Sharing and Automation for
Privacy Preserving Attack Neutralization

(H2020 833418)

D2.3.1 Visualisation requirements (M6)

Published by the SAPPAN Consortium

Dissemination Level: Public

H2020-SU-ICT-2018-2020 – Cybersecurity

Document control page

Document file: Deliverable 2.3.1
Document version: 1.1
Document owner: Christoph Müller (USTUTT)

Work package: WP2
Task: T2.3
Deliverable type: Report
Delivery month: M6
Document status: ☒ approved by the document owner for internal review
☒ approved for submission to the EC

Document History:

Version	Author(s)	Date	Summary of changes made
0.1	Tanja Blascheck (USTUTT), Christoph Müller (USTUTT), Michal Pavúk (USTUTT)	2019-04-29	First outline.
0.2	Tanja Blascheck (USTUTT), Christoph Müller (USTUTT), Michal Pavúk (USTUTT)	2019-10-27	Main parts finished.
0.3	Christoph Müller (USTUTT)	2019-10-29	Fixed issue with large display text being too short.
0.4	Michal Pavúk (MU)	2019-10-30	Partitioned the “Visualisation goals” section.
0.5	Christoph Müller (USTUTT)	2019-10-30	Subsections in section “Visualisation require-ments”; added paragraph on software and multi-display infrastructure.
0.6	Christoph Müller (USTUTT)	2019-10-30	Added details on the questionnaire used.
0.7	Tanja Blascheck (USTUTT), Christoph Müller (USTUTT)	2019-10-30	Partitioned the “Visualisation requirements” section.
0.8	Christoph Müller (USTUTT)	2019-10-30	Changed some subsections in “Visualisation goals” and added clarifying text.
0.9	Christoph Müller (USTUTT)	2019-10-30	Updated illustrations and added image credits.
0.10	Christoph Müller (USTUTT)	2019-10-31	Fixed issues brought up in Alexey's review.
0.11	Christoph Müller (USTUTT)	2019-10-31	Orthography check.
0.12	Christoph Müller (USTUTT)	2019-10-31	Added orthography and grammar fixes by Alexey and Tanja
1.0	Christoph Müller (USTUTT)	2019-10-31	Approved for submission to the EC.
1.1	Christoph Müller (USTUTT)	2019-10-31	Rewrite request from FSC.

Internal review history:

Reviewed by	Date	Summary of comments
Arthur Drichel (RWTH)	2019-10-29	Comments regarding spelling/grammar/punctuation.
Josef Niedermeier (HPE)	2019-10-30	Comments on HPE section.
Tomas Jirsik (MU)	2019-10-30	Technical content. Notes: 1. I would attach the survey to the deliverable. At least the questions (and if approved by partners, the responses as well, or some kind of their sta-tistical summary). Moreover, I believe that a brief

		description of the survey, it's intention, sectioning or the questions, etc. would be nice to include. 2. It is quite hard to distil the visualisation goals and requirements from the text. (Presentation in the form of a list with bolded names on the goals and requirements would be easier to read and get the content. Or the table with an overview.) 3. Further notes in comments in text.
Alexey Kirichenko (FSC)	2019-10-30	A number of typos fixed, a few fixes to statements, a few clarifications added to the FSC section. A couple of changes proposed and implemented by Christoph. A suggestion to add material about the questionnaire outcomes.

Executive summary

This document describes our efforts of collecting the expectations that security practitioners of four SAPPAN partners operating their own security operations centres (SOCs) have with regard to the visualisations implemented in the project.

The deliverable summarises the results of a questionnaire we designed to collect current practices in the SOCs and the general experience SOC analysts have with different kinds of data visualisation techniques. It also includes the findings from several site visits to the SOCs and discussions we had with the analysts there. From these sources, we derive the goals analysts in the SOC might pursue with visualisation. The most important ones we identified are the need to speed up decisions on false positives by providing the relevant contextual information for an alert in the presence of overwhelming amounts of data and conveying and correlating the temporal sequence of events. The deliverable also sheds light on the challenges we expect in achieving these goals, which are, most importantly, the sheer amount and the variety of data and the low level of adoption of any kind of visualisation in current SOCs.

Contents

1 Introduction

Handling cybersecurity incidents effectively requires assessing a large number of alerts based on an even larger amount of data, which need to be analysed to discern true incidents from false positives. This is one of the key areas where visualisation can excel, leveraging the capabilities of the human visual system to quickly find patterns and outliers. However, for cybersecurity operations visualisations are not yet sufficiently integrated. Currently, only simple charts like bar and line charts are used by analysts which are oftentimes integrated into web-based interactive displays. Therefore, we will rely on visual analytics (VA) methodologies that allow us to combine visual interfaces showing raw or aggregated data with interactive filtering and exploration capabilities. To define the specific goals and requirements we set out to achieve, this document contains results of a survey we conducted with security operations centre (SOC) operators from four project partners (CESNET, F-Secure, Hewlett Packard Enterprise, and Masaryk University), on-site visits and personal discussions with SOC operators. We define the areas, which we believe visualisation might be most useful and what security analysts hope to achieve when using a VA system. Namely, showing temporal data and conveying temporal relations between events, speeding up the handling of incidents by using dimensionality reduction of data, representing process trees to unveil hidden attacks, as well as a geographical representation of IP addresses. From these discussions we also extract the infrastructure of the SOC and tools commonly used during incident handling as well as common data (formats), machine learning methods applied, and prior experience security analysts have with interactive visualisation as well as typical incident handling workflows.

Some of the most common observations we extracted are the switching of different tools and sources of information, which are not integrated into a single solution, the manual documentation process, decisions that have to be made based on partial knowledge, the inability to rapidly distinguish normal form abnormal behaviour, as well as no natural point to obtain a global overview from the large amount of detailed data. Based on the results of this current state of the art of SOC operators, we define visualisation goals and requirements, some of which being the better communication to the public and customers, the development of a process tree to unveil hidden attacks, as well as the projection of high-dimensional vectors defining an attack to the 2D plane to reveal clusters of attacks that require similar handling to reduce response time.

2 State of the art

The findings described hereinafter stem from two sources: a questionnaire we handed to all the partners operating their own SOC and site visits to the SOCs of CESNET, F-Secure, Hewlett Packard Enterprise and Masaryk University as well as personal discussions with the SOC operators. In the visualisation questionnaire, we wanted to gain an overview of the security operations and currently used visualisation by the project partners in a structured way. Therefore, the questionnaire comprises a section with general information about the organisation and the size and structure of security operations. A second part focuses on the physical layout of the SOC, namely on the screen real-estate and use of large, high-resolution displays. The third part delves into the tools currently used by SOC analysts and how they are used to solve the most important types of incidents. We also asked the partners to specify what kind of data they regularly work with. The final part of the questionnaire evaluates the familiarity of SOC analysts with typical visualisation techniques ranging from standard (line, bar, ring, ...)

charts over uncertainty visualisation, timeline graphs, hierarchical visualisations and visualisations of correlations to techniques for high-dimensional data. Fig. 1 illustrates the sample images used in the questionnaire. The complete questionnaire can be found in the appendix of this document.

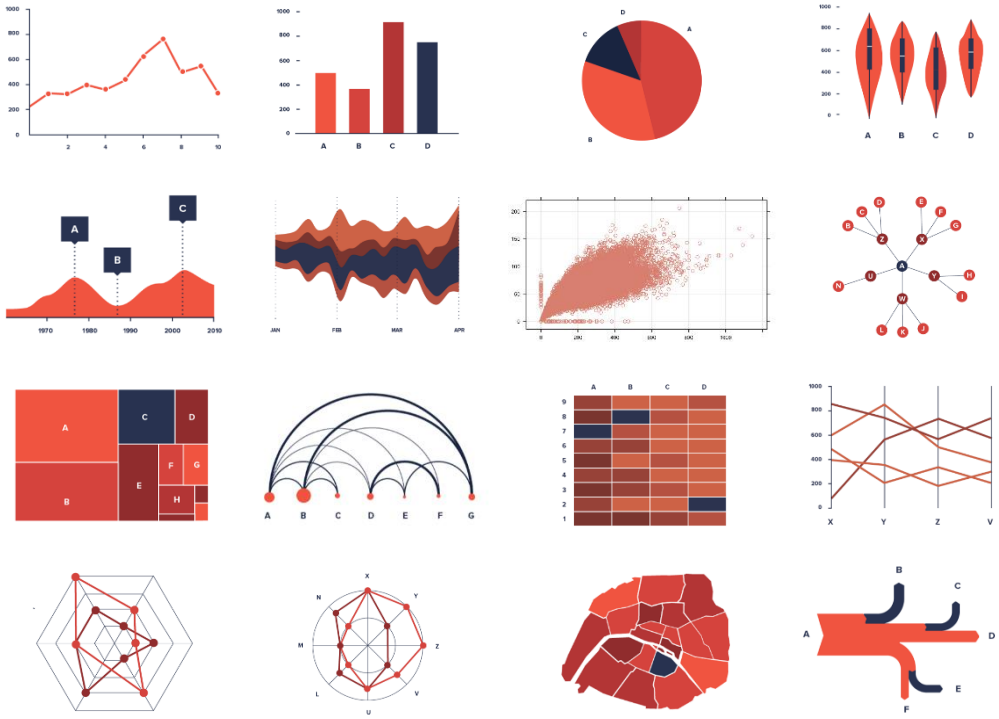

Fig. 1 Sample visualisations (from https://datavizproject.com) used to ask SOC agents about their familiarity with different kinds of visualisation.

The following section describes how visualisation is currently used in the SOCs by industrial and academic partners in the SAPPAN project. We also describe the typical process of incident handling as well as data and visualisations used during this process as demonstrated on the site visits. Afterwards, we summarise the results on how the partners use visualisations, interaction, as well as at which stages of incident handling visualisations are applied.

2.1 CESNET

CESNET as an association of Czech universities and the Czech Academy of Sciences is responsible for the development and operation of the national IT infrastructure for science, research and education. As such, it is in charge of the operation of the backbone network of the universities, but also pursuing research and development activities with respect to network technologies and applications. The CESNET Computer Emergency Response Team (CERT) is responsible for network monitoring, attack response and incident handling. Incident handling by CESNET CERT differs from the other project partners because many alerts it receives for endpoints are operated by the member organisations. Therefore, a significant share of the alerts are handled by finding the owner of the respective endpoint and re-directing them to the owner to solve the problem on their own.

Every one or two weeks CESNET CERT appoints one of its employees as the main incident handler responsible for the entire incident handling process, from reporting to recovery. During incident handling, they cooperate with system administrators and net-

work operators. After the administrator is informed about an incident, he or she is expected to resolve the problem on his/her own and report back to CESNET CERT. In more severe cases a specialist operator is dispatched, e.g., a backbone network admin, a member of the forensic analysis team, etc. The specialist then reports back to the main incident handler.

Despite the small number of team members, CESNET CERT manages to take care of a nation-wide infrastructure by leveraging a large variety of automated systems. The majority of network security incidents are escalated to the responsible parties, more precisely to the local administrator of the relevant network segments. Incident alerts are stored in an in-house system named *Mentat*. Alerts can originate either from internal monitoring or other collectors (*FTAS* – tool for flow and traffic analysis, network/server logs, honeypots, own FPGA-based hardware probes), or external partners sharing the alert using the platform *Warden*. *Warden* is a system for sharing security incidents, also developed by CESNET. Alerts are enriched using their IP reputation database.

While some of the tools deployed do provide basic visualisation functionality in the request tracking system *OTRS,* its use is not prevalent in actual workflows. Most of the data exploration and gathering is done using text-based search and filtering functions. Quantitative results of these can be visualised using simple bar charts, but the SOC analysts reported that such use cases are rare. So far, the only visualisation regularly used in the SOC is a schematic depiction of the state of fibre optic cables, which is, however, a visualisation out of the scope of the SAPPAN project.

2.2 F-Secure

F-Secure Corporation is a global provider of security services. Via its “Rapid Detection Center” (RDC), it offers customers a managed security service protecting against threats ranging from crimeware to targeted corporate cyberattacks. Defending F-Secure's own infrastructure is, therefore, only a small part of the activities of the RDC, with the main task being serving customers. For doing so, analysts at the RDC have to assess alerts generated by analytical engines applied to data from the sensors F-Secure deploys in the customers' infrastructure and then decide on whether those alerts are caused by serious incidents that must be reported to the customer. The reports made by the RDC staff need to be authored in such detail that the customer is able to understand and handle the incident.

RDC – F-Secure’s SOC – consists of individual offices as well as open-space offices with some having restricted access. The incident handlers mostly work on individual workstations with two to three displays with physical sizes ranging between 19 and 21 inches, with individual applications. Furthermore, large tiled displays and whiteboards are installed in the open-space offices. The tiled display is used for monitoring purposes and is a 3 × 2 configuration of 47-inch displays in the SOC room, but is not used permanently due to its high heat dissipation. Except of laptops and convertibles, no mobile devices are used in the SOC.

For incident handling and daily work, F-Secure uses *Atlassian JIRA* as their ticketing tool and *Kibana* on top of *Elastic Search* as their operational dashboard, which are also used for security information and event management. Knowledge management is done using *JIRA* and *Atlassian Confluence.* Software for network and endpoint monitoring (sensors) is developed by F-Secure itself.

Fig. 2 View of F-Secure's security operations centre.

The typical process for handling of an incident shown to us starts with a rule-based alert showing up on the *Kibana*-based dashboard. The alerts primarily originate from endpoint sensor data, but also from honeypot sensors and network sensors. Machine learning techniques at RDC are currently mostly used for finding the relevant contextual information to be shown to the analyst when working with an alert and for predicting significance of alerts. As a first step, an incident handler is assigned to classify the incident as false positive or whether it needs to be reported to the customer from which environment the alert originates. This process involves obtaining of contextual information by means of a variety of tools and approaches, including consulting open-source intelligence sources like *Virustotal* or even developer documentation like the *Microsoft Developer Network.* To improve this process, F-Secure also uses a variety of home-grown tools, for instance to investigate pairs of parent and child processes, which is useful because most attackers hide their traces by scattering their activities over multiple processes. Other tools include *CyberChef,* which assists the analysts in decoding many ways attackers obfuscate their code to avoid detection. Overall, the process involves switching among a variety of tools, most of which are browser-based and sometimes directly linked, until a final decision is made, which is logged along with an explanation of how the issue was resolved. If the decision confirms the incident, it is reported to the customer in question. The customer in turn needs to make sense of the textual description of the incident to further investigate or take corrective actions.

Handling of a specific incident by the SOC team is mostly done by a single analyst. To manage knowledge on how to handle incidents, analysts take personal digital notes, which are collected in a “log book”. A digital handler manual is available to analysts as a Wiki on *JIRA* or *Confluence*. Information is retrieved using keyword-based search. Analysts can add notes by themselves and notes are taken to be kept for later shifts. With respect to the context enrichment provided by machine learning techniques, the SOC agents can provide feedback by contacting the ML team or security analysts who write the rules.

From a data point of view, the input used in F-Secure's SOC comes mainly from endpoint (host) sensors and network traffic. The two most common routine tasks are de-

tection evaluation, which takes some minutes, and rule code development, which usually takes hours to perform. The two most common crisis tasks that the SOC handles are incident communication, which only requires a couple of minutes to handle, as well as proper detection evaluation (also takes only minutes). Approximately 50% of the incidents are false positives (ratio of 50/50). To minimise the number of false positives, ML scoring mechanisms and active whitelisting of clean behaviour are used. However, whitelisting is not always effective.

The visualisation currently used is limited to data visualisation provided by *Kibana.* Most prominent in the work sequence of the analysts are bar charts showing the distribution of events over time. The web-based interactive display of process trees can also be regarded as a visualisation in the broadest sense. The process tree is also the first potential area of improvement because it does not convey temporal sequences of events. In general, a proper visualisation of time allowing to correlate events was mentioned as an important area for improvement along with the ability to interactively retrieve the relevant contextual data at the right time. Finally, visualisation might be helpful for communicating incidents to F-Secure's customers, who currently receive only a textual report that might be easier to understand if visual elements would be included.

2.3 **Hewlett Packard Enterprise**

Hewlett Packard Enterprise (HPE) is a multi-national provider of information technology. Its department for cybersecurity operations is responsible for protecting the company against cyberthreats. The department operates two “Cyber Defence Centers” (CDC), one in the United States and one in Europe, providing their services 24/7 in a follow-the-sun way. HPE employs two kinds of security teams, the more relevant one for the SAPPAN project responding to automated alerts and the other one researching current threats without specific alerts (the “threat hunting team”). Besides relying on automated alerts, employees may report suspicious findings like potential phishing emails to the team. Handling of alerts is based on playbooks and extensive training of SOC analysts. Albeit playbooks play an important role in SOC operations, an estimate of 70% of the incidents cannot be expressed in playbooks, wherefore handling them heavily relies on the experience of the analysts. An important factor in the handling process is that decisions are almost always made on partial information.

HPE’s SOC is located in an isolated room comprising four rows of three workplaces for each analysts. The front wall is equipped with a 4 × 2 tiled display wall showing active alerts using ArcSight, HPE's security information and event management system, a histogram of event statistics, a digital message board and several browser windows. Furthermore, the tiled display wall is used for a video conference twice a day during the handover procedure from one of the CDCs to the other. However, HPE found that alerts displayed using standard desktop techniques on a large display are not salient enough to capture the analysts' attention, wherefore they have installed additional LEDs indicating important alerts by blinking lights in the room. The workplaces are organised so that the ones closest to the tiled display are manned by the analysts actually working on alerts, followed by people handling more abstract tasks like automating the response process and improving the tooling and by management staff like the information security manager. Each workplace is equipped with three 27-inch displays, accounting for the fact that the work of the analysts involves a lot of switching between different tools, which are not integrated into a single solution. Additional recurring challenges that the analysts reportedly face are the availability of a large amount of detailed data without a natural point to obtain a global overview and that decisions have to be made based on partial knowledge of the actual situation.

HPE already performs logging of activities in their SOC when analysts are using in-house tools, not for tracking analytical provenance as envisioned in the SAPPAN project, but for auditing and compliance reasons. In this context, the SOC needs to document who is responsible for the actions taken. The documentation process for SOC members themselves is mostly manual as it is a textual description of the incident and the resolution. Only a few aspects of the documentation are structured including MITRE-derived classifications, the temporal aspect of the incident and whether the incident needed to be reported according to GDPR regulations. The tool used to manage cases at HPE is *Atlassian JIRA.*

Visualisation is currently not widely used or, as one of HPE's employees coined it, "at the end, it all boils down to [the display of] a table". However, practitioners at HPE made some specific comments on what they want to achieve with visualisation and how such visualisations could look like. The first of their suggestions is the projection of the high-dimensional vectors defining an alert to the 2D plane, which ideally would reveal clusters of alerts that require similar handling and could be manifestations of the same attack that should be handled by a single analyst. The second suggestion involves creating a graphical depiction of the factors that lead to an alert, which would be automatically extracted from the definition of the rule of the alert. Such a visualisation could then be interactively refined following a VA approach in order to obtain contextual information for the alert, which is needed to decide whether it is a false positive or an actual incident. In addition, the site visit revealed a series of general requirements for visualisations, namely that it conveys temporal relations between events, allows for defining templates for alerts etc., which can be shared within the SOC, allows for interactive filtering for adding contextual information and ideally enables the SOC analyst to recognise and discriminate normal behaviour of a machine or network sensor from abnormal behaviour.

The overarching goals that HPE hopes visualisation can help to achieve are assessing the severity of an alert, prioritising the handling of alerts based on such assessment and reducing response times until corrective actions are taken as a result of this.

2.4 **Masaryk University**

CSIRT-MU is the certified cybersecurity team of Masaryk University. It is comprised of multiple groups of researchers and an Incident Handling Unit. The incident handlers use two 21-inch screens with a total resolution of 3840 × 1200 pixels. A large screen of 3.5 m width is also available, but rarely used as tools suitable for such a large screen estate are still in development. The majority of the tools used during the incident handling process are web-based. *Best Practical Request Tracker (RT)* is used as the ticketing system, *Flowmon Networks* as tooling for network monitoring and flow data collection and analysis as well as *Redmine* as an internal knowledge base system.

Incident alerts from e-mails and automated systems are stored in the ticketing system. These are then assigned to relevant queues, categories or directly to an incident handler. If the manual incident triage process decides that the incident is a security incident, the operator starts an investigation. The incident handlers reported that the mundane alerts (spam, auto-replies, verification of automatic reports) are handled within a couple of minutes. More severe incidents like a successful phishing campaign, cloud infrastructure abuse or ransomware campaigns may take multiple days to be completely resolved.

Handling of network-related incidents is centred around the *Flowmon Networks'* monitoring tooling (ADS, FMC, Dashboard). According to the incident handlers, adding a

better interlinking between visualisations provided by *Flowmon* and its filtering rules would provide a significant improvement of incident handling times. The rules in *Flowmon* tools are authored by hand and while CSIRT-MU does share its rules with a wider community, it does not use rules from others (aside from the ones directly from *Flowmon Networks*).

Fig. 3 Two of the workplaces in MU's SOC.

The R&D division of CSIRT-MU puts a lot of effort into research and development of tools for monitoring and tool-assisted incident handling. Projects focusing on visualisations are developed with the cybersecurity context in mind and are being integrated into incident handling processes of CSIRT-MU. The adoption rate, however, is still low and most visualisations are still in the development phase. While incident handlers are being consulted during the visualisation development cycle, it is difficult to make them use the visualisations, because they are used to a text-oriented analysis approach. Despite the slow adoption, incident handlers reported that they are interested in VA. Another area from the cybersecurity domain in which MU utilises visualisation is research. Techniques developed for researchers are often problem-specific and allow interactive data exploration and visual analysis. Some visualisations were created to address pitfalls of other visualisation tools or to directly support novel research.

2.5 Results of the questionnaire

The questionnaire we asked our project partners to fill in contains several sections asking for “demographics”, the working environment of the SOC team, tooling and data used as well as familiarity with visualisation and interaction techniques.

2.5.1 Organisation and workplaces

Overall, all of the partner organisations have more than 250 employees. They include public, academic and industrial institutions. The security operations are mostly organised in their own department with sizes ranging from small two-digit numbers to basically the whole organisation being a cybersecurity company. In most cases, the SOC services are an internal service for the organisation or its constituents with F-Secure being the main exception as a global provider of security products and services.

For the working environment, we see individual and open-space offices being used, the latter being more popular at the industrial partners. In the latter case, offices for security operations are in restricted areas with access control.

SOC analysts have at least two displays of at least 21" at their disposal, some organisations use three. These displays are typically used to organise data, i.e. the multi-display setup usually facilitates switching between different applications. If wall-sized displays are used, they are typically showing graphs for monitoring purposes. Mobile devices used in the SOCs are mostly the laptops of the team members, only one organisation also considers smartphones being a working device at their SOC.

2.5.2 Tools and data

The variety of tools used in the SOCs is large. All of the respondents have some kind of ticketing system in place with open-source solutions like *RT* and *OTRS* being popular at the academic partners while the industrial partners tend to use commercial products like *Atlassian JIRA.* The network monitoring solutions used vary from commercial solutions like *Flowmon* to in-house software. One of the partners reported the use of eight different tools for this purpose. Various solutions are also used for host-based monitoring, again with companies whose core business is security using their own technology. For the operational dashboard and security information management, the opposite is the case with industrial partners relying on off-the-shelf solutions like *HPE ArcSight, Kibana* and *Elastic Search* while CESNET, for instance, developed their own web-based SIEM system. On top of that, SOC analysts use a variety of other tools including source code management and issue tracking tools, mailing lists, custom web-based sharing platforms, *Office 365* and different types of Wikis and knowledge management platforms.

Knowledge management heavily relies on the operators' personal experience and personal notes in the academic organisations, the industrial partners have a stronger focus on the central, digital collection of knowledge. The documentation generated is typically semi-structured due to the variability of tasks being handled. The typical way of accessing the digitally collected knowledge is by keyword search.

For all organisations, events triggering an investigation by the SOC team are currently generated based on rules, though for some partners machine learning methods are used to provide input for the rules, such as event anomaly scores. Another application of machine learning is to automatically enrich the data provided with the alert with more context. The analysts can normally provide feedback by informally contacting the authors of the rules, only one respondent has integrated feedback directly into their SIEM system. None of the organisations provides means in their SIEM or operational dashboard to obtain contextual information except the full detail information stored during the generation of the alert, i.e. there is currently no support for VA. The data sources include network flows, IDS alerts, logs from network infrastructure, firewall logs, mail transport logs and system logs for all partners. All partners also rely on static information about the network and system infrastructure like network diagrams or host directories. Package captures, application logs and active scan results are only used by one of the partners. Another partner uses events from endpoints, such as file and registry operations, process creation events, module loads, etc.

2.5.3 SOC operations

The most important activity routinely performed in the SOC is the investigation of detection alerts, i.e. evaluating whether a report is an incident or a false positive. Besides

this common top task, the routine tasks reported greatly vary. They include developing new detection rules, performing proactive scanning for vulnerabilities, consulting users and handling spam. Actual crisis tasks are also different for the respondents and depend on the nature of the organisation. For instance, F-Secure as a provider of managed security services handles crises by communicating the issue to the customer. For CESNET, the most important crises are operational problems on the network level. The frequencies of such crisis tasks vary, but with a sufficiently large infrastructure being monitored by the SOC, they happen at a daily basis. Handling both, routine and crisis tasks, usually takes between minutes and hours. Only organisations that need to handle the incidents completely on their own rather than just delegating it to the responsible persons reported that some of the tasks take several days to complete.

The respondents generally did not know what the false positive rate of the detection methods is, with around 50:50 being the most specific number. The predominant measure to reduce false positives is whitelisting, which was at the same time described as a problem as it is not very effective.

We finally asked SOC analysts what the three incidents are that are most difficult to detect. Again, we received a variety of distinct responses with only data exfiltration being mentioned by multiple partners. The other incidents mentioned here are advanced persistent threats, malware spreading across the network, all problems that do not leave traces by interfering with network infrastructure, botnets, phishing and Metasploit Windows API calls.

2.5.4 Visualisation and interaction techniques

In order to assess the familiarity of SOC analysts with specific visualisation techniques, we asked them to rate their familiarity on a five-level rating scale with 0 indicating that the technique is unknown, 1 indicating that one heard about this technique, 2 indicating that the technique was used outside the SOC context, 3 indicating that the technique was sometimes used in the SOC and 4 indicating that the technique is regularly used in the SOC. Tab. 1 summarises the results. It is apparent that results are rather on the low side in general, reflecting the responses during the site visits that visualisation is not widely used in the first place. In fact, only one SOC indicated that they use timeline graphs on a regular basis. In general, more advanced information visualisation techniques conveying correlations and uncertainty like parallel coordinates, visualisations based on projections or violin plots were oftentimes barely known.

Visualisation technique	**Minimum**	**Median**	**Maximum**
Standard charts (line, bar, pareto, pie, ...)	2	2.5	3
Charts indicating confidence/uncertainty/ranges (box plot, violin plot, candlestick, ...)	1	1.5	3
Timelines/streamgraphs	1	2.5	4
Correlogram (autocorrelation plot for timeseries)	0	0.5	3
Scatterplots	1	2	3
Scatterplots of projected (PCA, UMAP, ...) high-dimensional data	0	0	1

Tree-based methods (node-link diagram, dendrogram, treemap, radial tree layouts, ...)	1	2	3
Graph-based methods (node-link diagram, adjacency matrices, chord diagram, ...)	1	1.5	2
Dense, pixel-based methods (e.g. heatmap)	1	2	3
Parallel coordinates	0	1	1
Radar chart/star chart/spider chart	1	1.5	2
Spatial/geographic visualisation (maps)	1	1.5	2
Sankey diagram	0	1	1

Tab. 1 Ranges and medians of responses to rate familiarity of the SOC team with visualisation techniques (higher is better).

We asked SOC members to rate typical interaction techniques in a similar manner to the visualisation techniques. Again, techniques that might be considered a standard in the visualisation community are oftentimes barely known. Furthermore, high ratings came specifically from the academic partners where researchers and the security team are in close contact and people tend to know each other.

Interaction technique	**Minimum**	**Median**	**Maximum**
Coordinate views/brushing & linking	0	0	3
Filtering (e.g. dynamic queries)	3	3.5	4
Visual Analytics (combination of interactive visualisation with automated data retrieval)	1	1.5	4
Semantic zoom	0	0	3
Focus & context views	0	1	3
Overview, zoom, filter, details on demand (information seeking mantra)	0	2.5	4

Tab. 2 Ranges and medians of responses to rate familiarity of the SOC team with interaction techniques (higher is better).

The results to the question where visualisation is typically used were widely inconclusive – most likely because visualisation is not used very often in the first place. In summary, most partners use it during detection, triage and analysis.

2.6 **Key findings**

From our questionnaire and direct communication with incident handlers during our visits to the SOCs, we can conclude that the use of visualisation is limited to simple data visualisation, mostly bar charts of volumes of events or other quantities and sometimes line charts. SOC analysts still prefer text-based interfaces to visual ones. This

point was often brought up during personal discussions. The interactions reflect this text-oriented mindset as all respondents replied that they mostly use keyword-based search. Experts were most familiar with the “filtering” interaction technique.

Unsurprisingly, the most known visualisation techniques in the questionnaire were the collection of “basic/standard charts” and “timeline-based charts”, because these are the ones most commonly deployed in standard monitoring systems. Tree-based visualisations, scatterplots and pixel-based techniques came in a close second. Advanced techniques like scatterplots of projected higher-dimensional data were completely foreign to the respondents. The respondents answered that the visualisations are mostly used during the initial “preparation” phase of the incident handling process. Corporate partners also make use of visualisation during the last phase of “post-incident analysis”, presumably when informing the upper management or customers about the incident and the issue resolution process. This assumption was confirmed during the site visits.

With respect to the tools currently in use, most partners use some kind of issue tracking system originally intended for software development, like *JIRA* in case of the industry partners and *Redmine* in case of Masaryk University, for documenting the activities of a SOC analyst. Management of security events uses different tooling ranging from a dedicated commercial SIEM like *HPE ArcSight* over *Kibana* on top of *Amazon S3* to custom in-house solutions in case of CESNET. If visualisation is available, it is mostly a part of these solutions and not used intensively. Many of the partners are aware of the fact that there is a lot of room for improvements with respect to visualisation with comments including “visualisation is not our strength” and “it's all complicated”.

The equipment used in all the visited SOCs is quite similar: all workstations of the SOC analysts have at least two displays, reflecting the above-mentioned approach of using a multitude of tools for handling one incident. In several instances, the machines used were laptops attached to a dock, which suffice the current needs as many applications used are web-based or otherwise remote. Interestingly, all of the SOCs are equipped with some kind of wall-mounted displays. At two of the locations, these displays are not used regularly due to their heat dissipation. Obviously, there is currently no useful overview visualisation that would justify the inconvenience. The overview display of events used at HPE is also rather suboptimal for large, high-resolution displays as *ArcSight* does not scale to it. The main use of the display during the site visit was video communication with HPEs second SOC. The only site using wall-mounted displays for situational awareness was CESNET, but not in the security context, but for monitoring their nation-wide backbone fibre network.

3 Visualisation goals

The paramount challenge of the SOC analysts of the project partners is dealing with a large amount of data they process and a large number of false alerts generated from this data, which needs to be filtered out by an analyst. While this application case seems to be a natural fit for the VA methodology, which combines visualisation techniques leveraging the visceral filtering capabilities of the human visual system and automated (machine learning) methods, the reality is that visualisation plays little to no role in day-to-day SOC operations, let alone an integrated VA solution being deployed. From our observations, there are two major reasons for that, the first being Rafael

Marty's "dichotomy of security visualization"[1] with security practitioners and visualisation practitioners being almost disjoint groups of people having difficulties in understanding the mutual interests. This results in research in security visualisation oftentimes addressing specific issues with specific solutions that are only marginally relevant for typical SOC operations. The other way around, visualisation researches often do not have access to the actual relevant data. The second issue we observed on all the site visits is the multitude of tools required to handle an incident. There is no single interface a SOC analyst interacts with, but jumping between browser tabs and specialised tools is omnipresent. To achieve holistic filtering, brushing, linking and feedback capabilities that is the idea behind VA, an enormous engineering effort to unify at least a significant part of the tools used would be required. However, being able to correlate data of different kinds and from different sources (like endpoint sensors and network sensors) was a goal almost all of the analysts explicitly expressed.

3.1 Contextual awareness

The industrial partners in particular need not only to correlate different kinds of data, but the right kind of data for the task, or in the words of one of the analysts they hope VA to provide "the relevant context at the relevant time". This indicates a more abstract goal to be achieved with visualisation: retaining the same informative value while working with data sets that evolve over time. One of the possible approaches to this problem is differential visualisation which - instead of visualising the state of the system in a single time frame - shows the differences between two or more time frames. As a specific goal with respect to the portrayal of time and sequences of events, the practitioners at F-Secure expressed the hope that a better visual representation of process trees would enable them to quicker understand what is happening and whether an attacker tries to hide traces or not.

3.2 Better overview

Another specific goal distilled from the site visits is speeding up the handling of incidents by assigning them to the right SOC agent. To achieve this goal, practitioners envision a data-driven clustering of alerts and embedding these clusters in a 2D scatterplot. The reasoning behind this is that each SOC agent will focus on a specific area of events, in which hopefully similar alerts will pop up that might belong to the same root issue or at least allow the analysts to work in a similar direction, thus decreasing the likelihood of time-consuming switches of cognitive context between the incidents. A potential issue with this goal is that the survey we conducted revealed that the required projection techniques are the least comprehensible and least known visualisation technique we asked for. This way, the approach might have the opposite effect of slowing down the handling process by requiring the operator to focus more on understanding the visualisation itself rather than the actual tasks. On a more abstract level, we understand the desire for this specific kind of visualisation as the wish for visualisations providing a better overview as most currently used visualisations are tied to the result of a query or simply showing data volumes etc., which per se do not provide an organised view on the whole data.

3.3 Enhanced communication

Although CESNET does not use many visualisations in general, they did acknowledge a need for better visualisation of geographical data stored in their *Mentat* system,

[1] Raffael Marty, *Applied Security Visualization,* 1st ed. Boston: Addison-Wesley Professional, 2010, p. 7.

namely IP addresses enriched using their reputation database. A geospatial visualisation would provide a better overview than a simple table, especially when it comes to real-time data. However, they also stated that such a visualisation would be most beneficial when communicating with the public about the scale at which they handle incidents rather than the incident handling process itself.

A similar application case was made by F-Secure, who do not want to address the general public, but their customers. The purpose of visualisation would be not only to improve the incident handling process but to better communicate the intricacies of the incident to non-technical personnel, often high management. The analysts acknowledged that the same visualisation can be used for both purposes.

4 Visualisation requirements

The following paragraphs summarise the requirements for a visualisation solution in SAPPAN we distilled from the questionnaire, the goals and the discussions with the SOC analysts. These requirements range from extensibility and integration, the use of appropriate metaphors, the support of current SOC workflows to leveraging the existing infrastructure of the SOC.

4.1 Extensibility and integration

To fulfil the promise of a comprehensive VA system fostering an explorative approach to data analysis and providing easy access to the relevant data at any time, the system would need to ingest and process all data available in the different kinds of security information and event management systems and in the variety of tools used by all the project partners. The development of such a system would be an enormous engineering effort that goes beyond the scope of this project. Nevertheless, the goal of the SAPPAN project is demonstrating the effectiveness of the VA methodology and of recent developments in the area of visualisation (like tracking analytical provenance and uncertainty visualisation) in a real-world environment. Therefore, the visual interface needs to be designed so that it can combine at least the most relevant data for the project partners (alerts from endpoint and network sensors) and that it is extensible for the case that a specific kind of data turns out to be relevant in later phases of the project.

In addition, the SAPPAN interface needs to provide at least limited feedback to the existing software infrastructure of an organisation. The reason for that is that if the prototype should demonstrate how VA can improve real-world SOC operations, it should not increase the need for switching tools by being another one. The minimum that should be possible is recording the final decision of an analyst on whether an alert is a false positive or an actual incident.

4.2 Appropriate metaphors

From a design point of view, the visualisations should meet the intended analysts where they currently are – even if this might be basic kinds of visualisations. It seems to be imperative to introduce and develop visual metaphors iteratively, for instance, starting by improving the aforementioned depiction of a process tree, integrating this into an interface that adds visual feedback (and possibly visual query options) to the traditional (text) search-based interfaces currently in use. The basic system can then be developed into a desired direction like adding further relevant data sources or new visualisation techniques like the aforementioned depiction of uncertainty. Likewise,

high-risk, experimental approaches like the 2D clustering of alerts for a better task assignment need to be an optional part of the interface that is not vital for the whole concept to work.

4.3 Support current workflows

Given the fact that we experienced frequent switching between tools and sources of information, we propose adding the ability to track analytical provenance to the system, which will keep track of actions taken and be able to replay them. Such a system will yield reproducible sequences of actions that can be used to improve user experience and may help to optimise the incident handling process. Provided a sufficiently large number of data sources to handle one or more kinds of incidents solely from the interface has been integrated, tracking analytical provenance might also be useful for the analysts themselves to keep better track of their action. However, because provenance tracking needs to record all of the interactions, it is necessary to implement such a system from the beginning.

The fact that a large fraction of the tools used at the partners' SOCs are web-based suggests implementing the user interface and the visualisations in the SAPPAN project in a browser as well. Following such an approach could also help to integrate existing software infrastructure by means of deep linking. However, browser-based solutions have the drawback of not scaling as well to the prevalent multi-screen scenarios as conventional desktop application, which can actively adapt to the screen layout available. As a requirement, any visualisation solution in SAPPAN should be designed with multi-screen setups in mind and support those to the best possible extent.

4.4 Leverage existing infrastructure

Finally, although large, high-resolution displays are available at all SOCs, there does not seem an imminent need for SAPPAN visualisation to support overview representations on such displays as this kind of visualisation is currently not used. However, we consider it reasonable to design visualisation and applications in a way that they scale to such displays to potentially exploit the benefits of an overview display that is already available.

5 Summary

This document summarises the current use of and expertise with respect to visualisation in the SOCs of four of the partners in the SAPPAN project. In general, it is safe to say that visualisation and VA are not widely (effectively not at all for the latter case) used in the field. However, we also found that SOC analysts generally do believe that visualisation and VA can solve some of their problems, the most pressing ones being able to filter the relevant context information for an alert from a vast amount of data, improving the understanding of temporal sequences of events and alerts as well as being able to communicate findings to managers and customers. Given the finding that current workflows at SOCs are very different from what visualisation and VA researchers ideally imagine them to be, we conclude that great care must be taken to meet the SOC analysts where they are when designing visualisations and to link new tools to existing ones. The latter is even more important as we cannot expect the SAPPAN user interface to subsume all functionality required by SOC analysts to do their daily work given the multitude of tools used right now even at a single SOC.