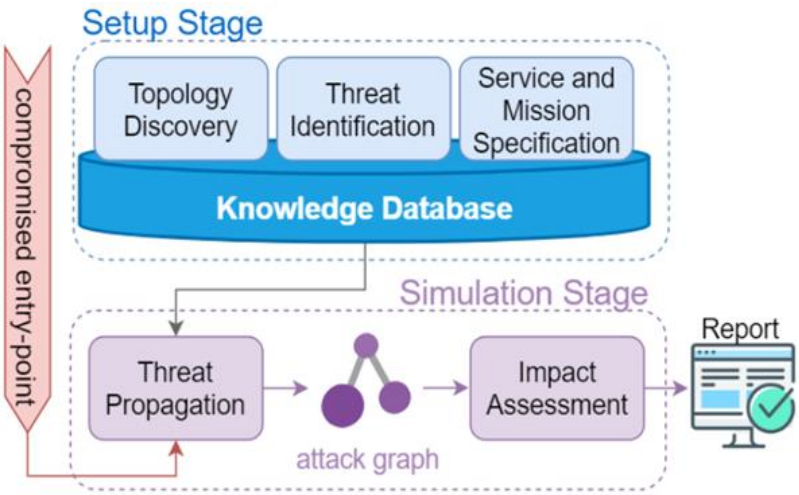

Figure 5: Automated cyber VAS data gathering method

The goal of BIA approach is to provide a solution to understand how cyber-threats can be leveraged to impact the organization's mission and identify the business-goals and processes compromised by an exploited threat. BIA is envisioned to be easily integrated with current approaches, tools and standards, and its design is two-fold: (1) to create a multi-layered evaluation model for MIA that can be easily integrated with current information sources and (2) to put forward a simulation platform that allows to reproduce how the impact of exploited cyber-threats propagate throughout the organization's infrastructure and to assess the impact on the organization's mission. To do so, this work proposes a two-stage approach for MIA and is architected as illustrated in Figure 5. BIA's general idea is to first create a knowledge database with the organization's cyber infrastructure and mission profile –the Setup stage –to then be used to simulate the impact of a user-chosen compromised entry-point on the organization's mission –the Simulation stage. The approach takes a set of three knowledge units as input during the Setup stage and a compromised entry-point during the Simulation stage to generate a MIA report as the output.

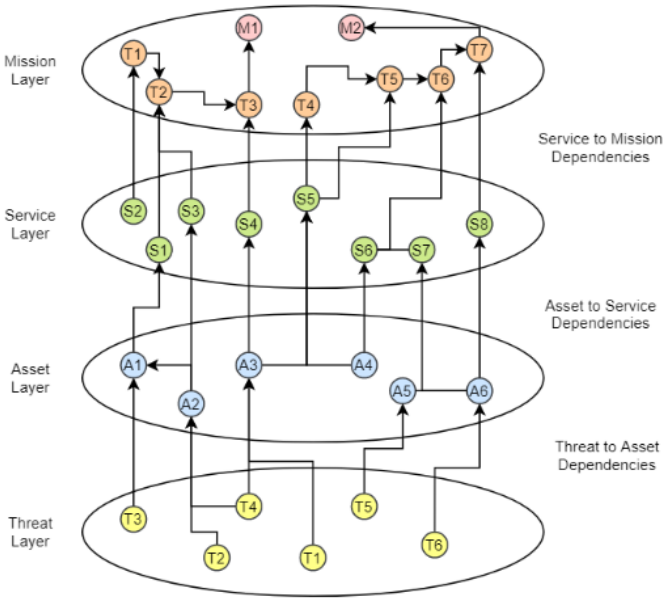

Figure 6: BIA evaluation model

4.2.1 Setup stage

The central idea of this stage is to capture the cyber infrastructure and business information and consolidate it in an integrated data representation to be interpretable by the simulation. The data representation proposed to map the organization's cyber infrastructure onto the business-objectives is based on a four-layer evaluation model, as depicted in Figure 6.

To populate the evaluation model based, BIA's Setup comprises three knowledge units that mine different data sources to extract the required information: a Topology Discovery unit, a Threat Identification unit and a Service and Mission Specification unit.

The **Topology Discovery unit** aims to gather information about the asset layer by receiving two types of inputs: (1) network packet captures and (2) firewall configuration, which are handled using two different components, a Network Discovery component, and a Connectivity Discovery component. At the end, this knowledge unit stores its findings about discovered assets and reasoned connectivity between assets, in the knowledge database to be used by the next knowledge units. The Network Discovery component resorts to a network analyser tool that receives packet captures containing network communications exchanged between the IT components of the infrastructure under evaluation. Using basic dissection techniques[1], those packet captures are parsed to extract information about the infrastructure's assets, such as Internet Protocol (IP) addresses, and their connectivity, such as network protocols and ports used. The Connectivity Discovery inspects firewall configuration, given as input, to infer missing allowed communications. To assess the communications that are effectively allowed by the firewall policy environment, the Connectivity Discovery component comprises two algorithms: the Comparing Algorithm to first assess allowed connectivity by each individual firewall, and a Filtering Algorithm to address firewall hierarchy and assess which rules survive the filtering action. Both algorithms work as follows:

- *Comparing Algorithm* - When a packet arrives at a firewall it is tested against each rule sequentially, meaning the firewall rules are order sensitive and the sequence of the firewall rule's list is to be taken into consideration when trying to understand which communication packets are effectively allowed. The proposed algorithm is designed to work as follows: first, take each deny-rule and compare it to the allow-rules that come next. It is possible to arrive to four possibilities, as suggested by previous work [70] and outlined in Figure 7; next, remove from the allow-rules the parts in common with the deny-rules (red zones in Figure 7). Applying this algorithm to all deny-rules in a firewall's configuration results in a list with only allow-rules (allow-list) that represent all the possible communications that may pass through the firewall.
- *Filtering Algorithm* - It assesses which rules survive the filtering hierarchy, and how. While traversing the network and the firewall infrastructure that constitute it, three distinct actions are proposed to be taken:(1) First, classify each allow-rule from the firewall allow-list according to its source and destination to understand which rules should be submitted to the other firewall's policies; (2) Next, rules with destination outside their firewall's domain are propagated to adjacent firewall's to be compared with their configurations and filtered accordingly; (3) Lastly, repeating this process to every firewall's rule list, results in a list of all allowed communications in the infrastructure
- *Combining both algorithms* - The inclusion of the missing connectivity reflects how the current firewall policy allows connectivity that may be leveraged by an attack to move within the network. This allow-list is translated to possible connectivity and is used to populate further the asset layer.

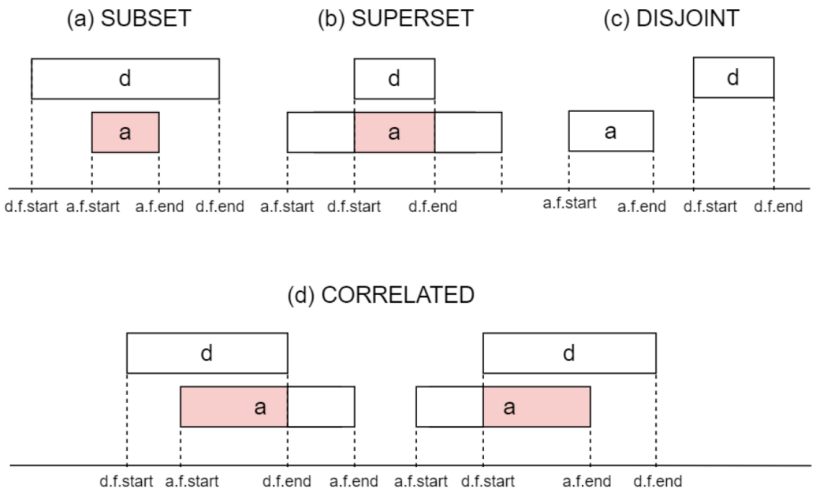

Figure 7: Comparing Algorithm applied to two firewall rules: deny-rule (f) and allow-rule (a).

The **Threat Identification unit** imports the information about the network topology already stored in the knowledge database and identifies the threats affecting the organization's cyber assets. With that objective in view, the unit's input is threefold: (1) a user chosen asset classifications to classify the organization's assets, (2) for each possible asset classification, a list of possible threats affecting that classification is provided and (3) to reduce the number of possible threats, a threat's classification is used to classify each identified threat, for instance, the STRIDE[5] framework for threat classification. The Threat Identification unit then proceeds to map threats with the corresponding assets, according to the user specified asset classification, threat identification and threat classification, and stores that information in the knowledge database.

The **Service and Mission Specification** unit creates identifies the organization's business critical operations associated with the assets found by the Network Discovery unit. Particularly, the organization's business-processes, represented by a collection of activities to be accomplished, which are provided by services running on assets. Thus, this unit receives business-processes specification and map this information to the assets already stored in the knowledge database.

4.2.2 Simulation Stage

Following the Setup stage, which results in a fully populated knowledge database based on the proposed layered model, the Simulation stage proceeds to simulate the impact of a user-chosen entry-point to the system and perform MIA. The Simulation stage was conceived as a simulation platform that leverages MulVAL [71] to perform MIA. Two components were implemented to achieve this purpose: a Threat Propagation component to convert the proposed Horn Clauses into MulVAL's knowledge base. These clauses are then used by MulVAL as rules to be validated by the organization's evaluation model and produce an attack graph; and a second component, the Impact Assessment component to extract relevant MIA information from the attack graph and present it to the user.

The **Threat Propagation** module takes the main stage for the impact propagation simulation, based on an attack graph model, where the goal is to determine whether a compromised asset is likely to deleteriously affect any of the business-goals of the organization. To this end, this module is designed as a simulation platform which is configured with the organization's infrastructure and mission identified and modelled by the Setup phase. The simulation begins with a user-chosen entry-point (a specific asset and exploited threat) and ultimately tries to determine which organizational business-objectives would be affected if that asset became unreliable or unavailable. Starting from that entry-point, the simulation performs a bottom-up analysis, searching for attack paths by leveraging the organization's model interdependencies to propagate the initial threat. If an asset is

[5] https://www.microsoft.com/security/blog/2007/09/11/stride-chart/

accessible and has a threat, then is exploitable and the simulation advances to that asset. Additionally, if an asset runs a service that has a role in the mission, then the threat's impact is propagated towards the mission's activity (or activities) the service supports, and, from there, to the business-process(es) that rely on those impacted activities. This threat propagation is achieved resorting to logic programming to express how the propagation advances with a set of series of Horn clauses, a logical formula that takes a particular rule-like form: $L_0 \leftarrow L_1,...,L_n$, where $L_i \forall i \in N$ are literals, and if $L_1,...,L_n$ are true then L_0 is also true. In the design of this module, the threat propagation was defined using the four following Horn clauses presented in Table 7.

#	Description	L_0	$L_{1,...,n}$
1	Entry-point compromised	*compromisedAsset(A)*	*attackerLocated(A), threatExists(A, Threat)*
2	Attack propagated to another asset	*compromisedAsset(A2)*	*compromisedAsset(A1), connectivity(A1,A2), threatExists(A2, Threat)*
3	Attack propagated to the service	*compromisedService*(S)	*compromisedAsset(A), runsService(A, S)*
4	Attack propagated to the business-process	*compromisedProcess(P)*	*compromisedService(S), runsActivity(S, A), runsProcess(A, P)*

Table 7: Horn calluses to define threat propagation in BIA

The procedure of combining the organization's multi-layered modelled entities and their dependencies and iteratively validating the clauses defined creates an attack graph depicting all the possible threat propagation paths found from the simulated entry-point to organization's business-processes. The resulting graph is then output by this module.

In the Threat Propagation module, MulVAL acts as a processor of Datalog rules to generate attack graphs, however, its original rules do not consider a threat and mission layer, hence, BIA reformulates MulVAL knowledge base by expressing the proposed four Horn Clauses for threat propagation as new Datalog rules, implemented as a part of interaction rules in MulVAL Interaction rules are based on primitive and derived facts to represent the preconditions and postconditions, respectively, of Horn Clauses. BIA transforms the organization's infrastructure and mission information identified in the Setup stage into primitive facts. MulVAL then applies the interaction rules towards the primitive facts and, if all preconditions are met, produces derived facts. In addition to facts and rules, MulVAL requires an initial point to start its verification process, and a target to direct and conclude that process. BIA defines MulVAL's target as the business-processes identified in the Setup stage that MulVAL will try reach, while the initial point is provided as an external input to this component and defined as the entry-point to the system by a<asset,threat>tuple. The entry-point is then transformed and combined with the rest of the primitive facts, which completes the required input to run MulVAL, and effectively triggers the start of the simulation. Furthermore, the entry-point is chosen by the user, which can choose to run the simulation several times for different entry-points independently of the Setup Stage. Here lies another main features of BIA's MulVAL extension, where every time the user chooses an entry-point, MulVAL's required input is automatically changed accordingly. The attack graph generated is output in PDF format (optional), together with two CVS files, one with the nodes and the other with all arcs present in the attack graph, and a TXT file with all this information combined. Since the graphically representation of the attack graph (in PDF) often results in an image difficult to digest at naked-eye, and it is the option that takes longer to produce

results, BIA's Threat Propagation component only outputs the two CSV files for the next component, to assess relevant MIA information.

The **Impact Assessment** module identifies the compromised assets and exploited threats, the explored connectivity between assets, and the business-processes compromised, and the propagation steps the simulation followed to advance throughout threatened susceptible assets towards the mission. This final analysis highlights relevant information and assembles it in a compact report for impact assessment. The module is implemented to parse the attack graph produced by the previous component and retrieve relevant information about the compromised performers, and the threats and connectivity exploited to that effect. This information is then presented to the user in JSON format, for its readability, and versatility to be further extended and integrated.

4.2.3 Visualization Analysis stage

The output of the simulation is a threat propagation graph describing the assets that may be impacted by the simulated threat, as can be observed in the example provided in Figure 8. BIA HMI displays a graph overview of the nodes involved in the simulated attack across all assessment layers: threat layer in red, asset layer in blue and the service and mission layer with yellow nodes.

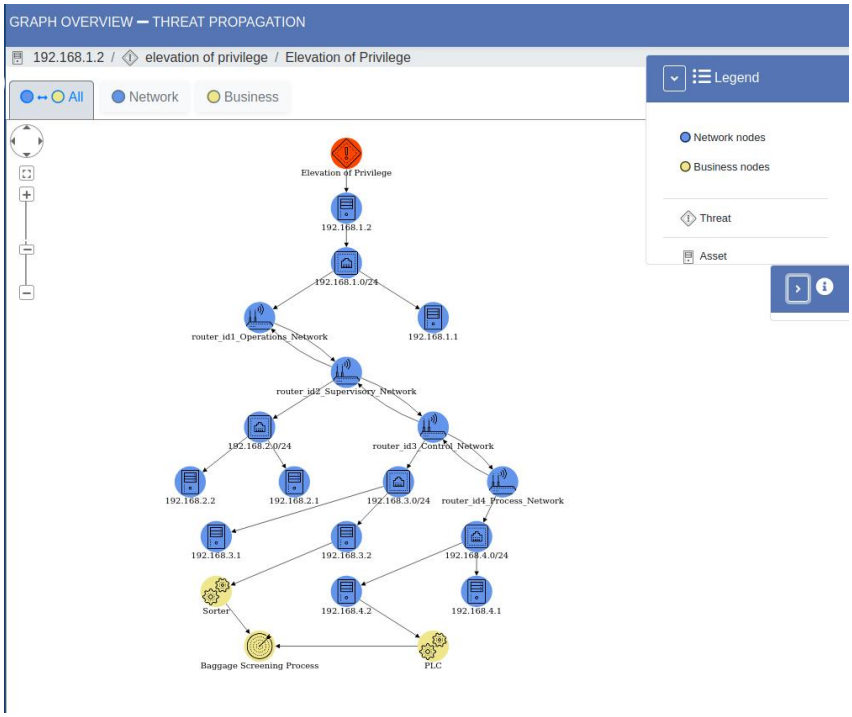

Figure 8: BIA dashboard simulation page for the BHS system

BIA allows to assess which organization's business processes are impacted and visualize how the threat propagated throughout the organization for impacting those processes.

To that end, BIA's simulation page includes three different views providing different perspectives of the simulated cyber-attack – *All*, *Network*, and *Business*. Although impact prioritization is ultimately dependent on the system, circumstances, and goals, BIA's different views allow visualizing keys aspects of the impact to support the decision-making process when addressing a cyber incident.

The *All* view presents the assets and business-critical processes affected by the cyber-attack offering a quick assessment of the impact on the system. From this view, the user can observe the main aspects of the impact, for instance:

- The business process that is closest to the entry-point, which can be perceived as the first business process that could be compromised and should be quickly addressed.

- The asset or service that supports the most processes and, if compromised, could lead to the most impactful attack.
- Using BIA's filtering properties that places focus on each business process, it is possible to conclude the process that is reached by the most propagation paths, which could be interpreted as to be the most vulnerable or the most certain to be compromised.

Although Figure 8 represents a particularly small example (with few assets and only one business process), BIA's results can aid in assessing the impact. For instance, using BIA's filtration properties to focus on propagation paths (Figure 9) it is possible to identify the shortest path it takes a cyber threat to impact a business process. In this example, the shortest path would be through the asset with IP address 192.168.3.2 that supports the Sorter critical service (yellow node "Sorter") of the Baggage Screening Process. This assessment can indicate where defensive measures could be most effective in stopping the attack from propagating further.

Furthermore, the *Network* view displays a filtered graph solely covering the affected assets (Figure 10), offering the SOC Operator a clear view of the affected infrastructure.

Finally, a *Business* view (Figure 11) presents a detailed view of all business-critical processes and operations affected by the cyber-attack while summarizing the infrastructure impact by only displaying the network nodes representing the entry-point and the impacted assets that directly support critical services.

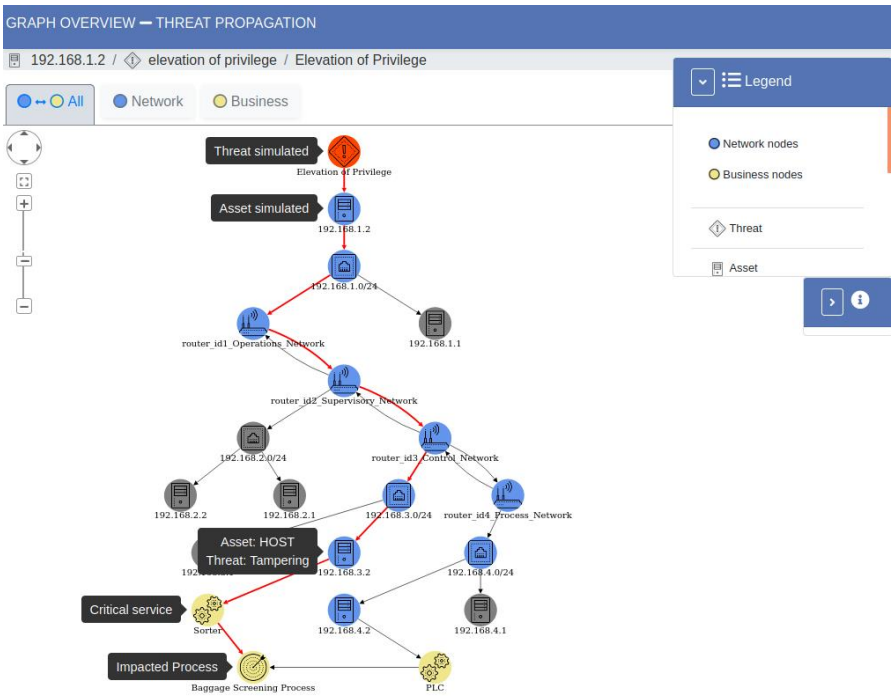

Figure 9: BIA simulation page using filters to focus on one propagation path

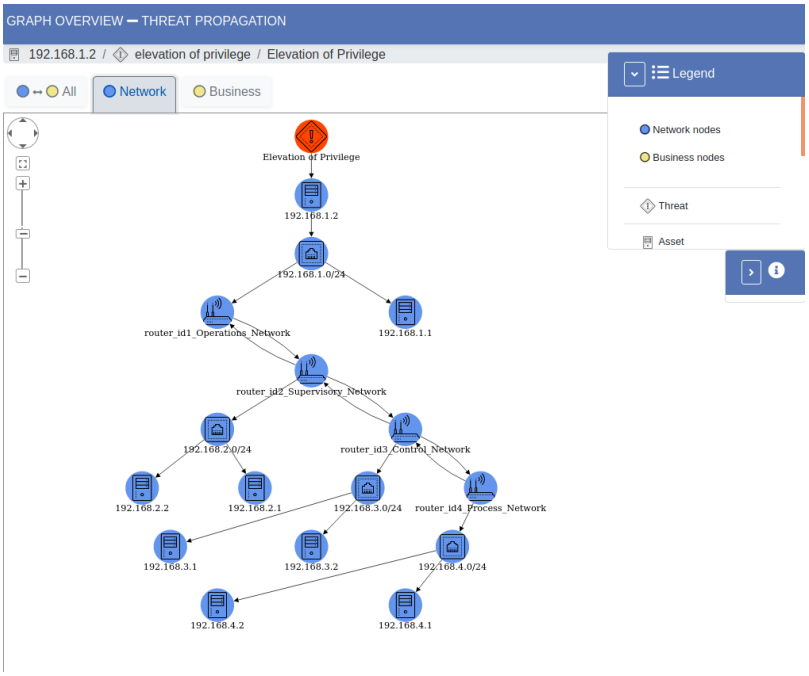

Figure 10: BIA simulation page - Network view

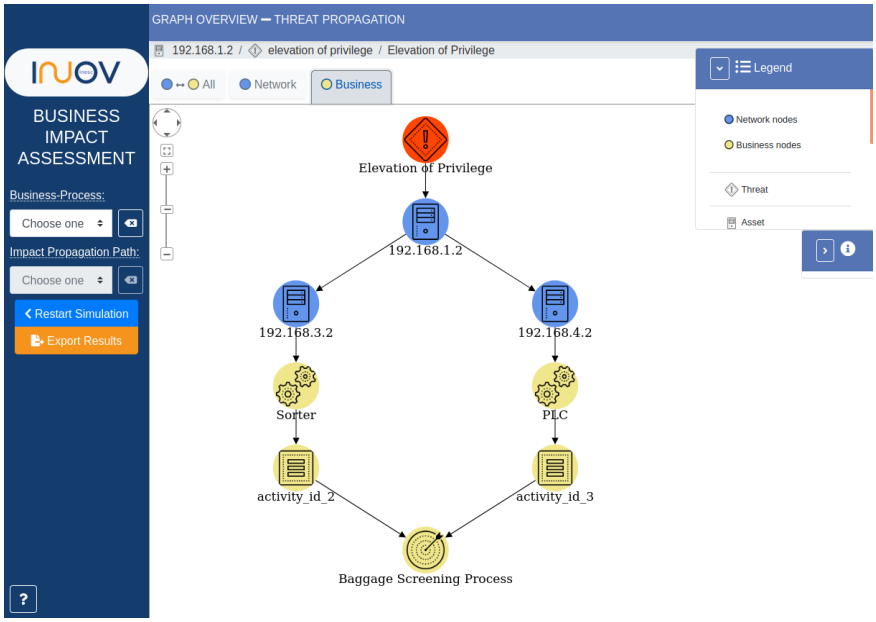

Figure 11: BIA dashboard simulation page - Business view

4.3 Linking VAS with others T-SHART project tasks

In this section, we propose basic concept of the Visual Analytics System (VAS) (WP4 T-shark project T4.2 task) and its relationship with other tasks of WP4.

Various threat intelligence data such as TTI, OSINT, HUMINT, GEOINT, FININT, SIGINT, CYBINT, SOCINT, CSTI and many others may be in the form of unstructured data. Such data may be used as inputs and for visualization to create objects for visualization in the VAS. Main concept and relationship between T4.2 task “Visual Analytics” Challenges Contest and others WP T-shark project tasks proposed by KTU depicted in Figure 12.

The main task is task **T4.1 "Staging and pilot management**", which contributes in staging and pilot management. Various threat intelligence tools collect lots of unstructured data from various sources and that process is analysed in task 4.3 "All Data based threat intelligence". After unstructured data are collected, it must be structured and transformed to the visualization object set for further processing. Task T4.2 "Visual analytics" gets visualization objects as input and creates output in the appropriate form for VAS visualization tools that produce visual output for cybersecurity expert-analytic, who must evaluate visualization results and make decisions (Figure 13). Finally, the results from tasks **4.3 "All Data based threat intelligence"** and T4.2 "Visual analytics" are shared for further processing to the task T4.4 "Information sharing and integration".

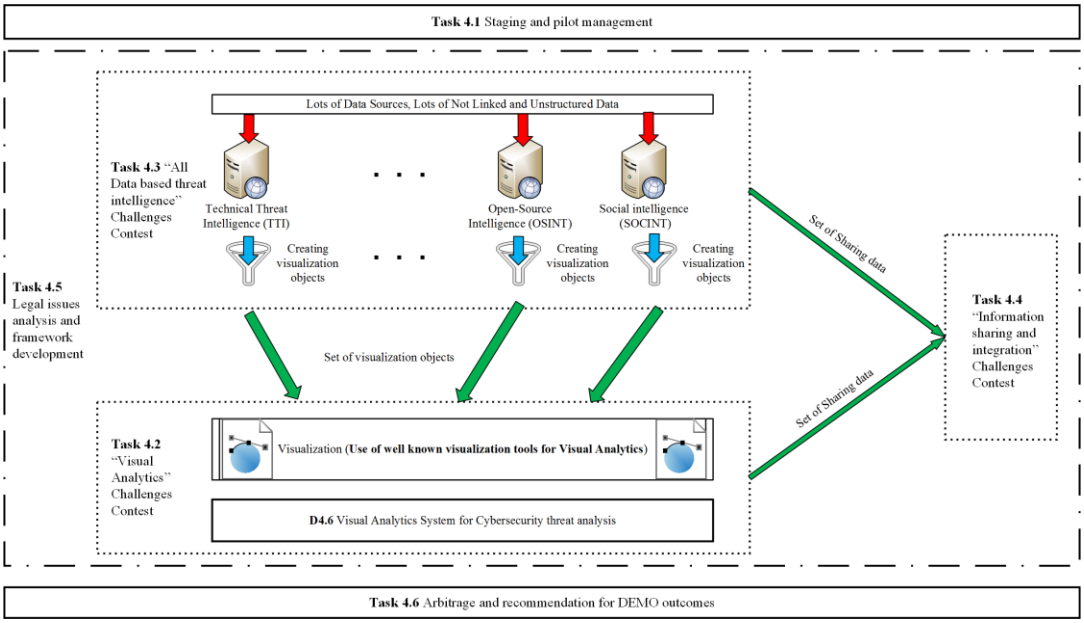

Figure 12: T4.2 task relation with others WP T-shark project tasks

4.4 Visual Analytics System for cybersecurity threat analysis

Methodology for detection of target content manifestations with the MISP and monitoring and analysis of their dynamics. Cyber threat intelligence analysis in the VAS.

KTU acted as task lead for T4.2 ("Visual Analytics" Challenges Contest). In this task was developed approaches to facilitate the gathered data and developed models by means of interactive visual interfaces to create solutions for knowledge generation, situational awareness, and decision-making support. KTU performed research into intrusion detection in computer networks using machine learning methods. KTU presented **LITNET-2020**, a new annotated network benchmark dataset obtained from the real-world academic network.

KTU proposed (1) a novel approach for network intrusion detection using multistage deep learning image recognition and (2) an ensemble classification-based methodology for malware detection.

The definition of the concept follows the well-known principle of starting by developing a *minimum viable product* (MVP).[6] We propose a MVP for the VAS flows concept based on four key concepts:

1. Threat intelligence – Security threat intelligence obtained from large-scale Internet monitors, the dark web, or SIEMS, are a major source of human-domain knowledge. For instance, it allows obtaining information about near-future attack campaigns, malware being sold, stolen personal data, etc.

[6] Moogk, Dobrila Rancic. "Minimum viable product and the importance of experimentation in technology startups." Technology Innovation Management Review 2.3 (2012).

2. Network monitoring – Supported in devices such as intrusion detectors, gateways, firewalls, programmable switches, SIEMs and others, this network function is critical for the security operation of networks.
3. Storage and visualization – They are important components of security management systems, including VASs.
4. Alerting – The generation of alerts is one of the main objectives of security monitoring frameworks, either visual or not.
5. The MVP can be later expanded to include more capabilities over time. Examples of additional functionalities are:
    - User activity monitoring, including Privileged User Monitoring and Audit (PUMA).
    - Log retention, in the sense of long-term archival.
    - Generation of reports.
    - IT compliance reports.
    - File integrity monitoring.
    - Support for log forensics.

4.4.1 Process model

VAS using logs from a log of syslog's server that is configured as a collector for a corporate website. Here are providing visualization in the form of an internal system activities, breaking down the hours of a day, geo-IP addressing for the total number of requests and bytes, and finally geo-IP source and destination. From internal / external logs database server that is used for different application able to visualizing the total number of a given log type, logs by level over time, and a list of recent logs. When we are talking about cloud services or microservices, which provides visibility into user activity. A threat analysis is used to determine which system components need to be protected and the types of security risks (threats) they should be protected (Figure 13). VAS for analytics can show a high-level summary of user activity levels, access requests, the top active users, and any resource groups that have been changed.

The integration of interactive visualizations and the previously mentioned data and models from other tasks may require the adoption of the Visual Analytics and data will be provided by The General Data Protection Regulation (EU) 2016/679 (GDPR) is a regulation in EU law on data protection and privacy for all individual citizens of the European Union (EU) and the European Economic Area (EEA). Every day, small and medium businesses receive an increasing number of warnings. They are constantly inundated with alerts, and the number continues to increase. Because analysts are unsure whether they signal a serious threat, they must be examined one by one to determine which ones are genuine.

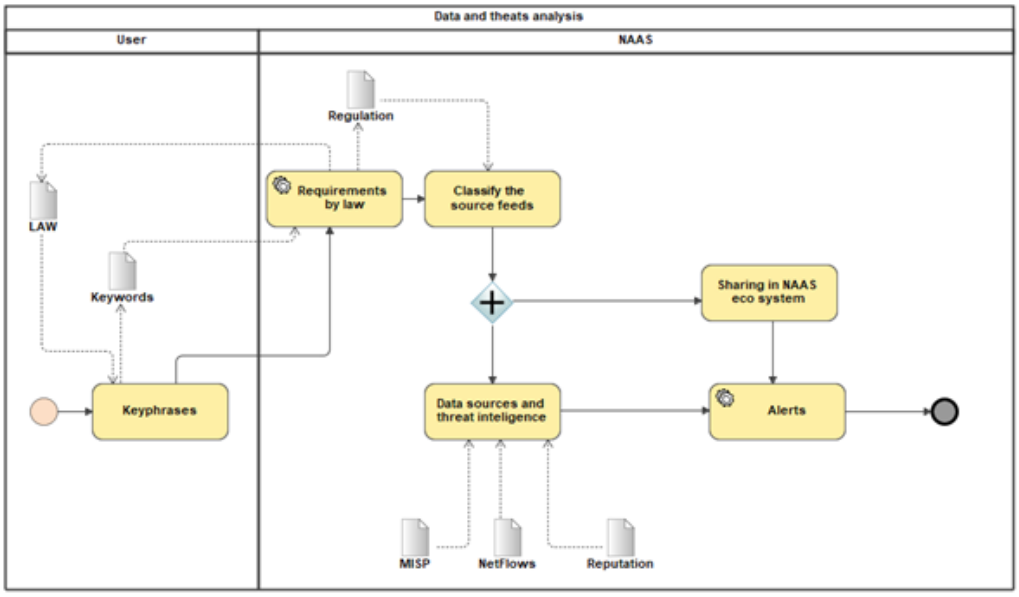

Figure 13: Data and threats analysis process

A data model (Figure 14) is represented by a capitalized key, whereas a non-capitalized key is only an attribute. There just is not enough time to attend to all of them. What if analysts could ignore the signals that are not a concern and concentrate on the ones that are? If they could do that, they might just have enough time in the day to finish them all. Threat Indicators are the answer to this question. Security analysts can estimate how likely an adversary in an alert is a threat or vulnerability based on historical behaviour as determined by Threat Indicators.

The data model:

1. The point is to count the risks in the information space.
2. Each information space (and there can be many, under different names) is necessarily associated with at least one location (city, state).
3. Each location can be associated with at least one (i.e., one or more) information spaces.
4. Each information space includes at least one source.
5. Each source can be included in at least one information space.
6. Each source must belong to at least one IP.
7. Each IP can have at least one source.
8. Each location can be associated with at least one IP.
9. Each IP is necessarily associated with only one location.
10. Each IP over a period of time can be described by a particular reputation rating.
11. Each reputation rating is specific to only one IP.
12. A certain score can be calculated for each IP for a certain period of time.
13. Each rating is calculated for a specific IP.
14. Every piece of malware can be distributed by a specific organization.
15. Each organization can distribute at least one malware.
16. Every malware (malware) can be targeted to a specific organization.
17. Every malware has a client, a server.
18. Every organization can fall victim to at least one malware.
19. Each threat indicator can be associated with at least one malware.
20. Each malware is associated with a single threat indicator.
21. Each flow has one IP source and one IP destination.
22. Each IP can be the source or destination of at least one traffic.
23. Each client has an IP source.
24. Each IP can be classified as: either Public Threat, or at risky Server, or high-risk Client.

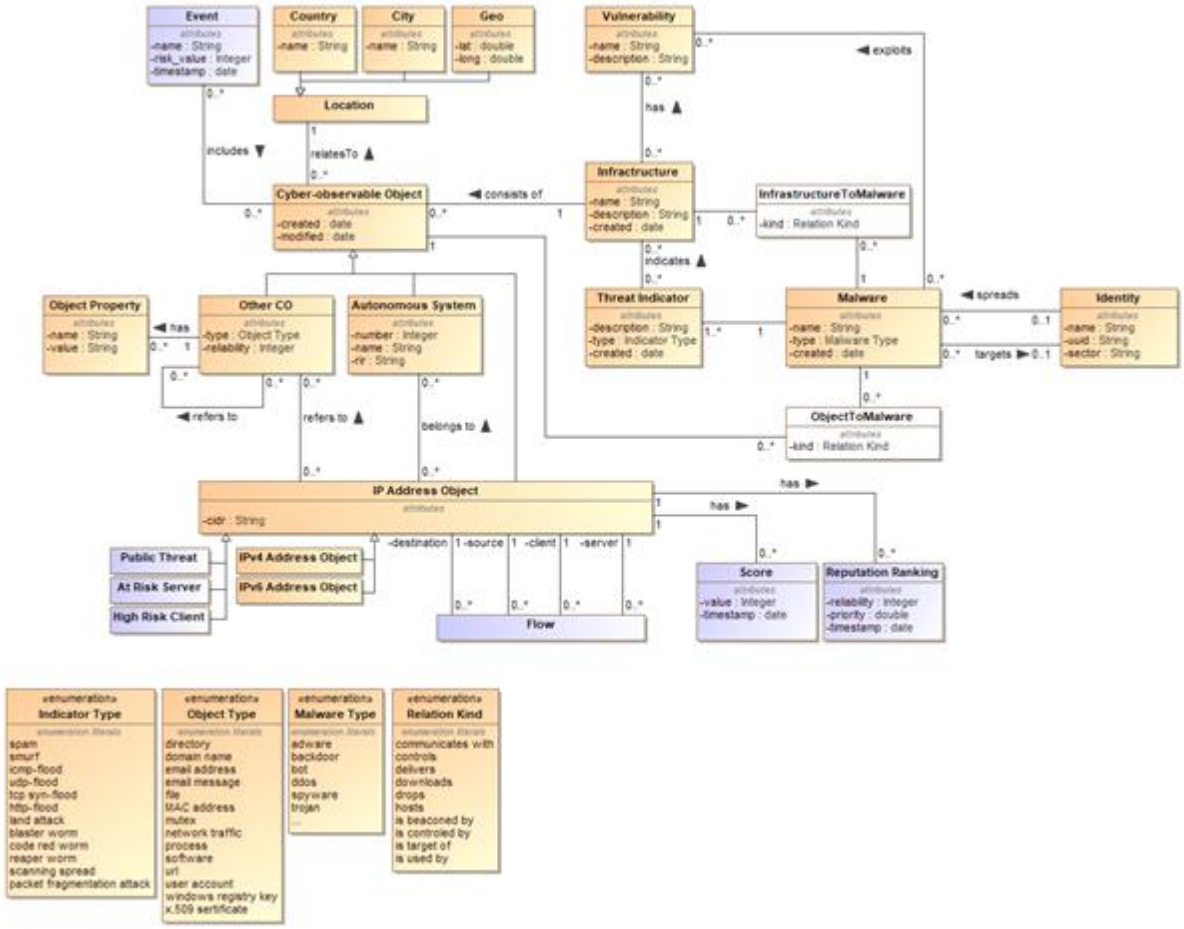

Figure 14: Data model

4.4.2 VAS flows diagram

The concept of VAS flows can be represented as a *data-flow diagram* (DFD). A DFD is a way of graphically representing flows of data through a process or an information system, including its inputs and outputs.

The VAS DFD is represented in Figure 15. The diagram is split into two layers:

1. *Human layer* – processes and components concerning human aspects, like flexibility, creativity, and background knowledge.
2. *Computer layer* – processes and components concerting IT aspects, i.e., capacities for data storage, processing, and visualization.

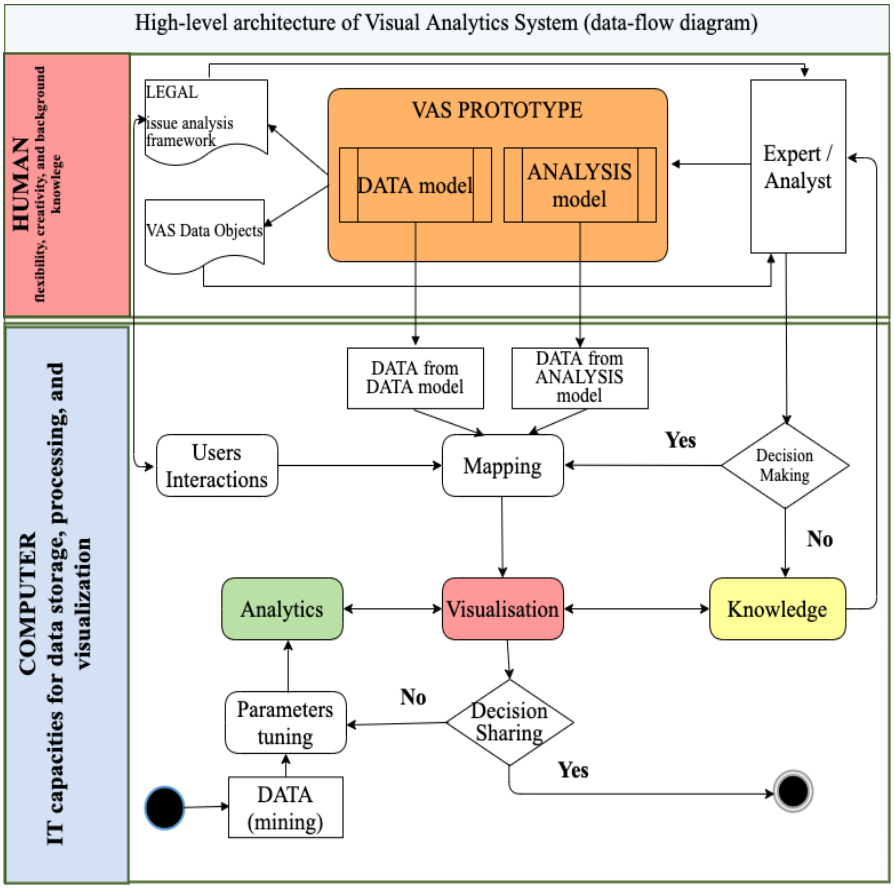

Figure 15: VAS data-flow diagram

4.4.3 Human layer

The human layer has several components, as seen in the Figure 15. The *data model (*Figure 14*)* is the unified data model presented in Deliverable D4.1. This model is based on MISP, a de facto standard for collecting and sharing cyber threat intelligence. More precisely, the model is a unification of data models in MISP. MISP has several components: a data-sharing platform, a protocol for synchronization between instances, and a selection mechanism that allows defining which records to import from other instances. In the context of D4.1, a unification process was defined and applied to the set of use cases related to election interference considered in T-Shark: DDoS backscatter, Twitter, Election interference, BP-IDS, and Malware.

The *analysis model,* or *threat analysis model,* is a representation of the threats to which a certain target is exposed or that exist in a certain scenario like the election interference situation. A threat analysis or threat modelling is a process that allows identifying such threats, in this sense, including both a vulnerability and an attack vector that allows exploiting it. In this report, a threat analysis process was defined and applied to the election interference case.

The data model and the analysis model provide two types of data: information artifacts, designated *VAS data objects,* and *legal* information. The human layer of the VAS data flow includes a feedback loop in which *cybersecurity* or *context experts* (e.g., election officials) consume the data provided by the models, process it, and feed new information back into the model.

4.4.4 Architecture

Visual Analytics System Objects are the data necessitate the use of visual analytics techniques, such as combinations of different private and public feeds.

All flow source event fields have been translated into a model with VASO-related Visual Analytics Incidents Information mapping model, allowing reporting and analysis from a set of common Kibana dashboards. Elasticsearch uses index templates to assign the relevant data types to each field based on the model. A Functional Architecture is an architectural model that identifies the system function and its interactions (Figure 16).

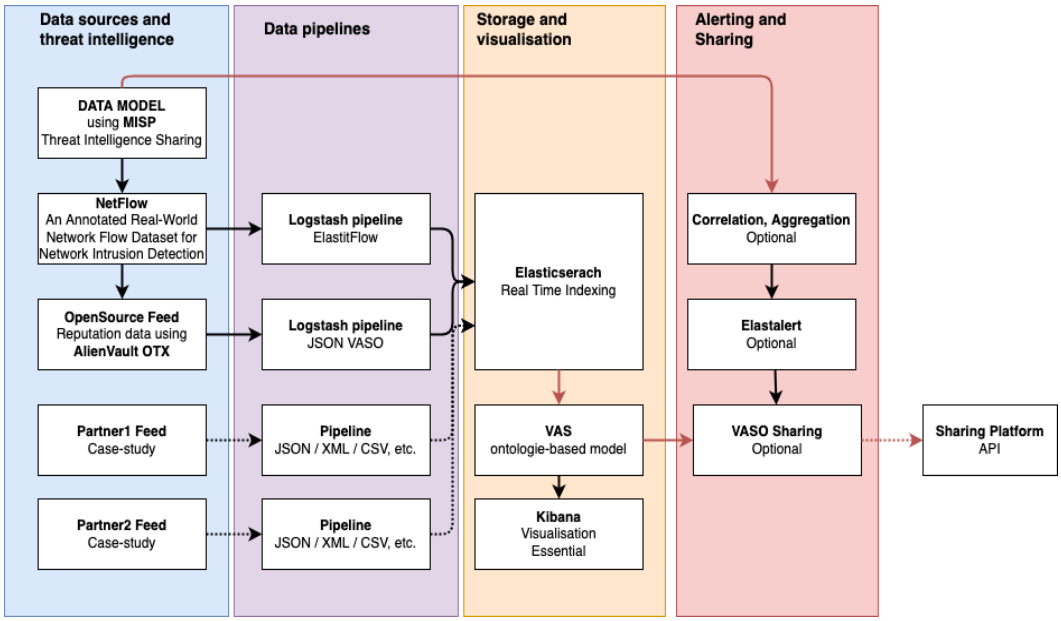

Figure 16: Functional architecture

VASO – Visual Analytics System Objects (Figure 17) is the data necessitating the use of visual analytics techniques, such as combinations of different private and public feeds. Event fields from all flow sources have been mapped to the model with VASO-related types, allowing reporting and analysis from a set of common Kibana dashboards. Index templates enable Elasticsearch to assign the correct data types to each field as dictated by the model (see Annex 2: Layers).

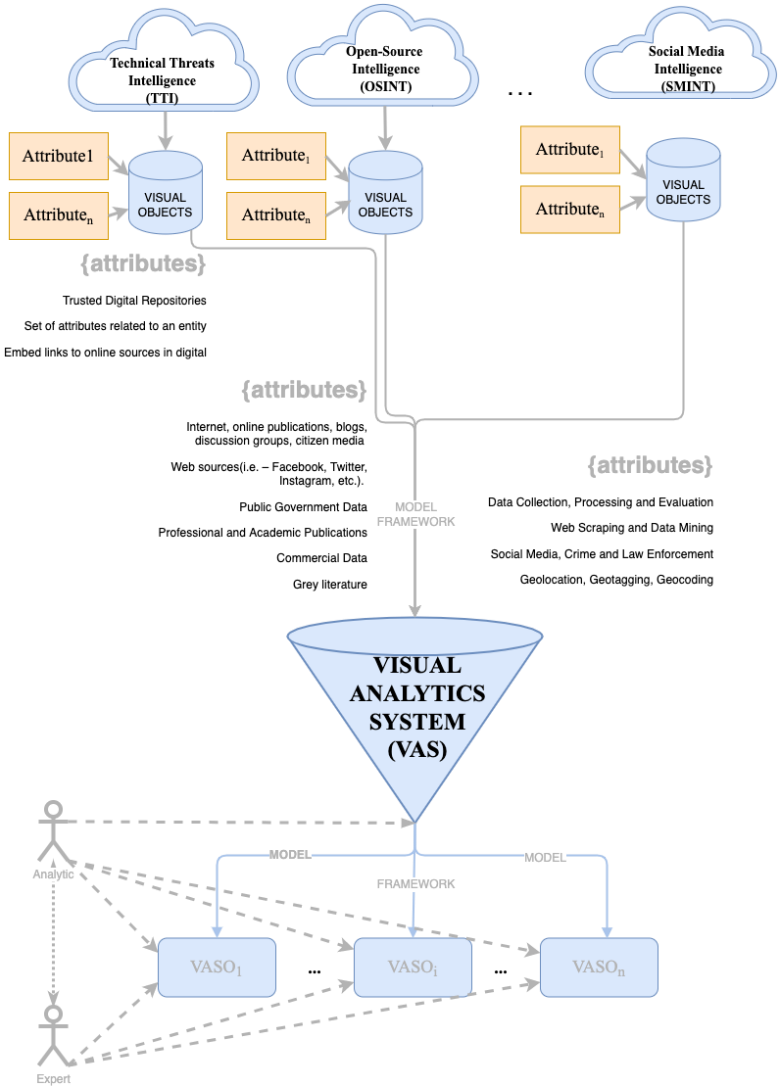

Figure 17: Visual Analytics System Objects

The VAS dashboards built within Kibana have been designed to work together as a single application for navigating through the collected VASO data feeds, allowing experts to drill into specific areas of interest for more detailed analysis.

For application search, site search, corporate search, registration, metrics, application performance monitoring, security analysis, and business analysis, Elastic makes use of the power of search and applications. Elastic Stack (also known as ELK Stack), Elastic Engineer Elasticsearch, Kibana, Beats, and Logstash enable users to obtain data from any source and in any format in milliseconds or less. Technical Architecture (Figure 18) is a form of IT architecture that is used to design computer systems.

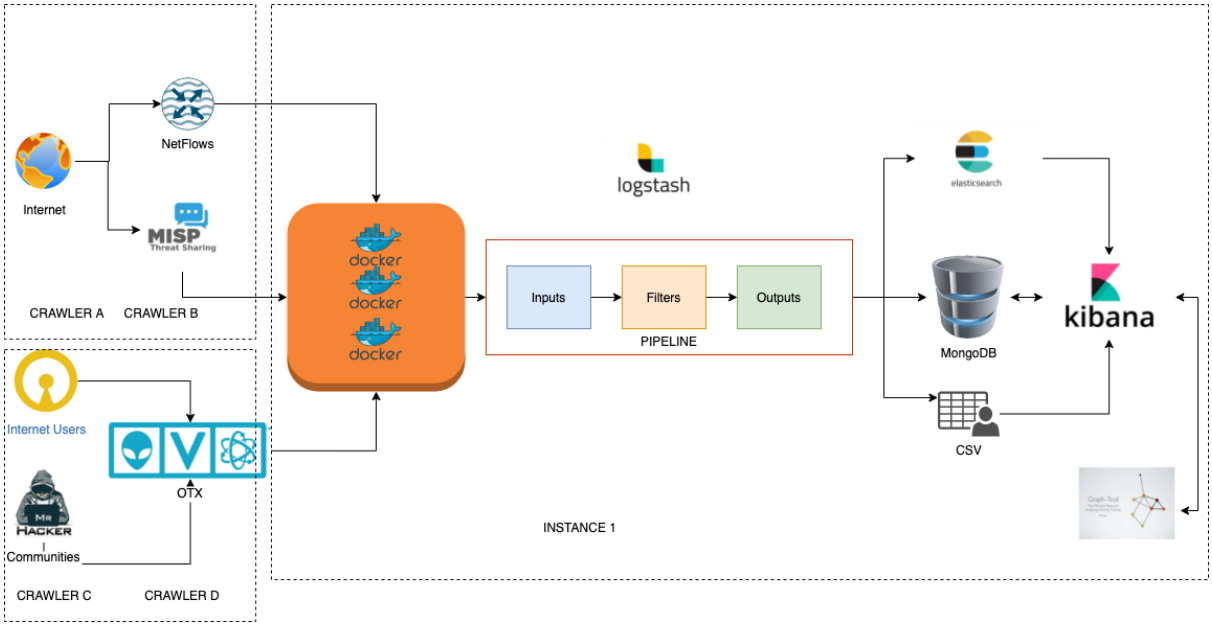

Figure 18: Technical architecture

4.4.5 Technical requirements

Instance 1

- description: CPU
- product: Intel Xeon Processor (Skylake, IBRS)
- vendor: Intel Corp.
- slots (cores): 4
- size: 2GHz
- description: System Memory
- size: 8GiB
- description: SCSI Disk
- product: HARDDISK
- size: 40GiB (42GB)
- capabilities: 5400rpm gpt-1.00 partitioned partitioned:gpt

4.5 Objectives of investigation

Data leaks and computer threats are unavoidable. It seems like a data breach occurs every day, and the victims of the data breach suffer as a result. Their personal information is either stolen or made public online. The breached organization or organizations learn a little from the attack and continue to provide credit monitoring as if nothing happened. The cyber risks analysis at the conclusion will inform you about the types of detection and mitigation areas you need to safeguard.

Threat Sharing is an open-source threat intelligence platform that allows you to create your own events made up of IoT's and then leverage these as a threat data feed. Out of the box also has support for many open-source threat feeds, and it can aggregate these and display them in a chosen standard. This can really help with centralizing threat data.

Alongside the WebUI for MISP, there is a powerful API engine running underneath. Again, we won't focus too much here on singing its praises.

Logstash is the powerhouse behind our enrichment setup. Logstash is split up into 3 sections. Input, filter, and output. The input section is where we define the source of the logging data with which we

want to work. The filter section is where we work with the logging data. This could be via parsing, normalizing, transforming, or multiple other methods to prepare the data for sending out to Elastic Search. The output section is where we define how to then send the data out of logstash, and this could be sent directly to Elastic Search or many other output options. So, we can map all logs up against the Elastic Common Schema via Grok parsing.

Network devices, such as routers and switches, acquire flow statistics by sampling network traffic. Data may be further processed by the collector or simply forwarded to a data repository. If the correct solution is used to gather and analyse flow data, it can reveal a lot of detail about the traffic that traverses a network, and it can yield a lot of insights. Any flow collecting system must be able to normalize data from many divergent flow sources without losing any of the original information in the flow record to extract the most value. This will allow similar reporting and analytics on be performed to all the data without sacrificing the ability to use subcase counter samples and other features.

The IP **reputation database** is refreshed every hour, although the caller can demand a refresh. The first-time call creates a cache directory and file in the user's home directory, then downloads and generates the data frame before writing it out as an R object. Unless refresh == TRUE, further calls will re-read this data frame. If refresh == TRUE, the function will refresh the database.

Data Attributes. Data table with IP & Reputation information (Figure 19).

- IP- IPv4 address
- Risk- how risky is the target (1-10)
- Reliability- how reliable is the rating (1-10)
- Activity- what type of host is it
- Country- what is the IPv4 country of origin
- City- what is the IPv4 city of origin
- Latitude- geolocated latitude of the IPv4
- Longitude- geolocated longitude of the IPv4

The score scale below and at the source Table 8.

Level	Score	Description
Not trusted source	1	It could have a bad design, broken links, and grammatical and spelling mistakes. They may be missing information such as author, date, and/or source. They will not be linked to legitimate institutions, organizations, or entities. They could contain erroneous or unbelievable information.
Partially reliable source	2	There is an option to link to the original (first) source. It appears reliable, so it is especially important to evaluate the credibility of the source very carefully.
Trusted source	3	A trusted source means a source recognized by the VAS ecosystem as a reliable source for verifying a piece of information.

Table 8: Source scores explained

4.5.1 Reputation data

Retrieves Alien Vault's IP reputation database. AlienValut refreshes every hour, but the onus is on the caller to force a refresh. First-time call will setup a cache directory & file in the user's home directory, download & generate the data frame then write the data frame out as an R object. Future calls will just re-read this data frame unless refresh == TRUE should the function refresh the database.

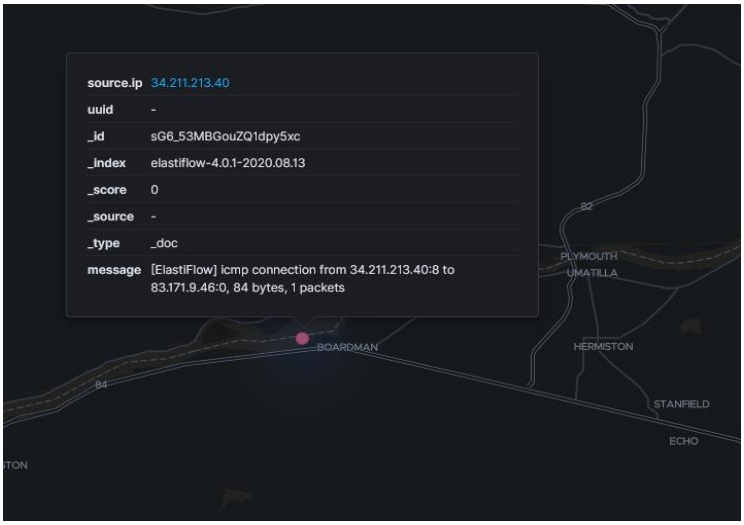

Figure 19: Reputation data leak address

Businesses and organizations can track the public IP and domain reputation of their own assets with the Reputation Monitor Alert service. If one of their IP addresses or domains appears in a hacker forum, a blacklist, or matches one in an IP reputation database, they will be notified. AlienVault additionally keeps an eye on the user's DNS registration and SSL certificates to ensure that no unexpected changes have occurred. Objects are the kind of data that require the application of analytics tools, such as combining private and public feeds. The ontologies-based model with associated kinds has been mapped to the event fields from all flow sources, allowing reporting and analysis via a set of standard Kibana dashboards. Elasticsearch uses index templates to assign the relevant data types to each field based on the model.

4.5.2 Collecting from Open Source

Elastic Stack is a collection of open-source solutions that Elastic created and maintains. While these technologies can be used in a variety of applications ranging from business intelligence to scouting professional sports teams, they are particularly well-suited to the area of Visual Analytics System Infrastructure Management.

The Elastic Stack combines the primary components listed below, which can be used to create a world-class Flow gathering and analytics system.

Elasticsearch – A horizontally scalable, JSON-based search and analytics engine with easy management. Your flow data will be kept in Elasticsearch.

Logstash – A dynamic data collecting pipeline with a robust Elasticsearch synergy and an extensible plugin ecosystem, Logstash will act as the solution's flow collector and provide additional processing power. It is one of the most used methods for gathering data and transferring it to Elasticsearch.

Kibana – It is the expandable user interface for customizing and managing all aspects of the Elastic Stack. It gives shape to your data and is the extensible user interface for configuring and managing all aspects of the Elastic Stack. This is how you look at your information.

4.5.3 Visualization

The fundamental traffic statistics is summarized in the overview dashboard. Before delving into the data, this is an excellent location to set any filters. By default, Kibana has been set up to pin filters so that they remain visible while switching between dashboards. The user will have a more

application-like experience as a result of this. Explore location data using Elastic Maps or get creative with custom layers and vector shapes (Figure 20).

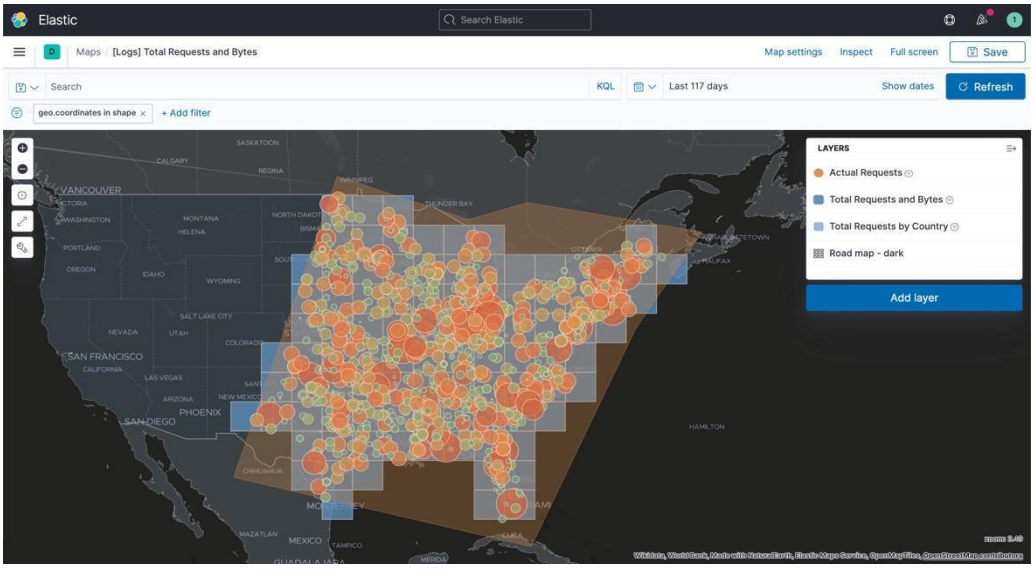

Figure 20: Location analysis

The top navigation pane allows the user to quickly navigate between dashboards. The concept of Traffic Locality is also introduced here. The solution is intended to determine whether a communication is taking place between systems on the target network or whether one or both participants are public addresses. This is beneficial for a variety of use-cases (e.g., KTU sub-case "Network flow-based threat intelligence (Figure 21) approach for visual analytics system"), particularly those involving security concerns. The Geo Location of all public participants is determined in relation to Traffic Locality. This data is summarized here and serves as the foundation for the Geo Analyzer dashboard.

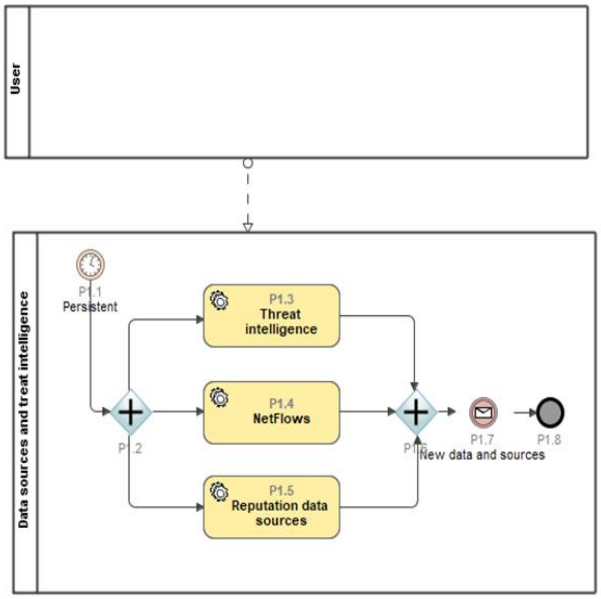

Figure 21: Data intelligence collection

The dashboards incorporated within Kibana will be designed to function together as a unified application for navigating across the data feeds of the collected objects, allowing specialists to drill down into specific areas of interest for further in-depth research.

Intelligence on threats. A model's foundation, centered on a central MISP instance (https://misp-sparta.liberouter.org/) linked to VAS (KTU infrastructure) Reputation data sources. The Open Threat Exchange is a computer-security platform that is based on crowdsourcing (Figure 22). It has over 80,000 members in 140 countries who share over 19 million possible threats daily. Netflows. The Logstash Netflow Module has been deprecated and replaced by the Elastic Common Schema-compliant Filebeat Netflow Module (ECS).

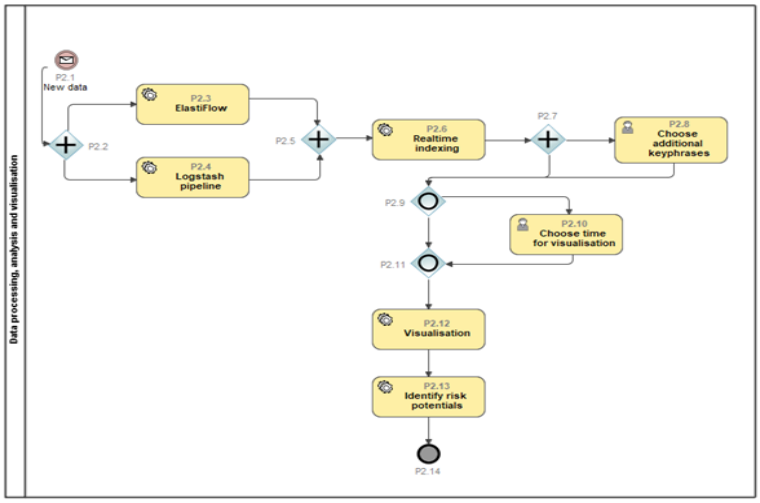

Figure 22: Data pipelines

4.6 Cybersecurity threat integration to VAS

It is designed threat intelligence method for multidimensional visual analytics system to help the expert in making decisions. The method is be based on network flows. The prototype solution demonstrated in the anonymized academic network demo environment and integrated to Visual Analytics System (Figure 23).

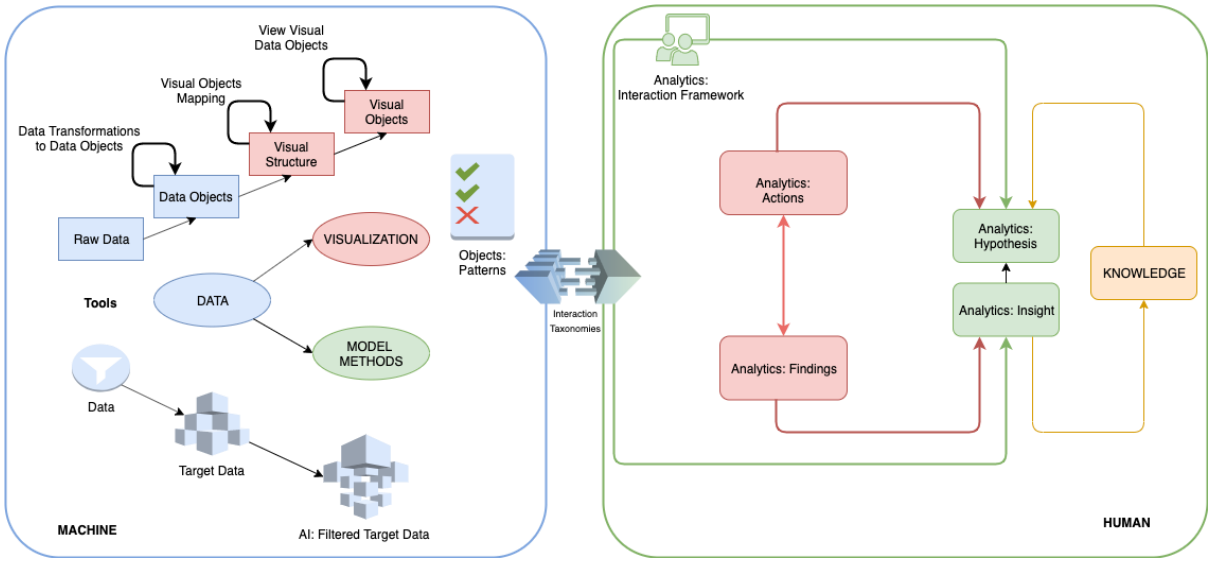

Figure 23: Analytics System - integration between Machine and Human

To uncover threat-related data that can be utilized to anticipate technical risks and produce proactive responses by constantly refining network flow-based methodologies. Intelligent cyber security technical threat identification visual analytics systems can find correlations between data instances and technological trends that would be impossible for an expert to detect due to their ability to handle and analyse large volumes of structured and unstructured data.

The goal of T-SHARK is to develop and validate methodological, organizational, and technological solutions that extend cybersecurity to a more comprehensive organization of security functions, focusing on threat prediction and full-spectrum cybersecurity awareness, providing high situational awareness, informing decision and policy makers on broad or long-term issues, and/or providing timely warning of threats (Figure 24). To enable an effective full-spectrum analysis, Task 4.2 focuses on the development of generic approaches to support the incorporation of expert domain knowledge into the data analysis process.

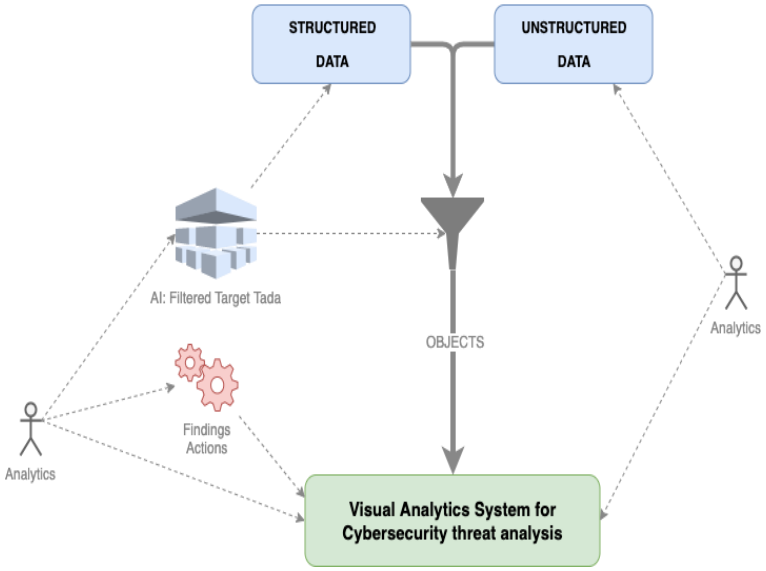

Figure 24: Visual Analytics' data objects

Effective intelligent cybersecurity flow-based threat identification for the visual analytics system. A tool would help the Partners to achieve their goals by proactively identifying, analysing risks, threats,

and vulnerabilities, and responding to risk in a repeating process, looking at all aspects of security both confidentiality, integrity, and availability, and for the technical infrastructure.

4.7 Visual Analytic Systems data flows

This section presents the *VAS flows concept*. Task 4.2 is concerned with the development of generic approaches to support incorporation of *human-domain knowledge* into *data analysis processes*. Specifically, the task is concerned with the use of VASs for that purpose. VAS flows are data flows, as in a data-flow diagram, that represent how this incorporation of human-domain knowledge is performed. Because the information contains sensitive data, access is restricted from the Internet (Figure 25).

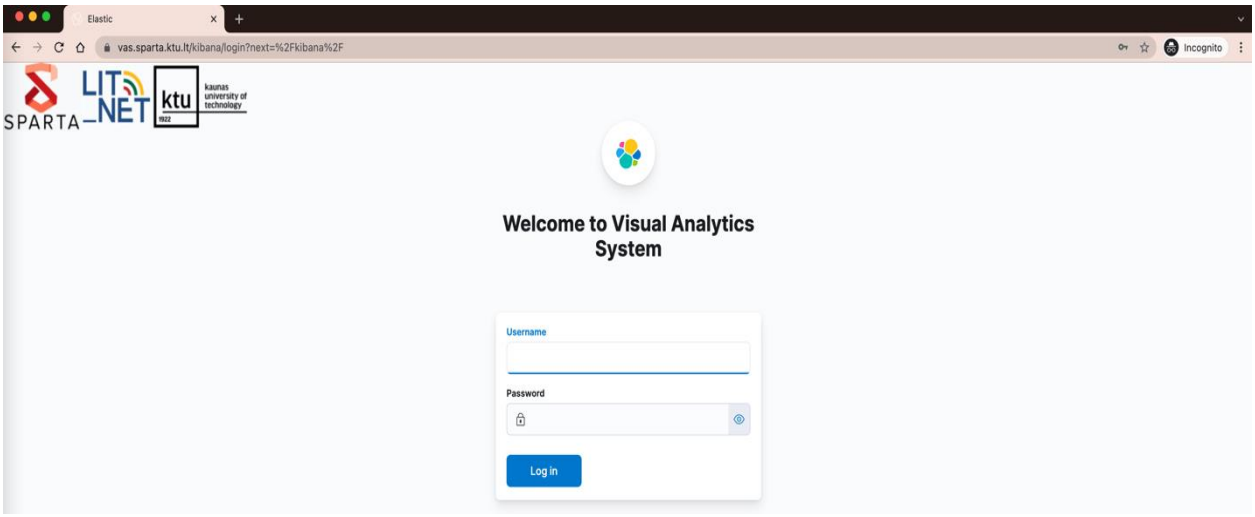

Figure 25: Login page at **https://vas.sparta.ktu.lt/**

Based on the volume of traffic per service in bytes or packets, the Traffic Analyzer helps users quickly discover the most important discussions between clients and servers. It is well within the realm of possibility to consume numerous terabytes of network flow data on a daily basis. Analysts can use the Graph Analyzer to visually explore the connections between targets both inside and outside their network. The circles represent servers, whereas the rectangles represent clients. The diameter of the circle denotes the amount of data in bytes, while the line thickness represents the number of connections (i.e. the number of flows).

The Geo Analyzer (Figure 26) gives you a look at how traffic moves between private networks and the public Internet by using Geo IP functionality (see Annex 3: GEO IP). The use of two maps is to provide a rapid visual indication of the volume of potentially harmful traffic.

Figure 26: Geo location analysis

The maps would be identical if all traffic were valid bi-directional communications, such as someone viewing a webpage or retrieving email. Drilling down and identifying potential malicious actors is simple with this dashboard. You can see a broad range of access attempts from all around the world after filtering for only traffic arriving from the public Internet to one of our publics facing IP addresses. Some of these connections are genuine because we provide some material from this site. The Kibana dashboards were created to function together as a unified application for navigating through the collected data, allowing users to drill down into specific areas of interest for further in-depth research.

4.7.1 IP Reputation Ranking Criteria

IP Reputation identifies IP addresses and domains worldwide that are crawled ed by the VAS ecosystem. IP Reputation (Figure 27) verifies them as either malicious or, at least, suspicious until more data comes in to increase their threat ranking.

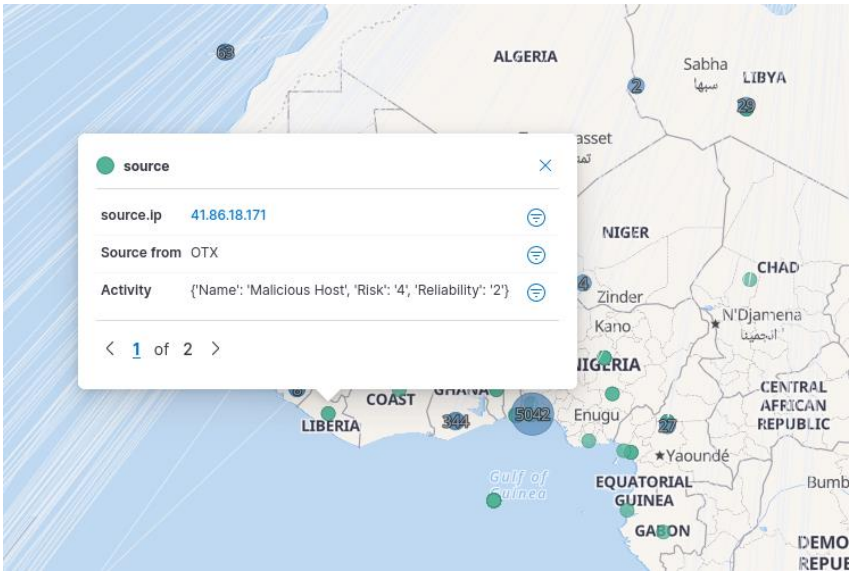

Figure 27: IP Risk and IP Reliability

IP Reputation uses ranking criteria based on IP Reliability and IP Priority that updates continuously to calculate changing assessments to risk level. This helps prevent false positives.

1. IP Reliability. IP Reputation data derives from many data sources of differing reliability. Ranking in this case is based on the relative number of reports regarding a malicious IP in relation to others reported. If, for example, VAS receives 10 reports on a given IP address versus 20 on another, it gives the IP with 10 reports a lower reliability ranking than the IP with 20 reports.
2. IP Priority. VAS ranks IP address priority, based on the behaviour associated with each IP address listed. For example, an IP address used as a scanning host receives a lower priority than an IP address known to have been used as a Botnet server.
3. Ongoing Ranking Reassessment. VAS constantly updates its IP Reputation data as new information emerges affecting IP reliability or priority criteria. Each update reprioritizes IP reliability and priority values and the threat level of an IP accordingly.
4. CIDR. Classless inter-domain routing. Specifies a range of IP addresses on a network that is suspected of malicious activity or attack.
5. Vulnerability. Standards group identification of Common Vulnerabilities and Exposures (CVEs).
6. Domain. A domain name for a website or server suspected of hosting or engaging in malicious activity. Domains may also encompass a series of hostnames.
7. Email. An email address associated with email address and email message.
8. File. (MD5, SHA1, SHA256, PEHASH, IMPHASH). A hash computation for a file that can be used to determine whether contents of a file may have been altered or corrupted. A SHA256-format hash that summarizes the architecture and content of a file deemed suspicious.
9. Filepath. Unique location in a file system of a resource suspected of malicious activity.
10. Hostname. The hostname for a server located within a domain, suspected of malicious activity.
11. IPv4 (and in the future IPv6). An IP address used as the source/destination for an online server or other device suspected of malicious activity.
12. Mutex. (Mutually exclusive access) or mutual exclusion object allowing multiple program threads to share the same resource. Mutexes are often used by malware as a mechanism to detect whether a system has already been infected.
13. URI. A uniform resource identifier (URI) that describes the explicit path to a file hosted online, which is suspected of malicious activity. It can be liked with URITypeEnum.
14. URL. Uniform resource locations (URLs) that summarizes the online location of a file or resource associated with suspected malicious activity.
15. IP Reputation receives data from a variety source.
16. Hacker forums - public and non-public.
17. Open-source intelligence - Public and private security research organizations.

4.7.2 Malware information exchange

On the Internet, as well as in the security and intelligence communities, sharing threat information has become a must. Compromise indicators, malicious file indicators, financial fraud indicators, and even extensive information on a threat actor are all examples of threat information. This nomenclature can be used to describe the VAS format in a different data structure. A simple meta structure scheme in which characteristics and metadata are incorporated to build a cohesive set of indicators is called an event (Figure 28). An incident, a security analysis report, or a specific threat actor analysis can all be used to create an event. The significance of an event is determined solely by the information contained in the security event figure.

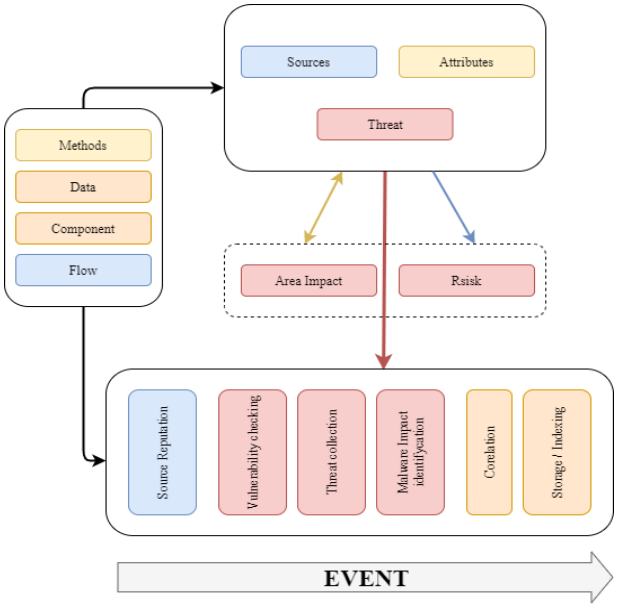

Figure 28: Security event

Find below a non-exhaustive list of parameters that can be used to filter data in VAS search (some parameters specific to given export formats are not mentioned):

- uuid: Restrict the results by uuid. includeEventUuid: Instead of just including the event ID, also include the event UUID in each of the attributes.
- timestamp: Use the timestamp to limit the results (last edit). Any event with a newer timestamp than the one specified will be returned. If you're using /attributes as a scope, the lookup will be based on the attribute's timestamp. A timestamp or a short-hand time description can be used as the input (7d or 24h, for example). You can also pass a list with two values to set a time range (for example ["14d", "7d"]).

In the value field of the attributes, look for the specified value. Depending on the scope, limit the number of results returned (for example 10 attributes or 10 full events). Sets the page to be returned if a limit is set. Limit 100 on page 3 will yield records 201–300).

If this option is enabled, the attachments / zipped malware samples are encoded as base64 in the data field of each attribute, and just the metadata (events, tags, relations) is returned, leaving attributes and proposals out. We proposed the Indicator Types to aid in the enhancement of existing indicating methods as well as the development of new ones [80]. They may be used to distinguish 12 different types of network attacks. The proposed data model was analysed and compared to classical and modern important features, and the use case described its benefits and drawbacks.

Set whether events should be returned as published or unpublished. If you want both, don't set the parameter. Remove any properties from the result that might result in a warning list entry being triggered.

4.7.3 Connection with a D4.1 Cybersecurity threat intelligence common data model

For communication between a source and a destination, network flow data is aggregated packet header data (but no content capture). The protocol-level information in the header and the proximity in time identify communications (i.e., a flow contains aggregated header information for all packets that use the same protocol settings within a designated time window) in Figure 29.

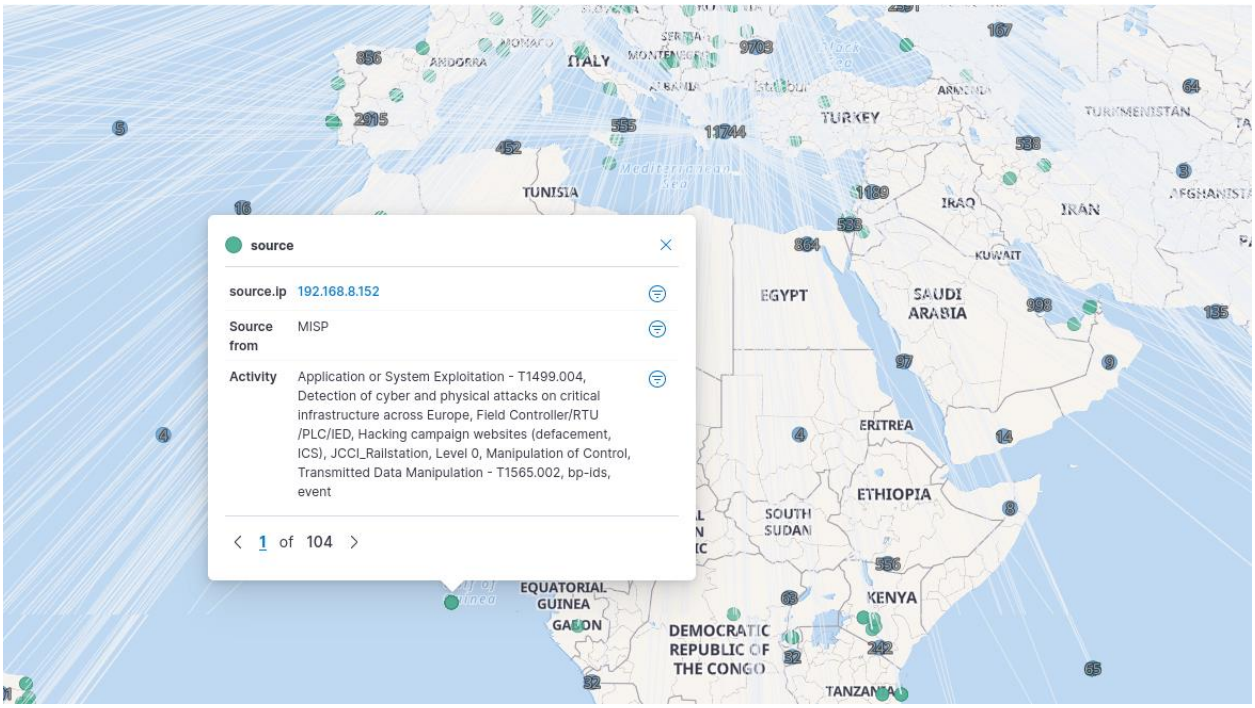

Figure 29: **D4.1** Cybersecurity threat intelligence common data model integration

1. **Flow.** The legacy Logstash-based solution in this repository has been deprecated since it became generally available. It will be kept here for historical purposes, but only minor modifications will be made. It is recommended that you test out the new collector.
2. **Threats.** This includes a dictionary of public IP addresses with a bad reputation. This dictionary is made up of data from various OSINT sources that have been normalized to a common taxonomy. The Threats dashboard uses this IP reputation information to indicate three threat/risk classes.
3. **Public Threats** - Public clients with a poor IP reputation that are reaching private addresses.
4. **At-Risk Servers** - Private Servers that are being reached by clients with a poor IP reputation.
5. **High-Risk Clients -** Private clients that are accessing public servers which have a poor reputation.
6. **Flows.** Client/Server, Source/Destination, and Autonomous System viewpoints all have their Sankey dashboards. The new visualization plugin is used to create the visualizations.
7. **Geo.** There are separate Geo Location dashboards for Client/Server and Source/Destination perspectives (see Annex 3: GEO IP).

4.7.4 Use of Neural Networks and Machine Learning Models

The study provides a malware detection mechanism based on ensemble classification. A stacked ensemble of dense (fully connected) and convolutional neural networks (CNN) performs the first-stage classification, while a meta-learner performs the final-stage classification. Was investigated and compared 14 classifiers for a meta-learner. K-Nearest Neighbors, Linear Support Vector Machine (SVM), Radial basis function (RBF) SVM, Random Forest, AdaBoost, Decision Tree, ExtraTrees, Linear Discriminant Analysis, Logistic, Neural Net, Passive Classifier, Ridge Classifier, and Stochastic Gradient Descent classifier are among the 13 machine learning methods used for the baseline comparison. Experiments on the Classification of Malware with PE Headers (ClaMP)

dataset yielded the following results. An ensemble of five dense and CNN neural networks, as well as the ExtraTrees classifier as ameta-learner, achieve the greatest results (Figure 30).

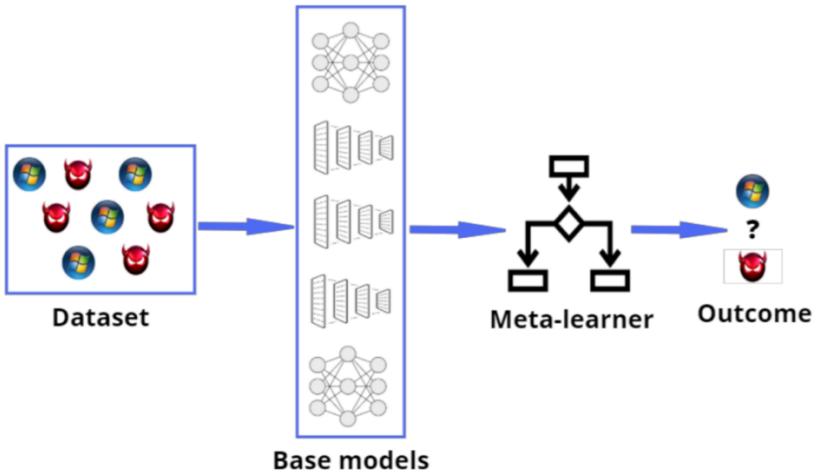

Figure 30: Ensemble classification approach diagrams

The results reveal that the ensemble stacking outperforms other machine learning methods, including neural networks, in terms of malware identification. This result was achieved by using data sets that were provided to all SPARTA T-SHARK partners to test and obtain the same results. KTU demonstrated that an ensemble learning framework based on lightweight deep models may be used to successfully solve the malware detection challenge. Ensemble learning approaches can be integrated and used as intelligent tools for malware detection, according to the findings. The classification system that used the Extra Trees algorithm as a meta-learner and an ensemble of dense ANN and 1-D CNN models to perform the classification operation had the highest accuracy value, beating other machine learning classification approaches. The presented approach can lead to very accurate malware detection models that are tailored to real-world Windows PE malware.

4.7.5 Information effect and extraction

Organizations have digitalized many aspects of their operations because of technological advancements. Because there is a lack of adequate measurements, methods, and frameworks to analyse and assess the harm companies experience from cyberattacks (Figure 31the cyberattack danger environment is continually evolving, and the potential impact of such attacks is unpredictable.

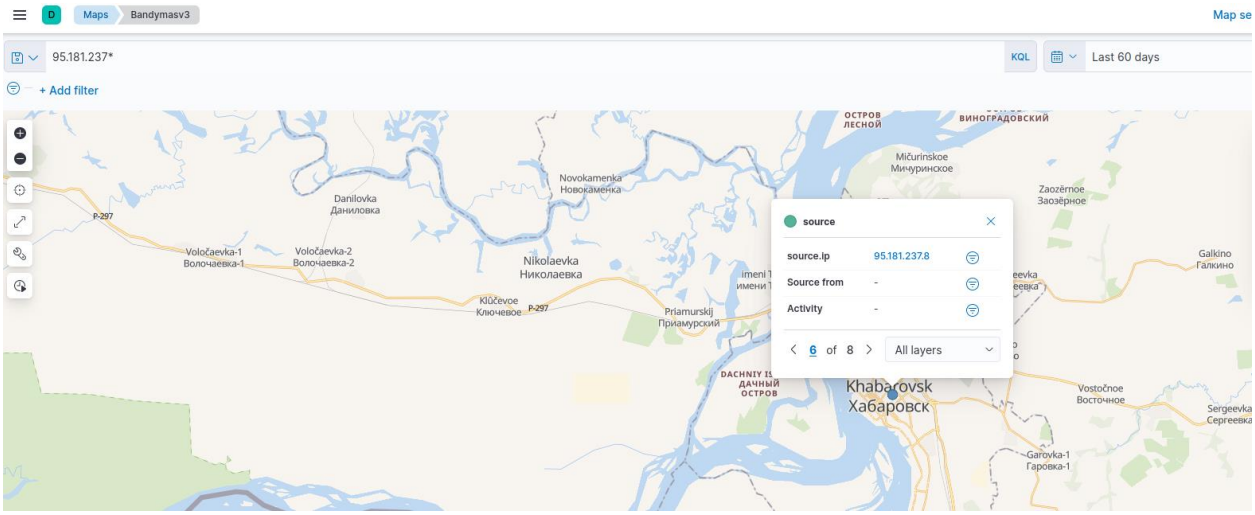

Figure 31: An attacker detection

This VAS use case detects several sorts of harm based on a comprehensive literature review and review of news stories and databases covering cyber-incidents, cybercrimes, hacks, and other attacks. This method includes practice areas that are outlined in the risk section. Several cyber-harms that can come from cyber-attacks are presented in each of these areas. This essay also analyses and derives insights from four real-world case studies covering Sony, JPMorgan, Cyber-partisans, and Ashley Madison to provide preliminary hints regarding how these distinct sorts of harm are connected and how cyber-harm, in general, may spread.

One of the great things about Elasticsearch is its extensive REST API (see Annex 1: API's Query Description) which allows to integrate, manage, and query the indexed data in countless different ways. Examples of using this API to integrate with Elasticsearch C3ISP Collaborative and Confidential Information Sharing and Analysis for Cyber Protection'' system and use cases. This category of APIs is used for handling documents in Elasticsearch. Using these APIs, for example, you will create documents in an index, update them, move them to another index, or remove them. The APIs detailed below are for handling single documents, but you can also make use of certain multi-document APIs for performing bulk actions (e.g. *multi get*).

We emphasize the need for analytical tools for organizational cyber-harm, which can be based on a VAS ecosystem such as the one we propose here. These would allow organizations to identify assets, link these to different types of cyber-harm, measure those harms and, finally, consider the security controls needed for the treatment of harm.

Search API. These API calls can query indexed data for specific information, as the name implies. Search APIs can be used globally, across all accessible indices and kinds, or inside a single index. The responses will include matches to the query. The Mustache language, which is implemented as a scripting language within Elasticsearch, is sometimes required by the Search API (see Annex 1: API's Query Description).

Indices API. Elasticsearch APIs of this type allows you to manage indices, mappings, and templates. This API can be used to create or delete new indexes, verify if a given index exists, and specify the new mapping for an index, for example (see Annex 1: API's Query Description).

catAPI. Use the cat API as much as feasible. The goal is to provide data in a more user-friendly manner than the standard JSON response. Cat aliases, cat allocation, cat anomaly detectors, cat count, cat data frame analytics, cat datafeeds, cat fielddata, cat master, cat nodeattrs, cat pending tasks, cat plugins, cat recovery, cat repositories, cat shards, cat segments, cat snapshots, cat task management, cat templates, cat thread pool, cat trained model, cat transforms are some of the other cat API options (see Annex 1: API's Query Description).

Ingest APIs. Elasticsearch provides a flexible RESTful API for connecting to client applications. REST calls are used to ingest data, perform search and data analytics, and manage the cluster and its indexes. Under the hood, all of the mentioned methods use this API to feed data into Elasticsearch (see Annex 1: API's Query Description).

4.7.6 Assessment of risk potentials

Most often, the risk is an already occurring fact. On the information field, it is measured with categorized threats from zero (only information), low, medium, high to critical risks. The three factors that feed into a risk are:

1. What is the threat?
2. How vulnerable is the system?
3. What is the reputational or financial damage if breached or made unavailable?

Using this simple framework, a high-level calculation of threats risk can be developed:

Risk = Threat x Vulnerability x Information Value

"The goal of Special Publication 800-30 is to provide guidelines for performing risk assessments of government information systems and organizations, expanding the guidance provided in NIST SP (Special Publication) 800-39," according to the National Institute of Standards and Technology (NIST). Risk management guidelines give a framework for risk management principles and a procedure. Regardless of size, activity, or sector, any organization can use it. ISO 31000 can assist organizations in increasing the possibility of meeting objectives, improving the identification of opportunities and threats, and effectively allocating and using risk treatment resources. Let's look at risk as described by standards. We can see that they were all utilized to detect, quantify, and prioritize risk to organizational operations, organizational assets, persons, other organizations, and the nation due to information system operation and use.

There are numerous methodologies and technologies to perform a risk assessment. Usually, that is where the methodology ends. We are missing one component, which is an impact. Our approach is to assemble the results of a Threat, Sources, and an Impact to determine a numeric value of Risk for each Practice areas (e.g., Information technology, Economic, Business) and threat pair.

$$Risk = T_v * V_v * I_v \ .$$

were

T_v - Threat Value (Table 9); V_v - Vulnerability Value (Table 10); I_v - Impact Value (Table 11).

We can summarize the entire process of Risk as:

1. Find the assets and groups that need to be protected.
2. Count a threat to identify and define the threats that could cause harm to the facility and its individuals.

Find assets and threats.

1. Conduct a vulnerability to find weaknesses that an aggressor might exploit.
2. Compute the risk using the results of the value, threat, and vulnerability.

4.7.7 Vulnerability Value

Vulnerability Criteria		
Value		
Critical	16	The likelihood of a threat being used against the IT system is imminent. Internal decision-makers and/or external law enforcement determine the threat is credible.
High	8	The likelihood of a threat being used against the IT system is imminent. Internal decision-makers and/or external law enforcement determine the threat is credible.
Medium	4	The likelihood of a threat being used against the IT system is imminent. Internal decision-makers and/or external law enforcement determine the threat is known but is not verified.
Low	2	The likelihood of a threat being used against the IT system is imminent. Internal decision-makers and/or external law enforcement determine the threat exist but is not likely.
Informative	1	The likelihood of a threat being used against the IT system is imminent. Internal decision-makers and/or external law enforcement determine the threat is non-sexist or unlikely.

Table 9: Vulnerability value

4.7.8 Threat Value

Threat Criteria			
Value	**Resources / Sources**	**Potential target (individuals or groups of individuals)**	**Effects (size)**
1	Readily available	Local incident. Occurred less than one year.	More than 50
2	Easy produce	Regional / Local incident. Occurred less than one year.	More than 100
4	Difficult produce	Regional / Local incident. Occurred less than one year. Affects a group of individuals.	More than 500
8	Very difficult produce	Regional / Local incident. Occurred less than two years. Affects a group of organizations and individuals directly.	More than 1000
16	Extremely difficult produce	Regional / Local incident. Occurred less than five years. Affects a group of organizations and individuals, and government.	More than 10000

Table 10: Threat value

4.7.9 Impact Value

Impact Criteria		
Value		
Critical	16	Loss or damage of assets has exceptionally grave consequences, such as extensive loss of primary services, core processes, and functions; property damage; and a catastrophic impact on the practice areas of the nation.
High	8	Loss or damage of assets has grave consequences, such as extensive loss of primary services, core processes, and functions; property damage; and a catastrophic impact on the practice areas of the nation.
Medium	4	Loss or damage of assets have moderate to serious consequences, such as impairment of core functions and processes; and functions; property damage; and a moderate impact on the practice areas of the nation.
Low	2	Loss or damage of assets have minor consequences or impact, such as impairment of core functions and processes; and functions; property damage; and a moderate impact on the practice areas of the nation.
Informative	1	Loss or damage of assets have negligible consequences or impact.

Table 11: Impact value

4.7.10 Risk Rating

Value	Level	Range	
		From	To
16	Critical	513	4096
8	High	65	512
4	Medium	9	64
2	Low	3	8
1	Informative	1	2

Table 12: Risk rating

The calculated risks and reliability can be seen immediately to compromise the target (Figure 32).

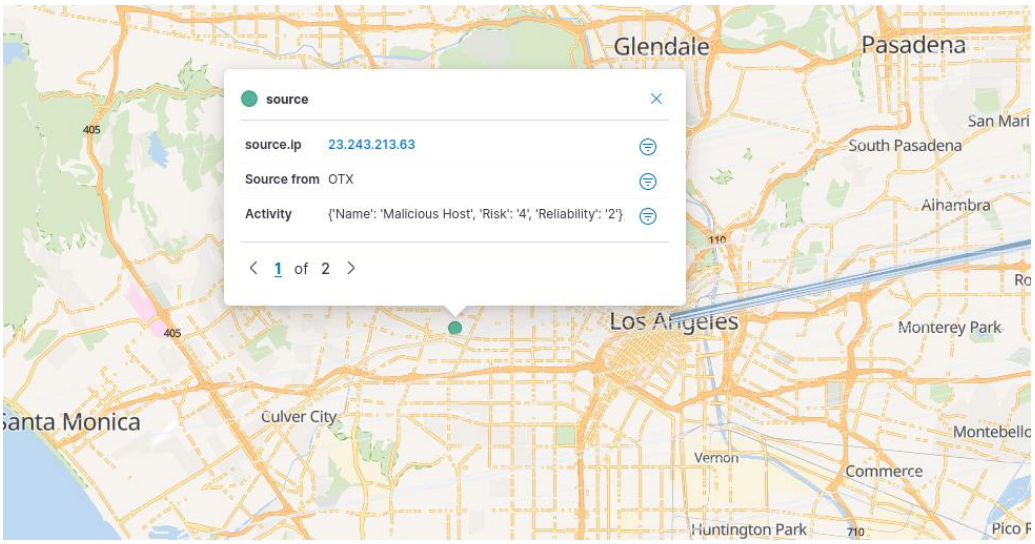

Figure 32: Risk Rating

The facility's Risk Rating against a specific threat based on the case will be compared to a quantitative range to determine the facility's risk against the threat. Based on this risk, cybersecurity designers, architects, and construction engineers can create mitigation methods to counteract the threat. Because it is impossible to eliminate risk and because every sector of information has resource constraints. Cybersecurity designers must have a thorough understanding of how risk mitigation techniques affect practice areas.

Chapter 5 Sub-Case Scenario for VAS

The umbrella case of the T-SHARK program is related to the electoral interference scenario, as mentioned (Figure 33). The case is presented in the form of a scenario, in which many attacks are discussed, each aimed at a different part of the election process.

Individual subcase output can compile information to generate hypotheses and select relevant bits of information from the massive general flow of information about cyber incidents, and the umbrella case serves as a platform to place individual subcases into the general content and identify links, interdependencies, and how individual subcase output can compile information to generate hypotheses and select relevant bits of information from the massive general flow of information about cyber incidents.

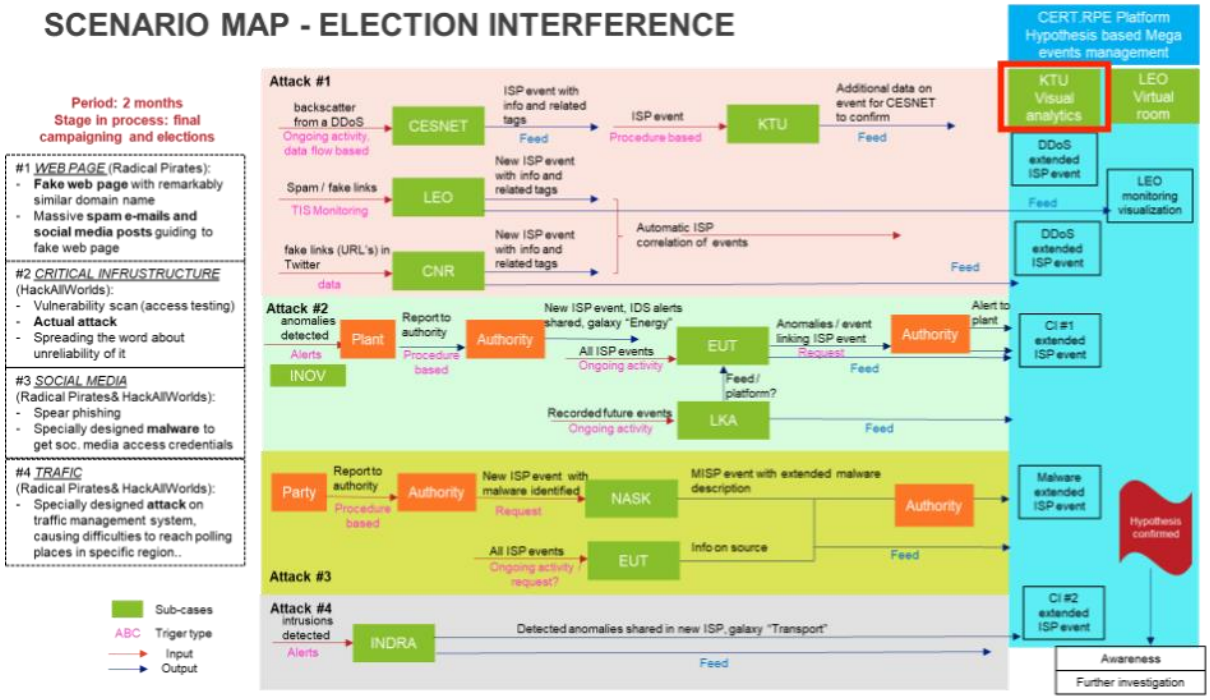

Figure 33: Visual Analytics place in Election Interference

5.1 Connection with a Sub-Case

For network intrusion detection, we present LITNET-2020, a new annotated network benchmark dataset obtained from the real-world academic network (Figure 34).

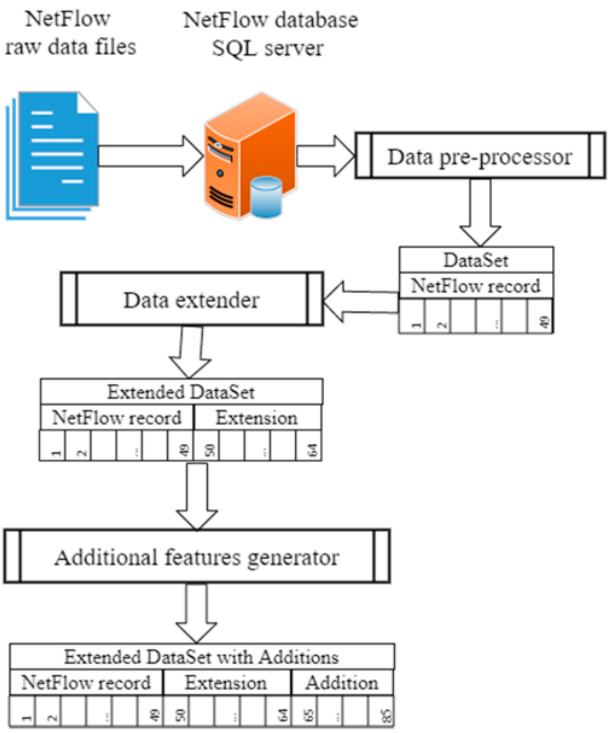

Figure 34: LITNET-2020: creation of VAS NetFlow extended dataset with additional features

Real-world examples of normal and under-attack network traffic are included in the dataset. The dataset's **85 network flow features** and **12 attack** types are described and analysed. Statistical analysis and clustering approaches are used to analyse the dataset's features.

The proposed feature set may efficiently identify different assault classes in the dataset, according to our findings.

Data pipelines. Logstash + Filebeat + JSON logs (e.g. Zeek) open server-side data processing pipeline that ingests data from a multitude of sources, transforms it, and then sends it to VAS stashes. Basic pipeline provided for VAS and send to dedicated pipeline for netflow/sflow (wok with Logstash + Elastiflow). A lightweight shipper for forwarding and centralizing log data. Installed as an agent on your servers, Filebeat monitors the log files or locations that you specify, collects log events, and forwards them either to Elasticsearch or Logstash for indexing.

Storage and Visualizations. Share deployment configs within a Elasticsearch group. For visualization it will share new data with dashboard processes (free and open user interface that lets you visualize your Elasticsearch data and navigate the Elastic Stack). Built on a foundation of free and open, Elasticsearch and Kibana pave the way for diverse use cases that start with logging and span as far as your imagination takes you. Elastic features like machine learning, security, and reporting compound that value.

Correlation scripts. A generalized version of scripts with the Risk rule-based alerts. Based on the traffic volume per service in bytes or packets, the Traffic Analyzer helps users to quickly discover the most important discussions between clients and servers. It is well within the realm of possibility to consume numerous terabytes of network flow data on a daily basis. Analysts can use the Graph Analyzer using Geo IP (see Annex 3: GEO IP) to visually explore the connections between targets both inside and outside their network. The circles represent servers, whereas the rectangles represent clients. The diameter of the circle denotes the amount of data in bytes, while the line thickness represents the number of connections (i.e., the number of flows).

The Geo Analyzer gives you a look at how traffic moves between private networks and the public Internet. The use of two maps is to provide a rapid visual indication of the volume of potentially

harmful traffic. The maps would be identical if all traffic were valid bi-directional communications, such as someone viewing a webpage or retrieving email. This dashboard allows you to quickly dive down and detect potential problematic actors. You can see a broad range of access attempts from all over the world after filtering for only traffic arriving from the public Internet to one of our public-facing IP addresses. Some of these connections are genuine because we provide some material from this site.

Global connectivity and the use of services to hold sensitive data and personal information are expanding the impact of cybersecurity. The chance of your firm being the victim of a successful cyber-attack or data breach is on the rise, thanks to poor service configuration and increasingly sophisticated cyber-criminals. Our cybersecurity strategy includes multiple levels of defence that include computers, networks, reputation, and alarms. As a result, in order to develop an effective defence against cyberattacks, an organization's people, procedures, and technology must all complement one another.

To sum up, listed above are some of the cyber-attacks you can face as a business owner or technological device user. The data, accounts, passwords, and sensitive information that can be lost, deleted, or made public by cyber-attacks is alarming.

5.2 Intervention measures and their impact in the VAS

Intervention strategies are based on the huge increase in cyberattack incidences, which frequently have devastating and grave repercussions. Malware is the most common weapon used to carry out malevolent intents in cyberspace, whether using existing vulnerabilities or the unique properties of developing technology. We begin by providing an overview of the most exploited vulnerabilities in existing hardware, software, and network layers to aid in the achievement of this intervention measure's purpose. Following that, critiques of the existing state-of-the-art mitigation approaches are offered, along with explanations of why they work or don't work.

Vulnerability and exposure are both fluid concepts. Because most of these new technologies provide services through the Internet, several typical attacks are increasingly relying on malware concealed inside extensions or vulnerabilities in scripting languages to gain access to sensitive information. The use of social engineering in scams is on the rise. Popular social networks such as Facebook, Twitter, and others are increasingly being used to trick unsuspecting users into installing or spreading malware. The usage of botnets has been used in more orchestrated attacks. Botnets are a significant source of concern, as their impact is much greater than individual attacks. According to recent statistics, the number of cyberattacks tailored to a specific system, such as a command-and-control system, employing insider information and employees is increasing. When disaster risk is realized, trends in susceptibility and exposure are important drivers of changes in disaster risk and its consequences. Individuals and groups are exposed and vulnerable in many ways, which can affect prosperity and even health. The best defence is to be proactive. We offer guidance on how to harness the VAS ecosystem to defend against new attack patterns in developing technologies such as social media, cloud computing, smartphone technology, and critical infrastructure.

Chapter 6 SPARTA Programmes Interconnection

The goal of the T-SHARK visual analytics systems was to improve understanding of pertinent amalgamated cyber-physical-cognitive analytical methodologies, techniques, tooling, and composite use to give integrated intelligence on how to detect and thwart new sorts of attacks. The JCCI (computation and communication services needed for a shared and distributed test bed), "MISP Threat Sharing" systems, and the "C3ISP Collaborative and Confidential Information Sharing and Analysis for Cyber Protection" system developed within the scope of T-SHARK serve as a testing environment to verify the relevance and applicability of the VAS under real-world conditions.

WP2 Responsible Innovation: Ethical, Legal and Societal Aspects

T-SHARK incident data sharing and collaboration systems (MISP, C3ISP) that allow SPARTA users to upload data and manage analysis execution necessitate selecting the appropriate privacy level and setting a sufficient privacy policy. We were able to examine the risks associated with personal data protection thanks to close collaboration with WP2 knowledge experts. Furthermore, we were able to identify and implement the appropriate technical and organizational procedures, allowing us to adhere to the essential principles related with the processing of personal data. T-SHARK helped organize a Moot Court competition in 2020, allowing researchers to get practical experience by imitating criminal trial proceedings of unlawful activity. This entailed attributing recognized risks to the actors behind the scenes, as well as considering the factual and legal aspects.

WP7 Program #4: SAFAIR – Secure and Reliable AI Systems for Citizens

The T-SHARK visual analytics system NetFlow sub-case includes Machine Learning (ML) techniques. In the last decade, machine learning has gotten a lot of attention and has become an important aspect of the threat analysis process, helping to deal with new threats, massive volumes of data, process classification, automated clustering, and the need to optimize classifiers. The WP7 SAFAIR Program presented a novel AI threat modelling method based on a thorough inventory to capture all relevant information about unique threats to AI system components. One of the main sources that contributed to the development of the T-SHARK visual analytics system was the Knowledge Base, which was implemented to capture all aspects of the organization and taxonomy of AI threats, as well as the selection and adjustment of defensive and reactive mechanisms (WP7 D7.1) to ensure resilience against new and complex cyber threats.

WP8 Clustering, platforms, and ecosystems activities

The SPARTA WP8 cooperation program was successful in bringing the community together on cybersecurity challenges. The established links between SPARTA WP8 and ecosystems that hold a vast portfolio of knowledge, practical experience, niche capabilities, and creative advances have opened new potential for the T-SHARK research program to benefit from the knowledge generated. It's worth noting that T-SHARK solutions, i.e. those based on MISP and C3ISP technologies, will be accessible via the SPARTA online application, which will give other organizations and specialists direct access to T-SHARK Program results.

WP9 Cybersecurity training and awareness

While the SPARTA Cybersecurity Skills Framework was being established, the T-SHARK WP actively collaborated with WP9. Adaptability was one of the factors used in the Framework's creation. This refers to the Framework's ability to incorporate and communicate new, emerging tasks that will become relevant to cybersecurity practitioners soon. As a result, the Framework can serve as an

early warning system for stakeholders about new developments, allowing for speedier integration of fresh elements into educational programs or professional training tools. Interviews with the T-SHARK comprehensive intelligence team were used to assess adaptability, identifying trends in highly sophisticated attacks and the capabilities required to defeat them. As a result of visual analytical systems are integrated with the training platform. Also, the results are disseminated through the JCCI platform (https://www.sparta.eu/JCCI/jcci_home.html)

WP12 Dissemination and communication

Inspired by the dissemination campaign “Cyber Shield” established close collaborative links with National Cyber Security Centre the development of the national cyber training programme. Intercommunication with WP9 (SPARTA Cybersecurity Skills Framework) through the JCCI platform (https://www.sparta.eu/JCCI/jcci_home.html) virtual instance become public available.

Presentations in Conferences:

1. Technical Threat Intelligence Analytics: What and How to Visualize for Analytic Process was presented in 24th International Conference ELECTRONICS 2020 and for publication in the IEEE Conference Proceedings.

Scientific articles published in Journals:

1. LITNET-2020: An Annotated Real-World Network Flow Dataset for Network Intrusion Detection; Visual Analytics for Cyber Security Domain: State-of-the-Art and Challenges.
2. Ensemble-Based Classification Using Neural Networks and Machine Learning Models for Windows PE Malware Detection.
3. A Novel Approach for Network Intrusion Detection Using Multistage Deep Learning Image Recognition.

Dataset (LITNET-2020: An Annotated Real-World Network Flow Dataset for Network Intrusion Detection) for benchmarking, testing, and validation cyber-attacks is publicly available at: https://dataset.litnet.lt/

Chapter 7 Summary and Conclusion

All of the significant results are that during the project, specifically 4.6 derivable research was conducted, and published scientific articles were published in journals with impact factors. A total of 8 scientific papers were published, with the help of which achieved the following results:

1. The key objective of the Visual Analytics System (VAS) is to provide the necessary information for system administrators to facilitate identifying:
   1.1. the IT systems subject to visual analysis.
   1.2. the vulnerabilities present on those systems.
   1.3. how cyber-attacks are exploiting such vulnerabilities.
   1.4. the actual impact of those attacks on the services and goals offered on the IT systems analysed.
2. To do so the VAS techniques gather knowledge from many cyber-security resources such as:
   2.1. reconnaissance tools for IT system discovery.
   2.2. vulnerability scanners and/or public vulnerability databases for vulnerability affecting organizations.
   2.3. intrusion detection systems and SIEMs for information about attacks.
   2.4. impact assessment mechanisms that identify services and goals affected by those attacks.
3. The VAS system must faithfully address and fulfil the following desired characteristics:
   3.1. **Scalability**. The large amounts of data available for analysis should be visualized effectively.
   3.2. **Report generation**. Report generation function must be provided to ensure effective facilitation of the work by the analyst.
   3.3. **Reliability**. Measures for uncertainty modelling to ensure reliability and trustworthiness of visual analytics and supporting making informed decisions should be provided.
   3.4. **Heterogeneous data analysis**. Security analysts often obtain information by analysing different sources of information such as operating system logs and social networks. Although a wide variety of techniques have been proposed to visualize data sources individually, the interaction mechanism between these views must be consistent and concise.
   3.5. **Collaborative intelligence**. Incident response teams often must work closely together when trying to resolve a system breach. The need for integration of visual displays of data between different tools and stakeholders should be ensured.
   3.6. **Situation awareness**. Large volume of data extracted from computer networks and high complexity of relations between data inhibits achieving situational awareness by an expert. The final aim is to present a bird's view on the security events to cyber security experts for decision support and situation awareness.
   3.7. **Usability**. To provide efficient decision support, visualization interfaces should have the general look of the system well designed and aesthetically pleasing while ensuring the needs of target audience of a system are satisfied Main user experience components including familiarity, learnability, responsiveness, performance, intuitiveness, efficiency, helpfulness, and satisfactoriness must be ensured.
   3.8. **Privacy**. Security visualization systems should ensure the protection of private user data when user privacy is top priority, while analysing user behaviour through checking network traffic or analysing logs.
   3.9. **Human Factors**. The effective use of colour, texture, shape, and other elements of visual representation should be ensured to effectively capture human attention, which is highly relevant for real-time monitoring tasks.

Finally, this article examined the current state of cyber threat models, methodologies, tools, and demonstrate Visual Analytics System capabilities. This Visual Analytics System was then verified and mapped on the subcases developed within T-SHARK, including D4.2 Cybersecurity Threat Analysis Model, D4.1 Cybersecurity Threat Intelligence Common Data Model, and T4.4 Information Sharing and Integration, as well as the umbrella scenario of election interference. As a result, the Visual Analytics System for Cybersecurity Threat Analysis will be innovative.

Chapter 8 List of Abbreviations

Abbreviation	Translation
AI	Artificial Intelligence
API	An application programming interface
BP-IDS	Business Process Intrusion Detection System
CSTI	Comprehensive Cybersecurity Threat Intelligence
CYBINT	Cyber Intelligence
COTS	A COTS (commercial off-the-shelf) product is one that is used "as-is."
ECOSSIAN	European COntrol System Security Incident Analysis Network
FININT	Financial Intelligence
GEOINT	Geospatial Intelligence
HUMINT	Human Intelligence
LITNET	Lithuanian Research and Education Network in Lithuania
MISP	Open-Source Threat Intelligence Platform & Open Standards for Threat Information Sharing
OSINT	Open-Source Intelligence
OWASP	Open Web Application Security Project
SIEM	Security information and event management
SIGINT	Signals Intelligence
SOCINT	Social Intelligence
TTI	Technical Threat Intelligence
VAS	Visual Analytics System

Chapter 9 References

[1] S. Noel, E. Harley, K. H. Tam, M. Limiero, and M. Share, 'Chapter 4 - CyGraph: Graph-Based Analytics and Visualization for Cybersecurity', in Handbook of Statistics, vol. 35, V. N. Gudivada, V. V. Raghavan, V. Govindaraju, and C. R. Rao, Eds. Elsevier, 2016, pp. 117–167.

[2] A. de Barros Barreto, P. Costa, and M. Hieb, 'Cyber-Argus: Modeling C2 Impacts of Cyber Attacks', GEORGE MASON UNIV FAIRFAX VA CENTER FOR EXCELLENCE IN COMMAND CONTROL …, 2014.

[3] S. Musman and A. Temin, 'A Cyber Mission Impact assessment tool', in 2015 IEEE International Symposium on Technologies for Homeland Security (HST), 2015, pp. 1–7.

[4] J. R. Goodall, A. D'Amico, and J. K. Kopylec, 'Camus: Automatically mapping Cyber Assets to Missions and Users', in MILCOM 2009 - 2009 IEEE Military Communications Conference, 2009, pp. 1–7.

[5] 'What's New with Microsoft Threat Modeling Tool 2016', Microsoft Security, 08-Oct-2015. [Online]. Available: https://www.microsoft.com/security/blog/2015/10/07/whats-new-with-microsoft-threat-modeling-tool-2016/. [Accessed: 08-Dec-2019].

[6] W. Van Der Aalst, Process mining: discovery, conformance and enhancement of business processes, vol. 2. Springer, 2011.

[7] K. Böhmer and S. Rinderle-Ma, 'Automatic Signature Generation for Anomaly Detection in Business Process Instance Data', in Enterprise, Business-Process and Information Systems Modeling, Cham, 2016, pp. 196–211.

[8] M. Wolfgang, 'Host Discovery with nmap', Explor. Nmaps Default Behav., vol. 1, p. 16, 2002.

[9] 'Moloch'. [Online]. Available: http://molo.ch. [Accessed: 26-Nov-2019].

[10] 'SiLK'. [Online]. Available: https://tools.netsa.cert.org/silk/. [Accessed: 26-Nov-2019].

[11] 'CVE - Common Vulnerabilities and Exposures (CVE)'. [Online]. Available: https://cve.mitre.org/. [Accessed: 26-Nov-2019].

[12] 'NVD - Vulnerabilities'. [Online]. Available: https://nvd.nist.gov/vuln. [Accessed: 26-Nov-2019].

[13] 'Download Nessus Vulnerability Assessment | Tenable®'. [Online]. Available: https://www.tenable.com/products/nessus. [Accessed: 26-Nov-2019].

[14] 'Nikto2 | CIRT.net'. [Online]. Available: https://cirt.net/Nikto2. [Accessed: 26-Nov-2019].

[15] 'ZAP'. [Online]. Available: https://www.zaproxy.org/. [Accessed: 26-Nov-2019].

[16] A. Tiwari, 'Archery - Vulnerability Assessment and Management Tool', Medium, 20-Sep-2018. [Online]. Available: https://medium.com/archerysec/archery-vulnerability-assessment-and-management-tool-ecbf5e92f717. [Accessed: 26-Nov-2019].

[17] 'OpenVAS - OpenVAS - Open Vulnerability Assessment Scanner'. [Online]. Available: http://openvas.org/. [Accessed: 26-Nov-2019].

[18] 'OWASP Threat Dragon - OWASP'. [Online]. Available: https://www.owasp.org/index.php/OWASP_Threat_Dragon. [Accessed: 26-Nov-2019].

[19] 'IBM QRadar Risk Manager', 24-Oct-2014. [Online]. Available: www.ibm.com/support/knowledgecenter/en/ss42vs_7.3.2/com.ibm.qradar.doc/c_qrm_ug_overview.html. [Accessed: 26-Nov-2019].

[20] 'DefectDojo | CI/CD and DevSecOps Automation'. [Online]. Available: https://www.defectdojo.org/. [Accessed: 26-Nov-2019].

[21] ‘ThreadFix Vulnerability Management Platform’. [Online]. Available: https://threadfix.it/. [Accessed: 26-Nov-2019].

[22] J. T. F. T. Initiative, ‘Guide for conducting risk assessments’, National Institute of Standards and Technology, 2012.

[23] R. Bray, D. Cid, and A. Hay, OSSEC host-based intrusion detection guide. Syngress, 2008.

[24] M. Roesch, ‘Snort: Lightweight intrusion detection for networks.’, in Lisa, 1999, vol. 99, pp. 229–238.

[25] V. Paxson, S. Campbell, and J. Lee, ‘Bro intrusion detection system’, Lawrence Berkeley National Laboratory, 2006.

[26] ‘Enterprise Intrusion Prevention (IPS) Software & Solutions’, Trend Micro. [Online]. Available: https://www.trendmicro.com/en_us/business/products/network/intrusion-prevention.html. [Accessed: 08-Dec-2019].

[27] ‘INOV / BP-IDS – Business Process Intrusion Detection System’. .

[28] ‘2018 Gartner Magic Quadrant for Intrusion Detection and Prevention Systems (IDPS) | Alert Logic’. [Online]. Available: https://www.alertlogic.com/resources/industry-reports/intrusion-detection-and-prevention-systems-providers/. [Accessed: 26-Feb-2019].

[29] K.-T. Cho and K. G. Shin, ‘Fingerprinting electronic control units for vehicle intrusion detection’, in 25th ${$USENIX$}$ Security Symposium (${$USENIX$}$ Security 16), 2016, pp. 911–927.

[30] H. Choi et al., ‘Detecting Attacks Against Robotic Vehicles: A Control Invariant Approach’, in Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, pp. 801–816.

[31] M. Kneib and C. Huth, ‘Scission: Signal Characteristic-Based Sender Identification and Intrusion Detection in Automotive Networks’, in Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, pp. 787–800.

[32] D. I. Urbina et al., ‘Limiting the impact of stealthy attacks on industrial control systems’, in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 1092–1105.

[33] C. M. Ahmed, J. Zhou, and A. P. Mathur, ‘Noise Matters: Using Sensor and Process Noise Fingerprint to Detect Stealthy Cyber Attacks and Authenticate sensors in CPS’, in Proceedings of the 34th Annual Computer Security Applications Conference, 2018, pp. 566–581.

[34] Y. Chen, C. M. Poskitt, and J. Sun, ‘Learning from mutants: using code mutation to learn and monitor invariants of a cyber-physical system’, in 2018 IEEE Symposium on Security and Privacy (SP), 2018, pp. 648–660.

[35] W. Aoudi, M. Iturbe, and M. Almgren, ‘Truth Will Out: Departure-Based Process-Level Detection of Stealthy Attacks on Control Systems’, in Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, pp. 817–831.

[36] D. Formby, P. Srinivasan, A. Leonard, J. Rogers, and R. A. Beyah, ‘Who’s in Control of Your Control System? Device Fingerprinting for Cyber-Physical Systems.’, in NDSS, 2016.

[37] L. Cheng, K. Tian, and D. D. Yao, ‘Orpheus: Enforcing cyber-physical execution semantics to defend against data-oriented attacks’, in Proceedings of the 33rd Annual Computer Security Applications Conference, 2017, pp. 315–326.

[38] C.-W. Ten, C.-C. Liu, and G. Manimaran, ‘Vulnerability assessment of cybersecurity for SCADA systems’, IEEE Trans. Power Syst., vol. 23, no. 4, pp. 1836–1846, 2008.

[39] C. Ten, C. Liu, and M. Govindarasu, 'Vulnerability Assessment of Cybersecurity for SCADA Systems Using Attack Trees', in 2007 IEEE Power Engineering Society General Meeting, 2007, pp. 1–8.

[40] A. Kott, J. Ludwig, and M. Lange, 'Assessing Mission Impact of Cyberattacks: Toward a Model-Driven Paradigm', IEEE Secur. Priv., vol. 15, no. 5, pp. 65–74, 2017.

[41] P. A. Porras, M. W. Fong, and A. Valdes, 'A mission-impact-based approach to INFOSEC alarm correlation', in International Workshop on Recent Advances in Intrusion Detection, 2002, pp. 95–114.

[42] N. Friedman, D. Geiger, and M. Goldszmidt, 'Bayesian network classifiers', Mach. Learn., vol. 29, no. 2–3, pp. 131–163, 1997.

[43] B. J. Argauer and S. J. Yang, 'VTAC: Virtual terrain assisted impact assessment for cyber attacks', in Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2008, 2008, vol. 6973, p. 69730F.

[44] S. Jajodia, S. Noel, P. Kalapa, M. Albanese, and J. Williams, 'Cauldron mission-centric cyber situational awareness with defense in depth', in 2011 - MILCOM 2011 Military Communications Conference, 2011, pp. 1339–1344.

[45] G. Jakobson, 'Mission cyber security situation assessment using impact dependency graphs', in 14th International Conference on Information Fusion, 2011, pp. 1–8.

[46] F. Balmas, 'Displaying dependence graphs: a hierarchical approach', J. Softw. Maint. Evol. Res. Pract., vol. 16, no. 3, pp. 151–185, 2004.

[47] C. Liu, A. Singhal, and D. Wijesekera, 'A layered graphical model for mission attack impact analysis', in 2017 IEEE Conference on Communications and Network Security (CNS), 2017, pp. 602–609.

[48] A. Motzek, R. Möller, M. Lange, and S. Dubus, 'Probabilistic mission impact assessment based on widespread local events', Assess. Mission Impact Cyberattacks, p. 1, 2015.

[49] Y. Yang, Z. Cai, C. Wang, and J. Zhang, 'Probabilistically Inferring Attack Ramifications Using Temporal Dependence Network', IEEE Trans. Inf. Forensics Secur., vol. 13, no. 11, pp. 2913–2928, 2018.

[50] D. Krahl, 'ExtendSim Advanced Technology: Discrete Rate Simulation', 2010, pp. 333–338.

[51] J. S. Banerjee, D. Goswami, and S. Nandi, 'OPNET: A New Paradigm for Simulation of Advanced Communication Systems', Int. J. Manag. Tech. Res., vol. 1, pp. 319–328, Mar. 2014.

[52] I. Kotenko and A. Chechulin, 'A Cyber Attack Modeling and Impact Assessment framework', in 2013 5th International Conference on Cyber Conflict (CYCON 2013), 2013, pp. 1–24.

[53] P. A. Porras, M. W. Fong, and A. Valdes, 'A mission-impact-based approach to INFOSEC alarm correlation', in International Workshop on Recent Advances in Intrusion Detection, 2002, pp. 95–114.

[54] K. Stouffer, S. Lightman, V. Pillitteri, M. Abrams, and A. Hahn, 'Guide to Industrial Control Systems (ICS) Security', National Institute of Standards and Technology, NIST Special Publication (SP) 800-82 Rev. 2, Jun. 2015.

[55] 'Now That's Smart! - IEEE Journals & Magazine'. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4408617. [Accessed: 12-Aug-2019].

[56] G. Frey and L. Litz, 'Formal methods in PLC programming', in Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics. 'cybernetics evolving to systems, humans, organizations, and their complex interactions' (cat. no.0, 2000, vol. 4, pp. 2431–2436 vol.4.

[57] B. Galloway and G. P. Hancke, 'Introduction to Industrial Control Networks', IEEE Commun. Surv. Tutor., vol. 15, no. 2, pp. 860–880, Second 2013.

[58] 'Chernobyl | Chernobyl Accident | Chernobyl Disaster - World Nuclear Association'. [Online]. Available: http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/chernobyl-accident.aspx. [Accessed: 22-Feb-2019].

[59] P. Y. Lipscy, K. E. Kushida, and T. Incerti, 'The Fukushima disaster and Japan's nuclear plant vulnerability in comparative perspective', Environ. Sci. Technol., vol. 47, no. 12, pp. 6082–6088, 2013.

[60] 'Tokaimura Criticality Accident - World Nuclear Association'. [Online]. Available: http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/tokaimura-criticality-accident.aspx. [Accessed: 22-Feb-2019].

[61] 'The Real Story of Germanwings Flight 9525 | GQ'. [Online]. Available: https://www.gq.com/story/germanwings-flight-9525-final-moments. [Accessed: 22-Feb-2019].

[62] N. Falliere, L. O. Murchu, and E. Chien, 'W32. stuxnet dossier', White Pap. Symantec Corp Secur. Response, vol. 5, no. 6, p. 29, 2011.

[63] Robertas Damaševičius, Jevgenijus Toldinas, Algimantas Venčkauskas, Šarūnas Grigaliūnas, Nerijus Morkevičius, Vaidas Jukavičius, 'Visual Analytics for Cyber Security Domain: State-of-the-Art and Challenges', In International Conference on Information and Software Technologies (pp. 256-270). Springer, Cham. (2019, October). https://link.springer.com/chapter/10.1007/978-3-030-30275-7_20

[64] Robertas Damasevicius, Algimantas Venckauskas, Sarunas Grigaliunas, Jevgenijus Toldinas, Nerijus Morkevicius, Tautvydas Aleliunas, Paulius Smuikys, 'LITNET-2020: An Annotated Real-World Network Flow Dataset for Network Intrusion Detection', Electronics, 9(5), 800. 2020, May https://www.mdpi.com/2079-9292/9/5/800

[65] Robertas Damasevicius; Jevgenijus Toldinas; Algimantas Venckauskas; Sarunas Grigaliunas; Nerijus Morkevicius, 'Technical Threat Intelligence Analytics: What and How to Visualize for Analytic Process', 24th International Conference ELECTRONICS 2020 and for publication in the IEEE Conference Proceedings (ISBN 978-1-7281-5868-6)). 2020, June https://ieeexplore.ieee.org/document/9141613

[66] Robertas Damaševičius, Algimantas Venčkauskas, Jevgenijus Toldinas, Šarūnas Grigaliūnas, 'Ensemble-Based Classification Using Neural Networks and Machine Learning Models for Windows PE Malware Detection', 2021, February "Electronics 10, no. 4: 485. https://doi.org/10.3390/electronics10040485, https://www.mdpi.com/2079-9292/10/4/485"

[67] Jevgenijus Toldinas, Algimantas Venčkauskas, Robertas Damaševičius, Šarūnas Grigaliūnas, Nerijus Morkevičius, Edgaras Baranauskas , 'A Novel Approach for Network Intrusion Detection Using Multistage Deep Learning Image Recognition', 2021 August "Electronics 10, no. 15: 1854 https://doi.org/10.3390/electronics10151854 "

[68] Corinna Köpke, Kushal Srivastava, Louis König, Natalie Miller, Mirjam Fehling-Kaschek, Kelly Burke, Matteo Mangini, Isabel Praça, Alda Canito, Olga Carvalho, Filipe Apolinário, Nelson Escravana, Nils Carstengerdes, Tim Stelkens-Kobsch, 'Impact Propagation in Airport Systems 2021. In: Abie H. et al. (eds) Cyber-Physical Security for Critical Infrastructures Protection. CPS4CIP 2020', Lecture Notes in Computer Science, vol 12618. Springer, Cham. https://doi.org/10.1007/978-3-030-69781-5_13

[69] Fabian Reuschling, Nils Carstengerdes, Tim H. Stelkens-Kobsch, Kelly Burke, Thomas Oudin, Meilin Schaper, Filipe Apolinário, Isabel Praça and Leonidas Perlepes 'Toolkit to Enhance Cyber-physical Security of Critical Infrastructures in Air Transport', Cyber-Physical Threat Intelligence for Critical Infrastructures Security by John Soldatos, Isabel Praça and Aleksandar Jovanovic´ (eds.). 2021. ISBN 978-1-68083-822-0. E-ISBN 978-1-68083-823-7, DOI: 10.1561/9781680838237.ch11, Published: 15 Sep 2021

[70] Abedin M., Nessa S., Khan L., Thuraisingham B., 'Detection and Resolution of Anomalies in Firewall Policy Rules', In: Damiani E., Liu P. (eds) Data and Applications Security XX. DBSec 2006. Lecture Notes in Computer Science, vol 4127. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11805588_2

[71] Xinming Ou, Sudhakar Govindavajhala, Andrew W. Appel,' MulVAL: A Logic-based Network Security Analyzer', 14th USENIX Security Symposium USENIX Security 05 (2005), https://www.usenix.org/legacy/publications/library/proceedings/sec05/tech/full_papers/ou/ou_html/index.html

[72] Bertin, J.: Graphische Semiologie: Diagramme, Netze, Karten; Translated from the 2nd French Edition (1973). Walter de Gruyter, Berlin, Germany (1974). ISBN 3-11-003660-6

[73] Störrle, H., Fish, A.: Towards an operationalization of the "Physics of Notations" for the analysis of visual languages. In: Moreira, A., Schätz, B., Gray, J., Vallecillo, A., Clarke, P. (eds.) MODELS 2013. LNCS, vol. 8107, pp. 104–120. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41533-3_7

[74] de Bruijn, H., Janssen, M.: Building cybersecurity awareness: the need for evidence-based framing strategies. Gov. Inf. Q. 34(1), 1–7 (2017). https://doi.org/10.1016/j.giq.2017.02.007

[75] Zheng, M., Robbins, H., Chai, Z., Thapa, P., Moore, T.: Cybersecurity research datasets: taxonomy and empirical analysis. In: 11th USENIX Conference on Cyber Security Experimentation and Test (CSET 2018), p. 2. USENIX Association, Berkeley (2018)

[76] He, J., Chen, H., Chen, Y., Tang, X., Zou, Y.: Diverse visualization techniques and methods of moving-object-trajectory data: a review. ISPRS Int. J. Geo-Inf. 8(2), 63 (2019). https://doi.org/10.3390/ijgi8020063

[77] Kotenko, I., Novikova, E.: VisSecAnalyzer: a visual analytics tool for network security assessment. In: Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L. (eds.) CD-ARES 2013. LNCS, vol. 8128, pp. 345–360. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40588-4_24

[78] Zhao, Y., Zhou, F., Fan, X., Liang, X., Liu, Y.: IDSRadar: a real-time visualization framework for IDS alerts. Sci. China Inf. Sci. 56(8), 1–12 (2013). https://doi.org/10.1007/s11432-013-4891-9

[79] Haggerty, J., Haggerty, S., Taylor, M.: Forensic triage of email network narratives through visualisation. Inf. Manag. Comput. Secur. 22(4), 358–370 (2014). https://doi.org/10.1108/IMCS-11-2013-0080

[80] Chen, V.Y., Razip, A.M., Ko, S., Qian, C.Z., Ebert, D.S.: Multi-aspect visual analytics on large-scale high-dimensional cyber security data. Inf. Vis. 14(1), 62–75 (2013). https://doi.org/.1177/1473871613488573

[81] Liao, Q., Striegel, A., Chawla, N.: Visualizing graph dynamics and similarity for enterprise network security and management. In: 7th International Symposium on Visualization for Cyber Security (VizSec 2010), pp. 34–45 (2010). https://doi.org/10.1145/1850795.1850799

[82] McKenna, S., Staheli, D., Fulcher, C., Meyer, M.: BubbleNet: a cyber security dashboard for visualizing patterns. Comput. Graph. Forum 35(3), 281–290 (2016). https://doi.org/10.1111/cgf.12904

[83] Wongsuphasawat, K., Guerra Gómez, J.A., Plaisant, C., Wang, T., Taieb-Maimon, M., Shneiderman, B.: LifeFlow. In: Annual Conference Extended abstracts on Human Factors in Computing Systems - CHI EA 2011. ACM Press (2011). https://doi.org/10.1145/1979742.1979557

[84] McPherson, J., Ma, K.-L., Krystosk, P., Bartoletti, T., Christensen, M.: PortVis. In: Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security - VizSEC/DMSEC 2004. ACM Press (2004). https://doi.org/10.1145/1029208.1029220

[85] Qiu, H.S.: Streaming data visualization for network security. Ph.D. thesis, Princeton University (2017)

[86] Goodall, J.R.: Introduction to visualization for computer security. In: Goodall, J.R., Conti, G., Ma, K.L. (eds.) VizSEC 2007. Mathematics and Visualization, pp. 1–17. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78243-8_1

[87] Hu, H., Zhang, H., Liu, Y., Wang, Y.: Quantitative method for network security situation based on attack prediction. Secur. Commun. Netw. 2017, 19 (2017). https://doi.org/10.1155/2017/3407642. Article ID 3407642

[88] Scheepens, R., Michels, S., van de Wetering, H., van Wijk, J.J.: Rationale visualization for safety and security. Comput. Graph. Forum 34, 191–200 (2015)

[89] Staheli, D., et al.: Visualization evaluation for cyber security. In: Eleventh Workshop on Visualization for Cyber Security, VizSec 2014. ACM Press (2014). https://doi.org/10.1145/2671491.2671492

[90] Shiravi, H., Shiravi, A., Ghorbani, A.A.: A survey of visualization systems for network security. IEEE Trans. Vis. Comput. Graph. 18(8), 1313–1329 (2012). https://doi.org/10. 1109/tvcg.2011.144

[91] Fink, G.A., North, C.L., Endert, A., Rose, S.: Visualizing cyber security: usable workspaces. In: 2009 6th International Workshop on Visualization for Cyber Security. IEEE (2009). https://doi.org/10.1109/vizsec.2009.5375542

[92] Khanh Dang, T., Tri Dang, T.: A survey on security visualization techniques for web information systems. Int. J. Web Inf. Syst. 9(1), 6–31 (2013). https://doi.org/10.1108/17440081311316361

[93] Tianfield,H.:Cybersecurity situational awareness. In: 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (2016). https://doi.org/10.1109/ithings-greencom-cpscom-smartdata.2016.165

[94] Marty, R.: Cyber security: how visual analytics unlock insight. In: 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013. ACM Press (2013). https://doi.org/10.1145/2487575.2491132

[95] McCaslin, M.L., Scott, K.W.: The five-question method for framing a qualitative research study. Qual. Rep. 8(3), 447–461 (2003)

[96] Angelini, M., Blasilli, G., Lenti, S., Santucci, G.: Visual exploration and analysis of the Italian cybersecurity framework. In: Workshop on Advanced Visual Interfaces AVI (2018). https://doi.org/10.1145/3206505.3206579

[97] Furmanova, K., et al.: Taggle: Combining Overview and Details in Tabular Data Visualizations, 14 p. (2019). arXiv:1712.05944v3 [cs.HC]

[98] Tillekens, A., Le-Khac, N.-A., Thi, T.T.P.: A Bespoke forensics GIS tool. In: 2016 International Conference on Computational Science and Computational Intelligence, pp. 987–992. IEEE (2016). https://doi.org/10.1109/csci.2016.188

[99] Aldwairi, M., Alsaadi, H.H.: FLUKES: autonomous log forensics, intelligence and visualization tool. In: Proceedings of ICFNDS 2017, Cambridge, United Kingdom, 19–20 July 2017, 6 p. (2017). https://doi.org/10.1145/3102304.3102337 Visual Analytics for Cyber Security Domain: State-of-the-Art and Challenges 269

[100] Tuncel, M.A., Francis, H., Taylor, M., Jones, D.L.: Visualdrives forensic tool. In: International Conference on Developments of E-Systems Engineering (DeSE), Burj Khalifa, Dubai, United Arab Emirates, 13–15 December 2015. https://doi.org/10.1109/dese.2015.68

[101] Hales, G., Ferguson, I., Archibald, J.: Insight: an application of information visualisation techniques to digital forensics investigations. Int. J. Cyber Situat. Aware. 2(1), 100–118 (2017)

[102] Olsson, J., Boldt, M.: Computer forensic timeline visualization tool. Digit. Investig. S78–S87 (2009). https://doi.org/10.1016/j.diin.2009.06.008

[103] Leschke, T.R., Nicholas, C.: Change-Link 2.0: a digital forensic tool for visualizing changes to shadow volume data. In: VizSec 2013, Atlanta, GA, USA, 14 October 2013, pp. 17–24 (2013)

[104] Catanese, S.A., Fiumara, G.: A visual tool for forensic analysis of mobile phone traffic. In: MiFOR 2010, Firenze, Italy, 29 October 2010, pp. 71–76 (2010)

[105] Goswami, A., Mohapatra, D.P., Zhai, C.: Qu antifying and visualizing the demand and supply gap from e-commerce search data using topic models. In: WWW 2019 Companion, San Francisco, CA, USA, 13–17 May 2019, pp. 348–353 (2019)

[106] Le, T.V.M., Akoglu, L.: ContraVis: contrastive and visual topic modeling for comparing document collections. In: Proceedings of the 2019 World Wide Web Conference (WWW 2019), San Francisco, CA, USA, 13–17 May 2019, 11 p. ACM, New York (2019). https://doi.org/10.1145/3308558.3313617

[107] Yang, F., et al.: XFake: explainable fake news detector with visualizations. In: WWW 2019, San Francisco, CA, USA, 13–17 May 2019. https://doi.org/10.1145/3308558.3314119

[108] Fittkau, F., Krause, A., Hasselbring, W.: Software landscape and application visualization or system comprehension with ExplorViz. Inf. Softw. Technol. 87(2017), 259–277 (2017)

Chapter 10 Annex 1: API's Query Description

Category	Query	cURL

index – Add (or update) a document	PUT /<<indexname>>
	Copy
		curl -XPUT
		'vas.sparta.ktu.lt:9200/twitter/my_index/my_type/1?pretty'
		-H 'Content-Type: application/json' -d'
		{
		"field : "value",
		...
		}
		'
		Copy

get – Retrieve a specific existing document	GET /<<indexname>> Copy	curl -XGET 'vas.sparta.ktu.lt:9200/my_index/ my_type/0?pretty' Copy
delete – Delete a document	DELETE /<<indexname>> Copy	curl -XDELETE 'vas.sparta.ktu.lt:9200/my_index/ my_type/0?pretty' Copy

reindex – Copies a document from one index to another	POST /_reindex
	Copy
		curl -XPOST
		'vas.sparta.ktu.lt:9200/_reindex?pretty'
		-H 'Content-Type:
		application/json' -d'
		{
		"source": {
		"index": "some_old_index"
		},
		"dest": {
		"index": "some_new_index"
		}
		}
		'
		Copy

multi get API (mget) – This lets you pull documents from multiple indices, specifying as many docs as	GET /<<targetindex>>/_mget Copy	curl -X GET "vas.sparta.ktu.lt:9200/_mget?pretty" -H 'Content-Type: application/json' -d' { "docs": [

Category	Query	cURL

necessary per index		{
		"_index": "index1",
		"_id": "1"
		},
		{
		"_index": "index1",
		"_id": "2"
		}
		]
		}
		Copy

bulk – This lets you perform multiple types of requests at once.	POST /<<targetindex>>/_bulk Copy	curl -X POST "vas.sparta.ktu.lt:9200/_bulk?pretty" -H 'Content-Type: application/json' -d' { "index" : { "_index" : "test", "_id" : "1" } } { "delete" : { "_index" : "test", "_id" : "2" } } { "create" : { "_index" : "test", "_id" : "3" } } { "field1" : "value1" }} Copy

delete by query	POST /<<targetindex>>/_delete_by_query
	Copy
		curl -X POST
		"vas.sparta.ktu.lt:9200/index1/_delete_by_query?pretty"
		-H 'Content-Type: application/json' -d'
		{
		"query": {
		"match": {
		"user.id": "gedalyahreback"
		}
		}
		}
		Copy

update by query – the parameter at the end tells the query to proceed in the event there is a conflict between versions of a document	POST /<<targetindex>>/_update_by_query Copy	curl -X POST "vas.sparta.ktu.lt:9200/myindex1/ _update_by_query?conflicts=proceed" Copy

Table 13: API multi get

Category	Query	cURL

Create a new Elasticsearch index	PUT /<<indexname>>
	Copy
		curl -XPUT
		'vas.sparta.ktu.lt:9200/indexname?pretty'
		-H 'Content-Type: application/json' -d'
		{
		"settings" : {
		"index" : {
		...
		}
		}
		}
		'
		Copy

Delete an index	DELETE /<<indexname>> Copy	curl -XDELETE 'vas.sparta.ktu.lt:9200/<<indexname>>?pretty' Copy
Open or Close an index	POST /<<indexname>>/_open Copy POST /<<indexname>>/_close Copy	curl -XPOST 'vas.sparta.ktu.lt:9200/<<indexname>>/_close?pretty ' Copy curl -XPOST 'vas.sparta.ktu.lt:9200/<<indexname>>/_open?pretty' Copy
Shrink	POST /<<indexname>>/_shrink/<<indexname>> PUT /<<indexname>>/_shrink/<<indexname>> Copy	curl -XPOST "vas.sparta.ktu.lt:9200/<<indexname>>/ _shrink/shrunken-indexname" Copy

Split	POST /<<indexname>>/_split/<<indexname>>
	PUT /<<indexname>>/_split/<<indexname>>
	Copy
		curl -XPOST
		"vas.sparta.ktu.lt:9200/indexname/_split/split-indexname"
		-H 'Content-Type: application/json' -d'
		{
		"settings": {
		"index.number_of_shards": 4
		}
		}
		'
		Copy

Category	Query	cURL
Clone	POST /<<indexname>>/_clone/<<clonedindexname>> PUT /<<indexname>>/_clone/<<clonedindexname>> Copy	curl -X POST "vas.sparta.ktu.lt:9200/indexname/_clone/clonedindex" Copy
Resolve	GET /_resolve/index/<<indexname>> Copy	curl -X GET "vas.sparta.ktu.lt:9200/_resolve/index/indexname" Copy

Rollover	POST /<<indextoroll>>/_rollover/<<newindex>>
	POST /<<indextoroll>>/_rollover/
	Copy
		curl -X POST
		"vas.sparta.ktu.lt:9200/indextoroll/_rollover/newindex"
		-H 'Content-Type: application/json' -d'
		{
		"conditions": {
		"max_age": "14d",
		"max_docs": 5000,
		"max_size": "15gb"
		}
		}
		'
		Copy

Table 14: Index Management

Category	Query	cURL

Add a new type to existing mapping	PUT /<<indexname>>/_mapping
	Copy
	PUT /_mapping
	Copy
		curl -XPUT
		'vas.sparta.ktu.lt:9200/indexname/_mapping/user?pretty'
		-H 'Content-Type: application/json' -d'
		{
		"properties": {
		"name": {
		"type": "text"
		}
		}
		}
		'
		Copy

Retrieve mapping for a specific field	GET /<<indexname>>/_mapping	curl -XGET 'vas.sparta.ktu.lt:9200/indexname/_mapping/

Category	Query	cURL
	Copy GET /_mapping Copy	my_type/field/my_field?pretty' Copy

Table 15: Mapping Management

Category	Query	cURL
Search – Enter a search query and return hits matching the query	GET /<<targetindex>>/_search POST /<<targetindex>>/_search Copy	curl -XGET 'vas.sparta.ktu.lt:9200/my_index/my_type/_count?q=field:value&pretty' Copy
Validate – Validate a potentially heavy query without actually executing it	GET /<<targetindex>>/_validate/<<query>> Copy	curl -XGET 'vas.sparta.ktu.lt:9200/my_index/my_type/_validate?q=field:value' Copy
Explain – Calculate a score for a query for getting feedback on whether a document matches the query or not	GET /<<targetindex>>/_explain/<<id>> POST /<<targetindex>>/_explain/<<id>> Copy	curl -XGET 'vas.sparta.ktu.lt:9200/my_index/my_type/0/_explain?q=message:search' Copy
Scroll	GET /_search/scroll POST /_search/scroll DELETE /_search/scroll Copy	curl -X GET "vas.sparta.ktu.lt:9200/_search/scroll?pretty" -H 'Content-Type: application/json' -d' {} ' Copy
Search Template	GET /_search/template Copy	curl -X GET "vas.sparta.ktu.lt:9200/_search/scroll?pretty" -H 'Content-Type: application/json' -d' } ' Copy

Storing a search template using the _scripts API	POST _scripts/<<templateid>>
	DELETE _scripts/<<templateid>>
	Copy
		curl -X POST
		"vas.sparta.ktu.lt:9200/_scripts/<<templateid>>?pretty"
		-H 'Content-Type: application/json' -d'
		{
		"script": {

Category	Query	cURL

		"lang": "mustache",
		"source": {
		"query": "{{some_template}}"
		}
		}
		}
		}
		Copy

Table 16: Search API

Category	Query	cURL
Cat Indices – Gives us access to info & metrics regarding our indices	GET /_cat/indices	curl -XGET 'vas.sparta.ktu.lt:9200/_cat/indices?v&health=red&pretty'
Cat Health – Overview of index health	GET /_cat/health	curl -XGET 'vas.sparta.ktu.lt:9200/_cat/health?v&pretty'
Cat Nodes – Info on Elasticsearch nodes *#Tip: You can use headers to retrieve only relevant details on the nodes.*	GET /_cat/nodes	curl -XGET 'vas.sparta.ktu.lt:9200/_cat/nodes?v&pretty'

Table 17: catAPI

Category	Query	cURL

Manage Pipelines	PUT /_ingest/pipeline/<<pipelineID>>
	GET /_ingest/pipeline/<<pipelineID>>
	GET /_ingest/pipeline
	DELETE /_ingest/pipeline/<<pipelineID>>
		#versioning example
		curl -X PUT "vas.sparta.ktu.lt:9200/_ingest/pipeline/<>?pretty" -H 'Content-Type: application/json' -d'
		{
		"description" : "my pipe does this",
		"version" : 24,
		"processors" : [
		{
		"set" : {
		"field": "names",
		"value": "bartholomew"
		}
		}
		]
		}

Simulate Pipelines	POST /_ingest/pipeline/<pipeline>/_simulate GET /_ingest/pipeline/<pipeline>/_simulate POST /_ingest/pipeline/_simulate GET /_ingest/pipeline/_simulate	

Table 18: Ingest APIs

Chapter 11 Annex 2: Layers

OTX – Opensource

60,267 hits

222.168.57.209#3#2#Malicious Host#CN##43.8800010681,125.322799683#3

_index

_score

_type

column1

column2

MISP

misp-*

1533 hits

Info:OSINT Trend Micro Exposes LURID APT Date:2011-09-22 timestamp:1432213272 uuid:555dcba2-bdd0-49d6-8c72-4e87950d210b Org:Kaunas University of Technology OrgC:CthulhuSPRL.be Source:MISP data:{ "type": "ip-dst", "IP address": "184.22.240.174", "comment": "" }, { "type": "ip-dst", "IP address": "184.22.251.12", "comment": "" }, { "type": "ip-dst", "IP address": "184.95.36.75", "comment": "" }, { "type": "ip-dst", "IP address": "204.12.197.70", "comment": "" }, { "type": "ip-dst", "IP address": "58.64.149.29", "comment": "" }, { "type": "ip-dst", "IP address": "106.123.126.151", "comment": "" }, { "type": "ip-dst", "IP address": "109.123.126.143", "comment": "" }, { "type": "ip-dst", "IP address": "109.123.126.151", "comment": "" }, { "type": "ip-dst", "IP address": "109.123.126.156", "comment": "" }, { "type": "ip-dst", "IP address": "109.123.126.157", "comment": "" }, { "type": "ip-dst", "IP address": "173.212.195.216", "comment": ""

ElasticFlow

556 hits

Aug 1, 2020 @ 08:26:00.000 - Mar 17, 2021 @ 09:27:05.372

_index

_score

_type

@timestamp

@version

agent. hostname

agent. id

agent. name
agent. type
agent. version
as. organization. name
client. as. number
client. as. organization. name
client. bytes
client. domain
client. geo. city_name
client. geo. country_iso_code
client. geo. country_name
client. geo. location
client. ip
client. packets
data
destination. as. number
destination. as. organization. name
destination. domain
destination. geo. city_name
destination. geo. country_iso_code
destination. geo. country_name
destination. geo. location
destination. ip
destination. port
ecs. version
event. category
event. dataset
event. duration
event. end
event. kind
event. module
event. severity
event. start
event. type
flow. client_rep_tags
flow. direction
flow. dst_mask_len

flow. dst_port_name
flow. dst_rep_tags
flow. input_ifname
flow. input_snmp
flow. next_hop
flow. output_ifname
flow. output_snmp
flow. rep_tags
flow. sampling_interval
flow. server_rep_tags
flow. service_name
flow. service_port
flow. src_mask_len
flow. src_port_name
flow. src_rep_tags
flow. tcp_flags
flow. tos
flow. traffic_locality
geo. city_name
geo. country_iso_code
geo. country_name
host. ip
host. name
log. level
message
netflow. dst_as
netflow. engine_id
netflow. engine_type
netflow. flow_records
netflow. flow_seq_num
netflow. in_bytes
netflow. in_pkts
netflow. sampling_algorithm
netflow. src_as
netflow. version
network. bytes
network. iana_number

network. packets
network. transport
network. type
server. as. number
server. as. organization. name
server. bytes
server. domain
server. geo. city_name
server. geo. country_iso_code
server. geo. country_name
server. geo. location
server. ip
server. packets
source. as. number
source. as. organization. name
source. bytes
source. domain
source. geo. city_name
source. geo. country_iso_code
source. geo. country_name
source. geo. location
source. ip
source. packets
source. port
Source_from
tags
uuid

Chapter 12 Annex 3: GEO IP

destination.geo.city_name
Kaunas
destination.geo.country_iso_code
LT
destination.geo.country_name
Republic of Lithuania
destination.geo.location
{
"lon": 23.9002,
"lat": 54.9002
}