Assessing a Decision Support Tool for SOC Analysts

JASSIM HAPPA, Information Security Group, Royal Holloway, University of London, UK
IOANNIS AGRAFIOTIS, Department of Computer Science, University of Oxford, UK
MARTIN HELMHOUT, CISO office, Royal Philips, NL
THOMAS BASHFORD-ROGERS, Department of Computer Science and Creative Technologies, University of the West of England, Bristol, UK
MICHAEL GOLDSMITH, Department of Computer Science, University of Oxford, UK
SADIE CREESE, Department of Computer Science, University of Oxford, UK

It is difficult to discern real-world consequences of attacks on an enterprise when investigating network-centric data alone. In recent years, many tools have been developed to help understand attacks using visualization, but few aim to predict real-world consequences. We have developed a visualization tool that aims to improve decision support during attacks in Security Operation Centres (SOCs). Our tool visualizes propagation of risks from sensor alert data to Business Process (BP) tasks. This is an important capability gap present in many SOCs today as most threat detection tools are technology-centric. In this paper we present a user study that assesses our tool's usability and ability to support the analyst. Ten analysts from seven SOCs performed carefully designed tasks related to understanding risks and recovery decision-making. The study was conducted in laboratory conditions with simulated attacks and used a mixed-method approach to collect data from questionnaires, eye tracking and semi-structured interviews. Our findings suggest that relating business tasks to network asset in visualizations can help analysts prioritise response strategies. Finally, our paper also provides an in-depth discussion on user studies conducted with SOC analysts more generally, including lessons learnt, recommendations and a critique of our own study.

CCS Concepts: • **Human-centered computing → Empirical studies in visualization**; *Empirical studies in interaction design*; *Visual analytics*; • **Security and privacy → Intrusion detection systems**; *Network security*;

Additional Key Words and Phrases: Business Process Modeling and Notation, Decision Support, Cyber Security, Assessment, Intrusion Detection Systems, Situational Awareness, Usability, User Study

ACM Reference Format:
Jassim Happa, Ioannis Agrafiotis, Martin Helmhout, Thomas Bashford-Rogers, Michael Goldsmith, and Sadie Creese. 2020. Assessing a Decision Support Tool for SOC Analysts. *Digit. Threat. Res. Pract.* 9, 4, Article 39 (July 2020), 36 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

In recent years, many visualization methods and tools have emerged for security and situational awareness applications. These are often designed to make network-traffic patterns and impact of

Authors' addresses: Jassim Happa, Information Security Group, Royal Holloway, University of London, UK; Ioannis Agrafiotis, Department of Computer Science, University of Oxford, UK; Martin Helmhout, CISO office, Royal Philips, NL; Thomas Bashford-Rogers, Department of Computer Science and Creative Technologies, University of the West of England, Bristol, UK; Michael Goldsmith, Department of Computer Science, University of Oxford, UK; Sadie Creese, Department of Computer Science, University of Oxford, UK, firstname.surname@cs.ox.ac.uk.

2576-5337/2020/07-ART39 $15.00
https://doi.org/0000001.0000001

attacks tangible to analysts [1]. Generally, these facilitate understanding particular elements of attacks. However, good situational awareness is difficult to achieve in Security Operation Centres (SOCs) as Intrusion Detection Systems (IDSs) are often flooded with false positives. Situational awareness is a military term, often used in SOCs to denote the perception and comprehension of the threat landscape in order to make well-informed decisions [2]. Analysts do not always possess enough insight about the monitored organisation to be able to formulate reasoning of wider attack consequences or harms across an enterprise. The full potential of attack ramifications may therefore not be understood until a SOC manager has escalated alerts to someone who can form a meaningful judgement at the enterprise level. This leads to time wasted when investigating attacks. This time could instead have been better spent prioritising response options. We developed a tool, *CyberVis* [3], designed to reduce time needed to make an incident response decision. In this paper, we examine how SOC analysts respond to such a tool in laboratory conditions.

Our tool visualizes propagation of risks from attacks by relating *Business Processes* (BPs) [4] (a formal description of organisation day-to-day tasks, akin to flow-charts) to *machine assets* (e.g. servers, clients and Bring-Your-Own-Devices (BOYDs)). The tool maps alerts on assets to enterprise critical tasks. It uses existing industry standards such as IDS and Anti-Virus (AV) alerts (e.g. Snort [5], Nagios [6] and ClamAV [7]), traditional computer network diagram icons and Business Process Modeling and Notation (BPMN) [4] to help form an easily-understood representation of the current situation. The tool is designed with the Human Visual System (HVS) in mind to make use of key principles from usability and visual-perception literature [8–10]. Severe attacks are visualized as salient objects, and the tool minimised navigation necessary to identify attack causes using graph abstraction. It was conceived as a decision-support tool for analysts and business owners alike. This study focuses on SOC-analyst users to limit its complexity and because we envisage SOC analysts as primary users. A large body of work exists in security visualization, see Section 2, but no works assess visualizations that predict mission risks to organisations based on attack data.

1.1 Paper Contributions

In this paper we present a user study and the methodology for the assessment of our visualization tool. Our assessment takes the form of a user study with ten SOC analysts from seven different SOCs. The study focuses on assessing the visualization component and usability of our tool. We use a mixed-method approach with two questionnaires (before and after using the tool), semi-structured interviews supported by eye tracking and video captures. The contributions of this paper are:

(1) a mixed-method user-study with domain experts (SOC analysts) as participants;
(2) an in-depth discussion on lessons learnt and recommendations from this study.

As no similar tool exists, the **value of this study is in demonstrating the feasibility of risk propagation visualizations for SOC analysts. In particular, showing that SOC analysts can understand uses of BPMN to obtain situational awareness and prioritise response strategies (for synthetic scenarios). Finally, we also provide an in-depth discussion on lessons learnt and recommendations for these types of user studies**.

Our aim is not to assess how well our visualization approach operates in production environments or against other visualization techniques, but instead whether the uses of visualization tools that aim to link network-level activities with business-process activities is helpful for SOC analysts. We believe it is necessary examine the feasibility of our approach first, prior to investigating its effectiveness in production environments (including against other visualization methods).

The paper is structured as follows: Section 2 summarises related work. Section 3 overviews the tool. Section 4, presents the study's motivation and design. Section 5 overviews the results, including key observations and feedback from participants, while Section 6 presents an in-depth discussion on the results, recommendations and future work. Section 7 concludes the paper.

2 BACKGROUND

Large SOCs process millions of IDS alerts per day. Visual analytics can help identify attack patterns. Obtaining a complete picture however, is rarely possible. Many works explore plotting of network topologies, traffic patterns, payload characteristics and event-logs [11–21], but no attempts have been made to deliver situation awareness to support incident-handling.

2.1 Security Visualizations

Visualizations can be classed as *data models* (mathematical abstractions such as time series, scatter plots, parallel coordinate plots etc.) or *semantic models* (visualizations that create reasoning structures from raw data). The former often include network statistics and graphs [22–24], geographic map overlays [25], plotting of activities (e.g., time series, histograms, parallel coordinate plots), while the latter presents some type of reasoning system as a visualization. Conti [23] and Marty [22] provide detailed discussions about the security-visualization field, and a large gallery of visualizations can be found on the SecViz.org website [24]. Examining the literature, we can see that common practices include making use of:

- **Abstraction**: Graphs and hierarchies are often used to abstract out more complex subsystems.
- **Colours**: Data type (e.g. TCP traffic is a different colour to UDP traffic) or severity levels.
- **Velocity (or motion)**: Intensity or freshness.
- **Location (on screen)**: Some form of unique ID (e.g. IP address, port number, real-world coordinates as two items cannot occupy the same pixel without overlapping each other).
- **Opacity/time on screen**: Freshness.
- **Shapes**: Data type (e.g. subnet, host or connection).
- **Sizes**: Amount of data belonging to the same category.

Data-model visualizations for network traffic analysis include Rumint [15] and Wireshark [26], while other tools analyse network usage patterns such as Best et al. [14], Kim et al. [16], Lau [17], Liao et al.[19]. Other approaches include attempting to understand the potential wider impact of attacks on network assets visualized by Chu et al. [20], and intrusion-detection event-correlations visualized by Rasmussen et al. [21] and Yelizarov et al. [18]. IDS Rainstorm [27] is an example of IDS pattern visualization over a 24h period. The commercial Arcsight tool [28] also provides mapping between event alerts, source IPs and business role.

Example semantic-model visualizations include: Tenable 3D [29] visualizes a network topology from vulnerability scans, encoding information about vulnerabilities, missing patches, firewalls, IDS alerts etc. SecureScope [30] addresses business impact of attacks by mapping clusters of potentially malicious network activity to business roles or organisational units. To our knowledge, no existing approach specifically focuses on the mapping of attacks to enterprise-critical tasks, and no decision-support tools facilitate real-time risk decision making from such a mapping. Finally, our tool aims to provide decision support through a risk propagation logic. The tool's requirements and specifications were derived from present day SOC capability gaps, as several of the authors have worked with and in SOCs in the past. Instead of focusing on a technology-centric view of intrusion detection, the authors developed requirements and specifications that have a business operation-centric view – often absent in SOCs. As mentioned, the purpose of this paper is to assess whether such an approach can be valuable in operational environments.

2.2 Visualization Assessments

A vast body of work exists on visualization assessment. Bertini [31] for instance overviews a list of more than 50 examples from the information-visualization community that provide insight into: component/system level evaluation, low-level components/perceptual studies, longitudinal

studies, case studies, metrics, benchmarks, model-based evaluation, frameworks, taxonomies, novel evaluation methodologies, non-conventional methods/parameters and reviews. Plaisant [32] outlines many of the key challenges in conducting evaluations and mentions how four thematic areas of evaluation exist:

- **Controlled experiments comparing design elements**. These studies compare specific widgets or compare mappings of information to graphical display. Examples include: Ahlberg et al. [33] and Irani et al. [34].
- **Usability evaluation of a tool**. These studies aim to provide feedback on the problems users encountered with a tool and show how designers went on to refine the design. Examples include: Sutcliffe et al. [35] and Byrd et al. [36].
- **Controlled experiments comparing two or more techniques**. These studies aim to compare a novel technique with the state of the art. An example includes the work by Plaisant et al. [37].
- **Case studies of tools in realistic settings**. These investigate users in their natural environment performing real tasks, showing a tool's feasibility and in-context usefulness. These can be time consuming to conduct, and their findings may not be reproducible or possible to generalise [38].

More recently, Staheli et al. [39] surveyed and categorised evaluation metrics, components, and techniques that have been utilised in the decade prior to its publication in the security visualization literature from 130 publications. They highlight common key techniques, which include *critique* (holistic assessment of a human reading and interpretation), *co-creation* (participants aiding the design of the visualization), *inspection* (informal methods in which experts review an interface to determine adherence to a set of best practices in relevant domains), *interviews* (structured or semi-structured), *usability testing* (measurement of effectiveness, efficiency and satisfaction), *questionnaires*, *simulation* (use of laboratory condition datasets as opposed to live data), *interface implementation* (extracting user interaction information), *psycho-physiological measurements* (indicative of a cognitive state, such as attention, cognition, or emotional response), *automated image analysis* (analysis of visual characteristics such as consistency of rendering), *application performance testing* (analysis of system load or response times).

Johnson and Onwuegbuzie [40] argue that mixed-method approaches are becoming increasingly important in modern research, especially for use case studies and usability evaluation for expert domain users. This is because it is not straightforward to measure why an participant makes an action in the way they did from quantitative data alone. It is necessary to ask participants directly, and have them reflect on their actions.

3 TOOL OVERVIEW

In this section we establish tool requirements, design, implementation and usability principles. The requirements of this tool include: real-time performance; showing areas of effect within an enterprise across the network and on the BPs; making use of saliency to guide users to see the most severely affected assets first; using best of breed standards (including Snort [5], Nagios [6] and ClamAV [7], BPMN 2.0 [4] as well as common network icons); using abstraction to facilitate ease of navigation; remaining consistent in both visuals and navigation; supporting users of both high and low technical proficiency, and finally; reflect uncertainty as threat intelligence may be from sources of varying provenance. Screenshots can be found in Section 4 and in the Supplementary Material at the end of this paper.

The tool was implemented in Java [41], MySQL [42] (intended as a cache) and Java OpenGL (JOGL) [43] and can run on Windows, Linux and Mac. It takes three types of file configurations: a BPMN XML file (BPMN 2.0), a network topology file, and a service-to-network mapping XML file. IDS and AV events are parsed to a local database via XML events.

3.1 Scene Layout

The tool's design takes inspiration from three bodies of work: 1) Bertin's work [9] aided in development of *visual variables*. 2) Shneiderman's work on user interface design [8] helped improve usability. 3) Itti and Koch's work on visual saliency maps [10] help guide users' vision to look at important objects at any time.

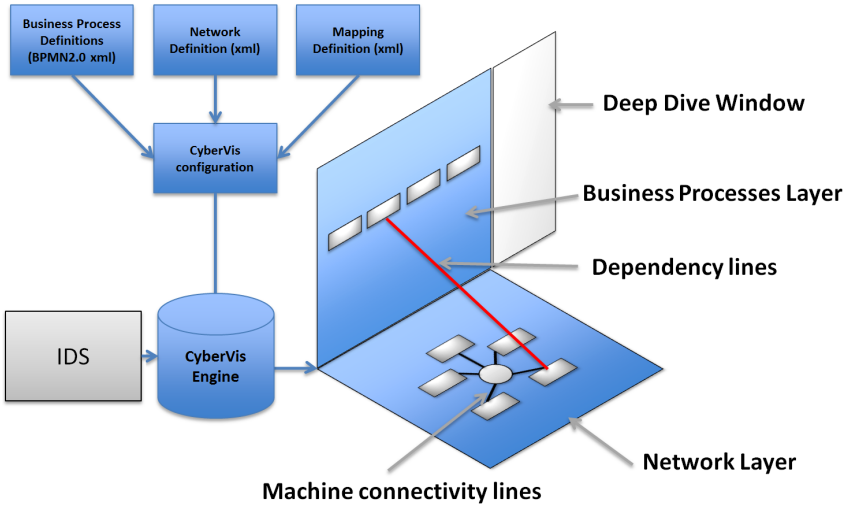


Fig. 1. Anatomy of the tool. Visuals are drawn in two main windows: a viewport (OpenGL) to render virtual space, and a deep dive window (Swing) to view and query text data in detail.

The visuals consist of three main constituent components: 1) **Network layer**: a graph in the virtual scene drawn on the XZ plane and represents the network assets in the enterprise, 2) **Business Process layer**: a graph in the virtual scene drawn on the XY plane that represents the BPs of the enterprise, 3) **Deep dive window**: a side window enabling textual data to be displayed.

The layout of the scene is computed during startup based on configuration input, see Figure 1. The tool then shows a 3D graph of the network and BPs monitored. The XZ plane holds the network (as a stacked star) topology, while the BPs are drawn on the XY plane (as boxes that can collapse and expand). The decision for drawing the scene in 3D was made to enable dependencies to be drawn between planes, i.e. so users can click either a BP element or network element and quickly identify how they depend on each other. Additional information is also displayed in a Heads-Up Display (HUD), such as time, selected item and an alert list.

3.2 Viewport Window

The viewport window renders a 3D scene populated with 3D-modelled objects. Each enterprise has its own set of subnets, which are comprised of machines that are clustered together (e.g., all machines affected with low-severity alerts are bundled to form one machine displaying appropriate severity information and a number to show how many assets of this class is affected). A palette of desaturated (non-vivid) colours has been chosen to texture the icons. The use of desaturated colours with only subtle warm and cold tones is used to lessen the saliency of these icons, see Figure 2. By employing shades of grey, white and black we aim to direct the user's visual attention to important elements of the scene at any given time. As alerts are revealed using salient colours, shapes and motion, they are more likely of being detected by the user.

3.2.1 Viewing and Navigation. The user of the tool expand aspects of the network or enterprise that is of interest to them after having seen an increase in risk colours at the most abstract level. This way, the screen is only comprised of content the user currently needs to be aware of (or has explicitly specified to see). The intention is to allow for the user to obtain coarse information fast using abstraction, but also to investigate aspects of interest in detail; whether it is the network layer or enterprise layer. Objects within each layer have clickable icons, boxes (platforms for network icons that belong in the same domain) and lines that display asset and mission task dependencies, see Figure 1. Once an object is clicked, textual data about it can be found in the deep dive window. Any unknown objects are represented with a question mark cube (alerts referring to unknown nodes on the network). Examples of such asset can include laptops or BOYDs.

Exploratory navigation is a key feature, and serves to keep the user engaged in order to not miss an attack. There are two main layers the user can navigate through: the *Network layer* (XZ-plane) and the *Business Process layer* (XY-plane). Users start at the top level of the Network layer and can swap to the BPs view at any time, or investigate a BP lane or the subnet of an enterprise. The basic control scheme is the same at any time. The mouse moves viewing direction and can click items, while the keyboard is optional. A single left-click selects a scene icon and displays its relevant information in the *deep dive* window, see Section 3.3. Only one icon can be selected at a time. A double-click allows for icons highest level to be expanded. If an organisation icon is double-clicked in the network layer, the user expands the subnet level of the network layer. Double-clicking a BP box opens up the set of tasks needed to complete a BP. Details about selected machines or BP tasks appear in the deep dive window once they are single-clicked. Our abstraction-approach aims to display pertinent information first, while removing unnecessary complexity for the viewer.

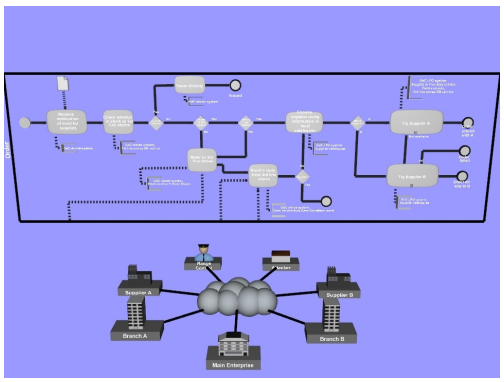
3.2.2 Presenting Alerts. Alerts draw the user's attention to risk. Unlike many data visualizations, we make use of a semantic structure to compute risk propagation. The risk is propagated up to the processes by assuming that any task dependent upon a piece of asset currently exposed to a potential attack carries the same level of alert. We then calculate the ability of the BP to terminate successfully. The red (high), orange (medium), yellow (low) colouring scheme was chosen as it is commonly found in SOCs today. We introduced the purple to signify *uncertainty* (i.e. machine or task that could be affected due to proximity). The alert colouring scheme is designed with four main principles in mind, to show: *when new alerts arrive* (i.e. spheres hovering upper right above an icon); *how many alerts exist* (i.e. cylinders appearing next to an icon); *dependencies between network and business tasks* (i.e. lines connecting icons); and *severity of alerts on nodes themselves* (i.e. colours are added to icon surfaces themselves). Our choice to not use other vivid colours (such as green, blue or white) was to avoid having analysts conflate the purpose of those colours. For instance, green may be interpreted as 'all clear', when, in reality, this is a difficult property to guarantee.

3.3 Deep Dive Window

The purpose of the deep dive window is to enable users to access detailed data about alerts in text form. The window allows viewing of text information about the currently selected object in the viewport window. When an icon is left-clicked, all machines relating to that icon are listed in a Table (for instance: an organisation, subnet or group of machines), and includes machine name, IP address and active alert count. If a machine is deemed to be cleaned, analysts can right-click that machine in the deep dive window, select "*Device cleaned*" and the whole risk propagation is recalculated. We have a separate forensic mode that allows for playback of past events using a series of filters which displays events according to user-defined parameters.

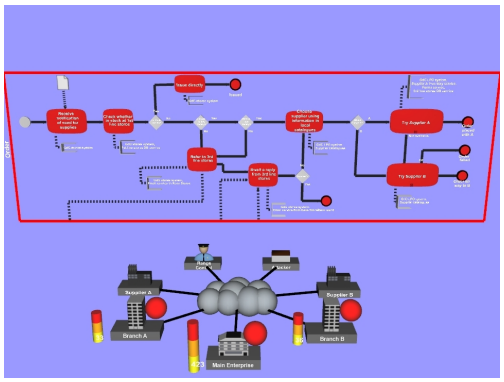
4 MOTIVATION AND STUDY DESIGN

Prior to designing our study, we used saliency maps [10] for ten views of the tool to identify likely viewing patterns of potential users. One of these views can be seen in Figure 2. We hypothesised that our colouring schemes, use of sharp edges and motions are appropriate from a saliency perspective. We also had a variety of informal trials and demonstrations with industry domain experts. The feedback they provided supported our hypothesis. However, using saliency maps and informal feedback alone is not sufficient. The use of saliency maps does not inform us about the tool's value and usability to SOC analysts. We therefore deemed it necessary to conduct an in-depth assessment in laboratory conditions to investigate whether our approach can support real SOC analysts' decision making and identify which features should be refined further.



(a) No Alerts

(b) Saliency Map - No alerts



(c) Alerts

(d) Saliency Map - Alerts

Fig. 2. Examples of how colours affect saliency [10]. In this example, we see red is more salient than black in our scene, and the addition of coloured shapes (spheres and cylinders) also have a saliency impact.

A major challenge in our research involves demonstrating decision support capabilities and identifying whether business processes can help decision-making. We therefore opted for a mixed-method design of our study to allow behaviour, reflection and opinions to make up our assessment. This helps us **determine whether analysts can understand uses of BPMN for SOC tasks, and if yes, whether visualizations that link BPMN and network asset alerts be used to offer better situational awareness in SOCs**? Note that we do not claim that SOCs will benefit from BPMNs as we made use of a synthetic scenario with controlled outcomes and attacks.

4.1 Study Overview

Our primary focus for this study was to get a first-impressions response (in the form of a cross-sectional study) from actual SOC analysts and obtain their usability and expert-opinion feedback. User tasks were designed to be as close to how we envisage the tool being used in a real-world SOC. Our assessment is based on: how analysts respond to given tasks; how they make their decisions and prioritise their actions using the tool; and their feedback about the tool itself.

The purpose of the study is to identify whether our prototype is able to facilitate decision-making with new modes of insight for SOC analysts. We propose a mixed-method methodology in order to understand the patterns of viewing and behaviour of analysts, but also make use of interviewing to be able to ask analysts why certain behaviour and viewing patterns were made in the first place. This follows several of the methodologies described by Staheli et al. [39]. The motivation for using several data collection methods in our study was because we had a limited number of expert domain users available, and thus wanted to collected more types of data to aid our investigation. Key research questions include:

- **Do the navigation and viewing patterns of SOC analysts match our intentions?** Our hypothesis is that changes in colour, shapes and motion should drive users to look at new alerts as they come in, in an abstraction-based manner.
- **Are analysts able to identify key concerns in situations using the tool?** Our hypothesis is that analysts will be able to identify key issues in a simulated setting and be able to suggest how to resolve those issues.
- **Is our tool able to tackle any challenges that SOC analysts face in real SOCs today?** This is expert-opinion feedback that we aim to obtain through interview and questionnaire answers.
- **Are any features considered useful (and not useful) by analysts?** This is also expert-opinion feedback that we aim to obtain through interview and questionnaire answers.

In the interest of clarity, we explicitly rule out the following, related research topics:

- *whether existing works by Bertin, Shneiderman and Itti & Koch function well together* in visualization tools, and assume this to be true.
- *whether our approach can function in production environments with actual attacks.* We deem it scientifically more important to remove compounding factors in our analysis and instead focus on conducting our research in laboratory conditions. This means our study makes use of synthetic data and takes place in our own offices, outside a SOC. We designed two scenarios for our experiment: a military scenario for training purposes, and a commercial scenario for the main assessment phase of the user study. The configuration of the organisations and how the attacks executed are detailed in Sections 4.3 and Section 4.4.
- *whether our approach performs better than other visualizations.* Before we can assess the tool's effectiveness against other tools, we need to make sure the novel usability and visualization concepts are valid for decision-making support in the first place. As no visualization today specifically focuses on the mapping of attacks to enterprise-critical tasks, this prevents us from doing a controlled experiment comparing our visualization against other tools or techniques (as described by Plaisant [32]).

4.1.1 Procedure. The four main parts of each user study session were (in order): **Introduction**, **Training Period**, **Main Scenario**, and finally **Reflection**, see Figure 3. The Consent Form can be viewed in Supplementary Material (SM) A. Questions posed in the Demographics Questionnaire can be seen in SM B. It includes questions about the analysts themselves, visualization preferences, current practices and standards in their SOC, and opinions about the state of tools and features available today. A transcript and screenshot from the video tutorial can be found in SM C. Details

about the Training and Main Scenario can be found in SM D and E respectively. SM F lists all the training scenario tasks that the participants had to complete during the training. A second questionnaire captured quantitative data about the participant's opinion in SM G. Finally, SM H outlines the key questions asked in the Reflections interviews.

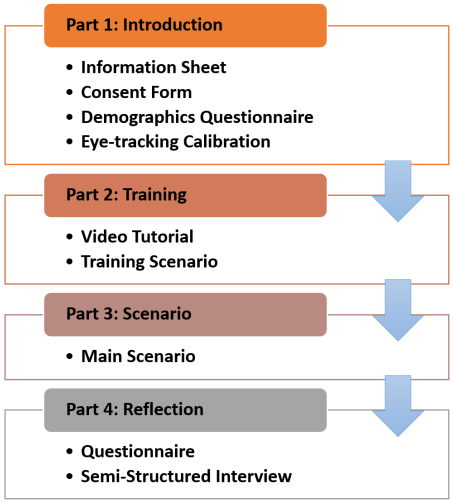


Fig. 3. Participant pipeline: the four main parts of the assessment were, in order: **Introduction**, **Training Period**, **Main Scenario**, and finally **Reflection**.

During each user study session, two experimenters and one participant were involved. The same two experimenters were present for all sessions, and had the same roles. The first experimenter led the session. He explained the tasks at hand and the agenda for the assessment. The second experimenter focused on the data collection at all times (notes, video, audio and eye tracking). Both experimenters made observations (as notes) about participants' behaviour during the Training Scenario and the Main Scenario, and asked additional questions from these observations during a Reflection part (in addition to a predefined list of questions).

It should be noted that the experimenters are also the developers of the tool. The appropriateness of doing so was carefully deliberated in the design of the study. We ultimately deemed it both appropriate and necessary to do so because: 1) the participants may have technical questions that only the developers can answer, 2) the developer could ask about unexpected, observed user behaviour, 3) a double-blind experimenter may fail to communicate nuanced elements in intention behind the design of the tool, especially for open-ended sessions such as the training scenario and the reflection portion of the study, 4) unexpected events (e.g. bugs) may happen during the user-study session in experimental tools, and the developers would be best suited to handle them (no such events occurred in our case). All participants were made aware that the experimenters are also the developers of the tool.

To reduce the risk of bias, the experimenters asked participants at the start of the study session to provide their honest and complete opinion, including any and all positive and negative feedback. There was no direct benefit or reward for the analyst to partake in the experiment. This was by design to ensure feedback was honest. Finally, prior to the ten participants, we ran two pilots with two other analysts to identify where potential bias might arise. The pilots allowed us to rehearse the execution of the study, so the actual sessions could run smoothly.

4.1.2 Recruitment. For our study, we believe with a smaller sample set using domain experts is more valuable than a larger pool of general population participants. This made recruitment challenging, particularly given the small population of SOC analysts in the UK. Using domain experts allowed us to ask participants about how our tool compares (opinion-wise) to the tools they typically have at their disposal, in order to – if only informally – ascertain some indication about the tool's value in a real SOC. Convenience sampling was used for recruitment of SOC analysts and was done through contacts of the university. All participants were required to have at least one-year experience in SOCs.

4.2 Introduction

The introduction (approx. 10 min.) consisted of an information sheet and consent form; a questionnaire about demography and SOC practices, and; eye-tracker calibration. This part provided introductory instructions to the participants and general information about the study. Participants were asked sign in a consent form, a questionnaire, and go through eye tracker calibration. In the questionnaire they were asked questions related to the SOC analyst's background, to common practises today and the state of the art. After filling in the questionnaire, participants went through an eye tracker calibration. During the calibration process we obtained a profile based of the participant's facial features to maximise eye tracking accuracy. This profile was deleted after the study completed. Participants were also informed that the experimenters were also the developers of the tool.

4.3 Training Period

The training period (approx. 35 min.) allowed participants to familiarise themselves with the tool and learn about the main features of the tool through a tutorial video and a series of tasks provided to complete in a training scenario instance of the tool. The Training Period part consisted of: **a video tutorial** detailing core the tool features, see SM C, and **hands-on Training Scenario** with a series of tasks to complete to obtain a first-hand experience of using the tool, before the Main Scenario). The tutorial video guided analysts through several different instantiations of the tool, illustrated the capabilities of the tool and explained the semantic structure to facilitate reasoning within the tool. The video lasted for 6 minutes and detailed the use of colours, the BPs, the deep dive window, the lines connecting BPs to the network. A separate training dataset (incl. topology, alerts and BPs) was created to gently introduce participants to the tool. The Training Scenario had a military-inspired network and organisation structure and consisted of three main organisations, a Main Operating Base and two Forward Operating Bases. The participants performed tasks related to navigation and interpretation of alerts. We did not assess their ability to perform these tasks during training, but made note of questions they raised as first-time users. The scenario began with straightforward navigation tasks, and later moved on to participant's understanding of interpretation of events. If their understanding was different than what was outlined in the tutorial video (with a different scenario), the experimenter would explain how to interpret the events. We answered any questions they had and let them use the tool to the point they felt comfortable with it.

4.4 Main Scenario

The Main Scenario (approx. 15 min.) consisted of a series of escalating (severity and complexity) attacks, see Figure 4 on a commercial enterprise scenario. This dataset was of a different set of alerts, network topology and infrastructure and BPs to that of the Training Scenario. The participants had to identify what attacks were happening and answer what could be done about it. At this stage, the participants were informed they could not ask for help unless they were truly unable to progress. We told them this was because: 1) we wanted them to focus their efforts as much on

the scenario at hand, and 2) at three different points they would be asked to briefly summarise the current situation. At the end of this scenario, they were asked to "*clean a machine*" and outline what they think would happen to the overall risk picture.

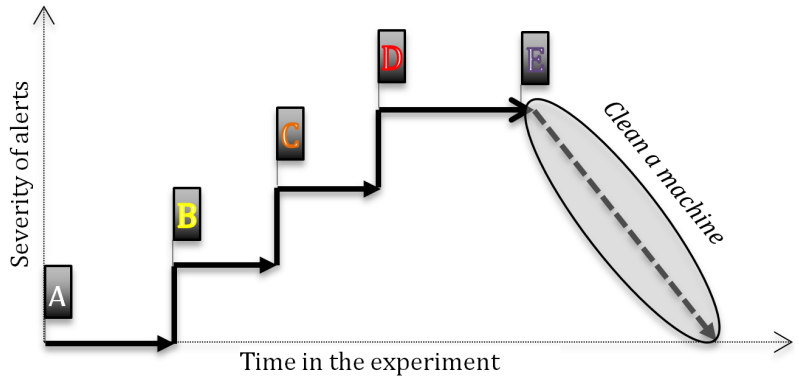


Fig. 4. Illustration of the main scenario alerts

The Main Scenario investigated whether participants obtained sufficient situational awareness and how well they were able to identify the cause of the risk picture shown. The starting point A had no alerts. After ten seconds the first wave of alerts started (point B), targeting all machines in the network with low severity events. This wave lasted for four minutes and all the alerts were of low severity. The amount of alerts per sub-network was proportional to the number of machines which this sub-network contained. The second wave of alerts was of orange severity (point C) and lasted for three minutes, targeting (again) all machines. The final wave of alerts (point D) was of red severity, it lasted for three minutes and specifically targeted machines that affected a majority of the enterprise critical tasks specified in the BPMN definition. One minute before the end of each wave participants were requested to outline what they think was happening. At the end of all the waves (point E) users had the opportunity to present a mitigation plan and execute it, by cleaning one machine. 1300 alerts were generated in total.

The data collected from the eye tracker in this part identified where users spent the most time looking. By selectively exposing machines to red alerts we were able to show whether users prioritise their actions based on information from BPs. More specifically, in a chronological order, there were four sub-networks flagged with red alerts, some partially critical to business operations (i.e. one or two dependencies existed), others were crucial (five or more dependencies on tasks that affect risk propagation significantly). By introducing red alerts to specific sub-networks, we aimed to examine whether participants consider the BP notations before prioritising their actions. By including the "*Stores and Services*", "*Servers*" and "*Logistics*" sub-networks, we investigated if participants dwell on BPs. For example, the "*Stores and Services*" and "*Servers*" only influence two BPs, but the "*Logistics*" sub-network influences all of them. If participants decide to clean a machine from the "Logistics" sub-network, it will result in most of the BPs being flagged as purple, allowing the organisation to function in the least risky fashion achievable.

We decided to create synthetic data in order to control the effects that alerts would have on the BPs and create scenarios where user's choices could be assessed. By introducing red alerts to specific sub-networks, we aimed to examine whether participants considered the BP notations before prioritising their actions. For example, having chosen to clean the machines in the "Admin and Monitoring" sub-network would not have changed the severity in the BPs, indicating participants'

preference to concentrate on specific type of alerts or on specific branches while remaining oblivious to the importance of BPs and how these represent the way the enterprise functions.

We make particular note of the key differences between the Training Scenario and the Main Scenario. During the Training Scenario, participants could ask as many questions as they wanted and were offered to repeat the training if they did not feel confident in using the tool. In the Main Scenario, they were asked to focus on a particular task, and not ask questions about how to operate the tool. This was by design to allow the cognitive processes of the participant to focus on the current task at hand. In the Training Scenario, this was to learn how to use the tool. In the main scenario we wanted them to focus their efforts on taking in the attack scenario and being a SOC analyst using our tool. We wanted them to use everything they had learnt from the Training Scenario in a new context, so they can form an opinion of the tool in use (as opposed to learning how to use the tool).

4.5 Reflection

In the final part, participants were asked to undergo an interview to clarify participants' thought process during the main scenario and enable us to obtain user-experience feedback. This was deemed necessary because eye trackers only record foveal vision data (i.e. where users look, not attention). A semi-structured interview enabled us to tailor some interview questions to observations about the participant actions specifically from the Main Scenario, for instance, recurring actions they would make (in the effort to understand why actions were repeated). Our aim was to gain an in-depth understanding of 1) the participant's thought-process during the Main Scenario, 2) features of the tool which they deemed useful (and not), 3) suggestions for changes, 4) strengths and weaknesses of the tool compared to other SOC tools.

Exploring the potential of gaining competitive advantage once the tool is deployed in such a context, we asked whether BPs contribute to situational awareness and how useful the information which BPMN provides is when handling emergency situations. Finally, we also asked if participants could foresee issues due to scalability problems and sought potential solutions. After the interview, they were also asked to fill in a second questionnaire asking them to rate their opinions, including: 1) our approach against tools that they currently use 2) its usability features, 3) major concerns, and 4) future improvements listed. The questions were not exposed to the participant during the interview as we did not want to influence their own, independent thinking, verbal feedback.

4.6 User Study Setup and Execution

The setup configuration is shown in Figure 5. A participant sits in front of a computer screen, with only a mouse at their disposal. The eye tracker is in front of the participant, and the output of the eye tracker is captured by a networked machine (as shown to the left of the participant).

The duration of each participant lasted between 1.5 hours to 2.5 hours. The experimenters let the participants spend as much time as the participants deemed necessary in order to maximise the participant's internal locus of control. Conducting the study at the participant's pace helped achieve this.

The data output from the study was produced by a Seeing Machines Facelab 5 eye tracker [44], video captures and a voice recorder. We recorded eye tracking as video and in CSV format. Video encoding was encoded at 10 Frames-Per-Second (to avoid any potential issues of video capture), while eye tracker CSV data was recorded at 60 Frames-per-Second. We recorded eye tracking at three separate instances during the assessment per participant; during the Tutorial Video, Training Dataset and the Main Scenario. A laptop with voice recording software was during the interviews.

Fig. 5. Example participant (blurred) in the setup

4.6.1 Demographics. Ten male SOC analysts from seven UK-based SOCs took part in our assessment. The first questionnaire from the introduction comprised of fourteen questions. The average age is 40 and their experience as analysts ranged from 1-2 years to 20 years, see Table 1 and Table 2.

Table 1. Range of participants' ages

Age	18-29	30-39	40-49	50-59	60+
	1	4	4	1	0

Table 2. Participants' years of experience as SOC analysts

Years	0-5	5-10	10-15	15-20	20+
	5	2	1	1	1

5 RESULTS

5.1 Questionnaire 1: Current Practices

Below follows the results from the first Questionnaire, presented in the Introduction part of the study, see SM B. Tools used by the analysts included (parentheses give a count of how many participants use these tools): Nagios(2), Snort(6), Arcsight(4), bespoke tools(2), Tipping Point(1), Splunk(2), Argus(2), Muntsman(1), Netflow(2), Qradar(1), McAfee IDS(2), Wireshark(1), Nessus(1), ISS(1), Quacys(1), WAF(1), Sitescope(1), nHp proxylogs(1), and Solonwinds(1). Five analysts consider visaulization "very important" for their work, and the remaining five consider it "somewhat important" in a five point likert scale. Two participants stated they prefer not to use visualizations to explore data in their tools, one prefers static visualizations and finally seven prefer interactive visualizations. From a degree of satisfaction point of view, results can be seen in Table 3. None of the ten analysts were very satisfied with the ability of the current tools to detect and monitor attacks, aid in incident handling of attacks or visualize attacks. Knowledge of BP semantics was scarce, and none of the analysts use BPMN in their SOC, but some were familiar with other standards for representing BPs, see Figure 6.

Visualization was considered as a very important feature in their line of work, and participants favoured software with interactivity, see Figure 6. It is of the opinion of the participants in the study that none of the IDS tools currently used in SOCs support effective visualizations of attacks. Users

Table 3. Degree of satisfaction for SOC tools' capabilities

	Very satisfied	Somewhat satisfied	Neither	Not satisfied	Not satisfied at all
Detect and monitor cyber-attacks	0	7	1	2	0
Incident handling of cyber-attacks	0	3	2	4	1
Visualize cyber-attacks	0	1	1	8	0

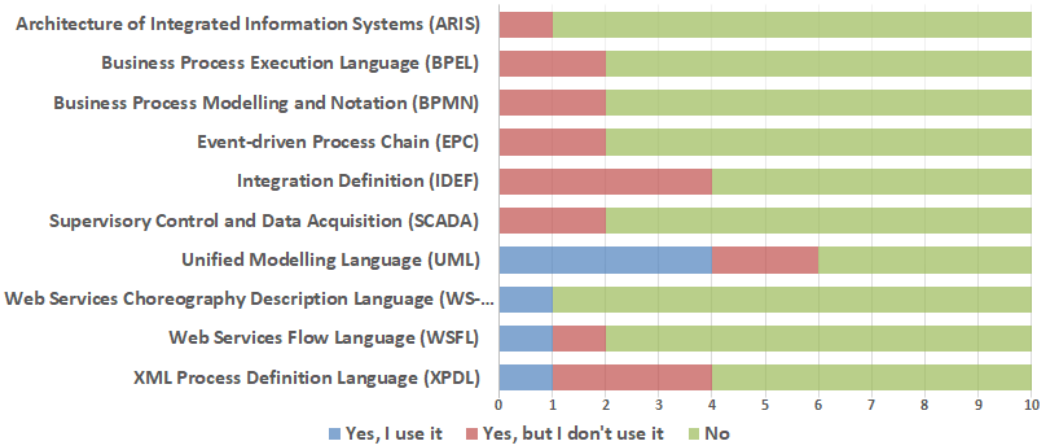


Fig. 6. BP standards familiarity. Yes is split into "Yes, I use it" and "Yes I know of it, but do not use it", and no simply meaning "No, I have not heard of the standard". Worth noting here is that most analysts in our sample set have not heard about several of the standards to describe BPs, with UMLs being the most widespread.

are on the other hand satisfied in general with the degree of detection and monitoring capabilities that current tools provide, see Figure 7.

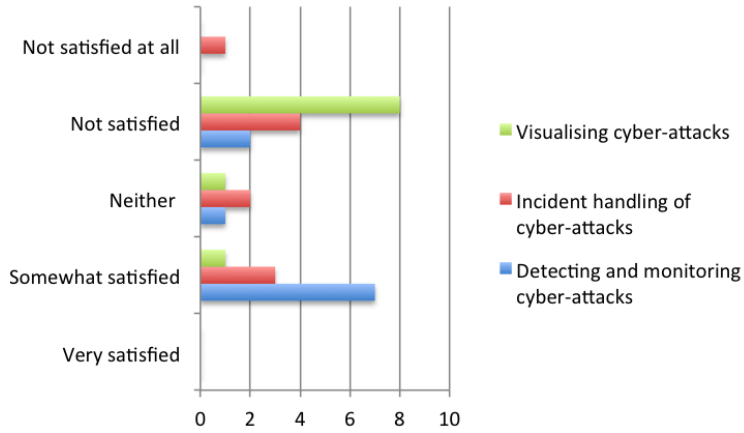


Fig. 7. Degree of satisfaction with current tools in use.

5.2 Training

None of the participants had questions after watching the tutorial video. During the training scenario, questions that participants had were noted. The recurring questions focused on clarifying and reminding them about how to perform a task such as navigation and interpretation of the visuals. Specifically, these included: reminding them about the purpose of purple, how to interpret dependency lines between the asset and BP layer, differences between single and double clicks, and finally, re-iterating a task to complete from the list of tasks asked of them, see SM F.

5.3 Main Scenario

In the main scenario, participants were requested, once they felt confident to make a decision and clean a specific machine. Nine out of ten participants were able to correctly identify the machine that was responsible for the most severe risk propagation at the organisation on their first attempt. The one who did not find the cause, assumed it was a DMZ-related machine at first, but identified the correct cause on his second attempt. The experimenters only stated that his first choice was incorrect and did not provide hints to the participant. Within two minutes he was able to identify the correct asset. Two participants highlighted that their typical intuition would be to focus their efforts on the DMZ instead of the Logistics. Although no participant had experience with BPMN, no one expressed concerns with regards to understanding the notation, and they were able to prioritise their actions to the same single machine and suggest it to be cleaned accordingly.

5.4 Interview Findings

We reviewed the interviews by using a thematic analysis approach [45], and classified respondents answers into different themes to provide some understanding of the participant statements. Based on the interview questions, we identified themes (e.g. scalability issues, SOC processes etc.) and allocated excerpts of the interview transcripts to these themes. The remainder of this section provides a summary of the findings from the thematic analysis.

All participants states that the tool was easy to use, without having difficulties navigating through the various functionalities of the tool or understanding the information provided by the BPs.

When participants were asked if the tool would fit into a SOC, most answers were in the positive regard. All participants mentioned that the tool has the potential to improve situation awareness, with one stating that it offers a *"completely different way of working"*. They stated have had a need for a tool to be able to link the importance of alerts to BPs since this will determine "*what is the most important aspect to protect*" and will allow them to provide much better information to their clients who may not always be *"as techy to understand a denial of service attack"*.

Another point raised by the participants was their belief that analysts would be able to cope with new processes that the tool introduces. One participant said: *"The level of detail presented to evaluation it would work for an analyst. Anyone reasonably intelligent would be able to find this out"*. Conversely, the biggest concern they all expressed however was that of scalability, whether the star-topology approach can represent large organisations, and whether a relational database like MySQL database will be able to handle millions of entries, even as a cache to a Hadoop cluster. Finally, it was also suggested that the tool could be used in SOC environments as a tool to facilitate learning for the new members of staff.

Seven participants suggested that analysts have already a good idea of the organisation's system which they protect and how these systems may be linked to BPs, albeit not formally documented. *"Most of this information is in their heads. The analysts inherently know what BPs are in there.* This tacit knowledge insight, appears to relate to the tacit knowledge discussed by Ahrend et al. [46].

The problem, however, is that analysts "*cannot remember all these different networks, whether it is a human resources system or a payment system [and what action they] do. There are more than 50*

different assets to remember". By using the tool, according to one participant, analysts "*would know the various functions and the in-depth details of the network*".

One participant noted: "*We need a way to link managers to analysts. They should both have the tool and use it. Not phoning up for updates and spending time on the phone*". Another participant pointed out that analysts may sometimes attempt to look at the attack as a technical puzzle to solve, and how visualization can point out issues that help analysts back to the practical, day to day considerations: "*There is a danger that a technical guy will go for something that is technically challenging but is simple to solve practically fast. Now, they can see the logic of where to focus: clean a machine that has no impact on the BPs?*".

Participants were also asked to provide further feedback on any personalised requirements. Three stated the benefits of options to modify time duration of how quickly alerts enter the system in, and adapt to how fast any situation is changing. One participant mentioned specifically either a clickable item or visual feature next to the cylinders or a numerical value that can be changed in the deep dive window. It was also suggested to provide analysts with the option to filter alerts based on description or severity and to choose whether spheres and cylinders would appear on screen. Additionally, two proposed colouring the table lines in the deep dive window to indicate the severity of the alert appropriately. At the moment we only provide a numerical value e.g. 1, 2, 3.

Participants were sceptical regarding the ability of the tool to integrate with other systems, primarily due to scalability concerns. It was noted that the organisations they monitor may have "*easily 10-20 sites in place*" and this network topology with coming alerts of real magnitude will result in *"a congested screen"*. Another stated: "*If there are more complex organisations, with multiple sub-organisations, I was thinking how it would scale. With anything that you are looking, the potential of false positives is huge*". Another issue was providing meaningful dependency models. Organisations possess critical infrastructures, and if *"anything happens to those then everything turns to red quickly"*. One participants noted however, that there is an opportunity to potentially use the tool to simulate attacks to provide mitigation policies or alternate more resilient model dependencies.

5.5 Questionnaire 2: Reflection

Questions in the second questionnaire were similar to those presented in the first, enabling us to compare participants' answers before and after they experienced the tool. Key comparisons included identifying whether the use of our tool changed the perception of tools in SOCs after having experienced it. The results of the first question, indicate a positive change in participants' attitude towards visualization. There is also a change in their opinion about incident handling, as all of them consider it now to be an important feature for a tool, see Figure 8. Participants stated that the strength of the tool lies in its ability to improve situational awareness. Participants also noted that our tool was easy to use and helped them to focus on pertinent information. They were however sceptical about integrating it with existing systems, see Figure 9. Regarding limitations of our tool, participants were concerned with the BPs and in particular how the tool would obtain and update such information. Another issue they foresee hindering its functionality is handling scalability issues with large volume of alerts being visualized, see Figure 10.

With respects to future improvements, participants concur that the ability to explore raw traffic data using alternative visualization methods is the most important step forward, followed by a multi-user support of the tool. Providing templates of BPs was deemed necessary, as well as a dynamic configuration of the network. Being able to correlate incoming alerts to known vulnerability exploits drew participants' interest, as well as the ability to predict attacks based on previous data. Participants also strongly suggested to design a variety of logics which would consider different aspects of risk, see Figure 11.

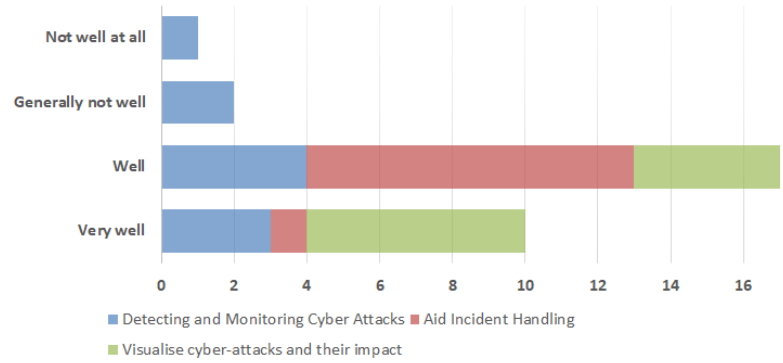


Fig. 8. Tool features

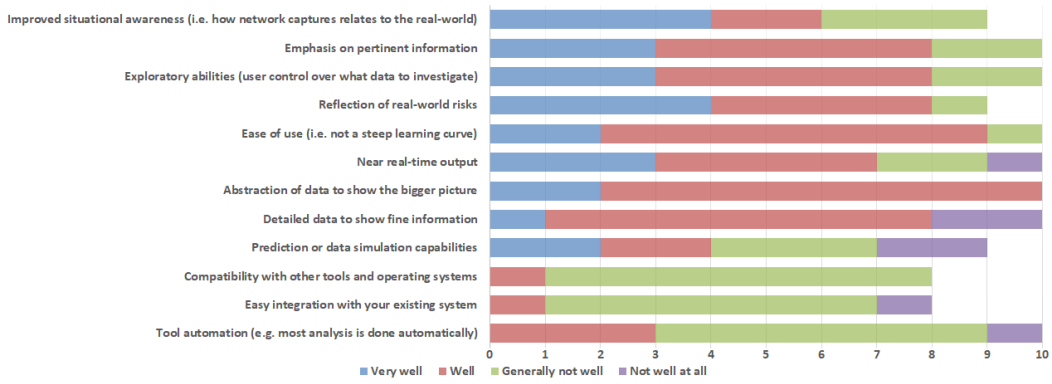


Fig. 9. Assessment of the tool's capabilities

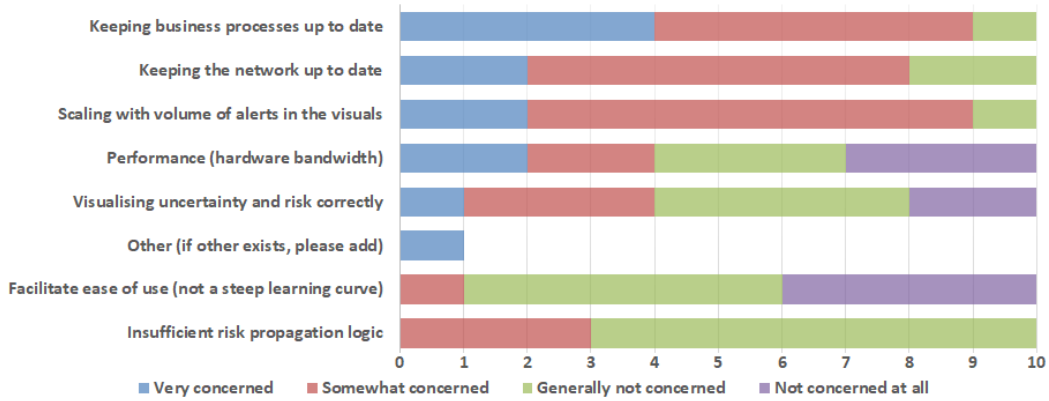


Fig. 10. Concerns regarding the tool.

We also asked participants to rate how crucial limitations would be to address. Most of the participants recognised that the ability to identify real-world consequences of an attack is very important,

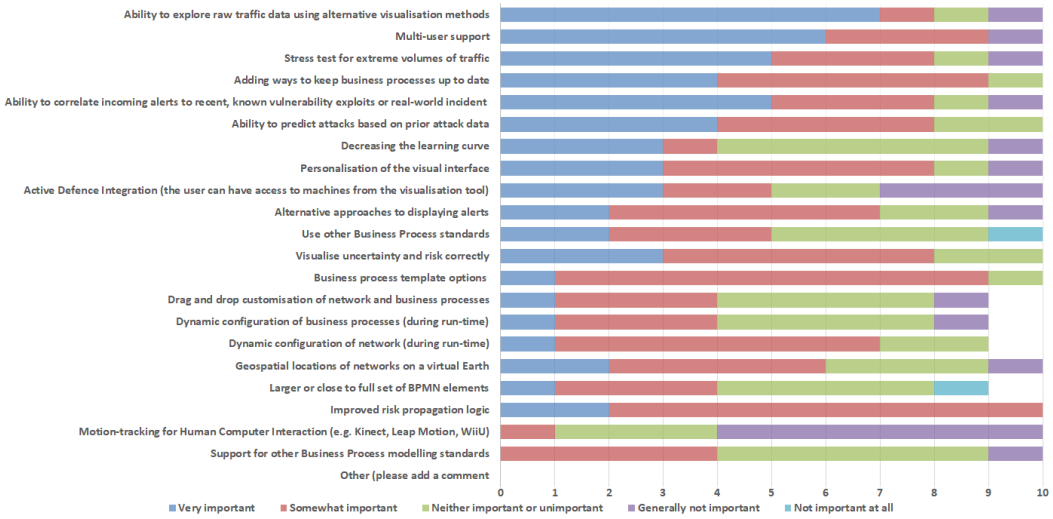


Fig. 11. Suggestions for the future

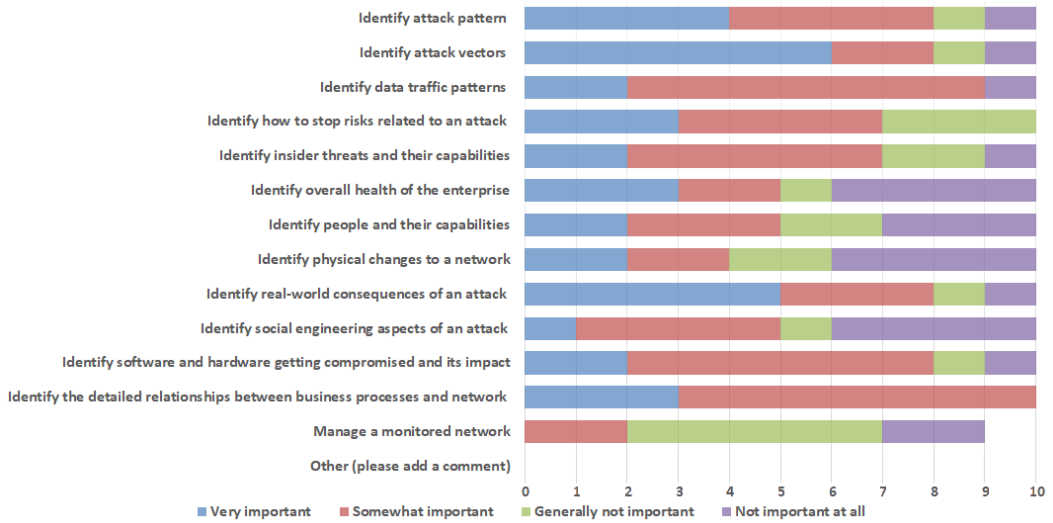


Fig. 12. Limitations of the tool. "Ability to:"

followed by the ability to identify attack vectors. Other areas, where our future endeavours to improve the tool should focus, are the refinement of the mapping between BPs and the network topology and the ability to identify data traffic patterns, see Figure 12.

6 DISCUSSION

Prior to discussing our findings, it is important to note that we do not claim that our findings can be generalised or that our approach has been fully validated. This is because we used a limited number of participants and conducted our experiment in laboratory settings. Our study should be

regarded as an examination on the feasibility of our visualization method to analyst audiences. Our findings suggest that our general approach is suitable for the tasks we exposed analysts to. Many more participants will be required to make any conclusive statements about the tool's usefulness with a general viewer in mind, and for a production environment.

We set out to determine whether analysts can understand uses of BPMN for SOC tasks, and if yes, whether visualizations that link BPMN and network asset alerts be used to offer better situational awareness in SOCs. We successfully documented the analysts' understanding in two distinct ways: firstly, through the Main Scenario as 9/10 responded correctly to the decision-making task we gave them. Secondly, we asked them questions through questionnaires and interviews. Eye tracker video gave us insight about viewing patterns, but asking analysts gave us the insight we needed to understand participant opinions and their opinions about the tool.

All participants said that navigation was straightforward. Overall, viewing patterns suggest that users viewed the tool in two alternating ways to traverse the scene. Users would either take a systematic approach in which they step through items of interest in an ordered fashion (this may be stepping through the flow chart in the BP); or, users would jump between objects in a (seemingly) unstructured order (possibly stimulus-driven vision). Generally, users would traverse the layers separately (BP or network), e.g. alternate between a BP-centric focus and a network-centric focus. A majority of viewing paths followed dependency lines or new changes to alert levels of assets. Three of the participants also expressed that the IDS alert count (next to the cylinders) also contributed a great deal to the paths they took. Our findings suggest that saliency guides vision, but other factors may also have an impact as well.

Four participants examined individual IDS events in the deep dive window for longer periods of time (> 30 seconds), more than three times. When asked about this, the analysts said this was because the deep dive window was not wide enough to display the alerts properly and said it was of minor annoyance.

Five participants highlighted that metadata about the events themselves would be of more use to analysts than many single events. The amount of time spent in the bottom right corner to view the single events was very short in comparison to the viewport window. We are of the opinion this section of the screen could be better spent on showing other information continuously (e.g. information about volume, general health of the system, most at risk BP task or machine or throughput of events). Three of the participants also requested a timer for the spheres or to be switched off entirely as they found the sphere motion distracting at times.

Participants only expressed issues in using the tool during the Training Scenario, and did not express concerns in using the tool during the interviews (after the Main Scenario). Each time participants expressed difficulty, we explained navigation controls again. As mentioned, nine of the ten SOC analysts were able to correctly identify the correct machine that was the major cause of the BPs being at high risk. All participants used BPs in their reasoning to (at least) some degree, while two participants focused on which assets affected the dependency lines (leading up to the BPs), the remaining eight also spent time opening up the BPs themselves to explore the flow charts within the swim lanes to greater extent.

Among the tool's stronger aspects (as mentioned by analysts) is the ability to improve situational awareness and guide the user when faced with large amounts of IDS data. This capability, was deemed by the participants to be important and innovative. This capability raised concerns as BP information about enterprise missions is scarce in real-world SOCs. Among the tool's weaker points is the inability to update BPs and network in an automated fashion. Some participants expressed concern about how scalable the tool would be for very large organisations. This is a topic we will have to address in future work. The participants suggested visualization of attacks is one of the main limitations of current tools used in SOCs today. Once they had experienced the

tool, their perception regarding the importance of visualization increased, as well as their opinion regarding the importance of incident handling (using visualization), two features that, according to the participants, the tool excels at. Regarding the tool's usability, all the participants indicated that it was easy to learn and use, and that its semantics were explained in enough detail.

6.1 Challenges in Conducting Assessments of Novel Visualizations

Proposing novel tools that aim to tackle security analytics in manners that have not been attempted previously poses its own set of assessment challenges. A large amount of literature in the SOC domain has been written by practitioners with a keen eye on new capabilities, but not necessarily focusing on scientifically assessing them. A large amount of literature in the SOC tool evaluation domain also often focus on quantitative findings, or very rudimentary, informal feedback. We argue more research efforts should go in to attempting to understand the human elements and technical elements of threat detection in SOCs. Our mixed-method approach outlines a framework for such an approach, one we designed carefully to attempt to capture more tacit feedback (i.e. things that are left unsaid or not immediately clear from quantitative data analysis alone) through "reading, watching and doing". Below follows example challenges faced during the design of our study, as well as insights into how these challenges were addressed.

6.1.1 Minimising confounding factors. To minimise confounding factors, we added two restrictions to the tool, we disabled: 1) forensic mode, and 2) keyboard navigation. This means that participants could not "rewind" the alerts as one can normally do in the tool. Also, participants can only hop between different levels of the network and BPs, but not move around in 3D space. This was done to focus the assessment on the visualization and usability elements of the tool by reduce any extra features. Our results are therefore only valid in the context of mouse navigation and not using forensic mode and a separate study would need to assess the integration of these component.

6.1.2 Designing datasets and realistic scenarios. Designing the training and main scenario posed a number of challenges that we addressed, including:

- **verisimilitude of scenario** – i.e. is the scenario (attack patterns and organisation configurations) *realistic enough*? The participant saw gradually escalating alerts as opposed to pseudo-randomness or scripted/storyboarded randomness in the attacks. In real SOCs, attacks at any severity-level can happen at any point in time. We argue that for laboratory conditions, our scripted approach was appropriate for several reasons: 1) it allowed the experimenters to observe how the participant would treat situations of escalating nature, 2) it was a different attack pattern than the training scenario, and 3) it allowed the experimenters to observe participants reactions to modular situational awareness changes, which allows experimenters to compare participant behaviour to model answers.
- **validation of attacks** – whereas verisimilitude refers to whether the analyst perceives the attacks as realistic/believable enough, there is also the question of whether the attack itself objectively could happen in the real world. Our attacks were scripted in terms of a simulated storyboarded red-team exercise. We built a story, developed an organisation configuration, and then the attacks according to the story and scripted the IDS alerts in a pseudo-random fashion to generate the alert data in a simulated network. We deemed our assumptions appropriate for laboratory conditions as our primary focus of the study is to investigate the value of our visualization and usability of the tool. Being able to have a model answer allowed for more control over the setup. We do recognise that a separate validation would be necessary for real-world attacks in live SOCs.

- **how to measure whether participants have *improved* situational awareness**? – at the end of the main scenario, participants were informed that they need to make a single decision: which machine they should clean to yield the most effective result overall. This doesn't measure improved situational awareness, but rather whether our approach is feasible in the first place. Answering this question helped us identify whether or not the participants were able to use the tool and make these considerations in making a (limited resource) decision: 1) what is the configuration of the network – i.e. which dependencies exist? 2) how have alerts propagated? – i.e. which assets or tasks with the most dependencies have been affected the most severely (based on severity, not amount of alerts)? 3) what are the intrinsic priorities of the mission? In our case, we explained that all assets and tasks are equally important. As participants could only clean one machine at the end, this was a way of putting a time restriction on their ability to clean machines – by "limiting" them to only "have time" to clean one machine.
- **ensuring a "representative set of BP-to-asset dependencies" exist (as seen from an experiment point of view) across assets and tasks**. If a significant number of dependencies exist on a single machine, this would make the dataset unfair, in favour of the analyst – as they would be able to identify which asset to clean quickly. The model answer machine in our scenario had as many dependencies as other machines on the network, the key difference was that this machine affected different types of BPs, and not only a single BP "*swim lane*" (terminology from the BPMN standard).
- **combating biases**. The key purpose of the study is to determine whether our approach can in principle be used in a SOC environment to provide decision support, given more research and development time and resources. We therefore consider the assessment as an approach to check feasibility of such a system using SOC analysts to trial the tool in laboratory conditions. In order to combat confirmation bias, the experimenters only answered direct questions from the participant, and rehearsed a script to describe the system at the various parts of the study. As we are primarily identifying the feasibility of such a system working, in principle, and our primary concern is to obtain first-impression commentary from different analysts, we deemed it necessary to present them with two simulations, in the same ordering. This allowed us to compare commentary across analysts and scenarios. For a larger study, we recognise that a pseudo-random ordering of training and assessment scenarios will be required to combat ordering effects.

We hope to see more user studies with SOC analysts and risk-owners in the future, as tools and studies to improve decision-making is not well-explored in the literature.

6.2 Recommendations for similar tools and studies

It remains to be seen whether the utility of semantic visualization can outperform the traditional data model visualizations often found in SOCs today. Our assessment suggests that novel visualization paradigms and how these can be applied in SOCs yield potential, and can work in tandem as opposed to in opposition to each other. The attack use cases in the assessment were captured in a red team simulations, leaving the approach to be validated in production environments with actual attackers. Based on our experiences, we have the following recommendations:

6.2.1 Developing visualization tools with semantic models. When developing new tools to enhance SOC analyst capabilities through semantic visualization, we envisage a number of key considerations should be addressed:

- **Build a concept of operations or a finite state machine to represent how users can navigate and run the tool** to be able to model intended behaviour with actual behaviour, this facilitates debugging and validation.

- **Simplify the tool configuration pipeline**. Each configuration instance of our tool requires three configuration files: BPs, network topology and the mapping between the two. Minimising the configuration requirements would lessen the burden on users and debugging further.
- **Familiarity through association in iconography**. Our tool introduced a new working paradigm for analysts to use visualizations. The fact that the iconography was heavily inspired by traditional network diagrams, we believe, helped ease the training and create a gentle learning curve.
- **Ensuring visuals metaphors follow easy-to-understand and consistent reasoning structures**. Our tool used a graph structure and abstraction to communicate hierarchies of network architectures. We believe this approach also helped facilitate participants' mental model of understanding the network topology.
- **Allow users to view the raw logs or alert events**. The deep dive window allowed for analysts to inspect individual alerts. Everyone were of the opinion that alerts should always be available for inspection, should the analyst ever need to review them.
- **User testing during implementation**. Testing with users during implementation is crucial. While SOC analysts did not show up until our study, other, non-expert testers provided basic usability feedback which helped us refine the tool, so the assessment could focus entirely on SOC analyst feedback, rather than addressing more fundamental usability concerns.
- **Regard initial stages as highly experimental** and be prepared to redesign or drop features.

6.2.2 Assessment design, setup and execution. When conducting similar assessments to the one we have presented, we have the following recommendations:

- **Run a pilot on the assessment prior to inviting analyst participants**. This ensures that the execution of assessment session will run smoothly and is well-rehearsed.
- **Weigh the advantages and disadvantages of having the developers run the experiment**. Psychology and medical studies often make use of a double blind study as best practice to limit bias during the study. We deemed it sufficiently important for the participants to be able to speak directly with the developers of the system, see Section 4.1.1.
- **Understand the Learning Curve**. When designing and implementing a new tool, developers are likely to become blind to seemingly obvious minor issues in learning how to use the tool. Through user-testing with new users, allows for the learning curve to be gauged. While every user will be different, and thus the learning curve will be different per person, having enough people trying the tool enabled us to identify whether a new feature noticeably increased the learning curve of not. We suspect more measures to quantify the learning curve can be created in the future.
- **Minimise factors by limiting navigation**. Our tool uses a navigable 3D scene by default. However, prior to the assessment, we locked the screen to fixed positions and disabled forensic mode in the interest of minimising factors for analysis. See Section 6.1.1 for more info on this.
- **Support for new attack vectors**. Our approach focused on proposing and assessing whether the core principles visualization show merit in situational awareness. Attack vectors have therefore remained relatively simple. When developing new tools, a wide range of attacks and reasoning of those attacks should be implemented.

6.3 Improvements made to the tool

According to the participants, several current tools used in SOC environments lack the ability to handle cyber-attack incidents or visualize attacks effectively. We showed feasibility of making use of semantic model visualizations for decision making, and deem it necessary for such visualizations be easy to use and straightforward to integrate with existing SOC systems. Keeping the BPs and

the network layout up-to-date, ensuring scalability (throughput and ability to visually represent large organisations) as well as new abilities to identify pattern of attacks were deemed important as well. Improvements made to the tool as a consequence of the study include the following:

- **Data window optimisations**. We now colour alerts in deep dive data table window so users can also determine which assets matter the most (severity level-wise) from the deep dive window alone.
- **Configuration exploration**. Users can now explore all available Assets, BPs, People, Vulnerabilities in a tree structure in the deep dive window to have easy access to all of the above.
- **Filtering Options**. Users can now filter out alerts of different severity levels.
- **Alert statistics**. We have since implemented a *statistics view tab* in our deep dive windows that summarises key activities in the last 24h. Some of these include: average rate of alerts, number of vulnerabilities, number of alerts (IDS and people), total alerts, alerts by severity.
- **Supporting prioritisation of BPs**. We added a prioritisation component of BP tasks. Each of them are now ranked by priority. These can be set manually or determined through computation based on the state of the organisation.
- **Greater support for BPMN syntax**. We now support subtask Activities, Task Types (e.g. manual, user, receive, send, script, service tasks etc.), and all Event types as described in the BPMN specification. We now also provide intrinsic priority values to the BPs, and display those to the users, enabling analysts to make faster prioritisation decisions based on these values. Automated evaluation of mission criticality vs. alert severity remains to be investigated.
- **Dynamic networks**. We assume that assets can be plugged in and out of the organisation. Should an IDS report on an asset that does not exist, we can now assign these aforementioned question marks to a department and specify its role in the department (client, server, router, switch or BYOD).
- **Sphere properties**. The spheres can now be set to a timer, so the spheres do not remain on screen indefinitely. We expect to add other properties in the future.
- **Personalisation options**. Users can now use different sets of iconography or import their own OBJ-format models.
- **Tool tips**. For a certain subset of items, we now support tool tips to show users what those items do.

6.4 Limitations and Future Work

Several research challenges remain, ranging from incremental improvements to the tool, implementing feature requests, and additional testing of the tool in production environments. Future research will also need to investigate how assessment methodologies can improve for visualization tools more generally in SOCs. There is a need to identify when visualization-to-visualization comparisons is appropriate, and when the value of novel ideas need to be examined in isolation to determine their use. In particular, this means determining when and what type of cross-sectional or longitudinal studies are appropriate to use, what datasets to use etc.

There are three major limitations of our study because it was executed in laboratory settings. Firstly, as mentioned, the scenario, including the network topology, BPs and alerts (albeit inspired by real organisations) were synthetically generated. Secondly, we would need to scale the tool for production environments (throughput, volume and complexity of attacks). We believe it will be necessary to conduct a longitudinal study in a real SOC. To minimise disruption of the SOC, we might collect usage statistics over time. Thirdly, our synthetic dataset is driven by heuristics and pseudo-randomness. In our synthetic datasets, we targeted specific victim assets at specific times, and pseudo-randomly created false positive alerts for all other assets. For the Training Scenario, alert

generation was entirely pseudo-random, however for the Main Scenario, as discussed, we wanted a gradual escalation of severity of alerts. We generated our synthetic alerts in this manner because, to the best of our knowledge, there is no work that proposes a best-practice for distributions of false positives in synthetic datasets for visualization user studies.

Finally, we believe future research could investigate applications of aspects of risk. For instance, instead of examining risk from severity levels alone, infer another aspect of that risk, using the cyber kill chain (e.g. how far along an attacker is?), identifying whether the asset is indeed exposed to a vulnerability (filtering away false positives), or computing risk aspects such as confidentiality, integrity or availability of the compromised asset.

6.4.1 Road map. We have several plans for our road map. These include adding facility to:

- **Straightforwardly update BPs and network configurations**. We are looking into how to best achieve this, and are considering implementation of Graphical User Interfaces (GUIs) to help users validate their proposed asset/BP list as well as their dependency mappings.
- **Prediction and what-if capabilities**. We have in mind to build a prediction module. This could be achieved through machine learning additions to detect trends in alert data and relate this back to visualized risks.
- **Geographical context**. We envisage this can be best achieved by mapping network-layer assets to a globe to identify which businesses are affected by an attack. Currently, this is still at the design stage.
- **Swapping between data model visualizations and semantic model visualizations**. We envisage that more traditional raw data visualization can complement our method. How to best achieved this remains to be seen.
- **Adding alternatives to BP modelling**. UML or UML-like languages was expressed as desirable to include.
- **Multi-user support to facilitate cross-SOC collaboration**. Presently, multiple users can use separate instances of our tool, but there are no specific multi-user features available. We can see value in such an approach, particularly with different levels of access control (e.g. one user being able to see more sensitive parts of the network), but we have no immediate plans to implement a multi-user version in the foreseeable future due to the challenges in validating such an approach.

7 CONCLUSION

In this paper we have presented an assessment of a tool aimed at enhancing situational awareness in SOCs. The assessment identified how well the tool is able to do so by exposing it to ten SOC analysts in laboratory conditions. There are still several research challenges ahead as the tool uses a very different working paradigm to existing tools. Our findings suggest that the tool's approach is a step in the right direction to help incident handling in SOCs, however, more research is necessary to better understand how enterprise-centric solutions (rather than traditional technology-centric data visualizations) can be used in SOCs.

ACKNOWLEDGMENTS

As research contains studies with human subjects we obtained ethical approval for our assessment from the University of Oxford Research Ethics Committee. The study itself did not collect personally identifiable information, only performance data (what participants did) and opinions/feedback. No names were recorded in the process and each participant's data was simply labelled P1, P2 etc. Many thanks to Nick Moffat for narrating the tutorial video. This research was funded by the UK Defence Science and Technology Laboratory (DSTL).

REFERENCES

[1] R. J. Crouser, E. Fukuday, and S. Sridhar, "Retrospective on a decade of research in visualization for cybersecurity," in *Technologies for Homeland Security (HST), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1–5.

[2] S. Jajodia, P. Liu, V. Swarup, and C. Wang, *Cyber situational awareness*. Springer, 2009.

[3] S. Creese, M. Goldsmith, N. Moffat, J. Happa, and I. Agrafiotis, "Cybervis: visualizing the potential impact of cyber attacks on the wider enterprise," in *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*. IEEE, 2013, pp. 73–79.

[4] BPMN, "Business process modelling and notation," http://www.omg.org.

[5] M. Roesch, "Snort-lightweight intrusion detection for networks," in *Proceedings of the 13th USENIX conference on System administration*, 1999, pp. 229–238.

[6] NAGIOS, "NAGIOS network monitoring software application," http://www.nagios.org/.

[7] ClamAV, "Clam anti-virus," http://www.clamav.net.

[8] B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Boston, MA: Addison-Wesley, 2004 4th edition.

[9] J. Bertin, *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin press, 1983.

[10] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1254–1259, 1998.

[11] G. Conti, *Security Data Visualization: Graphical Techniques for Network Analysis*. No Starch Press, 2007.

[12] R. Marty, *Applied security visualization*. Addison-Wesley, 2009.

[13] SECVIZ, "Security Visualization," http://www.secviz.org.

[14] D. Best, S. Bohn, D. Love, A. Wynne, and W. Pike, "Real-time visualization of network behaviors for situational awareness," in *Proceedings of VIZSEC 2010*. ACM, 2010.

[15] G. Conti, "Rumint," http://www.rumint.org.

[16] H. Kim, I. Kang, and S. Bahk, "Real-time visualizaton of network attacks on high-speed links," *IEEE Network*, vol. 18 Issue 5, pp. 30–39, 2004.

[17] S. Lau, "The spinning cube of potential doom," *Communications of the ACM*, vol. 47 Issue 6, pp. 25–26, 2004.

[18] A. Yelizarov and D. Gamayunov, "Visualization of complex attacks and state of attacked network," in *Proceedings of VIZSEC 2009*. ACM, 2009.

[19] Q. Liao, A. Striegel, and N. Chawla, "Visualizing graph dynamics and similarity for enterprise network security and management," in *Proceedings of VIZSEC 2010*. ACM, 2010.

[20] M. Chu, K. Ingols, R. Lippmann, S. Webster, and S. Boyer, "Visualizing attack graphs, reachability, and trust relationships with navigator," in *Proceedings of VIZSEC 2010*. ACM, 2010.

[21] J. Rasmussen, K. Ehrlich, S. Ross, S. Kirk, D. Gruen, and J. Patterson, "Nimble cybersecurity incident management through visualization and defensible recommendations," in *Proceedings of VIZSEC 2010*. ACM, 2010.

[22] R. Marty, *Applied security visualization*. Addison-Wesley, 2009.

[23] G. Conti, *Security Data Visualization: Graphical Techniques for Network Analysis*. No Starch Press, 2007.

[24] RMarty, "SecViz.org," http://secviz.org/, 2007.

[25] K. Gancarz and K. Prole, "Visual techniques for analyzing wireless communication patterns," in *Homeland Security (HST), 2012 IEEE Conference on Technologies for*. IEEE, 2012, pp. 341–347.

[26] The Wireshark team, "Wireshark," http://www.wireshark.org.

[27] K. Abdullah, C. P. Lee, G. J. Conti, J. A. Copeland, and J. T. Stasko, "Ids rainstorm: Visualizing ids alarms," in *VizSEC*, 2005, p. 1.

[28] Arcsight, "Arcsight enterprise security manager," http://www.arcsight.com.

[29] Tenable, "Tenable 3d tool," http://www.tenable.com.

[30] SecureDecisions, "Securescope," http://www.securescope.com.

[31] E. Bertini, "Background literature on evaluation for information visualization," http://www.diag.uniroma1.it/beliv06/infovis-eval.html, accessed: 2017-08-06.

[32] C. Plaisant, "The challenge of information visualization evaluation," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2004, pp. 109–116.

[33] C. Ahlberg and B. Shneiderman, "The alphaslider: a compact and rapid selector," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1994, pp. 365–371.

[34] P. Irani and C. Ware, "Diagramming information structures using 3d perceptual primitives," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 10, no. 1, pp. 1–19, 2003.

[35] A. G. Sutcliffe, M. Ennis, and J. Hu, "Evaluating the effectiveness of visual user interfaces for information retrieval," *International Journal of Human-Computer Studies*, vol. 53, no. 5, pp. 741–763, 2000.

[36] D. Byrd, "A scrollbar-based visualization for document navigation," in *Proceedings of the fourth ACM conference on Digital libraries*. ACM, 1999, pp. 122–129.

[37] C. Plaisant, J. Grosjean, and B. B. Bederson, "Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation," in *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on.* IEEE, 2002, pp. 57–64.
[38] J. G. Trafton, S. S. Kirschenbaum, T. L. Tsui, R. T. Miyamoto, J. A. Ballas, and P. D. Raymond, "Turning pictures into numbers: extracting and generating information from complex visualizations," *International Journal of Human-Computer Studies*, vol. 53, no. 5, pp. 827–850, 2000.
[39] D. Staheli, T. Yu, R. J. Crouser, S. Damodaran, K. Nam, D. O'Gwynn, S. McKenna, and L. Harrison, "Visualization evaluation for cyber security: Trends and future directions," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security.* ACM, 2014, pp. 49–56.
[40] R. B. Johnson and A. J. Onwuegbuzie, "Mixed methods research: A research paradigm whose time has come," *Educational researcher*, vol. 33, no. 7, pp. 14–26, 2004.
[41] J. Gosling, "Java programming language."
[42] Oracle, "Mysql," http://www.mysql.com/.
[43] JogAmp Community, "Java opengl (jogl)," http://jogamp.org/.
[44] Seeing Machines, "Facelab," http://www.seeingmachines.com/product/facelab/.
[45] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.
[46] J. M. Ahrend, M. Jirotka, and K. Jones, "On the collaborative practices of cyber threat intelligence analysts to develop and utilize tacit threat and defence knowledge," in *Cyber Situational Awareness, Data Analytics And Assessment (CyberSA), 2016 International Conference On.* IEEE, 2016, pp. 1–10.

A CONSENT FORM

Consent Form: CyberVis – Visualisation and Usability Assessment Study

Researchers: [redacted]

Description of project: This research project is developing a novel framework for visualising cyber-attacks on an enterprise and their potential consequences at an operational level. Our implementation, *CyberVis*, attempts to visualise said attacks in meaningful and easy-to-understand ways. This is achieved by combining traditional network diagram symbolism with Business Process Modeling and Notation (BPMN). Instead of overwhelming a user with excessive amounts of information, the application abstracts the visuals to only show noteworthy differences in the attack data and indicates potential impact both across the network and on enterprise tasks using easy to understand colouring schemes. Severe events stand out compared to other components populating the scene. The project has so far developed a prototype that we wish to evaluate with domain expert users in mind.

Description of experiment and instructions: This is a four-part experiment intended to assess the visual approach and usability of CyberVis. Security Operation Centre analysts are asked to go through a set of parts in which we measure participant usage, performance and opinions allowing us to assess how well they are able to use CyberVis.

5. **Introduction**: Short questionnaire about their job tasks and eye tracking calibration.
6. **Training period:** Video briefings, and initial 'training/testing session' with training tasks to complete. The participant is free to ask any questions they may have about navigating and operating CyberVis.
7. **Main experiment:** The participant will be asked to complete a set of tasks in CyberVis. At this stage, the participant may not ask for help unless they are unable to complete their task.
8. **Interview:** The participant will be asked some questions about their experience with CyberVis.

What will be used (and when) to record data?
- Eye tracking (during video briefing, training period and main experiment)
- Key logging (during training period and main experiment)
- Voice recording (during training and interview, once the interview is transcribed, these will be deleted)

The interview aims to gather feedback on their opinion of CyberVis, suggestions for improvement and other comments they may have. The whole experiment will be conducted in our laboratory facilities. Total time is estimated to be approximately 1.5 hours.

General points:
- Our assessment will be used to improve CyberVis and may be published in academic research literature. It will not be affiliated with the participant or their organisation.
- Participation in this research is **entirely voluntary and participants are free to withdraw at any time** without giving any reason and without being penalised or disadvantaged in any way.
- **All data will be anonymised** immediately following its collation, and all **information will be treated as confidential**.
- Participants are not paid for contribution. If of interest, a white paper version of the results will be forwarded to their organisation. Experiment completion is not a requirement for this to be delivered.

If you have further questions, you can contact [redacted]

Declaration:
I have been informed about the aims and procedures involved in the experiments. I reserve the right to withdraw at any stage in the proceedings, and note that the information that I provide as part of the study will be destroyed or my identity removed unless I agree otherwise.

Name (print): _____

Signed: _____ Date:_____

Fig. 13. Consent form from the user study

B QUESTIONNAIRE PART1

1 Questionnaire Part 1

1. What is your age range?

☐ 18-29
☐ 30-39
☐ 40-49
☐ 50-59
☐ 60+

2. What is your gender?

☐ Male
☐ Female

3. How many years have you been involved in cyber security work?

☐ 0-5 years
☐ 5-10 years
☐ 10-15 years
☐ 15-20 years
☐ 20+ years

4. Please list up to the 5 cyber-attack detection and analysis tools (i.e. AIDE, Bro NIDS, NetFlow, Nagios, OSSEC, P-Best, Samhain, Snort) you use the most for your work. (Note: if bespoke, please write broadly what the tool does e.g. "visualization", "data filtering")

1.
2.
3.
4.
5.

5. Are you familiar with Business Process Modeling and Notation (BPMN)?

☐ Yes, I use it
☐ Yes, but I don't use it
☐ No

6. Write 1, 2 or 3 next to the standards below if you are familiar enough to use
1 = Yes (I use it), 2 = Yes (but I don't use it) 3 = No, not familiar:

☐ Architecture of Integrated Information Systems (ARIS)
☐ Business Process Execution Language (BPEL)
☐ Business Process Modelling and Notation (BPMN)
☐ Event-driven Process Chain (EPC)
☐ Integration Definition (IDEF)
☐ Supervisory Control and Data Acquisition (SCADA)
☐ Unified Modelling Language (UML)
☐ Web Services Choreography Description Language (WS-CDL)
☐ Web Services Flow Language (WSFL)
☐ XML Process Definition Language (XPDL)

Fig. 14. Part 1 of the first questionnaire.

8. When viewing your data: do you prefer static dashboard visualisations or interactive ones?

☐ Do not use visualisation
☐ Static (e.g. plots that are updated from system feeds)
☐ Interactive (e.g. I am able to explore the data)

9. How satisfied are you with the state of tools today overall in relation to:

- detecting and monitoring cyber-attacks on an organisation's network? ……..
- incident handling of cyber-attacks on an organisation's network? ……..
- visualising cyber-attacks on an organisation's network? ……..

Answer: 1= very satisfied, 2 = somewhat satisfied, 3 = neither satis, 4 = not satisfied, 5 = not satisfied at all

10. Rate the five most important usability features for your work:
Answer: 1= most important, 2 = second most important, 3 = third most important, 4 = fourth most important, 5 = fifth most important

☐ Abstraction of data to show the bigger picture
☐ Compatibility with other tools and operating systems
☐ Detailed data to show fine information
☐ Ease of use (i.e. not a steep learning curve)
☐ Easy integration with your existing system
☐ Emphasis on pertinent information
☐ Exploratory abilities (i.e. user control over what data to investigate)
☐ Improved situational awareness (i.e. how network captures relates to the real-world)
☐ Near real-time output
☐ Prediction or data simulation capabilities
☐ Reflection of real-world risks
☐ Tool automation (e.g. most analysis is done automatically)

11. How often do you play computer and video games?

☐ Never or rarely
☐ 2-3 times a year or less
☐ Once a month or less
☐ Once a week or less
☐ Almost every day

12. Are you colour-blind?

☐ Yes
☐ No
☐ Don't know

13. Do you wear glasses or contact lenses to correct for eyesight?

☐ Yes
☐ No

14. To your knowledge, are you otherwise visually impaired?

Fig. 15. Part 2 of the first questionnaire.

C TUTORIAL VIDEO TEXT

Below follows the text narrated in the tutorial video. The purpose of including it here is to show what participants listened to as part of their training. Screenshots from the video are also included in this section.

This is CyberVis, a tool for showing operators the impact of network attacks on business processes. At its core CyberVis takes in sensor data, such as intrusion detection system alerts, and shows the impact on a monitored organisation. In this short video we will describe how to operate CyberVis first without alerts to explain navigation, and then with alerts to explain how to interpret attacks.

In the foreground we see an abstract representation of networks being monitored by the tool this is the 'network layer'. This network is represented in a star topology with item icons for an example enterprise HQ, a branch of this enterprise, and some sister enterprises and the HQ's suppliers. The cloud in the centre represents the Internet.

Towards the rear of the screen we see the 'business process layer', which in this case shows 5 business processes currently being monitored. These are: Order, Take Delivery, Issue Existing Stock, Receive and Issue Stock, and Maintain Stock.

On the right-hand side of the screen we see the Deep Dive window. It shows the operator which network assets support the currently selected business process or activity, plus supplementary information about these assets.

In terms of navigation CyberVis can collapse and expand items of interest in the scene. We can do so independently in the business process layer and the network layer. By double-clicking a business process, we expand it to reveal activities and gateways. Likewise, by double-clicking an organisation we expand it to reveal the subnet view of that organisation.

At the bottom right-hand corner the whole star topology is always available for navigation back to the organisation level. If the Internet is double-clicked, users return to the first screen. Other organisation subnets can be accessed if single-clicked while the subnets are in view.

Lines in CyberVis show different types of relationships between items. Lines in the network layer show connectivity relationships and are always black. Lines between the business process layer and the network layer show dependency relationships. A dependency line signifies that the indicated business process or activity depends on at least one device in the indicated network layer item.

By single-clicking an item in the business process layer or the network layer we see the business process dependencies that involve that item. A wireframe box shows which item is currently selected. Selecting HQ shows that all business processes depend on this item.

(slight pause)

CyberVis is now loading in alerts. We can see from the coloured pulsating spheres that HQ, Enterprise B and Branch A have each received high severity events.

Colours show severity levels: Red is high severity, Orange is medium and Yellow is low. Purple reflects uncertainty and indicates that CyberVis suspects the item has been affected by an attack.

Cylinders represent the total numbers of attack events, per severity level, received for indicated subnets of devices. They use a logarithmic scale. In each case the total is displayed in front of the cylinder.

Dependency lines are coloured to indicate the highest severity alert affecting any supporting devices. They are black when there are no alerts at the supporting devices.

Single-clicking HQ shows that high severity attacks are affecting a number of items in the subnets. Selecting Human Resources shows that this subnet is not impacting the business processes. In contrast, selecting Internet and DMZ shows that devices on this subnet are impacting all five of the business processes. The overall impact on business processes is serious as we can see that several of the five business processes have high severity alerts.

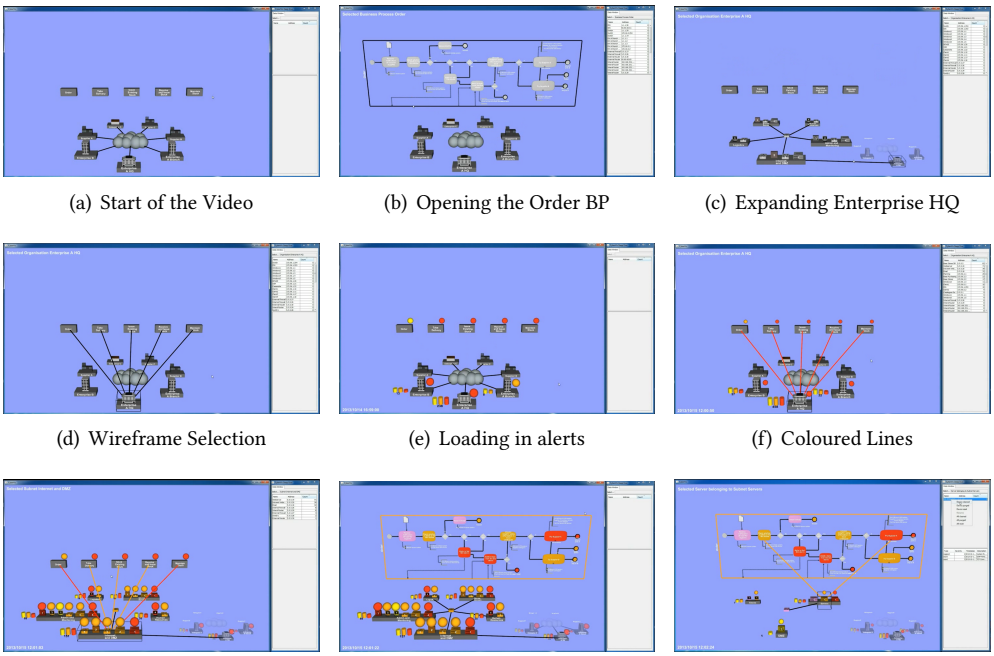
In this example, the operator double-clicks the Order business process because it is critical to operations. The Order process is shown to have medium severity issues overall; there are routes through this process that end at successful end points and avoid the highly impacted red activities.

If we click Check whether in stock or at 1st line stores we see that HQ has no dependencies on it. However, if we click Branch A, we can see a dependency that has been affected with orange severity. One item is affected in the Servers subnet.

Alongside monitoring, the operator can request sysops staff (via telephone or email) to investigate the machine and determine whether it has actually been compromised. Sysops could then either clean it - that is, recover it from the alerted attack - or put in additional defences to protect against future attacks.

At a suitable time in the future the sysops staff will notify the operator that the machine has been cleaned. At this point the operator is able to mark the device as cleaned by right-clicking it in the Deep Dive window and selecting Device Cleaned. The tool recalculates the current attack situation, and in this case shows that the activity and device are now clean. The item is still purple because it is on the same subnet as a high severity machine.

Overall, CyberVis aims to empower IDS analysts and business managers to quickly identify where the key attacks are taking place, investigate these key attack events, and deal with them as quickly as possible.



(a) Start of the Video (b) Opening the Order BP (c) Expanding Enterprise HQ

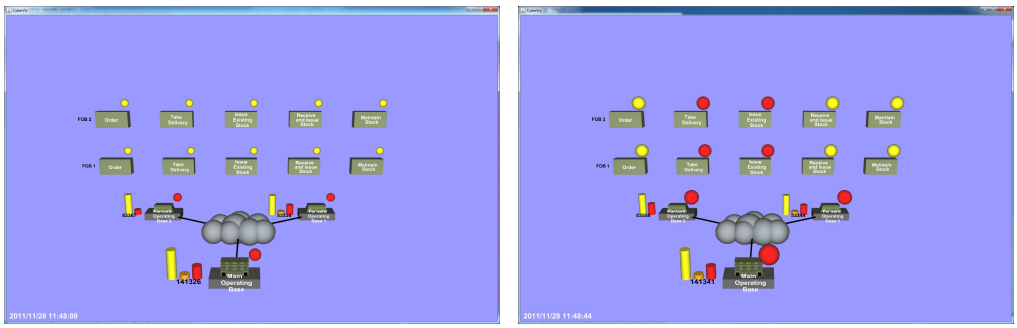
(d) Wireframe Selection (e) Loading in alerts (f) Coloured Lines

(g) Exploring the network (h) Exploring the Order BP (i) Cleaning a Machine

Fig. 16. Screenshots from the tutorial video.

D TRAINING SCENARIO

The Training Scenario was designed to mimic a military scenario primarily focused on logistics missions. It is comprised of 66 networked assets spread across three organisations, using IP addresses in the ranges: (Main Operating Base, 36 assets) 192.168.10.1/24, (Forward Operating Base 1, 15 assets) 192.168.20.1/24 and (Forward Operating Base 2, 15 assets) 192.168.30.1/24 respectively. Each organisation has an IDS, a web server, email server, alternative web server, and client machines. Each subnet is configured to be comprised of DMZ, Admin and Monitoring, Servers and Logistics departments. The tool is configured to monitor five key mission processes in the two Forward Operating Bases: Order, Take Delivery, Issue Existing Stock, Receive and Issue Stock and Maintain Stock. Clicking each BP box will open individual business process tasks relevant in BPMN notation, comprised of 60 BP tasks. Each Forward Operating Base carries out logistical functions that mirror each other.



(a) Early in the training (b) NOP Shellcode attack

Fig. 17. Screenshots from the training session.

E MAIN SCENARIO

The Main Scenario was designed to mimic a generic logistics company. It is comprised of 89 monitored networked assets spread across three organisations: Enterprise HQ (39 assets), Enterprise Branch A, (18 assets), Enterprise B, (17 assets), Supplier A (3 assets), Enterprise C (3 assets), Supplier B (4 assets) and Supplier C (5 assets). Enterprise A Branch, Enterprise A HQ and Enterprise B all have an IDS, a web server, email server, alternative web server, and client machines. The other organisations have assets whose purposes are 'unknown' to the analyst and are labelled client machines with pseudo-randomly allotted IP addresses. The analyst still sees alerts pertaining to those assets, but no BPs are associated with them.

Similar to the military scenario, the tool is configured to monitor five key mission processes in the Enterprise A Branch and Enterprise A HQ. These are also: Order, Take Delivery, Issue Existing Stock, Receive and Issue Stock and Maintain Stock. However, unlike the military scenario, the BPs are not mirrored, analyst should be able to observe that the BPs in the visualization only relate to Enterprise A Branch and Enterprise A HQ. Clicking each BP box will open individual business process tasks relevant in BPMN notation, comprised of 30 BP tasks.

F TRAINING SCENARIO TASKS

Table 4. Training Tasks. White = Instruction, Yellow = Question, Orange = Fact to learn, Purple = Task to re-emphasise the dependency approach of the tool if the participant gets it wrong the first time around.

Task no.	Task Description
1	Double-click Main Operating Base
2	Single-click MOB1
3	Single-click FOB1 - Maintain Stock
4	Double-click FOB 1 Maintain Stock
5	Single-click check stock shortages
6	Look at the data window on the right hand side
7	Select MOB1-ALTWEB
8	Double-click the Internet Cloud
9	Open Forward Operating Base 1
10	Select the yellow Client Machine Icon on the FOB1 platform
11	How many BPs depend on client machines on this subnet?
12	How many alerts are on FOB1-WS1 client machine?
13	How many alerts are there on FOB1 in total?
14	Select the red client.
15	Why do you think the red client has no lines to the BPs?
16	Select the FOB2 Order BP
17	Select the FOB2 Take Delivery BP
18	Select the FOB2 Issue Existing Stock BP
19	Select the FOB2 Receive and Issue Stock BP
20	Select the FOB2 Maintain Stock BP
21	As you can see, there is no dependency from FOB2 BPs to FOB1 network assets.
22	Next to the Internet Cloud, select Forward Operating Base 2
23	Select the FOB2 Order BP
24	The FOB2 BP now shows the dependency to the FOB2 network
25	Open the FOB2 Order BP
26	How many end points do you think this BP has?
27	Why do you think the border of this BP has yellow colour?
28	Double-click the Internet Cloud
29	Select FOB1 - Take Delivery
30	Open FOB1 - Take Delivery
31	Select Await News of whether supplies have been issued by 3rd line stores
32	Why do you think FOB1 Take Delivery is of red severity, but there is still yellow lines connecting this BP with Forward Operating Base 1 and Main Operating Base
33	Double-click the Internet Cloud
34	Open Forward operating base 2
35	Select FOB1 - Order
36	Why do you think the BP is illustrated with yellow severity, subnet assets dependencies are shown with red severity
37	Open FOB1 Order
38	Select the red server
39	Why do you think the BP still functions with yellow severity, when there are red activities?

G QUESTIONNAIRE PART 2

Questionnaire Part 2

1. Based on your experience with CyberVis today, rate how do you think the tool will be able to:
Answers: 1= very well, 2 = well, 3 = generally not well, 4 = not well at all

☐ Detecting and Monitoring Cyber Attacks
☐ Aid Incident Handling
☐ Visualise cyber-attacks and their impact

2. How well do you believe CyberVis would be able to deliver these features?
Answers: 1= very well, 2 = well, 3 = generally not well, 4 = not well at all

☐ Abstraction of data to show the bigger picture
☐ Compatibility with other tools and operating systems
☐ Detailed data to show fine information
☐ Ease of use (i.e. not a steep learning curve)
☐ Easy integration with your existing system
☐ Emphasis on pertinent information
☐ Exploratory abilities (user control over what data to investigate)
☐ Improved situational awareness (i.e. how network captures relates to the real-world)
☐ Near real-time output
☐ Prediction or data simulation capabilities
☐ Reflection of real-world risks
☐ Tool automation (e.g. most analysis is done automatically)

3. Rate initial concerns you have about CyberVis:
Answers: 1= very concerned, 2 = somewhat concerned 3 = generally not concerned, 4 = not concerned at all.

☐ Facilitate ease of use (not a steep learning curve)
☐ Insufficient risk propagation logic
☐ Keeping business processes up to date
☐ Keeping the network up to date
☐ Performance (hardware bandwidth)
☐ Scaling with volume of alerts in the visuals
☐ Visualising uncertainty and risk correctly
☐ Other (if other exists, please add) ……

4. Below is a list of possible future features of CyberVis. Rate each of them according to how important you believe them to be?
Answers: 1= very important, 2 = somewhat important, 3 = generally not important, 4 = not important at all.

☐ Ability to correlate incoming alerts to recent, known vulnerability exploits or real-world incident
☐ Ability to explore raw traffic data using alternative visualisation methods
☐ Active Defence Integration (the user can have access to machines from the visualisation tool)
☐ Alternative approaches to displaying alerts
☐ Ability to predict attacks based on prior attack data
☐ Business process template options
☐ Decreasing the learning curve
☐ Drag and drop customisation of network and business processes
☐ Dynamic configuration of business processes (during run-time)
☐ Dynamic configuration of network (during run-time)
☐ Adding ways to keep business processes up to date
☐ Geospatial locations of networks on a virtual Earth
☐ Larger or close to full set of BPMN elements
☐ Motion-tracking for Human Computer Interaction (e.g. Kinect, Leap Motion, WiiU)
☐ Multi-user support
☐ Use other Business Process standards
☐ Personalisation of the visual interface

Fig. 18. Part 1 of the second questionnaire.

☐ Stress test for extreme volumes of traffic
☐ Support for other Business Process modelling standards
☐ Visualise uncertainty and risk correctly
☐ Other (please add a comment) .....

5. Rate importance of limitations you believe the CyberVis team should improve in the future. Improve CyberVis's ability to:
Answers: 1= very important, 2 = somewhat important, 3 = generally not important, 4 = not important at all.
☐ Identify attack pattern
☐ Identify attack vectors
☐ Identify data traffic patterns
☐ Identify how to stop risks related to an attack
☐ Identify insider threats and their capabilities
☐ Identify overall health of the enterprise
☐ Identify people and their capabilities
☐ Identify physical changes to a network
☐ Identify real-world consequences of an attack
☐ Identify social engineering aspects of an attack
☐ Identify software and hardware getting compromised and its impact
☐ Identify the detailed relationships between business processes and network
☐ Manage a monitored network
☐ Other (please add a comment) .....

6. Overall, how interested are you in using such a tool for your work?
.....
.....
.....

Answers: Not at all, Not very, Somewhat, Very interested.

7. Is there anything about your experience you would like to highlight? If yes, what?
.....
.....
.....

End of Experiment, Thank you for participating!

Fig. 19. Part 2 of the second questionnaire.

H REFLECTION – INTERVIEW QUESTIONS

The purpose of these questions is to ask about the positive and negative aspects of your experience with CyberVis, plus map which improvements and feature requests you believe would be most valuable to your line of work.

Question about the tasks:

- *Could you explain your thought-process during the main scenario?*

About CyberVis:

- How would you describe your experience with CyberVis?
- What were the negative and positive aspects of using CyberVis and why?
- How could CyberVis be useful in your operational environment?
- What kind of scalability issues do you foresee with CyberVis?
- Did Business Processes add value to your understanding of a network attack and why?
- To what degree do you believe you would be able to model business process more generally?

Received July 2020; revised XXXX 2020; accepted XXXX 2020