# 99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms

Bushra A. Alahmadi
*University of Oxford*

Louise Axon
*University of Oxford*

Ivan Martinovic
*University of Oxford*

## Abstract

In this work, we focus on the prevalence of False Positive (FP) alarms produced by security tools, and Security Operation Centers (SOCs) practitioners' perception of their *quality*. In an online survey we conducted with security practitioners ($n = 20$) working in SOCs, practitioners confirmed the high FP rates of the tools used, requiring manual validation. With these findings in mind, we conducted a broader, discovery-orientated, qualitative investigation with security practitioners ($n = 21$) of the limitations of security tools, particularly their alarms' quality and validity. Our results highlight that, despite the perceived volume of FPs, most are attributed to benign triggers—true alarms, explained by legitimate behavior in the organization's environment, which analysts may choose to ignore. To properly evaluate security tools' adequacy and performance, it is critical that vendors and researchers are able make such distinctions between types of FP. Alarm validation is a tedious task that can cause alarm burnout and eventually desensitization. Therefore, we investigated the process of alarm validation in SOCs, identifying factors that may influence the outcome of this process. To improve security alarm quality, we elicit five properties (**R**eliable, **E**xplainable, **A**nalytical, **C**ontextual, **T**ransferable) required to foster effective and quick validation of alarms. Incorporating these requirements in future tools will not only reduce alarm burnout but improve SOC analysts' decision-making process by generating interpretable and meaningful alarms that enable prompt reaction.

## 1 Introduction

In 2013, Target was hit by the most prominent retail hack in U.S. history, in which attackers infiltrated Target's network, installing malware designed to steal customers' credit card data. Months before the breach, Target had installed a new $1.6m malware-detection technology by FireEye. This tool detected the malware, generating an alarm that was picked up by Target's Security Operation Center (SOC) [1] in Bangalore. However, when the alarm was escalated to the Minneapolis SOC, it was ignored, and no action was taken [37].

The malware used in Target's breach was far from sophisticated. Nevertheless, it was able to bypass a large and resourceful organization's security controls and procedures, suggesting that the problem goes beyond SOC's technological capabilities. Security monitoring is a human-centered process with security tools to support the work of analysts, triggering alarms on possible intrusions, and presenting the analysts with the information needed to investigate a potential threat. Although prioritization computations embedded in security tools (e.g., Security Information and Event Management—SIEM) can do much of the heavy lifting, SOC practitioners face the difficult task of figuring out which alarms are False Positives (FPs) and which indicate something dangerous.

Security tool vendors have been competing on the ability of their tools to detect threats, hence, focusing on metrics such as False Negatives (FNs)—*failure to detect security events when they occur*. In contrast, more focus needs to be placed on reducing the FPs—*flagging a security event when it is not a threat*, which is equally critical [11]. Analysts often spend time manually going through alarms to determine their validity, as well as performing monotonous tasks to reduce FPs. Such tasks include reconfiguring security tools, baselining normal behavior, and filtering out noise, when time could be spent on detecting more sophisticated attacks (i.e., threat hunting). Analysts are also under constant pressure to close tickets, as some SOCs evaluate their practitioners' performance accordingly [40]. This process not only leads to human error [13] but causes analyst fatigue and burnout [40].

An excessive number of alarms, in any system, contributes to alarm desensitization, mistrust, and lack of hu-

[1] We provide a background on SOCs in Appendix A.1

man responsiveness [12]. In Target's case, although the technology detecting the malware gave a high-risk alarm, the alarm might have faced skepticism from Target's security team at the time [37], who would not have prioritized the alarm. The deployed technology could have automatically responded by deleting the malicious file, a capability that had been disabled by Target's security personnel. This configuration is not unusual as, due to a lack of trust caused by the prevalence of FP alarms created by detection tools, most SOCs want to avoid any automated decision that could result in business disruption.

In recent work, Kokulu et al. [24] have found that SOC security practitioners do not consider FPs in automatic malicious activity detection to be a significant issue in SOC operations. Such findings contradict academic beliefs on the prevalence of FPs in SOCs and pose the following interesting research questions: *How do SOC analysts distinguish real alarms from false alarms? What do analysts perceive to be False Positives, and how can we establish a more precise definition? What are the shortcomings of alarms produced by current security tools? How can we design better tools that provide higher quality alarms to improve the process of alarm validation?*

In this paper, we address the questions above. We applied a multi-step empirical approach, first conducting a quantitative survey ($n = 20$) to understand practitioners' perceptions of the tools they use in the SOC. We then conducted a qualitative study with 21 SOC practitioners from seven SOCs. We analyzed these interviews using Template Analysis [23], a common thematic analysis approach in qualitative research.

We elicit from the interviews the alarm validation reasoning, and the factors that might impact practitioners' decision-making. This process was found to be human-centric, relying on humans' intelligence and reasoning not only in validating alarms, but in their configuration and review. Our findings also bring insights on the importance of establishing an understanding of the definition of FPs, as we will discuss in Section 6. The term *False Positive* is found to be broad and vague. For example, analysts expressed a distinction between what they call *false alarms* and *benign triggers* when evaluating security tools' performance. False alarms are used to describe an alarm generated without a true security-related event (the boy who cried wolf). In contrast, benign triggers are true alarms; meaning they match an existing signature (e.g., vulnerable java version) but the organization chooses to ignore it (e.g., due to legacy systems). Although Kruegel et al. in [25] defined detections of failed attacks in Intrusion Detection Systems (IDS) as *irrelevant positives*, they are nevertheless alarms generated due to malicious activity, and not a benign business-justified activity.

In our analysis, we found that analysts start their alarm-validation process by looking first at tools they "trust" or consider "reliable." The perceived limitations of alarms generated by either existing network-security tools or SIEMs can be distilled into four classes: (1) unreliable alarms (Section 7.1), e.g., due to loosely written signatures; (2) lack of customizability in traditional systems (Section 7.1); (3) black-box alarms and lack of explainability (Section 7.2); and (4) lack of context on the networks and systems, process, or business (Section 7.3).

To address these limitations, in Section 8 we define five concrete requirements for more useful and actionable alarms: **R**eliable, **E**xplainable, **A**nalytical, **C**ontextual, **T**ransferable. Using Machine Learning (ML)-based tools as an example, we discuss how adopting these requirements can help improve alarms for analysts to (1) make an informed decision about the validity of the alarm and (2) expedite the analysts' **REACT**ion to it, improving the SOC's performance overall.

## 2 Related Work

There has been an increased focus from the research community on developing security tools to automate operations in SOCs, such as data triage [46, 47], log aggregation [34], log mining [45], and SIEM-alert filtering [32].

Akinrolabu et al. found that current IDSs are inadequate in detecting multi-stage stealthy attacks [1]. Goodall et al. conducted semi-structured interviews with IDS experts to understand how they use IDSs [19]. The study shows that IDS tasks are collaboration-driven, and require a combination of common knowledge (e.g., network and security) and situational knowledge (e.g., of normal network behavior). Dietrich et al. used both quantitative and qualitative methods to investigate system operators' perspectives on security misconfigurations, identifying the factors that operators perceive to be their root causes [13].

On improving FPs in security systems, several contributions (e.g., [4, 25, 44]) have focused not on improving the quality of alarms themselves but on developing automated solutions to reduce the alarm volume. These contributions use techniques such as alarm mining, alarm correlation, and elimination of "irrelevant" or "uninteresting" alerts in IDS/IPS systems—a process called alert verification [21, 25, 35]. For example, Kruegel et al. [25] defined "irrelevant positive" alarms as correctly identified attacks by an IDS that failed to meet their objectives. In such work, the researchers aim to verify alerts (i.e., identify alarms relating to successful attacks) by combining/correlating multiple data sources [38] or adding context to alarms [2]. However, such work is based on an underlying assumption that produced alarms are of "quality" (i.e., interpretable, contextual, and meaningful

alarms such that analysts can quickly take informed action), which is not necessarily true. To reduce these FPs, we need to understand the shortcomings of the alarms themselves, improving their quality so that automated alarm-verification solutions can be more efficient.

Sundaramurthy et al. [40–43] takes an anthropological approach, studying three SOCs in educational institutions and making several observations regarding the people, processes, and technology. Specifically, in [41], they made remarks related to operational tools/teams, workflow, and how teams come together in solving security incidents. Factors that lead to analysts' burnout in a SOC were identified, providing a model that explains the burnout phenomenon [40]. However, this model focuses on burnout resulting from managerial issues (e.g., analysts' performance assessments), while we focus on alarm burnout due to a prevalence of FPs. In [42], they presented a Pentagon model for improving the SOC operations by identifying tasks that can be automated to resolve conflicts. One finding that we continue to explore in our study is the importance of customizability in tools.

Kokulu et al. [24] used a qualitative method to identify technological, human and operational issues in SOCs across sectors. Their work highlighted SOC issues related, for example, to low visibility of assets, poor tool usability, lack of analyst training, and communication. They expressed the need for research to define improved security metrics. Our work is a start in this direction: we seek to establish a clearer definition of what constitutes an FP. One of the most interesting findings by Kokulu et al. [24] was that analysts do not perceive FPs to be a significant issue, in contradiction with beliefs among academics. Hence, we also identify the strengths and weaknesses of security tools to explore this contradiction and the limitations (e.g., lack of context) leading to such confusion on tool adequacy. Although work on the importance of contextual knowledge for alert verification exists [19, 25], it is mostly focused on technical context (e.g., network topology), while we also consider environmental context (e.g., work hours).

## 3 Methodology

SOCs are diverse with distinct setups and goals, thus the people, technology, and processes will be unique as well. For this reason, we followed an inductive approach and used a quantitative study as a starting point for our qualitative research. The quantitative and qualitative studies are broader in scope, addressing topics such as SOC data presentation, which fall outside the scope of this paper.

**Ethical approval for this study was granted by the University of Oxford Central University Research Ethics Committee (R48822/RE001).** In both survey and interviews, informed consent was obtained from participants: participants were presented with information on the purposes of the study, the handling and anonymization of the data, the processes to withdraw from the study, and the researchers' contact information. At the beginning of the interviews, the participants were provided with an information sheet containing this information, and signed a written consent form before continuing. For the online survey, this information was provided on the first page, and participants were asked to indicate their consent by continuing to the next page of the survey.

Table 1: Interview participants: *(Expertise level: High(H), Medium(M), Low(L))*

| ID | Job Title | Expertise | SOC ID |
|---|---|---|---|
| *A1* | Analyst | - | A |
| *A2* | Engineer | - | A |
| *B3* | Lead Analyst | H | B |
| *B4* | Lead Analyst | H | B |
| *C5* | Incident Responder | H | C |
| *D6* | Engineer | H | D |
| *E7* | CISO | H | E |
| *E8* | SOC Manager | H | E |
| *E9* | Analyst | L | E |
| *E10* | L3 Analyst | H | E |
| *E11* | Analyst | M | E |
| *F12* | SOC Manager | - | F |
| *F13* | Engineer | - | F |
| *F14* | Unix, UTM Coding | L | F |
| *F15* | Engineer | M | F |
| *F16* | SIEM Engineer | - | F |
| *F17* | Analyst | M | F |
| *G18* | Lead Analyst | H | G |
| *G19* | Manager | H | G |
| *G20* | Monitoring Analyst | M | G |
| *G21* | L3 Analyst | H | G |

**Participant Recruitment—** Recruitment is a challenge; participation in such studies can burden analysts, making them take time away from their tasks when assessment is based on the daily number of closed tickets [40]. Hence, our sampling was not random. We found that obtaining senior management engagement and access to participants was more important than a random sample which might have reduced bias in the results. We leveraged researchers' institutional relationships and contacted senior-level security professionals in organizations that have an SOC. We asked these professionals to forward the online survey to security practitioners within their SOC. Some organizations asked to review the survey questions before forwarding them to their practitioners. We then liaised with these senior-level professionals to arrange interview dates with a convenient sample of their security practitioners. Gaining acceptance at a senior level helped in establishing trust with the participating analysts (essential in ensuring data validity [40]), and encouraging their participation.

Since responses to the survey were anonymous, we do not know how many distinct SOCs were represented. Our interview recruitment reached seven different SOCs

Table 2: Demographics of SOCs in interviews: *(Organization size: Large(L), Medium(M), Small(S))*, *Gov:* Government SOC)

| ID | Type | Size | Sector | Gov? | Country |
|---|---|---|---|---|---|
| *A* | MSSP | L | Aerospace | ✓ | UK |
| *B* | MSSP | M | Security | ✓ | UK |
| *C* | MSSP | L | Security | | Bulgaria |
| *D* | MSSP | M | Security | | UK |
| *E* | Internal | L | Defence | ✓ | UK |
| *F* | Internal | M | Security | | India |
| *G* | MSSP | M | Security | | UK |

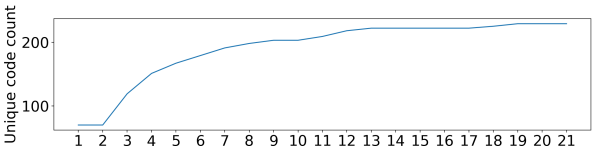


Figure 1: Saturation analysis of new concepts with each additional interview, across all codes

of various sizes serving government and non-government organizations, mostly in the security sector, with one in Defense and one in the aerospace industry. We show the demographics of our interview participants and SOCs in Tables 1 and 2, respectively, and survey participants in Appendix A.2. In both studies, there was a fair level of distribution of participants across SOCs, meaning that our results are not dominated by any particular organization; however, as we note in Section 9, some biases may exist as a result of the high concentration of SOCs within particular sectors and regions. Compensation for participation was not provided. Our anonymization of online survey results prevents us from determining the overlap between its participants and the interviewees.

Saturation analysis of new concepts with each additional interview is shown in Figure 1. This plot relates to all codes collected across a wider data collection effort that included the themes reported in this paper along with other themes such as data presentation in SOCs, and recruitment and skills challenges, for example. After the 13th interview, we observed only very few new codes, and none after the 19th (across all codes, and therefore also in the specific codes we focus on in this paper). We therefore decided not to schedule new participants.

## 3.1 Quantitative Method: Online Survey

We used an online survey to identify areas of focus for the interviews. The online survey enabled remote participation, improving our reach. We cannot claim statistical significance, but were able to identify areas of focus for the semi-structured interviews (we show how we derived the interview research questions from the survey findings in Section 4).

We presented survey participants with a set of assertions on the human-in-the-loop, the prevalence of FP alarms, and the importance of maintaining awareness of the network. Participants were asked to indicate their level of agreement with each assertion using a Likert scale (1: "Strongly Disagree", 2: "Disagree", 3: "Neutral", 4: "Agree", 5: "Strongly Agree"). Mode or median values higher than three constituted overall agreement with an assertion. We also calculated a comparison of non-neutral scores (CNNS), which represents the ratio of scores less than and greater than the neutral value (3). We discuss the survey questions in Appendix A.3.

We drew on existing literature to design our online survey questions, identifying theories requiring validation and constructing questions based on them. Examples of such literature includes work by Garcia et al. [17] and Goodall et al. [18], which explored challenges in intrusion detection systems such as the prevalence of FPs. Contradictory findings on security alert accuracy between [1], [24] led us to use Assertions (A) 1-4 to explore FP/FN perceptions. Assertions A (5-8) were devised based on reports of analysts' use of intuition in detection and decision-making [41], and A-9 and A-10 were devised to explore participants' interactions with security tools, supported by prior literature on the subject [42].

We followed best practice in online survey design [36], and incorporated feedback from subject-matter experts (a security analyst, senior security professional, and an SOC manager). We sought to keep the survey to a reasonable length (it took approximately 15 to 20 minutes to complete). We consulted with other survey-design experts to ensure we avoided ambiguity or double-barrelled questions through careful wording, and to ensure the response categories chosen were appropriate.

## 3.2 Qualitative Method: Semi-Structured Interviews

The qualitative study aimed to investigate further the processes involved in SOC work (including the kinds of activities security practitioners engage in daily, the tools they use, and the skills required), as well as the factors that influence these processes, in more depth. We drew on the findings from the quantitative study to derive research questions, and designed interview questions to address these research questions. In Section 4 we present these research questions and explain how they were derived from the survey results.

We chose to conduct semi-structured interviews to extend discussions based on the flow of conversation. The interview questions were discussed with three subject-matter experts (who worked, or had previously worked,

Table 3: Online Survey Results: Responses to Assertions (Resp, Ordered from "Strongly Disagree"(=1) - "Strongly Agree"(=5)) Mode, Median, and Comparison of Non-Neutral scores - Disagree (1-2):Agree (4-5)(CNNS: D:A)

| Assertion | Resp | Mode | Median | CCNS |
|---|---|---|---|---|
| **A-1:** The monitoring tools I use frequently produce false positive results (they detect a security event when there was not actually a security event) | 0,2,4,11,3 | 4 | 4 | 2:14 |
| **A-2:** The monitoring tools I use frequently produce FNs (fail to detect a security event that occurs) | 0,6,9,3,2 | 3 | 3 | 6:5 |
| **A-3:** I believe that current IDS are inadequate in detecting attacks | 2,5,7,5,1 | 3 | 3 | 7:6 |
| **A-4:** The number of alerts generated by most IDS are overwhelming | 0,4,4,8,4 | 4 | 4 | 4:12 |
| **A-5:** It is important to have a human in the loop for the detection and preliminary analysis of potential security events. This process cannot be carried out by automated systems alone | 0,2,0,7,11 | 5 | 5 | 2:18 |
| **A-6:** Human analysts monitoring the network are capable of detecting network anomalies missed by automated systems | 0,1,5,10,4 | 4 | 4 | 1:14 |
| **A-7:** I am often required to make decisions on the accuracy of alerts produced by automated systems | 0,0,5,8,7 | 4 | 4 | 0:15 |
| **A-8:** I sometimes rely on my experience and intuition to detect attacks rather than monitoring system alerts | 0,2,7,7,4 | 3 | 4 | 2:11 |
| **A-9:** Maintaining awareness of the network security state is important in enabling me to make decisions on the accuracy of alerts produced by automated monitoring systems | 0,0,3,13,4 | 4 | 4 | 0:17 |
| **A-10:** Keeping up with changing configurations in the network is difficult, but necessary to provide the context needed to analyze and diagnose an alert | 0,1,4,9,6 | 4 | 4 | 1:15 |

in SOCs), and their feedback was incorporated to ensure face validity [29]. We show the interview questions in Appendix A.4. We ran pilot interviews with a security analyst, gathering feedback on the questions to ensure their clarity and suitability for the target audience.

The majority of the interviews were carried out face to face at the practitioners' organizations, in rooms outside of the SOC. Two participants were interviewed through live video chats. The interviews were audio-recorded and lasted approximately one hour each; however, due to the nature of the job, one interview was interrupted multiple times for the analyst to deal with an incoming incident and therefore lasted longer. In addition, one organization (SOC ID: F) opted to have researchers interview multiple security practitioners at once. We started the interviews with a brief introduction of ourselves and the study objectives. Participants were then provided with the participant information sheet, and indicated their consent by signing the consent form. Two researchers conducted the interviews to ① enable consistency of the data collection process, ② mitigate the risk that an interviewer would bias participants by asking leading questions, ③ obtain multiple perspectives enabling peer reflection at a later date, and ④ ensure that all questions were covered. However, not all questions were discussed in depth, depending on the participants' position in the SOC.

**Data Analysis—** The audio-recorded interviews were transcribed, resulting in textual data of 105,523 words. We ensured the ethical handling of the data by preserving the anonymity of the participants and their organizations, anonymizing transcripts before analysis and stored with appropriate security protections. We applied Template Analysis (TA) [23], starting with an a priori set of themes we were interested in, allowing the code to evolve with the addition of newly arising themes. The Template Analysis approach was chosen over Grounded Theory as TA themes are less prescriptive, providing the ability to identify and add new concepts if discovered while allowing us to have preconceived theories [23]. Related work has applied TA [42] to analyze the qualitative data of their study on SOCs. As there are still few studies on SOCs, TA is useful due to our partial understanding of the concepts that need to be identified in the data [42].

Two researchers initially coded five interviews and identified parts of the transcriptions that were relevant to the specified themes, assigning an a priori theme code to them. When an interesting part was encountered that did not have a matching theme code, a new theme code was created or an existing theme was broadened. We used the codes that arose from the subset of the data to produce the initial template. The template was hierarchical, with additional sub-themes included within each theme.

The lead researcher then applied this initial template to the rest of the interviews, modifying the codes as necessary until a final template was generated. In developing the final template, the lead researcher engaged in frequent discussions with other team members, to ensure the quality of analysis and that personal beliefs and biases did not affect interpretations. Using the final template, we interpreted the data and wrote our findings.

## 4 Quantitative Findings

Twenty analysts completed our survey, as shown in Appendix A.2. The survey was not intended to identify statistical significance; instead, it was used to focus the semi-structured interviews, identifying pain points and priorities that might have been missed in prior work, to improve the comprehensiveness of our contributions. We present in this section a summary of the findings that drive the qualitative study design.

**Prevalence of FPs**— Our pilot study experts ex-

pressed a distinction between *alerts* and *alarms* in SOCs; security tools produce the former while SIEMs generate the latter as a result of the correlation of multiple alerts. We clarified this distinction in the survey questions.

90% of the participants reported that they use IDS, 80% use a SIEM and 55% use data/log Aggregation Tools (e.g., Splunk). Only two participants reported using machine learning-based tools. 45% of analysts reported that they receive less than 5K alerts daily. Participants receiving over 100K alerts were from large enterprises with fewer than 20 analysts working in the SOC, 40% of which served government customers. From these alerts, some analysts (n = 10) shared the number of alerts they processed daily and the proportion of these that they found to be legitimate. For example, one participant indicated that one out of every 100 alerts investigated is an actual threat, while another stated that in every 200 alerts, 50 were found to be legitimate.

**Assertions**— We show the responses to the Assertions in Table 3. 55% of analysts reported reliance on their tacit knowledge and experience (more than 51% of the time) and understanding of the monitored network in their job (A-8). Participants agreed that they are often required to make decisions about the accuracy of alerts produced by tools (A-7). Based on this, we expected analysts to be unhappy with tools such as IDSs; yet the results show that respondents were undecided on IDS adequacy (A-3). They did agree that they find the number of alerts generated by most IDSs overwhelming, however (A-4). Analysts strongly agreed on the importance of the human role in detection, filtering FPs, preliminary analysis of events and detecting network anomalies missed by automated systems (assertions A-5, A-6). 58% indicated that they process the security alerts based on awareness of regular network activity while 47% said they did so based on the alert severity rating.

### 4.1 Deriving Interview Research Questions

Researchers have used FPs as a metric for evaluating system performance when proposing security tools, seeking few FPs for optimal performance. In recent work, Kokulu et al. [24] found that SOC analysts do not consider FPs in automatic malicious activity detection a significant issue in SOC operations. In contrast, our quantitative analysis indicated that practitioners do, in fact, experience an overwhelming number of alerts which include frequent FPs (A-1, A-4). We therefore focused our interview Research Questions (RQs) on exploring this topic. Assertions A-3, A-7, A-8, A-9, A-10 indicated perceived limitations of alerts, meaning analysts need to decide their accuracy; RQ3 and RQ4 explore these limitations further and how the quality of tool alerts and SIEM alarms could be improved.

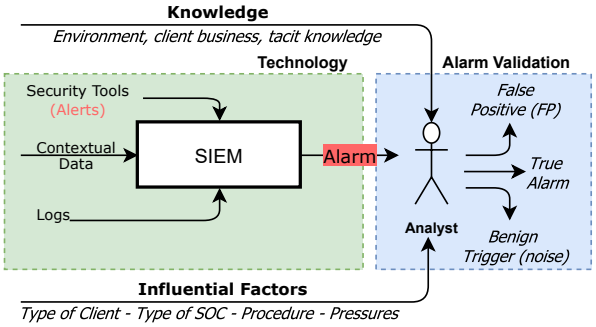


Figure 2: Alarm Validation in SOCs

- **RQ1:** How do SOC analysts distinguish true alarms from false alarms?
- **RQ2:** What do analysts perceive to be FPs, and how can we establish a more precise definition?
- **RQ3:** What are the limitations of alarms produced by existing SOC tools?
- **RQ4:** How can we design better tools to improve the alarms' quality and filtering of FPs?

## 5 RQ1: Understanding Alarm Validation

Our interview analysis revealed a similar process across the participating SOCs in alarm validation, shown in Figure 2. SOC operations rely on alerts from security technologies (e.g., NIDS, HIDS), logs derived from various organizations' systems and networks (e.g., proxy logs), and contextual data stored in the Knowledge Base (e.g., vulnerability scan results) collected into one system—the SIEM. Using these data sources, the SIEM generates an alarm evaluated by the analyst based on their knowledge of the organization, network, and tacit knowledge built from experience. The analyst then needs to make a decision on whether it is a true alarm or a FP, a process called alarm validation.

Similar to our survey findings (assertions A-5 to A-8), the interviews revealed the magnitude of human involvement throughout SOC processes. Although technology might detect and generate the alarm, it is still the analyst who needs to evaluate that alarm to determine its validity (A1, A2, B3, B4, C5, E7, E10, G18, G19, G21). In doing so, analysts rely on their human cognitive abilities (e.g., pattern matching, association, reasoning, and computation). As A2 remarked: *"I think that that's where the human element still remains because even when you get an alert, the alert will have to be sent to a human to make that intelligent decision."*

Although tools such as SIEMs may assist analysts in correlating alarms, in most cases analysts are more capable of connecting these patterns themselves to detect threats than SIEMs. For example, when discussing their validation of an incoming alarm, analysts mentioned analytical questions that are a result of their reasoning, such as "Is the activity ongoing?" and "What processes were running on the server?", the answers to which they would find using the existing technologies (e.g., SIEM).

There is also reliance on practitioners to configure the security technologies, such as defining the SIEM use cases/correlation rules (A1, A2, E7, E8, F13, F14), identifying data sources to collect into the SIEM (B3, B4, E8), and base-lining and tuning (A2, F14, G19). The SOC practitioners' configuration of the security technologies dictates the threats it will detect and the alarms it will generate. SOC practitioners spend time understanding the monitored environment, and its potential threats, to develop use cases. A use case is defined as a "*Specific condition or event (usually related to a specific threat) to be detected or reported by the security tool*" [9]. Analysts design use cases configured into the SIEM to generate an alarm on the detection of particular scenarios. A1 explained: *"The kind of rules, 'what is it I'm looking for?' [...] still initially has to be set up by a human, so there's definitely room for the human still."*

Alarms are periodically reviewed by the analysts to *tune* the defined alarms further and eliminate FPs. Similarly, through understanding the organization's regular environment use (i.e., base-lining), security tools' parameters (i.e., thresholds) are configured accordingly. G19 explained how SOC practitioners review alarms with the customer to check that they are within the customer's tolerance. G19 explained: *"We'll then dig into those and look for, 'is there noise in there?' We will look at the tickets that were generated for the customers, and how many of those will come back from the customers as 'well we sort of want to know about this but we sort of don't."*

Such configurations and tuning are critical to ensure that only true alarms are reported. Unfortunately, most of such tedious configurations fall on the analysts themselves.

Similar to any decision-making process, our analysis revealed this reasoning is impacted by multiple factors, hence, affecting their decision-making during alarm validation. We discuss such aspects in the following.

**Type of Customer—** Participating SOCs that serve public/government customers reported that the budget is a limitation (A1, A2, B3, E9, E11). Consequently, this introduces challenges in acquiring the latest security monitoring technologies or changing existing, expensive monitoring solutions. Participant A1 explained when prompted about why their SOC does not have a SIEM: *"We are a public sector body who move very, very slowly, so when new developments and technology come along, they don't necessarily get deployed straightaway."* Analysts in that SOC need to access the technology logs directly rather than have a SIEM to aggregate them, slowing down their detection and triaging. Hence, such limitations that hinder the adoption of new or more reliable tools may lead to untrue alarms.

As we found in the survey results (A-9, A-10), the analyst's knowledge of the monitored environment and its typical behavior impacts how they triage alarms. Such knowledge is built over time. Due to the sensitive nature of some customers, working in a SOC that serves such customers requires the analyst to have a *security clearance*. As our participants highlighted, this introduces many challenges to SOC operations, particularly skill recruitment (A1, E7), tools used (A1), information access (A2), and how incidents are reported and handled (E8). For example, obtaining access to a customer's network diagrams that may be "*classified*", but are vital to validate any alarms, adds to the complexity of the practitioner's job as explained by A2: "*It depends on the security classification as well, because we're working in a security tier system [..], you could be an analyst but they are setting their diagrams, you will never ever see it. So, that adds complexity into them.*"

**Type of SOC—** The type of SOC (internal or MSSP) has a significant impact not only on the procedures and processes it follows but also on analysts' decision-making (A1, B4, D6). For example, as B4 explained, MSSPs monitor and report any detected alarms to the customer while it is the customer's job to decide how to handle the incident. *"From our perspective, we're really monitoring, alerting and notifying customers. We don't have the authority to shut down communications or do any interaction on the network."*

However, monitoring for such a customer introduces pressure on the analysts to raise every possible suspicious alarm to avoid appearing incompetent or receiving possible fines (A1, A2, D6, E7). As D6 remarks: *"Being analysts, most of them are afraid not to raise them to the customer. They do tend to raise quite a lot of alarms to the customers, obviously to be on the safe side."* A1 also stated: *"The security teams are there to maintain availability above all else because it's when it's not available to the [customer] that you start seeing fines..."*.

Most SOCs have predefined procedures called a "*playbook*", which details steps the analyst follows in dealing with a security alarm. When monitoring multiple customers, MSSP SOCs rely on such documentation that describes each customer's requirements for dealing with a security event as explained by G19: *"So we have a playbook for internal events just like we have a playbook for external events, and the people that we'll interact with*

*are primed in the same way, so they know what we will be telling them during an event, and the same for escalation paths as well."* However, in in-house SOCs, such documentation may not be as critical as the SOC analysts monitor one network/customer. The analyst can directly communicate and discuss with colleagues and senior members on triaging an alarm as stated by E7: *"It's much better, quicker because our SOC is only the size of that bit of the room [approx 8x6m] Tom [analyst] turns around, and John [SOC manager] sat 3 foot away and says John, what do I do about this?' "*

**Knowledge of Monitored Environment—** Analysts' knowledge of the monitored environment is vital for proper validation of the SIEM alarms, which is achieved by experience and obtaining access to logs (A2, B3, B4, D6). As explained by A2: *"To be able to make a decision as to whether an event or an incident is a false positive or not, it comes down to knowing your environment"*.

Although obtaining the data sources from an organization for an in-house SOC might be achievable, an MSSP SOC is restricted by the customer's data provided to them according to the service level agreement (SLA). Limited access to important logs hinders the analysts' ability to make an informed decision on an incoming alarm. As B4 noted: "*For the most part it's about understanding the network topology, [...] It depends on how much the customer wants to share but the more they share, the better a job we can do.*"

Oftentimes an alarm may result from a benign change that occurred in the network/system. In such cases, the analyst needs to be aware of these changes to avoid tedious false alarm triaging (A1, A2, E7, G19). Adequate *Change Management* processes, where changes in systems/networks are documented, need to be in place. Such changes may not be properly communicated to SOC practitioners, an issue found in both internal and MSSP SOCs. A1 provided an example of how not communicating network changes led to wasting SOC team efforts. A1: *"We were seeing a huge amount of ICMP traffic that we don't normally see [...], actually, the networking team sit right next to us so you'd think they would be all over anything to do with ICMP, they have no idea. It turned out that they deployed a new security tool into the network and nobody had told us."* Similarly, G19 highlighted that some customers are not usually aware of the change within their network/systems until the MSSP SOC questions them, as G19 remarked: *"So usually we will highlight something to the customer saying "this server has gone quiet, have you changed it?" And they will go "let's go and check" because we will engage with the security side of the business"*.

**Summary—** Our findings highlight the overwhelming number of FPs generated by security tools and the reliance on humans (analysts) in the tool configurations, alarm tuning and base-lining. Alarms generated by these tools are then validated manually by analysts, a tedious process that is impacted by multiple factors (e.g, type of SOC, client). SOC operations are far from being automated, and humans use their cognitive abilities and knowledge to determine if an alarm is an FP or a true alarm. However, what do analysts perceive to be a FP? Is an alarm due to a misconfiguration considered to be a FP? We discuss the definition of FPs in the following.

Table 4: Strengths and Weaknesses of IDS and SIEMs as Reported by Interviewees

| Features | Participants |
|---|---|
| **IDS Strengths** | |
| Good first indicator for an attack | B3, D6 |
| Ability to deal with high volumes | C5, E7 |
| Easy and fast signature writing | A1, A2, G19 |
| **IDS Weaknesses** | |
| Mostly false alarms | A2, D6, B3, G19, C5 |
| Unreliable signature | A1, B3 |
| Black-box | B4, C5, D6 |
| Loosely written | A1, A2, G19 |
| Lack of detection of new threats | A1, B3, E10, G21 |
| Lack of context | B4, C5, G18, G19 |
| **SIEM Strengths** | |
| Visibility | B3, E8, F15 |
| Custom SIEM correlation rules | B3, C5, D6, E7, E8, G18 |
| Cross-event, Cross-platform | A2, B3, C5, E7, E10, F15 |
| Normalization | B3, D6, E10, F12, F15, G18, G19 |
| Prioritization | A2, A1, B4, D6, E10, F15 |
| **SIEM Weaknesses** | |
| Overwhelming data amount | B3, B4, E8 |
| Reliance on analysts for FP filtering | B3, B4, E8, F17, G18, G19 |
| Use of structured datasets, Time to retrieve query results | E9, E11, G21, G18 |
| Cost | F15, C5, A1, E8 |

## 6 RQ2: Definition of FPs

When describing the overwhelming number of alarms received, B3 quantified it as 99%, stating: "*We know 99% of the alarms we generate are false positives, but we still have to look at them.*" Such dissatisfaction of the number of alarms were expressed by multiple analysts, as found in our survey results (A-1, A-4).

On the other hand, analysts were neutral on IDSs' inadequacy in detecting attacks (A-3). Therefore, we followed up with interviewees on these conflicting assertion results, which prompted an interesting discussion on the perception of FPs. As a result, a number of themes emerged during coding, describing an alarm as being "noisy", "ignored", or the result of a "benign trigger". For example, if an alarm was produced by a security tool, but the customer is aware of them and its origin but chooses to ignore it for a business reason, should it be classified as a FP? Such alarms' classification as FP

depends on the analyst's perception of a FP, as D6 explained: *"Whether they're accurate and a FP, now, it depends on how you perceive a FP. So if we raise an alarm to the customer, and they're aware of it, they will just call out, we are aware of it, but it's a FP in that sense."*

This phenomenon was referred to by C5 as a "benign trigger." False alarms are used to describe an alarm being generated without an actual security-related event (the boy who cried wolf). In contrast, benign triggers are real alarms, meaning they match an existing signature but the organization chooses to ignore them for a "business-justified" purpose. C5 remarked: *"You have a very high number of events, and it doesn't mean they're a FP, but it means many are again triggered by benign triggers. Which means the condition is perfectly matched, and the filter as such works, but the circumstances are completely legitimate. This is benign, because the purpose is business-justified, and it's not malicious, it just manifests in the same way as particular malware would."*

C5 also noted that most alarms produced in an SOC are benign triggers: *"The benign trigger probability is usually very high, from hands-on experience. So is the volume itself, across the IPS/IDS vendors."* G19 also described such an occurrence where tools are performing as they were designed, but generating unactionable alarms as the tool being "noisy." G19 provided an example of a benign trigger due to outdated Java versions, explaining: *"Snort signatures, we have particularly noisy — well I say they're noisy, they're always doing exactly what they should do, they're always identifying vulnerable versions of Java, but a lot of companies have a lot of vulnerable versions of Java, so we get a massive influx of it."*

**Summary—** Participants showed a distinction between what they consider an FP and what they consider to be *noise* or a *benign trigger*. The former is a metric to describe false alarms due to the tool's low performance, while alarms that organizations choose to ignore for a business justification or due to how the network or system is configured are benign triggers/noise. Hence, when evaluating the performance of a system deployed in a real-world setting, using the term False Positive gives the impression that the technology itself is fundamentally flawed. When the analyst reported a 99% FP rate, this is found to be mostly benign triggers and not necessarily a measurement of the performance of the technology itself.

# 7 RQ3: Quality of Alarms

During the analysis of the interviews, we identified a consistent theme emerging where participants, when prompted to discuss security tools' strengths and limitations (summarized in Table 4), would express their frustration at the quality of the alarm. We discuss these limitations in the context of alarm quality in the following.

## 7.1 Alarm Reliability & Customizability

When asked about the limitations of security tools, the main faults reported by interviewees for IDSs is the high volume of FPs (A2, D6, B3, G19, C5). As F17 explained: *"I think the tedious part is FPs that we deal with"*.

IDS accuracy, and therefore the usefulness and reliability of alarms that help analysts take action, is highly dependent on how IDS filters, rules and signatures are written. For example, in the case of malware signatures, those that incorporate IP addresses as a filter are not reliable, as malware is prone to changing its domain. B3 explained: *"A lot of the network-based stuff I don't find very reliable, just because a lot of the IOCs are based on IP addresses which change, based on domains which get shut down."*

Likewise, A1 explained that a poorly written signature would result in many alarms that analysts can not review, rendering the signature not useful: *"It's kind of an internal battle of having a signature for the latest threats and having a useful signature because if it's just constantly firing, nobody's got the time to review all of them, so it has to be well written and they're not always."*

Signatures are not designed to consider benign triggers. Some organizations have certain benign conditions that could be easily flagged as malicious by security tools, as G19 explained: *"Some of the signatures are very poor, so they will look for the words 'select' and 'from,' in clear text, in a packet, but it could be 'from' and 'select' in a packet, there's no context applied, or 'drop tables' a classic one, so we have a customer who sells a [drop table], I'm assuming it's a fold-down side table, every time they sell one of them it fires an alert, but it fires it on the SQL injection alarm, so you can't turn that signature off, as poor as it is."*

Some analysts reported that signatures are "loosely written", meaning written to be general to capture a variety of activities. This results in a high volume of FPs, as A2 explained: *"Some signatures are written very loosely in order to allow it to capture a wide range of activities and those are some of the downfalls of using it."*

As the malware evolves and generates various attack behaviors, signatures can be written to be too broad to capture all these possible behaviors. D6 also explained how malware signatures in some tools are engineered to capture only part of the malware activities. As a result, when that malware signature is triggered, the analysts are left confused about the kind of malware that triggered it. A1 remarked: *"Malware can change so quickly now that, yeah, a lot of their signatures can be very loose and very broad to try and capture every possibility."*

The correlation of multiple data sources to generate an alarm is a strength of SIEMs, as reported by participants:

B3, D6, G18, C5, E7, E8. For example, D6 described the SIEM correlation capability to *"help paint a better picture of a customer"* and to *"help perform a better analysis"*. Such correlation alarms were carefully designed by the analysts to fit a specific client use case. B3 also highlighted that alerts generated by IDS/IPS tools are not reliable on their own, but, correlating them with alerts generated by other platforms provides more confidence in its validity. B3 remarked: *"Any one source on its own, I don't find that useful. That is why most SOCs now, the core bit of technology is a SIEM..."*

SOCs usually have a relatively extensive set of custom alarms (i.e, correlation rules), which the analyst is expected to react to immediately. The analyst finds these rules effective, as they design them specifically for the monitored organization, as B3 explained: *"We do correlation rules like that. It proved quite effective, just because you are generating that event yourself."* Such alarms are "respected" by analysts as explained by C5: *"It's about crosslatform, cross-event source correlation, that we can actually have high respect for."* SIEM correlation rules and signature-based systems' customized signatures are strengths, reported by A2 and G19. For example, G19 explained that if there are specific known traits within the environment that they want to monitor, they define a signature for it: *"In terms of the good features, they have capabilities for you to deploy your own customized signatures."*

**Summary—** Participants described alerts produced by traditional security tools (e.g., IDS/IPS) as unreliable for many reasons: (1) their reliance on features that change (e.g., domain names); (2) signatures written to deal with new threats quickly, thus not adequately designed and not reviewed later due to lack of time; or (3) signatures loosely or too broadly written, to cover multiple kinds of attacks. Analysts find SIEM correlation rules more useful as they incorporate multiple indicators to generate an alarm. However, as SIEMs rely on alerts generated by traditional security tools in correlation rules, the limitations of those tools will also impact the reliability of the generated SIEM alarm. Each organization's networks and systems are unique, so the customizability of the alarms to fit the monitored environment is a desirable feature. For example, SIEM alarms are described as more useful and effective, specifically correlated alarms due to their customizability where analysts design these alarms themselves specifically for the monitored environment.

## 7.2 Alarm Explainability

One of the disadvantages of commercial security tools is that they are closed systems (i.e., black-box (B4, C5, D6)), which means that the analyst receives an alert but does not know the reason for the detection. In cases where the security tool produces a "reason" or description of the alarm, they are incomprehensible or are short descriptions that do not *explain* why the alarm was produced, leaving analysts to decipher it themselves. This requires them to research and gather more information about the alarm using, for example, the filter name to understand the reason behind an alarm. C5 remarked: *"Sometimes it's a generic filter, you have a one-sentence description, this is very often the case, and you don't know what logic they put inside, because that's proprietary enclosed information."*

Such a lack of clear description of the reasoning behind an alarm forces analysts to spend time evaluating its validity, reducing their productivity. Moreover, the lack of explainability causes alarm desensitization. As C5 explained, analysts lose their "trust" in the validity of the alarm: *"So that's one issue of traditional IPS/IDS, that we also don't really have high respect for, because if you don't tell me the reason you fire, you're not ready to open up the reason you produce an event for me, and I have only a very high level of description that may not be usable, it's not usually usable, let alone actionable, then I can't have really high respect on that."*

Similarly, B4 explained that ambiguity and an inadequate description of alarms leads to uncertainty about the existence of an attack, resulting in no action from analysts. Hence, analysts are required to inspect the raw data (e.g., network packets) to determine the cause of the alarm. B4 remarked: *"Because without it you're pretty blind because if you see the content that a signature wants to match on plus a few bytes after [...] you can't really confirm whether something bad has happened or not."* Such a manual inspection of the packet might not be possible with encrypted traffic, as G19 noted: *"If that traffic is then encrypted, you might get the alarm but you can't see the raw text."*

One of the strengths of SIEM platforms is their ability to provide analysts with visibility on the monitored environment, as the more information they can view the better their situational awareness [6]. Although attractive for comprehensive monitoring, this might result in a vast collection of data and alarms that overwhelm the analysts. B4 remarked: *"Downsides may be sometimes too much information."* For this reason, it is best practice when deploying SIEMs to design properly the "correlation rules" and collect data sources needed only for these rules to avoid the problem of "too much information", as noted by E8, F13, F14.

**Summary—** Commercial security tools oftentimes are black-boxes, generating unexplainable alarms or those with short descriptions that require analysts to research further in order to make an actionable decision. In situations where analysts are evaluated by their per-

formance and the number of tickets they close [40], such a monotonous task reduces their productivity. Analysts described such unexplainable alarms as "untrustworthy", "unuseful", and "poor". Although SIEM systems are a repository of data that allows analysts to "fast query" based on their own human analytical reasoning, they still provide too much data that overwhelm analysts. Alarms generated from correlation rules, designed by the analysts to fit the organization's environment, include only the "bare minimum" of data sources needed, and are thus more respected by analysts.

### 7.3 Alarm Contextuality

The participants identified a lack of context in alarms generated by IDS (B4, C5, G18, G19). As we discussed in Section 5, context such as knowledge of the asset and network helps practitioners eliminate FPs and determine the path of investigation. This knowledge spans, for example, the network topology and devices, what these devices are used for, and their location and owner. For example, knowing an asset is a Windows machine helps when receiving an alarm for a Linux signature. As D6 expressed: *"Again, the topology of where the server is based is important"*.

Not only do analysts need to know the network they are defending, but they also need to be aware of the customer's business. Knowledge of the customer is critical in many aspects in the SOC operations, from ruling out FPs to deciding on how to respond to a threat. D6 remarked: *"It's not just about what you see in the security events, or in the tools, it's about what you know of the customer and the customer's nature—business nature."*

One participant provided an example of how valuable such knowledge could be for determining an FP. After detecting substantial outgoing connection from the customer to another company and spending time investigating, they researched whether there was any business between these two companies. They found out that their customer had recently acquired the other company. Similarly, E10 remarked on how knowing the customers' working hours helped in filtering an FP: *"I use open-source intelligence to build up a view of the customer first, [...]. So I'll figure out like if they are in Dubai or wherever that they might not be doing the same working hours [...], because we have been caught out by that one before, there is no traffic on a Friday."*

Analysts might acquire knowledge from third parties, such as ISPs, security vendors, and the security community to investigate threats. Security vendors are responsible for maintaining their products and providing Indicators of Compromise (IoC) and signatures for new threats. The security community can also assist the SOC process through intelligence sharing, writing blogs about how a threat behaves and spreads. As we found in our survey, 35% of our participants triage alarms according to newly announced vulnerabilities in security blogs. B3 remarked: *"WannaCry was great, because straight away people were publishing blogs, this is how it spreads. This is what you need to look for, and you could go and look for it before there were any signatures for it. We were finding things quite quickly."*

**Summary—** Contextual knowledge about the network and systems can improve alert validation [8, 25] significantly. This was confirmed in both our quantitative and qualitative findings. However, we found that analysts not only rely on technical knowledge (i.e., network and assets) in FP filtering but also knowledge about the business (e.g., work hours) and knowledge acquired from third parties (e.g., ISP and vendors). Incorporating this knowledge with the alarm provides context to expedite the alarm validation process.

## 8 RQ4: Designing Better Tools

One of the study findings is the lack of adoption of ML technologies in our participating SOCs, especially those that serve government clients. As shown in Section 4, only two survey participants (10%) indicated they use AI/ML security monitoring tools. Hence, we focused the interview questions on the most-used SOC tools (e.g., IDS, SIEM), but still probed analysts during interviews on why AI/ML security tool adoption is low. Analysts reported reasons such as: ① their skepticism of ML applications in commercial tools (B4); ② the absence of accreditation bodies attesting the correctness of the ML model design (e.g., not authorized to be deployed in government networks) (A1); ③ Absence of historical organization data to train the ML models (E7); ④ The prevalence of FPs (B4). As B4 explained: *"Machine learning itself would learn what normal is but humans aren't exactly going to be normal and do the same stuff every day, so I think that platforms like that themselves have a lot of false alarms."*

The latter reason is supported by the 2019 study by SANS [11], which found an overall dissatisfaction with AI/ML tools due to their frequent FPs and the high levels of involvement required by knowledgeable and skilled analysts. Despite ML systems' high prediction performance, they may not adequately explain the reasons behind these predictions, resulting in a deficit in the quality of the alarm.

Using our qualitative findings on the limitation and strengths of used tools, we derive the following recommendations to address how alarms produced by tools can be improved. As the prevalence of FPs was also a concern for AI/ML security tools, and to foster the better

future design of these tools, we frame these recommendations in the context of AI/ML based tools.

## 8.1 Recommendations

In response to our findings of the high number and low quality of security alarms rendering them unactionable, particularly in identifying FPs, we propose the REACT model. The REACT model proposes five properties that need to be present in a security system to produce an alarm that is "REACTable" by analysts: **R**eliable, **E**xplainable, **A**nalytical, **C**ontextual, and **T**ransferable.

**Reliable—** One of the main findings reported in Section 5 is that both vendors and the researchers should use *false alarms* or *benign triggers* as an alternative metric to FPs, which is general and vague. To increase the ML model performance, incorporating the analysts' feedback on the generated false alarms into the model itself can strengthen future predictions and alarm prioritization. In case of benign triggers, the model can then learn that the prediction made, although true, is not suitable for the organization at hand, adjusting itself accordingly.

Improving on limitations concerning alarms' reliability (Section 7.1), ML models need to be designed with attention to the features used. For example, avoiding features where values frequently change (e.g., malware domain names), or features that are too broad (e.g., Java version). Models should be trained to detect specific malicious activities (e.g., scanning) or a specific type of malware (e.g., worm), avoiding being too broad and attempting to detect everything (e.g., malware).

Promoting an ML system capable of detecting all types of malware might be more commercial, but each malicious behavior has a set of unique features that describe it, and broadening the detection capability of a single model might lead an increase in FPs. However, using limited features to train a model might result in adversaries finding ways to bypass detection (e.g., adversarial attacks). Therefore, using Ensemble learning [33] combines the predictions from multiple niche behavioral models, reducing the variance of predictions' generalization errors while detecting all forms of malicious behavior. Such an approach will help increase system detection *explainability*, a property we discuss in the following.

**Explainable—** Analysts' involvement in SOCs is indispensable, formulating and evaluating hypotheses about security observations based on their domain knowledge, intuition, and knowledge of the monitored environment. Proposed solutions should therefore promote AI-human collaboration by communicating the AI's explainable decisions to humans that provide feedback to the AI model. Black-box AI models do not provide human-understandable insights on their outputs, leaving security personnel unable to evaluate the reason behind these predictions.

Explainability of artificial intelligence (XAI) is a set of methods and techniques that tackle the interpretability problem of AI predictions, promoting the production of predictions that can be understood by human experts. Explainable AI systems can provide explanations for decisions in a human-comprehensible manner, thus keeping humans in the loop. Caution is needed when building AI systems in general; as AI is based on statistics and probability, analysts actually should not trust the system completely. Based on system explanations, the analyst should know when to trust the system's predictions and when to apply their own judgment. Analysts' trust needs to be calibrated using properly designed prediction explanations. Such explanations would help security personnel determine whether the AI's findings merit trust, based on their expertise of the threat landscape and the monitored environment.

Current implementations of AI models are unintelligible to non-AI experts, and technical solutions proposed in the field of XAI require very specific technical expertise [16]. Recent research suggests that implementations of symbolic systems based on semantic technologies such as Knowledge Graphs (KGs) to be a promising solution in providing AI prediction traceability and explainability to non-AI experts [26, 28, 39]. Research on explainability in AI security technologies are still in its early stages. More research is required on security systems that offer explanatory capabilities to non-AI experts, optimize decision-making, an enable a human-machine collaborative environment where security personnel cooperate with AI security systems to detect threats.

**Analytical—** As discussed in Section 5, analysts rely on their tacit knowledge and knowledge of the monitored environment, built through experience and documentation of the customer's system and network-activity norms. Such knowledge allows them to develop analytical questions and reasoning that help them investigate the issue at hand. Research is still far from integrating human cognitive and social abilities into SOC solutions. One emerging field to attempt to tackle this challenge is Cognitive Security [3], a concept that leverages multiple forms of AI, including ML and deep-learning, to uncover human cognitive ability.

KGs can be also be used to capture the knowledge about a security threat landscape, integrating and linking data from different sources, or different types of representation [20]. KGs are natively built to be queried, allowing users to interact with the symbolic system that provides explanations, then using their own analytical reasoning and knowledge [14]. Further research is needed in developing security human-machine solutions where humans and machines work together (e.g., alarm

validation), utilizing humans' analytical capability rather than using full automation. Our study emphasized the reliance of analysts on their experience and knowledge of the monitored environment in their job. However, developing such systems that embody analysts' experience can introduce biases, especially when considering a non-representative sample of security expertise.

**Contextual—** As discussed in Section 7.3, context is important. For example, a SOC tool that monitors for abnormal network activity should consider the customer's work hours to eliminate benign triggers such as the absence of traffic on a Friday. Such context is currently incorporated into technology to some extent. However, challenges such as the use of structured storage, where the volume of data delays query retrieval, limit its implementation. One research opportunity is in investigating how logs and other context data sources can be represented as KGs. In such graphs, knowledge about the customer, network, and knowledge from external entities can be incorporated to provide more context to alarms. Hence, contextual data can be incorporated into the graph to produce contextual alarms. Moreover, such a structure can be used by analysts during the FP filtering process or even hunting. Privacy is indeed a concern in any organization. Therefore, such SOC systems that incorporate context, are usually customized per client, with contextual data contained within the clients' SOC, similar to existing SIEM systems.

**Transferable—** One of the strengths of current SOC tools, discussed in Section 7.1, is customization (e.g., SIEM use cases). Such capabilities allow analysts to tailor alarms to the monitored organization. Each organization's network traffic and asset usage are different, as are the networks and systems governing them. Accordingly, although SOCs may deploy the same technologies (e.g., SIEM), they are configured and used differently and should not be designed as *one size fits all*.

The success of incorporating AI/ML systems in SOCs relies on their capacity to adapt the model to the monitored environments. Such an AI system can be built, for example, using Transfer Learning [31]. Network and system ecosystems change quickly, and therefore labeled data collected at a specific time will also change. Transfer Learning can also help in circumstances where data can easily become outdated, as with network traffic. Due to privacy concerns, Transfer Learning can help organizations reuse a vendor's pre-trained model, building a new model that incorporates contextual features into the commercially trained ML model feature space.

For an AI-based malware network detection system, the data features or data distributions of the organization's network in which it is deployed may be different, or there may be a lack of labeled training data. As a result, directly applying the malware network detection system to the new network might not be straightforward, but transferring the classification knowledge into the new organization's network would be useful. Using Transfer Learning, security vendors can transfer knowledge on malware behavior; organizations can then incorporate contextual information to them and build a customized model that fits their network, producing more effective alarms and reducing the number of benign triggers.

## 9 Limitations and Future Work

We employed a qualitative method, and as a result, our sample size is small ($n$ = 21). SOCs are distinctive, in their size, structure, operations, and personnel, and our study only involved seven SOCs. Most of our participants are based primarily in the UK, with some serving UK-based public and private sectors. Due to our non-random method of recruitment that leveraged institutional relationships to ensure participation and engagement, the participating SOCs were mostly in the security sector located in Europe. Given that many of the SOC processes described by our results are defined by the organization, such results are biased towards this industry/region. Future work should seek to explore practice in SOCs in other industries/regions, identifying inter-sector and inter-regional similarities and differences.

The decision to use semi-structured interviews as the qualitative method meant that the level of detail in which different participants discussed each question varied. Self-reported data also have limitations such as recall and observer bias. Furthermore, we report only those themes highlighted by the participants in this study. These limitations prevented us from extracting generalizable conclusions, and therefore, further validation of the study findings is required. In future work, we will evaluate this study's findings with further SOCs and participants residing in other countries or serving organizations of various sector types. In this work, we plan to conduct a quantitative study to measure analysts' perception of the definition of false alarms and benign triggers.

## 10 Conclusions

In this paper, we focused on understanding how SOC practitioners validate an alarm (i.e., determining whether it is a true positive, false positive, or a benign trigger). We found that adoption of the term False Positive has proved to be vague and a clear distinction when evaluating systems needs to be used. We also investigated the analysts' perception of limitations of the alarms produced by existing security tools. They were found to be unreliable, difficult to interpret, and lacking in the context needed by analysts to filter FPs from genuine

alarms. We elicit from these limitations recommendations for tool vendors and researchers to help improve the quality of alarms, reduce FPs and alarm burnout, and ultimately foster analysts' trust in security tools.

## References


[1] AKINROLABU, O., AGRAFIOTIS, I., AND EROLA, A. The challenge of detecting sophisticated attacks: Insights from SOC Analysts. In *Proceedings of the 13th International Conference on Availability, Reliability and Security* (2018), ACM, p. 55.

[2] ALEROUD, A., AND KARABATIS, G. Beyond data: contextual information fusion for cyber security analytics. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing* (2016), pp. 73–79.

[3] ANDRADE, R. O., AND YOO, S. G. Cognitive security: A comprehensive study of cognitive science in cybersecurity. *Journal of Information Security and Applications 48* (2019), 102352.

[4] ÅRNES, A., VALEUR, F., VIGNA, G., AND KEMMERER, R. A. Using hidden markov models to evaluate the risks of intrusions. In *International Workshop on Recent Advances in Intrusion Detection* (2006), Springer, pp. 145–164.

[5] AXON, L., ALAHMADI, B., NURSE, J., GOLDSMITH, M., AND CREESE, S. Sonification in security operations centres: what do security practitioners think? In *Workshop on Usable Security (USEC)* (2018), Internet Society.

[6] BHATT, S., MANADHATA, P. K., AND ZOMLOT, L. The operational role of security information and event management systems. *IEEE security & Privacy 12*, 5 (2014), 35–41.

[7] BOLLINGER, J., ENRIGHT, B., AND VALITES, M. *Crafting the InfoSec Playbook: Security Monitoring and Incident Response Master Plan*. O'Reilly Media, Inc., 2015.

[8] CHABOYA, D. J., RAINES, R. A., BALDWIN, R. O., AND MULLINS, B. E. Network intrusion detection: automated and manual methods prone to attack and evasion. *IEEE security & privacy 4*, 6 (2006), 36–43.

[9] CHUVAKIN, A., AND BARROS, A. How to develop and maintain security monitoring use cases. Tech. rep., Gartner, 2015.

[10] CICHONSKI, P., MILLAR, T., GRANCE, T., AND SCARFONE, K. Computer security incident handling guide. *NIST Special Publication 800*, 61 (2012), 1–147.

[11] CROWLEY, C., AND PESCATORE, J. Common and best practices for security operations centers: Results of the 2019 SOC survey. *SANS, available at https://www. sans. org/media/analyst-program/common-practices-security-operations-centersresults-2019-soc-survey-39060. pdf (accessed 7th November, 2019)* (2019).

[12] CVACH, M. Monitor alarm fatigue: an integrative review. *Biomedical instrumentation & technology 46*, 4 (2012), 268–277.

[13] DIETRICH, C., KROMBHOLZ, K., BORGOLTE, K., AND FIEBIG, T. Investigating system operators' perspective on security misconfigurations. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (2018), ACM, pp. 1272–1289.

[14] DORAN, D., SCHULZ, S., AND BESOLD, T. R. What does explainable AI really mean? A new conceptualization of perspectives. *arXiv preprint arXiv:1710.00794* (2017).

[15] DUKES, C. Committee on national security systems (CNSS) glossary. Tech. rep., Technical report CNSSI, 2015.

[16] FUTIA, G., AND VETRÒ, A. On the integration of knowledge graphs into deep learning models for a more Comprehensible AI—three challenges for future research. *Information 11*, 2 (2020), 122.

[17] GARCIA-TEODORO, P., DIAZ-VERDEJO, J., MACIÁ-FERNÁNDEZ, G., AND VÁZQUEZ, E. Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security 28*, 1 (2009), 18–28.

[18] GOODALL, J., LUTTERS, W., AND KOMLODI, A. The work of intrusion detection: rethinking the role of security analysts. *AMCIS 2004 Proceedings* (2004), 179.

[19] GOODALL, J. R., LUTTERS, W. G., AND KOMLODI, A. I know my network: collaboration and expertise in intrusion detection. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (2004), ACM, pp. 342–345.

[20] HEATH, T., AND BIZER, C. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology 1*, 1 (2011), 1–136.

[21] HUBBALLI, N., AND SURYANARAYANAN, V. False alarm minimization techniques in signature-based intrusion detection systems: A survey. *Computer Communications 49* (2014), 1 – 17.

[22] JACOBS, P., ARNAB, A., AND IRWIN, B. Classification of security operation centers. In *2013 Information Security for South Africa* (Aug 2013), pp. 1–7.

[23] KING, N. Doing template analysis. *Qualitative organizational research: Core methods and current challenges 426* (2012).

[24] KOKULU, F. B., SONEJI, A., BAO, T., SHOSHITAISHVILI, Y., ZHAO, Z., DOUPÉ, A., AND AHN, G.-J. Matched and mismatched SOCs: A qualitative study on security operations center issues. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (2019), ACM, pp. 1955–1970.

[25] KRUEGEL, C., AND ROBERTSON, W. Alert verification determining the success of intrusion attempts. *DIMVA 2004, July 6-7, Dortmund, Germany* (2004).

[26] LECUE, F. On the role of knowledge graphs in explainable AI. *Semantic Web*, Preprint (2019), 1–11.

[27] MAJ, M., REIJERS, R., AND STIKVOORT, D. Good practice guide for incident management. *European Network and Information Security Agency (ENISA)* (2010).

[28] MUNCH, M., DIBIE-BARTHELEMY, J., WUILLEMIN, P.-H., AND MANFREDOTTI, C. Interactive causal discovery in knowledge graphs. In *PROFILES/SEMEX@ ISWC 2019* (2019), vol. 2465, CEUR-WS. org, pp. 78–93.

[29] NEVO, B. Face validity revisited. *Journal of Educational Measurement 22*, 4 (1985), 287–293.

[30] OLTSIK, J. SOC-as-a-service for midmarket and small enterprise organizations. Tech. rep., The Enterprise Strategy Group, mar 2015.

[31] PAN, S. J., AND YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering 22*, 10 (2009), 1345–1359.

[32] PECCHIA, A., COTRONEO, D., GANESAN, R., AND SARKAR, S. Filtering security alerts for the analysis of a production saas cloud. In *Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on* (2014), IEEE, pp. 233–241.

[33] POLIKAR, R. Ensemble learning. In *Ensemble machine learning*. Springer, 2012, pp. 1–34.

[34] RAFTOPOULOS, E., EGLI, M., AND DIMITROPOULOS, X. Shedding light on log correlation in network forensics analysis. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (2012), Springer, pp. 232–241.

[35] RAMAKI, A. A., RASOOLZADEGAN, A., AND BAFGHI, A. G. A systematic mapping study on intrusion alert analysis in intrusion detection systems. *ACM Computing Surveys (CSUR) 51*, 3 (2018), 1–41.

[36] REDMILES, E. M., ACAR, Y., FAHL, S., AND MAZUREK, M. L. A summary of survey methodology best practices for security and privacy researchers. Tech. rep., 2017.

[37] RILEY, M., ELGIN, B., LAWRENCE, D., AND MATLACK, C. Missed alarms and 40 million stolen credit card numbers: How target blew it. *Bloomberg.com* (2014), 1.

[38] SALAH, S., MACIÁ-FERNÁNDEZ, G., AND DÍAZ-VERDEJO, J. E. A model-based survey of alert correlation techniques. *Computer Networks 57*, 5 (2013), 1289–1317.

[39] SEELIGER, A., PFAFF, M., AND KRCMAR, H. Semantic web technologies for explainable machine learning models: A literature review. *PROFILES 2019* (2019), 30.

[40] SUNDARAMURTHY, S. C., BARDAS, A. G., CASE, J., OU, X., WESCH, M., MCHUGH, J., AND RAJAGOPALAN, S. R. A human capital model for mitigating security analyst burnout. In *Symposium on Usable Privacy and Security (SOUPS)* (2015), pp. 347–359.

[41] SUNDARAMURTHY, S. C., CASE, J., TRUONG, T., ZOMLOT, L., AND HOFFMANN, M. A tale of three security operation centers. In *Proceedings of the 2014 ACM workshop on security information workers* (2014), pp. 43–50.

[42] SUNDARAMURTHY, S. C., MCHUGH, J., OU, X., WESCH, M., BARDAS, A. G., AND RAJAGOPALAN, S. R. Turning contradictions into innovations or: How we learned to stop whining and improve security operations. In *Symposium on Usable Privacy and Security (SOUPS)* (2016), pp. 237–251.

[43] SUNDARAMURTHY, S. C., WESCH, M., OU, X., MCHUGH, J., RAJAGOPALAN, S. R., AND BARDAS, A. G. Humans are dynamic-our tools should be too. *IEEE Internet Computing 21*, 3 (2017), 40–46.

[44] VALEUR, F., VIGNA, G., KRUEGEL, C., AND KEMMERER, R. A. Comprehensive approach to intrusion detection alert correlation. *IEEE Transactions on dependable and secure computing 1*, 3 (2004), 146–169.

[45] YEN, T.-F., OPREA, A., ONARLIOGLU, K., LEETHAM, T., ROBERTSON, W., JUELS, A., AND KIRDA, E. Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks. In *Proceedings of the 29th Annual Computer Security Applications Conference* (2013), ACM, pp. 199–208.

[46] ZHONG, C., LIN, T., LIU, P., YEN, J., AND CHEN, K. A cyber security data triage operation retrieval system. *Computers & Security 76* (2018), 12–31.

[47] ZHONG, C., YEN, J., LIU, P., AND ERBACHER, R. F. Automate cybersecurity data triage by leveraging human analysts' cognitive process. In *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)* (April 2016), pp. 357–363.

[48] ZIMMERMAN, C. Ten strategies of a world-class cybersecurity operations center. *MITRE corporate communications and public affairs. Appendices* (2014).

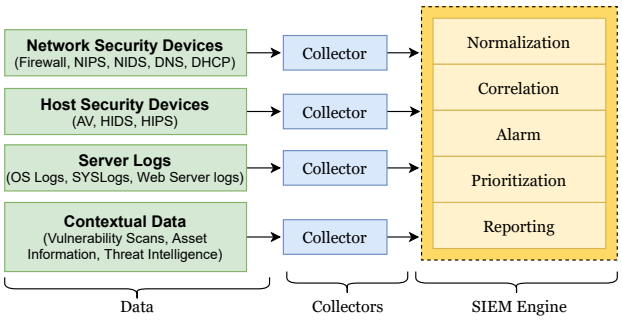


Figure 3: SIEM Architecture.

# A Appendix

## A.1 Background: Security Operations Centers (SOCs)

Security Operations Centers (SOCs) are a centralized unit providing monitoring capabilities for the detection, escalation and recovery of security incidents on an organizational and technical level. Once a security incident is detected, the SOC aims to contain the attack as soon as possible, to limit the potential damage, saving the organization money, data exfiltration or reputation damages.

There are different types of SOCs as discussed in [48], which could be classified by the services they provide, their capabilities or maturity [22]. Moreover, they can mainly be categorized as (1) in-house SOC, meaning the organization builds and staffs the SOC for its organization; (2) Managed Security Service Provider (MSSP), where an organization hires a third party, outsourcing the threat monitoring, detection, and response. Some customers combine these two approaches, building their own SOC but also hiring a MSSP to uplift their skill set by doing joined monitoring. There are several reasons why an organization would use one over the other. For example, their budget, their lack of security expertise, or to avoid the setup costs of a SOC [5, 30].

SOCs consist of complex processes and technology and involves multiple people from within the SOC, organization, customer and other third parties. We discuss these SOC domains in the following.

### A.1.1 People

The SOC team's goal is to detect, analyze, and respond to cybersecurity incidents using a combination of technology solutions and a strong set of processes. To do that, analysts have to maintain Situational Awareness (SA) of events from the systems and networks they monitor. Situational Awareness is defined as: "*Within a volume of time and space, the perception of an enterprise's security posture and its threat environment; the comprehen-*

*sion/meaning of both taken together (risk); and the projection of their status into the near future.*" [15].

SOC practitioners' team structure and responsibilities varies [48]. The SOC team is typically staffed with security analysts, engineers, incident responders, hunters, contractors, as well as managers who oversee security operations. SOC engineers are responsible for providing and supporting the SOC with the required software (e.g., SIEM scripts or configurations). Incident responders handle events that are escalated by the analysts that need in-depth investigation and forensics..

### A.1.2 Process

The SOC process involves the various workflows SOC security practitioners follow in their everyday tasks. For example, this includes process followed for SIEM monitoring and alarming, event management process, security incident ticket management, incident handling, reporting and escalation process. All these processes are documented in a Wiki Portal or a Knowledge Base, share point or share drive for team reference.

There are standard guidelines that direct analysts in SOC operations. For example, The European Network and Information Security Agency (ENISA) [27], and the National Institute of Standards and Technology (NIST) [10] have published guidelines for incident response teams. To provide a rapid automated response for incident identification, detection, response to SOC team communications, procedures are documented in a *playbook* [7], a document prepared by experienced SOC analysts that details steps an analyst should follow to deal with a security alarm.

### A.1.3 Technology: SIEMs

SOCs deploy various technological solutions such as Asset Discovery, Vulnerability Assessment, Behavioral Monitoring, Host/Network Signature-based Intrusion Detection/Prevention Systems, and SIEMS. We discuss in the following the main platform deployed in SOCs.

One of the most frequently chosen tools in a SOC is Security Information and Event Management (SIEM)—e.g, ArcSight [2], AlienVault [3]. We show the architecture of a SIEM in Figure 3.

SIEMs are systems that combine SIM (security information management), and SEM (security event management) functions into one security management system. SEM deals with real-time monitoring, correlation of events, notifications and console views. SIMs provide long-term storage as well as analysis, manipulation and reporting of log data and security records of the type collated by SEM software.

SIEMs replace the need for analysts to access traditional security tools directly. Instead, the SIEM aggregates the logs from the multiple data sources, and processes them to detect threats. Then, the first step is to parse the data collected from the data sources and normalize it to a standard format produced as a security event. Multiple security events may be correlated to create a correlation rule or alarm. When the rule is triggered, an alarm is fired and prioritized. Then, it is up to the analyst to determine if the alarm is a FP or it needs to be escalated. We discuss each step in further detail in the following.

**Data Sources—** The SIEM has data source plug-ins called collectors where data sources are fed into it. These data sources could be either raw logs or security events generated by security devices. Such data sources could be, for example, network-based security tools (e.g., firewalls, IDS), host-based security tools (e.g., Anti-Virus), logs (e.g., operating systems logs, web server logs). Contextual data provided by threat intelligence platforms and other processes (e.g., vulnerability scans) could also feed into the SIEM.

Ideally, these data sources are received from Security Device Event Exchange (SDEE) enabled devices/hosts. SDEE is a standard proposed by the International Computer Security Association (ICSA)[4] that specifies the format of messages and protocols used to communicate events generated by security devices. SDEE enables devices to collect logs in the device itself, and the SIEM retrieves these logs. If the raw logs match a specific criterion, then part of the message is inserted into the SIEM database as a security event. Other devices send the logs directly to the SIEM to store.

**Normalization—** Logs/messages received from SDEE enabled devices are intrinsically suited to the SIEM. They do not require manipulation because they are in the right format. However, some applications/devices were never designed to generate logs. Therefore, they have to be heavily edited by scripts to produce a log that will fit the SIEM's requirements.

**Correlation—** Most alarms in a SIEM are directive alarms, also known as correlation rules. Using correlation rules, SOCs can identify potential security threats by detecting behavior patterns in disparate yet related events. Directive alarms link these different events to generate an alarm that is more useful than any event seen in isolation. If the organization has a particular threat use-case, then they can create their own directives.

**Alarming and Prioritization—** Its worth noting that

[2] https://www.microfocus.com/en-us/products/siem-security-information-event-management/overview

[3] https://cybersecurity.att.com/solutions/siem-log-management

[4] https://www.icsalabs.com/

there is a distinction in the definition of alarm, alert, and event in a SOC operation. Security tools and networking devices produce alerts when they detect a threat. Similarly, these threats might be written in logs as an event. The alerts and events are then aggregated through the SIEM to produce an alarm. When multiple alarms are flagged, a frequent scenario in a SOC, what alarm is looked at depends on what is monitored. For example, if an organization is only concerned about where data is going, then network traffic and IDS alerts alone might be sufficient. However, receiving multiple alerts from different sources related to the same asset provides assurance of the validity of the alarm. How they choose the alarms they investigate depends on several factors, which could be defined using the prioritization module in the SIEM engine.

**Reporting and Visualizations—** Alarms produced by the SIEM are presented to the analyst through visualization (e.g., dashboards). In addition, the SIEM provides reporting capabilities, meaning analysts can autogenerate reports.

## A.2 Online Survey Participants

We show the demographics of our survey participants in Table 5. 20 people filled out the survey. The participants had a mix of demographics (e.g., expertise, years of experience, and position). Most of our participants were from large organizations. Moreover, most participants were from SOCs in the security sector, with three from government and two working in the aerospace industry.

Table 5: Survey Participants: *(Expertise Level: Very High(H+), High(H), Medium(M), Low(L)), (Organisation Size: Large(L), Medium(M), Small(S))*

| ID | Years of Exp. | Job Title | Expertise Level | Gov? | No. analysts in SOC | Org. Size |
|---|---|---|---|---|---|---|
| S1 | 0 - 3 | Analyst | H | | 1 - 9 | S |
| S2 | 0 - 3 | Analyst | L | | 1 - 9 | L |
| S3 | - | Architect | H | | 20 - 29 | M |
| S4 | 0 - 3 | Analyst | L | | 20 - 29 | M |
| S5 | 3 - 5 | Manager | H | | 10 - 19 | L |
| S6 | 5 - 7 | Engineer | H | | 10 - 19 | L |
| S7 | 7 - 10 | Engineer | H+ | ✓ | 10 - 19 | L |
| S8 | 10 - 15 | Engineer | H | ✓ | 10 - 19 | L |
| S9 | 0 - 3 | Analyst | L | ✓ | 1 - 9 | L |
| S10 | 0 - 3 | Architect | H | | - | M |
| S11 | 0 - 3 | Analyst | M | | 10 - 19 | L |
| S12 | 0 - 3 | Analyst | M | | 1 - 9 | L |
| S13 | 3 - 5 | Analyst | H | | 1 - 9 | L |
| S14 | 3 - 5 | IR Manager | H | | 200 + | L |
| S15 | 0 - 3 | Manager | H | | 200 + | L |
| S16 | 3 - 5 | Analyst | M | | - | L |
| S17 | 0 - 3 | Analyst | L | | 1 - 9 | L |
| S18 | 3 - 5 | Analyst | M | | 1 - 9 | M |
| S19 | 0 - 3 | Analyst | M | | 10 - 19 | S |
| S20 | 3 - 5 | Analyst | H | | 1 - 9 | M |

## A.3 Online Survey Questions

Prior to starting the survey, participants were presented with an information sheet detailing the study objectives, methods of contact with researchers and how to withdraw from the study. We detail the survey questions and participant information sheet in the supplemental document found here https://bit.ly/3BhjDtI. As the study was broader in scope, not all questions were analyzed as part of this paper.

## A.4 Semi-Structured Interview Questions

The semi-structured interview questions focused on the following themes:

### A.4.1 Interview Part 1

The first part of the interview was designed to encourage the practitioners to talk about their daily tasks, tools they used, and the challenges they face relating to tools and processes. In addition, we capture the security practitioners' perspectives on the importance of the human in the loop in SOC operations and the potential of automation. We present the interview questions in the following.

1. Please describe your position/job role/level.
2. Which network monitoring tools do you use in your monitoring work?
3. Looking back on past events, have there been times during your use of network-monitoring systems when they performed particularly well?
4. Can you describe an incident that was detected well?
5. Looking back on past events, have there been times during your use of network-monitoring systems when they could have performed better? Can you describe an incident that was not detected as well as it should have been?
6. What is your view on the strengths and weaknesses of the network-monitoring systems you use?
7. What is your view on the accuracy of the network-monitoring systems you use?
   (a) Are there events they do not detect that they should be detecting (false negatives)? If so, is this a frequent occurrence?
   (b) Do they detect false positive events? If so, what proportion of events detected are false positives?

(c) Can you comment on the balance between false positives and false negatives in your systems? Which are there more of?

8. What is your view on the usability of the network-monitoring systems you use?

9. We are interested in the balance between attack detection by automated systems, and work performed manually by analysts. Can you describe this balance? (e.g., are there times when you as an analyst use your own experience to explore the data and make decisions, rather than or alongside using automated system alerts, and how do you do this?)

10. Do you feel that you as an analyst monitoring the network are capable of detecting attacks that might be missed by automated systems?

    (a) How does your existing monitoring setup enable this?

    (b) Can you give an example?

    (c) How do you maintain situational awareness of your network?

11. Do you feel the level of Situational Awareness you are able to achieve currently is sufficient?

    (a) Can you comment on the usability of the systems through which you maintain SA? How easy is it to stay "in the loop"?

    (b) Is it necessary for you to maintain SA while performing other tasks, and how do you do so?

12. How do you decide which events to prioritize?

13. Would you like to share any more views on the network-monitoring systems you use that have not been covered by the questions so far?

#### A.4.2 Interview Part 2

During the interview, the participant was presented with a problem scenario and was asked to talk through the steps they will use to address the problem. We presented the analysts' with a number of scenarios in a semi-structured format, criticizing and inspecting scenarios in walk-throughs. Hence, we asked: For the following scenarios, could you describe step-by-step how you would investigate? Please discuss in detail the following.

1. Processes, roles, workflow between SOC team members. Which are the first tasks you would carry out?

2. Which tools would you use?

3. Which data would you gather to help in your investigation? What are the first information sources you would check, and what indicators would you look for?

4. How would you prioritize and assess severity?

5. Which tasks would you automate, and which data sources would you explore manually?

**Scenarios—** We then presented the analysts with a one randomly chosen scenario from the scenarios below.

1. You are monitoring the organization's network traffic, and observe an increase in the traffic from a server to an outside entity.

2. You are monitoring the organization's network traffic, and observe an increase in the traffic from a server to another device on the network.

3. You are monitoring the organization's network traffic, and observe an increase in the network outbound traffic (how does the analyst find the malicious internal host?).

4. You receive an alert from the SIEM of unauthorized privilege escalation attempt.

5. Scanning the network, you find an IoC such as a malware MD5 hash.

6. Your web server is receiving SYN flood requests.