Security incident filtering theory – a model of log analysts' prioritizations

Teodor Sommestad[1], Henrik Karlzén[1], Max Landauer[2], Markus Wurzenberger[2], and Florian Skopik[2]

[1] Swedish Defence Research Agency, 164 90 Stockholm.
{teodor.sommestad, henrik.karlzen}@foi.se
[2] Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria
{max.landauer, markus.wurzenberger, florian.skopik}@ait.ac.at

Abstract. Many organizations invest considerable resources in monitoring security-related events in their computer networks, e.g., by setting up or hiring security operation centers. Best-practice guidelines and qualitative analyses based on interviews with practitioners suggest that it is complex and difficult to make decisions related to prioritizations in monitoring and subsequent analyses. In this paper, we introduce *Security Incident Filtering Theory*. This theory suggests that three variables influence the priority that an event receives in log analysis processes: the operational context the event occurs in, how unusual the circumstances of the event are, and if the event is believed to be malicious or not. The model is tested on hypothetical scenarios with answers from 57 practitioners. It is found that all three hypotheses hold and that the model explains 46% of the variance in event prioritizations.

Keywords: Intrusion Detection, Security Log Analysis, Event Prioritization.

1 Introduction

Cyber threats are prevalent and sophisticated, and organizations need large amounts of security events in order to detect and mitigate potential attacks. In many organizations, Security Operation Centers (SOCs) have been created specifically to address this problem. Yet, identifying events that arise because of cyber security threats is a difficult task. This difficulty is due to the vast volumes of events and the problem of classifying them as either benign events or threatening events.

The literature on intrusion detection is vast and dense. There are thousands of scholarly papers on the topic of intrusion detection and security log analysis. Yet, the practical problem of detecting cyber security threats remains to be solved. A recent study by Alahmadi et al. [1] reported that 99 percent of the alerts processed by some security operation centers are false alerts. They further found that practitioners tend to think that the alarms produced by contemporary tools are "unreliable, difficult to interpret, and lacking in the context needed by analysts to filter false positives from genuine alarms". Earlier qualitative analyses have found that the process of intrusion detection is highly

collaborative, may require the development of own tools, and rely on tacit knowledge concerning the monitored environment [2].

Despite the large number of studies on intrusion detection, clearly articulated theories on how to detect cyber security threats are hard to find in the literature. This is not because thoughts, ideas or approaches are missing. On the contrary, cyber security professionals or scholars have many clear and carefully tested ideas on how to collect and analyze events. For instance, practitioners have produced the signature databases of tool suites such as Sigma[1], Splunk[2], and Elastic[3]. They have also developed data schemas such as Elastic Common Schema[4] and Open Cybersecurity Schema Framework[5] informing on the assets and events of relevance. Scholars have proposed many different ways to solve the problems. For instance, many studies on intrusion detection select a machine learning technique suitable for a particular detection problem and decide what data or events to train and test it on. Reviews of this type of technology-focused research often generalize research results and present testable universal statements concerning technology advantages and disadvantages. For instance, it has been stated that anomaly-based detection can be used to detect new attacks while signature-based detection has simpler design and is superior in detecting known attacks [3][4]. While scattered ideas like these are presented in many places, no widely accepted coherent model or theory explains what intrusion detection systems or security log analysts focus on.

Clearly articulated theories have many benefits. First, they can guide researchers on what questions to ask, what variables to examine, and how to interpret findings. Second, clearly articulated theories allow researchers to build on each other's work in a coherent way by testing, refining, or challenging existing theories and their hypotheses. Third, articulated theories make it easier to explain why phenomena occur or why a technology works, e.g., to explain research results for practitioners. Because of benefits like these, this paper revisits established ideas to formulate the *Security Incident Filtering Theory* (SIFT). In short, the theory rests on three hypotheses concerning how security events are prioritized by security log analysts:

H1: Analysts prioritize events according to organization-specific thresholds.

H2: Analysts prioritize events caused under unusual circumstances.

H3: Analysts prioritize events caused with maliciousness.

As will be described in section 2, the SIFT model is well aligned with established ideas and practices in the field. Section 3 describes how the model was operationalized and tested by surveying how 57 practitioners assessed different event types. Section 4 presents the results that validate the SIFT model. Section 5 discusses the results and what they imply for practitioners and researchers. Section 6 concludes the paper.

[1] https://github.com/SigmaHQ/sigma

[2] https://github.com/splunk/security_content/tree/develop/detections

[3] https://github.com/elastic/detection-rules

[4] https://www.elastic.co/guide/en/ecs/current/ecs-field-reference.html

[5] https://github.com/ocsf

2 Theoretical background and literature review

The *Security Incident Filtering Theory* (SIFT) is a theory presented in this paper and has a broad scope. It aims to predict and explain how teams in SOCs and individual log analysts will prioritize their investigation efforts. The SIFT model consists of four variables, where three variables influence the fourth variable, as illustrated in figure 1.

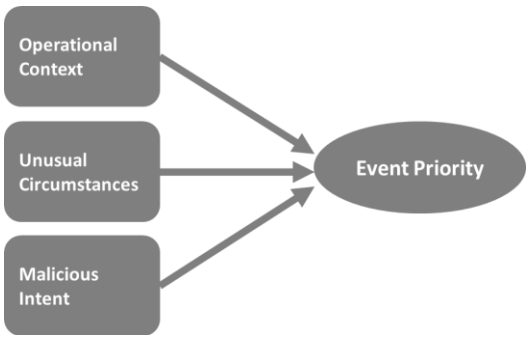

Fig. 1. Hypothesized relationships between variables in Security Incident Filtering Theory.

This model is a coarse simplification of how practitioners prioritize their analysis efforts, and the three predictor variables have complicated relationships to each other. Nevertheless, it is our belief that SIFT captures the essence of SOC decision processes. For instance, tuning of detection systems to the *operational context* is common practice, as intrusion detection alerts sometimes fire off when events happen under *unusual circumstances*, and detection efforts tend to target events believed to be caused with *malicious intent* rather than by random system crashes or user mistakes. The sections below elaborate further on the variables of the model and motivate their importance. Last, the model is compared to selected taxonomies and to other detection models.

2.1 Event priority

As mentioned above, a recent study found that practitioners tend to think that the alarms produced by contemporary tools are "unreliable, difficult to interpret, and lacking in the context needed by analysts to filter false positives from genuine alarms" [1]. Similarly, it has been reported that the literature on log analyst challenges often reports challenges such "volume of alerts", "false positives", and "workloads" [4]. Accordingly, a recent literature review of research on alert prioritization by Jalalvand et al. [5] argued that effective prioritization "is crucial for SOCs to effectively identify critical threats among a high volume of alerts, respond to them quickly and reduce analyst fatigue." Thus, what events log analysts are to focus on is an important practical issue worth researching, and the dependent variable of the SIFT model is an important one.

Prioritizations can be done in many ways and on different levels. For instance, an individual working within a SOC may prioritize the investigation of one suspicious event over another, or one of two machines when the same event occurs on both.

Additionally, the SOC as a whole may prioritize monitoring some types of events or some parts of their enterprise network. In a SOC, the work is often divided into tiers, where tier 1 focuses on triage and spends little time on each event, while other tiers spend more time on fewer events [4]. The SIFT model is designed to cover all these abstraction levels: priorities in all tiers depend on the *operational context*, are higher for events with *unusual circumstances*, and are higher for events with *malicious intent*. Sensors will be configured based on these variables, and all tiers will use them.

2.2 Operational context of the event

The process of intrusion detection and security log analysis is recognized as complex and contingent on many factors related to the organization. This has been recognized for decades. For instance, Denning's classical intrusion detection model is about recognizing the normal behavior of subjects using a statistical model [6], and this normal behavior of subjects tends to vary between organizations. When discussing issues with machine learning for intrusion detection, Sommer and Paxson [7] conclude that operational networks differ and one cannot simply assert that a machine learning solution trained on one site within an organization will work well on other sites. For instance, specific challenges exist in organizations with industrial control systems, and considerable effort needs to be put into configuring detection solutions [8]. The operational context is also recognized as important when signature-based solutions are used. Interviews with practitioners suggest that setting up and tuning solutions is complicated [9] [10] [11] [12]. For instance, Goodall et al. [9] found that it requires situated knowledge concerning the interaction between sensors and the environment. Werlinger et al. [2] found that it requires tacit knowledge about the organization's IT systems and services.

There is a myriad of variables associated with the organization and its operations that may influence if an event is to be prioritized or not. On a high level of abstraction, the risk tolerance of the organization can be expected to be important. For instance, those in military organizations tend to have a low threshold for anomalies, those in manufacturing may be somewhat worried with availability disruptions, and those in universities may be more relaxed in general. The SIFT model simply posits that the baseline rate at which events are prioritized will vary with organizations.

2.3 Unusual circumstances of the event

The notion that event priorities are influenced by circumstances related to the events is in many ways trivial and uncontroversial. Events' circumstances are central in virtually all detection solutions described in the literature, and circumstances are typically expressed in event logs and alerts. For instance, the classical Intrusion Detection Message Exchange Format [13] offers details about source, target, time, and more. More complex circumstances have also been suggested. Kotenko et al. [14] review methods that correlate events with each other, e.g., to identify unusual frequencies of events. In the review by Jalalvand et al. [5] five criteria are covered. The criterion most often used in the research was "alert", which concerned things such as the alert type or source IP-address. One example of an alert-based criterion is the one by Shah et al. [15], who

suggest a prioritization solution using attributes related to the user causing the event and the source of the event. Thus, the circumstance of an event should be expected to influence how events are prioritized.

There are many potential circumstances that may influence priorities, and they may differ with event type. For instance, a new user session to a server may be prioritized if the user is of a type that seldom accesses the machine in question, the session is established from an unusual source address, or if the session is created at an unusual time of day. Events are sometimes prioritized automatically based on circumstances like this, and generate alerts with different priority labels. While the relevant circumstances are likely to vary with the event, the SIFT model posits that four circumstances capture the bulk of scenarios: 1) the subject causing the event (e.g., the user), 2) the object of the event (e.g., the machine), 3) the timing (e.g., the time of day), and 4) the frequency (e.g., the events per minute).

2.4 Maliciousness behind the event

In general, there are clear links between intentions and risk [16], and it makes sense to consider these links in prioritization processes. Accordingly, the intentions behind actions are central to cyber security and to cyber security log analysts. First, the focus on incidents that occur because of intentional attacks are the primary difference between a SOC and a network operations center [17]. Thus, it should follow that cyber security log analysts will focus on events that are caused with some maliciousness, and leave other types of events (e.g., disk failures) to other organizational functions. Second, Vielberth et al. [18] explain that the first tier of a SOC should determine if the event is a false positive or not, before they move on to assess criticality. Meanwhile, Zimmerman [19] categorizes false positives as when "[t]he system alerts, but the activity but the activity was not actually malicious." It follows that a SOC will need to determine if events are caused with maliciousness or not when they prioritize.

Thus, the SIFT model predicts that events are prioritized based on judgments concerning the intentions, maliciousness, and the fact that the threat is serious. Put differently, the SIFT model assumes that log analysts prioritize events that seem to be true positives. These judgments are complex and surrounded by a great deal of uncertainty. The SIFT model does not depend on the accuracy of the judgments. It only posits that events believed to be malicious are given higher priority. Such beliefs can be measured with direct questions to analysts.

2.5 Relationship to other theories and models

As noted above, it is difficult to find ideas that are explicitly presented as theories in literature on security log analysis and intrusion detection. However, there are many ideas on how to detect and filter events to inform decision makers about the cyber security state of a system. The idea of using statistical models for anomaly detection is one such idea. The idea is often attributed to Denning [6], who said that the model is based on the hypothesis that "security violations could be detected from abnormal patterns of system usage." This anomaly-based model is often contrasted and compared to

the signature-based model, in which security violations are enumerated beforehand and associated with specific patterns in logs [3]. SIFT is aligned with both these basic ways of detecting security threats. In the anomaly-based detection model, SIFT recognizes that unusual circumstances are important input to anomaly detection systems, and they need to be adapted based on operational context. In the signature-based model, SIFT recognizes that unusual circumstances and events with triage is often needed even when threats have been detected using signatures.

There are many more fine-grained models on how to prioritize events. Jalalvand et al. [5] recently reviewed the literature on intrusion detection to identify criteria used to prioritize alerts in scholarly research and in practice. They put alert prioritization criteria into five categories:

- Organization (15 studies), e.g. number of analysts or business objectives.
- Analyst (39 studies), e.g., the expertise of operators or fatigue.
- External environment (11 studies), e.g., regulations or cyber threat intelligence.
- Alert (77 studies), e.g., IP-address or file category.
- Asset (29 studies), e.g., criticality or applicability of attack.

These five categories could be seen as candidate variables for SIFT, and there are relationships between SIFT and these categories. The operational context in SIFT concerns the context of the organization and its operations, and is related to the criteria types of *analyst*, *organization*, and *external environment*. The unusual circumstances of SIFT are directly related to the category *alert*. The maliciousness behind an event can be linked to *alert*, and to some extent to the threat intelligence associated with *external environment*.

SIFT does not cover the *asset* criterion of Jalalvand et al., which mainly concerns business impact and potential costs. These aspects of events are excluded from SIFT because they are believed to be difficult for SOC analysts to appraise and use in prioritizations. The reasons are as follows. First, suspicious events may start to unravel large incidents or problems that are difficult to deduce based on the assets involved in the event. For instance, uncovering a trojanized version of SolarWinds' Orion software unraveled a large campaign involving multiple organizations [20]. Second, predicting the assets that an attack may compromise in the near future is difficult. Husák et al. [21] conclude that "there is still no definite answer on how to effectively and precisely predict cyber attacks". Third, the business impact of a compromised asset is very difficult to appraise for a log analyst, especially in a large organization or when the log analyst works for a managed security service provider. For instance, elaborate process models may be required to properly assess impact [22], and the reasoning about attack chains may need to be further developed [23].

An unusually elaborate model for alert prioritization, covering more than one category by Jalalvand et al. [5], is the *composite alert risk-score model* [15]. This model aims to filter significant alerts by assigning different weights based on user context, alert source, severity, mission criticality, and details associated with the time of the event. There is partial overlap between this composite alert risk-score model and the SIFT model. SIFT does not explicitly cover the source or severity of the alert, but instead considers how unusual the circumstances are and how likely the event is to be malicious. The composite alert risk-score model has not yet been shown to align with

how log analysts prioritize, or how they should prioritize. However, it has been demonstrated that the model prioritizes differently than the severity-ratings of the network-based detection system Snort and that it is different from first-in-first-out principles [15].

Finally, researches on intrusion detection and security log analysis sometimes refer to theoretical frameworks from other domains. For instance, a recent review [24] found that the most commonly referenced theory for cyber situational awareness is Endsley's model of situation awareness, originally from aviation. Endsley's model [25] describes the process used to gain awareness and form decisions, e.g., that user interfaces and workloads influence situation awareness. Unlike the SIFT model, Endsley's model does not describe what information elements a security log analyst would use to gain cyber situation awareness.

3 Methods and materials

As stated above, this research surveys practitioners to test three hypotheses of the SIFT model. A survey with repeated measures design was used. Each respondent was presented with up to 17 security-related events and was asked to respond to questions related to priority, circumstances, and maliciousness. Responses of each respondent were clustered to capture the effect of the operational context. Section 3.1 describes the measurement instrument in detail, section 3.2 describes the data collection, and section 3.3 describes the analysis method.

3.1 Measurement instrument

A questionnaire was used to measure the priorities. The questionnaire consisted of five parts. The first part presented information about the study and questions concerning consent. The second part asked about the respondent. These questions were primarily aimed at supporting assessment of the respondents' fit with the sample frame, e.g., by asking about professional experience from various log analysis tasks. The third part asked about the organization(s) the respondent monitors. Again, these primarily serve as a way of validating the respondents' fit to the sample frame, e.g., by asking about the size of the networks they monitor. The fourth part presented the respondents with 17 event types in random order and asked questions for each of these. This was the main part of the survey and the one used to support the analysis. Finally, the fifth part asked some questions about the quality of the questionnaire, in relation to comprehensiveness and comprehensibility, e.g., asking how easy it was to answer the questions.

Part four, with the 17 event types, is the central part of the measurement instrument. Designing this part involved two steps: selecting the event types to include and formulating the questions to ask for the event types.

A first draft of the event types to include was determined based on discussions among the researchers. Candidate event types were identified based on:

- Personal first-hand interactions with practitioners, e.g., insight gained by training security log analysts in cyber ranges operated by the authors.

- Previous research on issues with intrusion detection systems, e.g., the interviews and questions used by Alahmadi et al. [1].
- Analyzes of signatures often employed in practice. In particular, we used the signature database Sigma6 to identify events that may be associated with analysis efforts and false alerts, such as /rules/windows/process_creation/proc_creation_win_schtasks_creation.yml.
- Common attack techniques and potential benign behavior. For instance, we considered attack techniques in the MITRE ATT&CK framework7 where the techniques had many procedure examples, such as T1053.005 Scheduled Task/Job: Scheduled Task.

The questions were designed to measure the variables of interest for each event type. These questions asked the respondent:

- The event priority, e.g., if the event occupies the log analyst's time. (4 items).
- To what extent the events occupying analysts' time are associated with unusual circumstances (4 items).
- To what extent the events occupying analysts' time arise with malicious intent (3 items).

Likert-scales were used because of the difficulty of formulating questions applicable to all in the target population, e.g., time ranges relevant to people monitoring networks of different size. *Operational Context* was not asked about directly, but inferred by comparing responses between respondents.

After settling on a set of candidate events and question items, the questionnaire was subject to review by three domain experts representing the target population. Each review was conducted during a one-hour interview where the questionnaire was presented and the domain expert was asked about validity, inconsistencies in questions, and general feedback. After these expert reviews, the survey was extended with another event type ("outbound connections") and the questions about the respondent were made more concrete. No other issue was identified, and no other change was made. The event types and questionnaire items are detailed in the appendix.

3.2 Data collection

The sample frame of this study is security log analysts working in operational networks. Eliciting survey responses from this audience is associated with some obstacles. First, security log analysts are occupied with analyzing logs and may not find 30 minutes to respond to a questionnaire with a weak relationship to their day-to-day operations. Second, the process of security log analysis is associated with confidentiality that may prevent potential respondents from responding to the questionnaire even if they are personally willing to. To receive as many relevant responses as possible, the researchers' personal contact networks were utilized. The data collection effort included invitations during conferences involving log analysts in Sweden and Austria,

[6] https://github.com/SigmaHQ/sigma
[7] https://attack.mitre.org/

personal emails to practitioners of the target population, invites via email lists (e.g., for CERTs), and invites via LinkedIn-posts.

The survey was open between September 17 and December 23 in 2024. During this period, 30 respondents completed the whole questionnaire and 76 respondents provided partial responses and responded to at least one event type. Screening based on the questions pertaining to the respondent suggests that all these responses are from the target population. Because of the sampling method used, the responses are biased towards the personal network of the authors, with most respondents located in Austria (N=38), Sweden (N=23), and Norway (N=5) (others N=10). Among the respondents who provided details on the network they monitor, 38 respondents monitor networks with more than 1000 endpoints, 11 monitor networks with 251-999 endpoints, 8 monitor networks with 50-250 endpoints, and 8 monitor small networks with 1-49 endpoints. Respondents were only included in the data analysis if they consented to each of the three consent questions and had at least one professional experience in a relevant job topic. It was also required that they monitor at least 1-49 endpoints, 1-49 network security devices, or 1-49 users. 57 respondents met these criteria.

3.3 Data analysis

The data collected contains information about between 1 and 17 events for each respondent. There are potential dependencies between these responses. First, it is possible that the order of the events influences how respondents respond, e.g., because the responses are anchored on the first event type. This was handled by randomizing the order that events were presented to the respondents. Second, it is highly likely that a respondent's assessment of the events is influenced by the overall cybersecurity priorities in the respondent's operational context. This is the first hypothesis of this study, and the effect is estimated by clustering responses to the respondent in a mixed model with random intercept. More specifically, the priority of alerts for event *i* by respondent *j* is predicted by the following regression model:

$$\text{Priority}_{i,j} = (\beta_0 + \beta_{0,j}) + \beta_1 \times \text{Circumstances}_{i,j} + \beta_2 \times \text{Malicious}_{i,j} + \text{Error}_{i,j}$$

In this model, $\text{Priority}_{i,j}$, $\text{Circumstances}_{i,j,}$ and $\text{Malicious}_{i,j}$ are calculated as the mean of the items associated with these variables for event *i* and respondent *j*. The random intercept is represented by the term $\beta_{0,j}$, which allows the constant in the regression model to vary with respondent *j*. The model will thus create regression models that predict each event's priority based on a respondent-specific constant and global (non-respondent-specific) weights related to unusual circumstances and malicious intent. With this model, the hypotheses are tested as described in Table 1.

Table 1. Statistical tests used to test the hypotheses.

Hypothesis	Statistical test
H1. Operational context.	The weight β_{0j} varies for different *j*. This variable explains variance in the priority (i.e., reduces the error term).
H2. Unusual circumstances.	The weight β_1 is larger than zero. This variable explains variance in the priority (i.e., reduces the error term).
H3. Malicious intent.	The weight β_2 is larger than zero. This variable explains variance in the priority (i.e., reduces the error term).

In addition to this analysis, conventional methods to assess reliability and validity of the measurement were performed. These include reliability analyses for the measured variables using Cronbach's alpha and correlation matrices. Priority had a Cronbach's α of 0.81; *Unusual Circumstances* had a Cronbach's α of 0.83; and *Malicious intent* (with one item inverted) had a Cronbach's α of 0.72.

4 Results

A total of 57 respondents answered the questions for at least one event, with 30 responding for all 17 events. Altogether there were 582 responses for the events. Some of these are missing data required for the analyses. The degrees of freedom in the analyses are in the range 441-466. The respondents who reported their years of experience had 1-30 (mean 10.0) years of experience in the cyber security field and 1-30 (mean 8.4) years of experience in analyses of cyber security events or logs. Thus, the typical respondent is a seasoned log analyst who has made many prioritizations.

Table 2 summarizes the results in terms of different models and parameters. The measure R^2 marginal corresponds to the explained variance without considering the differences between respondents' operational contexts. R^2 conditional corresponds to the explained variance when information about operational context is used in the prediction. As such, the marginal effect is at the population level, while the conditional effect considers individual differences.

Table 2. Models tested and the variance in event prioritizations they explain.

	Circumstances		Maliciousness		R^2 marginal	R^2 conditional
Model	β_1	p	β_2	p	(excl. context)	(incl. context)
Only intercept	-	-	-	-	0.00	0.20
U. circumstances	0.42	< 0.01	-	-	0.21	0.42
Malicious intent	-	-	0.48	< 0.01	0.19	0.31
Full model	0.34	< 0.01	0.39	< 0.01	0.31	0.46

As can be understood from the table, the explained variance increases from 0% to 20% when information about the *operational context* is added to an empty model. When the predictor variable *unusual circumstances* is used, the model explains 21% of the variance. With both *unusual circumstances* and *operational context* this number increases to 42%. If *malicious intent* is the only variable used in the model, 19% of

variance is explained. This increases to 31% when information about the *operational context* is added. The full model explains 46% of the variance in prioritizations. Of these 46%, 31% come from the two variables *malicious intent* and *unusual circumstances*, while 15% are associated with the respondent's *operational context*. Thus, all three hypotheses hold true.

All variables contribute to the full model. The full model only explains 4 percentage points (0.46-0.42) more variance than the model with *unusual circumstances* and *operational context*, which shows that *malicious intent* does not add a lot of explanatory power. The full model explains 15 percentage points (0.46-0.31) more variance than the model with *malicious intent* and *operational context*, which shows that *unusual circumstance* adds quite a lot of explanatory power. This shows that *malicious intent* overlaps with *operational context*, and more than *unusual circumstances* does. One interpretation of this overlap is that respondents in an *operational context* with a high security priority (i.e., high constant $\beta_{0,j}$) also see a lot of *malicious intent* behind the events (i.e., put high values on $Malicious_{i,j}$).

5 Discussion

As presented in Section 4, all three hypotheses hold true. In other words, it is possible to predict what events that practitioners prioritize based on:

- Operational context of the event (via the respondent identifier).
- Unusual circumstances of the event (through four questions).
- Perceived maliciousness of the event (through three questions).

Thus, the SIFT model seems to hold.

In section 5.1 the validity and reliability of these results are discussed. In section 5.2 the implications for future research are discussed. In section 5.3 the implications for practitioners in security log analysis are discussed.

5.1 Validity and reliability issues

The hypotheses were tested with data obtained by surveying practitioners. Several potential validity and reliability issues are associated with survey research of this kind. Potential concerns related to the present study are addressed below.

First, this survey is based on the presentation of 17 broad event types to practitioners. The selection of events may influence the results. For instance, events with fewer details will allow more room for circumstances, which could inflate the variable for *unusual circumstance*. No such bias was intended when the events were produced. We leave it to the reader to judge if such bias is present using the list of event descriptions in the appendix. A bias that was introduced was due to our desire to only include events that could be described without explicitly stating that they were attacks, such as listing examples of network commands executed on a machine corresponding to typical network reconnaissance activities. Some candidate events were discarded or combined with others due to the lack of suitable neutral descriptions. For example, a candidate event concerning "unusual network protocols" was included as an example in an event

concerning "outbound connections." In the final questions, respondents were asked how much of their total investigative effort that the 17 events covered. The mean answer (N=27) was 67% of the events they prioritized. Thus, while some unintended bias in the selection of events may exist, the 17 events cover most of the time that the respondents spend on log analysis.

Second, it is challenging to construct questions that are relevant and comprehensible. In the final part of the survey, respondents rated the survey on a scale from 1 to 5. The mean ratings (N=27) were 4.1 for ease of navigation, 4.1 for understanding the event types, 3.7 for ease of answering the questions, and 3.4 for providing accurate and confident answers. Thus, the main issue seems to be a natural uncertainty concerning the events, such as whether the analyses always determine if an event is harmful.

Third, in the absence of established scales, the questionnaire contained new scales for the measured variables. Domain experts screened these scales for face validity beforehand, but none were pilot-tested with statistical methods. Yet, all scales exceeded the commonly used threshold for inter-item reliability of 0.7 [26]. We suspect the low reliability for *malicious intent* (just above 0.7) is due to the natural uncertainty concerning the intent behind events, even in retrospect. Reliability issues with the scales may hamper the model's explanatory power. Nevertheless, the explanatory power of the model is good. Since the model predicts 0.46 of the variance of responses, the correlation (or square root of the variance) between the model's predictions and the observed values is 0.68. This demonstrates a clear association. With reliability scores around 0.8, the true correlation is around 0.85 if corrected for attenuation (i.e. an explained variance of 0.72). Thus, if all measurement errors were removed, the predictions of the models would be highly correlated to the observed priorities.

Fourth, a limitation of this study is the small convenience sample and the incomplete responses. Although the inclusion criteria were domain experts with recent experience in security log analysis, we cannot guarantee that the respondents are representative for security log analysts in general. For instance, other respondents might prioritize circumstances and maliciousness differently or focus on aspects that we did not measure. The demographics of the respondents are skewed towards the authors' professional networks (in Sweden and Austria). Differences in security log analysis practices across regions may affect the generalizability of these results. This is also connected to the context variable, which could have been operationalized with some combination of demographic items. Future studies with larger, more rigorously defined samples, will provide further insights.

Fifth, the survey items asked the respondents about event priority in multiple ways and focused on events that they spend time on for other items in the questionnaire. These are clearly related and is aligned with the hypotheses and the aims of this study. However, an even tighter connection between dependent and independent measures would likely help explain even more variance in the dependent variable.

5.2 Implications for researchers

Alert prioritization, or security event prioritization, is an active research field. Thousands of studies have been published on intrusion detection and related tasks performed

by security log analysts. The five types of alert prioritization criteria identified in [5] mostly covered the three criteria of the SIFT model. A conservative interpretation of our results is that:

- The operational context explains at least 15% of the variance in alert prioritization.
- The unusual circumstances explain at least 14% of the variance in event prioritization.
- The malicious intent explains at least 4% of the variance in alert prioritization.

A less conservative estimate, adjusting for measurement errors, suggests that more than 70% of the variance in alert prioritization can be explained with these three variables alone. These results suggest that SIFT can explain what practitioners prioritize.

These estimates are descriptive and based on current practice. Prioritization methods may evolve in the future. However, with two thirds of the variance explained, the SIFT model leaves little room for additional variables with strong explanatory potential, and is likely to capture the essence of log analyst concerns. That said, the scales and variables can most likely be refined further to predict prioritizations even better. More specifically, the variable *unusual circumstances* does not cover all aspects of the "alert" criterion in [5]. Additionally, the criterion "asset" of [5] is not well covered. Future research may develop solutions that consider the circumstances of events (e.g., additional circumstances to consider) and their potential impact on business (e.g., based on a crown-jewel analysis). Since the operational context is also important, future research could identify the key contextual variables that are hidden behind the identifier of the respondents in this study. For instance, it could be investigated if differences between respondents is due to varying experience, geography, organizational type, or something else.

A side result of this research is the survey and the 17 event types monitored by practitioners. These events cover approximately two-thirds of the events they spend time on. This list may be useful for related research, such as when investigating overlaps between benign behavior and malicious behavior.

5.3 Implications for practitioners

This survey aggregates responses from many practitioners and reveal clear variations in how practitioners prioritize events. The variable *operational context* shows that analysts from different organizations have different base levels for how much they prioritize. Variations in how they prioritize individual events also affirm previous findings and widely held beliefs about the need for tuning, adaptations, and differences between organizations. In other words, there is no one-size-fits-all solution for security log analysis.

The circumstances of an event appear to significantly influence prioritization. Much of current practice already uses these circumstances in rules that generate alerts, such as those triggered by certain geographical locations, user types, or file paths. Systemizing these circumstances may be worthwhile. Considering that circumstances may vary between organizations, networks, and user types, it may be beneficial to use anomaly-based intrusion detection approaches such as AMiner [27], which sets acceptable

circumstances based on historical records. The suspicion of *malicious intent* also significantly affects prioritization, but is less important than *unusual circumstances* and *operational context*.

Commercial products such as Splunk [28] and Elastic [29] have risk scores with clear links to the SIFT model. Given our results, it is advisable to configure these features, such as setting impact scores for different rules and assets based on the operational context. We suspect that some log analysts, particularly managed security service providers, may not know the operational context they monitor well enough to assign accurate impact scores or differentiate between legitimate use and threats. We also suspect that assessments of *malicious intent* are often uncertain or outright guesses, and that is why it is a less important variable in prioritizations. Together with differences between organizations in how things are to be prioritized, this may hamper the possibility of automating impact assessments and assigning priority scores to events.

The list of events produced in this research aims to cover important events for log analysts without assuming that the events are malicious. An extended and refined version of this list may aid in classifying false alerts in SOCs. Currently, most false alerts are labeled with vague categories such as "sysadmin activity" or "unusual user behavior." Our 17 event types, included in the appendix, may offer better ways to classify benign events and help SOCs improve rule quality.

6 Conclusions

All three hypotheses of the SIFT model were successfully tested in this study and shown to hold. With this data, a model with only circumstances explains 21% of the variance in priority, a model with only maliciousness of intentions explains 19%, and a model with only the respondent's operational context explains 20%. When all predictor variables are added to a full model, they explain 46% of the variance in alert prioritization. Thus, current practice in event prioritization focuses on events that have unusual circumstances and events that seem malicious. Additionally, some analysts and organizations will prioritize events higher in general, e.g., because of threat level, the number of analysts, or available threat intelligence.

7 Funding

This work was supported by the European Union through the European Defence Fund under GA no. 101103385 - AInception. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

References

1. Alahmadi, B.A., Axon, L., Martinovic, I.: 99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms. In: Proceedings of the 31st USENIX Security Symposium, Security 2022. pp. 2783–2800 (2022).
2. Werlinger, R., Muldner, K., Hawkey, K., Beznosov, K.: Preparation, detection, and analysis: the diagnostic work of IT security incident response. Information Management & Computer Security. 18, 26–42 (2010). https://doi.org/10.1108/09685221011035241.
3. Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J.: Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity. 2, (2019). https://doi.org/10.1186/s42400-019-0038-7.
4. Agyepong, E., Cherdantseva, Y., Reinecke, P., Burnap, P.: Challenges and performance metrics for security operations center analysts: a systematic review. Journal of Cyber Security Technology. 4, 125–152 (2020). https://doi.org/10.1080/23742917.2019.1698178.
5. Jalalvand, F., Baruwal Chhetri, M., Nepal, S., Paris, C.: Alert Prioritisation in Security Operations Centres: A Systematic Survey on Criteria and Methods. ACM Comput. Surv. 57, (2024). https://doi.org/10.1145/3695462.
6. Denning, D.E.: An Intrusion-Detection Model. IEEE Transactions on Software Engineering. SE-13, 222–232 (1987). https://doi.org/10.1109/TSE.1987.232894.
7. Sommer, R., Paxson, V.: Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. Proceedings of the IEEE Symposium on Security and Privacy. 305–316 (2010). https://doi.org/10.1109/SP.2010.25.
8. Anthi, E., Williams, L., Burnap, P., Jones, K.: A three-tiered intrusion detection system for industrial control systems. J. Cybersecur. 7, (2021). https://doi.org/10.1093/cybsec/tyab006.
9. Goodall, J.R., Lutters, W.G., Komlodi, A.: I know my network. In: Proceedings of the 2004 ACM conference on Computer supported cooperative work. pp. 342–345. ACM, New York, NY, USA (2004). https://doi.org/10.1145/1031607.1031663.
10. Werlinger, R., Hawkey, K., Muldner, K., Jaferian, P., Beznosov, K.: The challenges of using an intrusion detection system. In: Proceedings of the 4th symposium on Usable privacy and security - SOUPS '08. p. 107. ACM Press, New York, New York, USA (2008). https://doi.org/10.1145/1408664.1408679.
11. Werlinger, R., Hawkey, K., Beznosov, K.: Human, organizational and technological challenges of implementing IT security in organizations. In: Human aspects of information security. pp. 35–10 (2008).
12. Werlinger, R., Muldner, K., Hawkey, K., Beznosov, K.: Towards Understanding Diagnostic Work During the Detection and Investigation of Security Incidents. In: Steven M Furnell and Nathan L Clarke (eds.) Proceedings of the Third International Symposium on Human Aspects of Information Security & Assurance (HAISA 2009). pp. 119–134 (2009).
13. Debar H., Curry D., Feinstein B.: The Intrusion Detection Message Exchange Format (IDMEF); RFC 4765, (2007).
14. Kotenko, I., Gaifulina, D., Zelichenok, I.: Systematic Literature Review of Security Event Correlation Methods. IEEE Access. 10, 43387–43420 (2022). https://doi.org/10.1109/ACCESS.2022.3168976.
15. Shah, A., Ganesan, R., Jajodia, S., Cam, H.: A Two-Step Approach to Optimal Selection of Alerts for Investigation in a CSOC. IEEE Transactions on Information Forensics and Security. 14, 1857–1870 (2019). https://doi.org/10.1109/TIFS.2018.2886465.

16. King, Z.M., Henshel, D.S., Flora, L., Cains, M.G., Hoffman, B., Sample, C.: Characterizing and measuring maliciousness for cybersecurity risk assessment, (2018). https://doi.org/10.3389/fpsyg.2018.00039.
17. Shahjee, D., Ware, N.: Integrated Network and Security Operation Center: A Systematic Analysis. IEEE Access. 10, 27881–27898 (2022). https://doi.org/10.1109/ACCESS.2022.3157738.
18. Vielberth, M., Bohm, F., Fichtinger, I., Pernul, G.: Security Operations Center: A Systematic Study and Open Challenges. IEEE Access. 8, 227756–227779 (2020). https://doi.org/10.1109/ACCESS.2020.3045514.
19. Zimmerman, C.: Ten Strategies of a World-Class Cybersecurity Operations Center. The MITRE Corporation, Bedford, MA (2014).
20. Vijay A. D'Souza: SolarWinds Cyberattack Demands Significant Federal and Private-Sector Response. GAO (2021).
21. Husák, M., Komárková, J., Bou-Harb, E., Čeleda, P.: Survey of attack projection, prediction, and forecasting in cyber security. IEEE Communications Surveys and Tutorials. 21, 640–660 (2019). https://doi.org/10.1109/COMST.2018.2871866.
22. Musman, S., Temin, A.: A Cyber Mission Impact assessment tool. In: 2015 IEEE International Symposium on Technologies for Homeland Security (HST). pp. 1–7 (2015). https://doi.org/10.1109/THS.2015.7225283.
23. Spring, J.M., Hatleback, E.: Thinking about intrusion kill chains as mechanisms. J. Cybersecur. 3, 185–197 (2017). https://doi.org/10.1093/cybsec/tyw012.
24. Ofte, H.J., Katsikas, S.: Understanding situation awareness in SOCs, a systematic literature review. Comput. Secur. 126, (2023). https://doi.org/10.1016/j.cose.2022.103069.
25. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Hum. Factors. 37, 32–64 (1995).
26. Tavakol, M., Dennick, R.: Making sense of Cronbach's alpha. Int. J. Med. Educ. 2, 53–55 (2011). https://doi.org/10.5116/ijme.4dfb.8dfd.
27. Landauer, M., Wurzenberger, M., Skopik, F., Hotwagner, W., Höld, G.: AMiner: A Modular Log Data Analysis Pipeline for Anomaly-based Intrusion Detection. Digital Threats: Research and Practice. 4, (2023). https://doi.org/10.1145/3567675.
28. Splunk: Risk Analysis framework in Splunk ES, https://dev.splunk.com/enterprise/docs/devtools/enterprisesecurity/riskanalysisframework/, last accessed 2025/04/28.
29. Elasticsearch: Entity risk scoring, https://www.elastic.co/docs/solutions/security/advanced-entity-analytics/entity-risk-scoring, last accessed 2025/04/28.

Appendix: Questionnaire

Table 3 describes the 17 event types used in the survey, as they were presented to the respondents. Table 4 details the items presented for each of these event types. The examples of circumstances are adapted to each event type. All items have the response alternatives: [1-Never], [2-Rarely], [3-Sometimes], [4-Often], [5-Always], and [Not applicable].

Table 3. The 17 event types used in the survey.

Event type	Description
executions of custom user scripts	Events when someone executes some script or other automation of tasks. For example: Scrapes internal sites to gather information. Sorts files using powershell or bash. Automates tasks in MS Office using its record-features.
changes of login credentials	Events when someone changes the state of a password or credential. For example: Changes the password of an account. Resets the password for a web service account.
installation of new software	Events when someone installs new software or updates existing software. For example: Runs apt-get to install software on a Ubuntu machine. Runs Windows Update. Presses the update-button of software on the system. Runs an msi-file or exe-file to install some software.
complex system commands	Events when complex system commands are executed on a machine. For example: Utilities are executed, e.g., tools in Windows Sysinternals. The event viewer on a Windows machine or journalctl on a Linux machine is used. Execution with "sudo -i" on a Linux machine or execution as administrator on a Windows machine. Various settings software-related settings are fiddled with, e.g., whitelists or antivirus.
changes of operating system configurations	Events when someone changes the configuration of the operating system of a machine. For example: Manipulates scheduled tasks on the machine. Changes settings for services in a windows machine. Makes changes to mechanisms for system recovery and backups. Changes preferred applications of file extensions.
checking and cleaning things in servers	Events when someone checks or cleans things in a server (e.g. domain controller, web server, or mail server). For example: Monitors server performance metrics such as CPU usage. Cleans temporary files with Windows Disk Cleanup. Sifts through system logs from the last week.
logons to servers	Events when someone logs on to a server (e.g. domain controller, web server, or mail server). For example: Uses remote desktop to look at some setting, e.g. the network interface. Uses SSH to a machine to look at some setting, e.g. the network interface. Executes PsExec to look at some setting, e.g. the network interface.
changes of file permissions	Events when someone changes the permissions for a file. For example: Sets a file as executable, e.g., for a file in the downloads folder. Makes a folder accessible to everyone on a machine. Creates a shared folder in a Windows domain.

[Continuation of table 3.]

failed logons	Events when someone repeatedly fails to logon. For example: Enters the wrong password enough times that the account is locked. Forgets their username and tries different alternatives to get in.
outbound connections	Events when certain outbound connections occur. For example: Data is uploaded to external repositories. P2P protocols are used. Legacy software is connecting to remote servers.
many copy-operations	Events when someone copies a considerable number of files within a short time period. For example: Copies plenty of files to a folder and compresses them. Sorts documents and files in a new file structure. Selects a folder to synchronize to some enterprise file server. Makes a backup copy of an entire disk or a database.
new device in the network	Events when someone adds a new machine or device to the network. For example: Adds a device with a new MAC address to the enterprise network. Adds a new machine to a Windows domain.
inbound connections	Events when someone initiates connections to services or machines in the enterprise network from an external address. For example: Creates a VPN connection to the enterprise network. Connects to a terminal server.
certain software being executed	Events when certain software being executed in the network. For example: Never-before-seen software is executed. A software crashes over and over again. Debugging tools are used.
complex network commands	Events when someone executes complex network commands. For example: Runs software that captures traffic, e.g. using pcap. Runs tools and commands such as ifconfig, ping, tracert, and netstat. Adjusts various settings, e.g. related to network interfaces, iptables, and the local firewall.
rebooting of a server	Events when someone reboots a server (e.g. domain controller, web server, or mail server). For example: Reboots after a system update. Reboots caused by a system crash.

Table 4. Questions asked for each event type.

Item	**Text**
	In your experience, events related to [EVENT TYPE] …
Event priority 1	… are important to monitor.
Event priority 2	… cause security alerts.
Event priority 3	… occupy analysts' time.
Event priority 4	… lead to investigations.
	In your experience, are the following circumstances required for analysts to occupy their time with events related to [EVENT TYPE]?
Unusual Circumstances 1	Unusual subject, e.g., [EVENT-RELATED EXAMPLE].
Unusual Circumstances 2	Unusual object, e.g., [EVENT-RELATED EXAMPLE].
Unusual Circumstances 3	Unusual timing, e.g., [EVENT-RELATED EXAMPLE].
Unusual Circumstances 4	Unusual frequency, e.g., [EVENT-RELATED EXAMPLE].
	In your experience, events related to [EVENT TYPE] and occupy analysts' time, are finally determined to be …
Malicious intent 1	… malicious to the organization.
Malicious intent 2	… caused with good intentions.
Malicious intent 3	… a real and serious threat.