Measuring Metrics in Incident Response

Wahab Ahmed, University of Twente, The Netherlands

ABSTRACT

In today's cybersecurity landscape, Incident Response plays a critical role in mitigating the impact of increasingly sophisticated cyber-attacks. To measure the effectiveness of Incident Response, organizations deploy several metrics. However, these metrics often face limitations and challenges which will be covered in this study. The contribution of this paper is to identify some metrics that have been well-defined and explain their method of measurement, and any challenges associated with them. The goal is to serve as an educational resource for analysts or beginners to better understand how these metrics function. This will be done by using academic literature and real-world reports to extract well-known metrics. Additionally, this study will develop a prototype of a Security Information and Event Manager (SIEM) to demonstrate how different scenarios can impact the measurement of a metric.

1 INTRODUCTION

In today's interconnected digital world, Incident Response plays a critical role in cybersecurity as cyber-attacks become increasingly problematic.[13] Numerous metrics have been employed to help measure the effectiveness of Incident Response practices, which will be examined in detail further in the study. However, these metrics often face significant challenges which prompt analysts to view these metrics as misleading.[2]. Organizations adopt varied metrics tailored to their internal structures, resulting in numerous different metrics overall. This inconsistency poses a challenge to aligning academic literature and real-world practice. Although organizations may share some metrics in common, such as Mean Time To Detect (MTTD), they may be measured differently, such as time from initial compromise until alerted or time of first anomaly until intervention. Or, they may measure two metrics the same way, such as one organization may measure the Time Between Security Incidents the same as the Mean Blind Spot metric. This study will cover a few sections, such as the related work to this topic covering what metrics exist and what this study aims to highlight and bring attention to. It will also include a methodology for the SIEM demonstration and the metrics highlighted in order to show in depth the method used to answer the research questions and obtain results. The next sections will cover the metrics, why they were chosen, how they answered the research questions and the challenges faced by them. The SIEM prototype will demonstrate a way of measuring a singular metric, as well as support the challenge faced by these analysts. Finally, the conclusion will conclude these results as well as mention future work that can be done. To tackle this problem, this study aims to solve these three research questions:

- 1. What are some metrics that real-world practice mention?
- 2. What are some metrics that academic literature mention?
- 3. How are some of these metrics measured?

By focusing on answering these three research questions, the study aims to extract a set of metrics and highlight some of them.
To answer these research questions, academic literature and real-world practice will be mentioned and used, often studying multiple literature in order to extract a set of metrics, with some used for the Security Information and Event Management (SIEM) demonstration. This application will use one metric for the demonstration.

2 RELATED WORK

Numerous metrics have been defined[10, 9, 8, 12, 20, 7, 6, 14, 17, 2, 11, 22] such as mean time to detect, ability to handle a large volume of data, mean time to respond, financial impact, time to eradicate, risk assessment, and mean time between security incidents. While these metrics exist, organizations often adopt tailored combinations of these metrics based on their specific needs. Many of these metrics face challenges which will be examined later in the study for educational purposes. Furthermore, there exist similar metrics, such as Scalability/Resource Consumption[12] and the ability to handle a large volume of data[20] which are measured in similar ways. They are, however, classified as different metrics by

the organization. In light of this, existing studies have mainly focused on metrics for the individual, and this study aims to highlight these metrics and explain how some of them are measured. Many studies often forego the challenges the analyst faces in their methods or with establishing an objective metric[1].
Academic literature tends to emphasize metrics that may not be adopted by organizations, hence creating a gap between academia and real-world practice. Academic literature may focus on metrics such as the ratio of blind spots metric[11] which is rarely used in real-world practice. With this gap, it may be difficult to compare metrics between the state of practice and the academic literature, which adds to the difficulty of establishing an objective metric that takes several aspects of their work into consideration[1]. This challenge of establishing an objective metric can be caused by the different functions and tasks expected from analysts[1]. These specific tasks, such as tasks assigned to an analyst by their managers, are dependent on the functions expected of the analyst and may vary from one Security Operations Center to another, thus making these metrics vary.[1] This study aims to highlight some of these metrics, how they are measured, and the challenges they can face for educational purposes.

3 METHODOLOGY

3.1 For Metrics

To extract relevant metrics, this study will conduct a literature review focused on identifying metrics that are currently being used by organizations or literature, the method of measurement for these metrics, and what challenges these metrics face. The review will include both academic literature and real-world practice, as each may emphasize metrics that differ. Given the large number of available metrics, in order to decide which metric is suitable for this study, popularity will be the main criterion adopted. This criterion is based on the assumption that widely used metrics are more likely to be well-defined and easier to adopt. Popular metrics, due to their frequent adoption, are considered sufficient for this study as an educational source to highlight these metrics. These metrics are more likely to be encountered by users which further justifies the choices of using these metrics in this educational source. With this aspect prioritized, a set of metrics would be extracted and produced by reviewing the literature.
To find real-world practice sources to use, several blog posts[6, 17, 14, 9], industry reports[4, 5, 3] are used to identify metrics in use. These findings were then compared with academic literature to identify overlapping and contrasting metric preferences. After these metrics are extracted, they will be filtered using the defined popularity criterion to a subset of metrics for in-depth exploration. The corresponding measurement methods for some of these metrics will also be shown. This includes the challenges associated with these metrics (which will be discussed in later sections). Mainly, real-world literature and academic literature that mentions metrics that fit the popularity criteria will be used in this study. Additionally, some surveys[15] are used just to mention a couple of metrics that are being used, as well a study that proposes a new metric[11] to view the method of measurement for that metric. With the help of keywords, papers will be found through various search mediums and some academic literature will be taken as this will fit the educational purpose of this paper.

3.2 For the SIEM demonstration

In order to create a SIEM demonstration, the application was decided to be a prototype, as a full-fledged product is beyond the goal of this study. One metric was decided to be used. In this case, it is time to respond. This is because it is a flexible metric, suitable for different scenarios, such as a simple scenario and a complex scenario. Additionally, it is easier to set up in the simulation because the start time and the end time can be controlled. Only one metric was used for the SIEM demonstration because, as it is a prototype, it is also meant to be simple to simulate very quickly, rather than set up complex interactions that suit a full-fledged product more. Two scenarios were decided to be set up: a simple brute-force scenario and a more complex phishing scenario where an attacker gains access to an employee's computer, and this would be detected by checking the logs and flagging them as suspicious. The SIEM demonstration would be created using Flask and basic HTML, with open-source SIEMs as inspiration for this task[21, 16]. Flask was used to make a simple backend, while HTML with some JavaScript was used for the front end.

4 METRICS

4.1 Metrics Identified

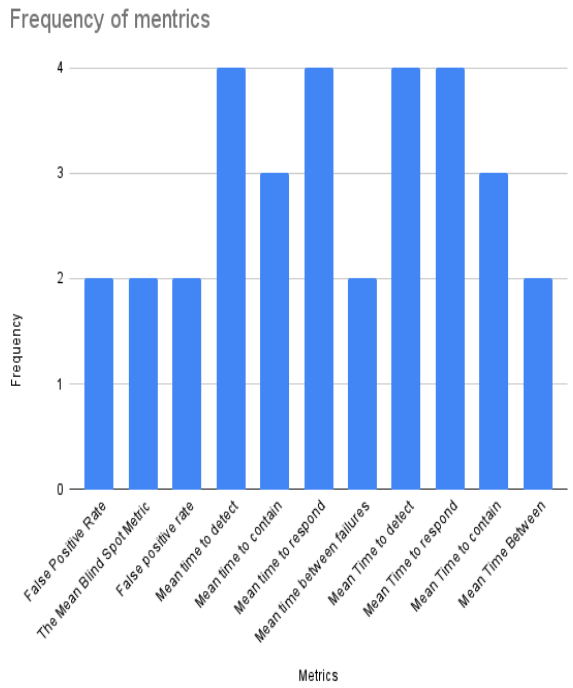

Figure 1: Popularity of each metric

When delving into both academic literature and real-world practice, a multitude of metrics have been proposed. Additional metrics that were less popular were response time[20, 12], reputational impact [10], formal interfaces for conducting agency incident management activities [7], Time measurement[19], The Mean Blind Spot Metric (MBS)[11], Performance Capability [18], false positive rate [22]. Furthermore there is mean time to acknowledge[9], reaction time for detections[8], Customer Impact [14], Incident Escalation Rate[17] to name a few. Metrics such as mean time to detect and mean time to respond are mentioned multiple times in many sources[17, 14, 9, 6, 10, 20, 12]. With this in mind, the most popular metrics and the metrics chosen are Mean Time To Detect, Detection Accuracy, Mean Blind Spot Metric and Mean Time to Respond.

- 1. What are some metrics that real-world practice mention?

Some of the metrics that real-world practice focuses on are Mean Time to Detect, Mean Time to Respond, Mean Time to Contain.[9, 10, 6] as well as mean time between failures, shown in figure 2.

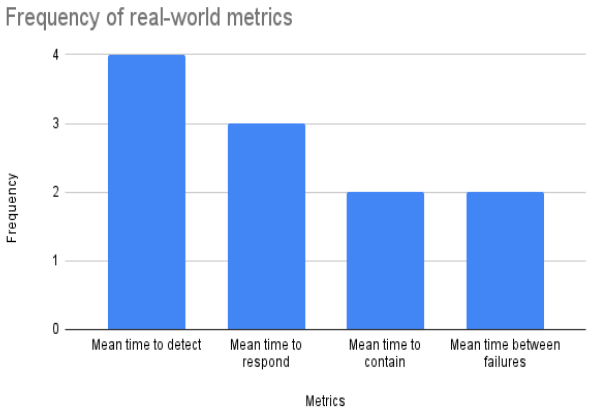

Figure 2: real-world metrics

With this trend it can be concluded that the most frequent metrics explored, are time-based metrics.

- 2. What are some metrics that academic literature mention?

The metrics that academic literature additionally mentions are the Mean Blind Spot Metric[11], Number of Critical Incidents Detected over a rolling day[2]. Figure 3 displays that the metrics used in real-world practice and academic literature can differ, with some metrics such as false positive rate not being visible in the real-world practice metrics that have been covered in this study.

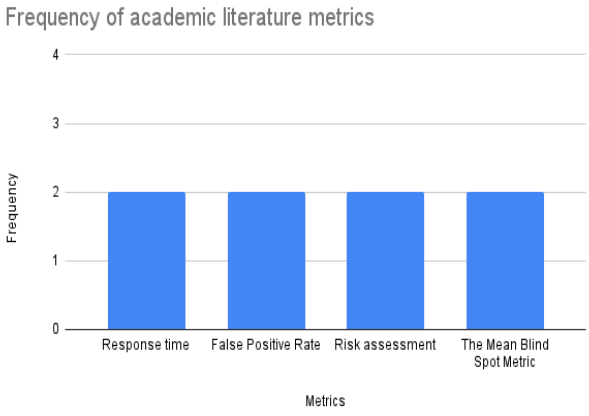

Figure 3: Academic Literature metrics

With the trend shown in Figure 3, it can be concluded from the literature used that there is a mix of time-based metrics along with other types of metrics, such as impact-based (risk assessment) or false positive rate (how many scenarios are flagged as attacks incorrectly). Academic literature has used simulated data[12, 20] using a SIEM or machine learning. It has also used surveys[2, 15] for the metrics they have used in their studies to show the use of metrics.

An interesting note is that time-based metrics are not used in the surveys and, instead, metrics such as parameterized password guessability [15] or Number of High Priority Alerts Analysed over a rolling day period [2]. It is interesting to note the difference between the metrics used in real-world practice and academic literature. The reason is suggested to be the challenges that time-based metrics run into[2] which is further discussed in the upcoming section.

- 3. How are some of these metrics measured?

Mean Time To Detect (MTTD) - time to detect is measured by

$$MTTD = \sum_{i=1}^{n} \left(\frac{T_i}{n} \right) \tag{1}$$

where T_i is the detection time for the i-th incident.[12] MTTD can be measured by measuring the time from when the incident begins (an attacker sends an attack) until it gets detected, and then taking an average of these repeated time to detects.

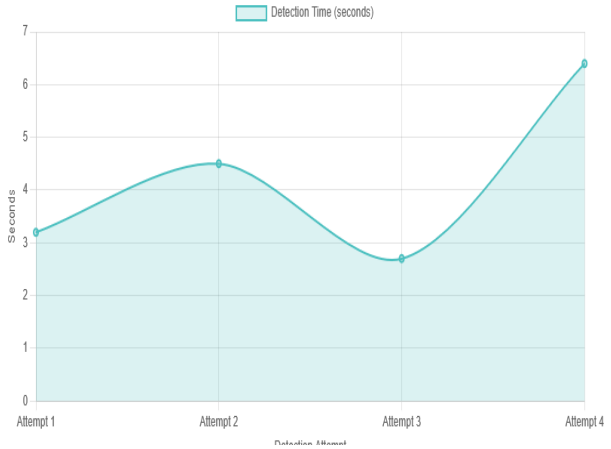

Figure 4: Metric detection times

As Figure 2 shows, a series of repeated incident with detection time, with MTTD being calculated by taking the average of these 4 to get a mean.
Mean Time To Respond (MTTR) - time to respond is measured by

$$MTTR = \sum_{i=1}^{n} \left(\frac{R_i}{n} \right) \tag{2}$$

where R_i is the response time for the i-th incident.[12] MTTR can be measured by measuring the time from when the incident was successfully detected, until the incident has been successfully responded to (such as investigated and flagged as an attack).
Detection Accuracy - this metric focuses on how many attacks are identified correctly, which can be calculated with

$$DetectionAccuracy = \left(\frac{(TP + TN)}{(TP + TN + FP + FN)} \right) \tag{3}$$

which essentially is the number of correctly identified attacks divided by n (number of scenarios used for detection).[12]
Mean Blind Spot Metric - this metric focuses on the time between security incidents - when the previous one got resolved and when the next one starts. It can be calculated by

$$MBS = \left(\frac{(sumofoccurrence - dateofrecovery)}{n - 1} \right) \tag{4}$$

[11].

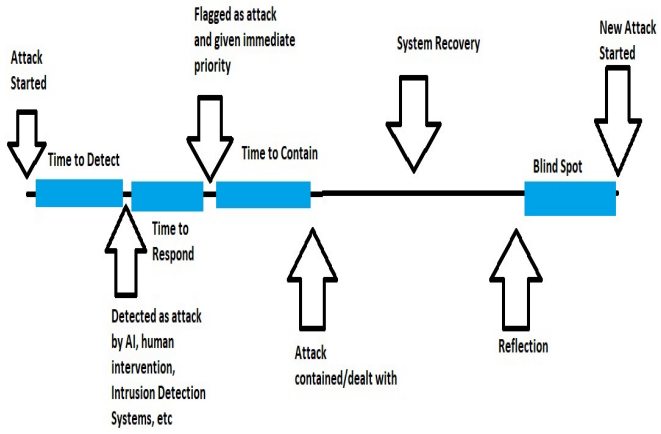

Figure 5: Timeline of attacks with metrics

4.2 Challenges

From the data that we have gathered, there are challenges associated with the metrics explored. Time-Based metrics such as Mean Time To Detect (MTTD) and Mean Time To Respond, analysts view these as misleading due to factors outside their control, such as reliance on third parties.[2] Additionally, the specific scenario in which a metric is applied significantly influences its value, as demonstrated in the SIEM example later in this study. The complexity of the scenario and the tool used are all factors that can directly impact time-based metrics, potentially resulting in values that do not fully reflect performance. This is explored more in-depth in the SIEM demonstration section of this study, using two scenarios to show a time of response, and illustrate that a shorter response time does not necessarily indicate better performance. Although the scenario may not significantly impact the Mean Blind Spot (MBS) metric—since it specifically measures the time between security incidents—it is still subject to the same external challenges as other time-based metrics, with factors outside their control.[2, 1] For example, a large volume of attacks can cause issues in the reaction time of an analyst.[1] Furthermore, met-

rics such as detection accuracy, while are good metrics, authors often mention analysis of performance metrics; but do not discuss the challenges of attaining this work[1] thus making it difficult to improve. This creates challenges for the performance metrics used, which may be detrimental to the performance of these metrics, thus creating future work in order to address these challenges and improve upon their performance.

5 SIEM DEMONSTRATION

To create the SIEM, with inspiration drawn from two open-source SIEMs [21, 16] the tool was designed to replicate the view of a data analyst working within a security operations center. One metric was sufficient to be used for this SIEM demonstration, as a singular metric can be explored more thoroughly. The single metric selected for the purpose of simulating this was time to respond, as it is a flexible metric that can be used in different scenarios, and is relatively easy to compare and simulate. The SIEM demonstration itself is simple, serving to be a prototype to showcase the metric. The SIEM demonstration will start with a page that will prompt the user to begin the simulation, which then leads to an alert dashboard. The dashboard presents a table that highlights two users, an admin and a regular user, who have been flagged for suspicious behavior which needs to be investigated.

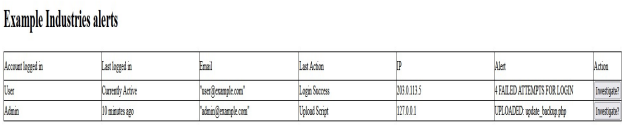

Figure 6: Two suspicious activities

Figure 6 illustrates the two situations that triggered alerts requiring investigation. The alerts are displayed in a table, showing a brief description of each user with made-up IPs and suspicious behavior that triggered the alert. There were two scenarios selected to be suitable, with each alert being a different level of complexity, with one being simple and the other being complex. This difference allows for a meaningful comparison fulfilling the goal of this simulation. One scenario was decided to be a simple brute-force scenario, whilst the other scenario is a complex compromised account into uploading script scenario. Once the user is on the page, they are presented with both flagged cases to proceed by selecting either the simple scenario or the more complex one to investigate further. Investigating either one leads to a page with different tailored scenarios.

5.1 Simple Scenario

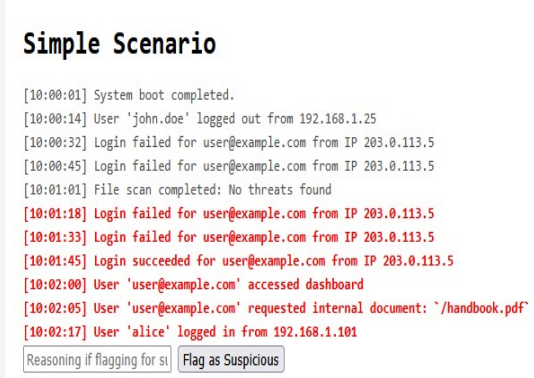

Figure 7: Simple Scenario

Figure 7 shows a scenario that showcases a simple brute-force scenario, where an attacker gains access to an employee's account via brute-force using static logs. This is shown by multiple failed logins, with the third one triggering the alert system for the data analyst being highlighted. Additionally, the scenario contains a text box that asks the analyst to type reasoning if the behavior is suspicious, and then flag it as suspicious. This is to simulate a real analyst experience, as incidents should have the motivation to flag them as an attack since this could just be an employee forgetting their login, or mistyping their password multiple times. To differentiate the two, in the logs a couple of IP addresses were shown, with the original user logging out from a static IP, and then another static IP trying to gain access, failing the login multiple times, triggering the alert after the third and finally logging in the 5th time. Once logged in, the attacker tries to access some company documentation before a different user logs in with the same IP. As an analyst, this would raise some concerns and hence be viewed as a brute-force attack.

5.2 Simple Scenario Calculation

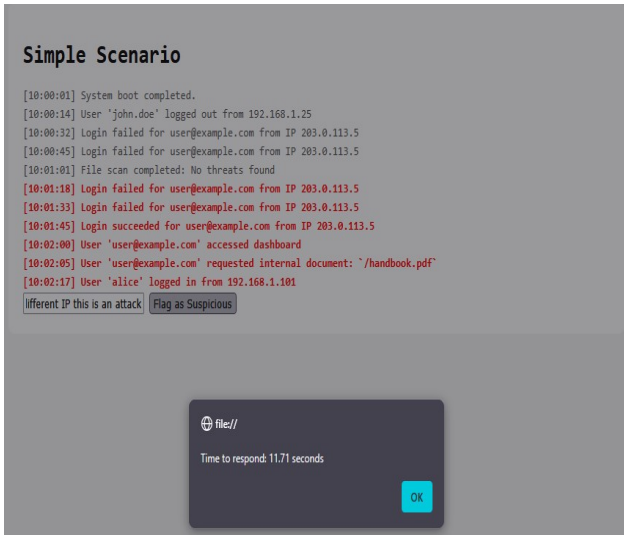

Figure 8: Simple Scenario Detection

The time to respond starts as soon as the analyst loads the page shown in Figure 7. Once the analyst logs in the reason and flags it as suspicious, an alert pops up showing the time to respond, which is shown in Figure 8. This time to respond starts from when the attack started (loaded onto the page) and ends once it is flagged as an attack, showing the response time. Unfortunately, since it is a simulation, this constraint means the time to respond starts when the analyst clicks investigate, and not right after the time to detect, since we cannot see the time to detect, as the time to detect should begin when the user successfully entered the account. With this, we only get to view one metric in a vacuum, without having the time to detect to approximate a time to respond more accurately, causing it to be shorter than it should be in a real scenario. Nevertheless, this gives us a metric to evaluate and compare with.

5.3 Complex Scenario

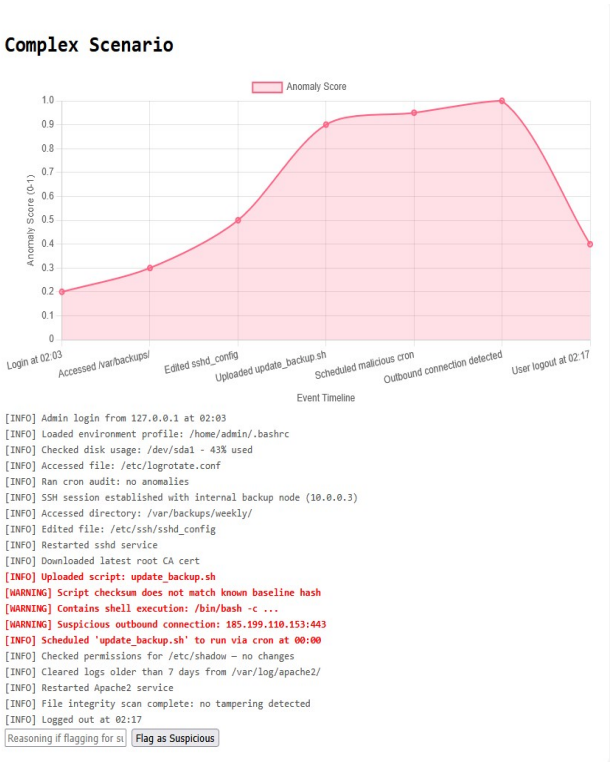

Figure 9: Complex Scenario

This scenario is a more complex scenario, which showcases an attacker gaining access to an employee with admin privileges, through phishing or via other means, and then adding in some noise to seem like a regular admin, as if nothing is suspicious. Using this noise, they try to disguise and upload a script, with a chart showing some arbitrary anomaly scores for a visual view to help analysis. Once the script is uploaded, some suspicious activity is highlighted and flagged. The time to respond works identically in this scenario, justifying its use due to its suitability and flexibility for this application. The data analyst must look through the graph and logs to decide whether this is an attack with the added noise, or if this is an admin. With the script showing, through deeper search, it can be revealed that this is an attack, and can be flagged as such after thorough reasoning by the analyst, thus giving a time to respond to this scenario.

5.4 Complex Scenario Calculation

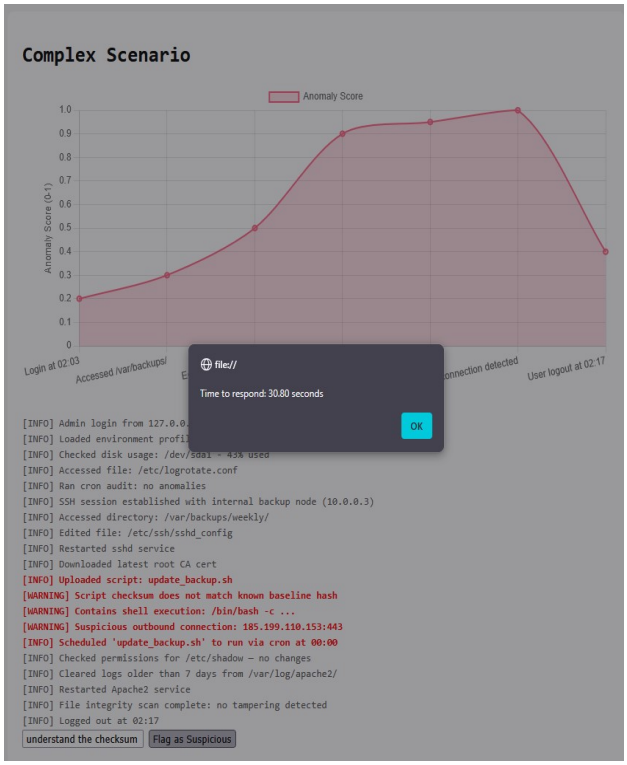

Figure 10: Complex Scenario Detection

Figure 10 shows a greater time to respond in the complex phishing scenario compared to the simple brute-force scenario's time to respond shown presented in Figure 8. This highlights how the various complexities of attacks can directly impact the time to respond. Thus, while we can compare these two different times to respond, they don't accurately display which scenario got solved faster and more aptly. The time to respond may be greater in this scenario. However, the other scenario is simpler and easier to solve, thus making it difficult to place one over the other. Nevertheless, this metric can be used to compare and measure different scenarios and cases, showcasing points of improvement. This comparison highlights that metrics must be interpreted in context - considering factors such as the complexity of the attack, experience with the attack (0-day vulnerabilities may be harder to detect), and available tools. By simulating these two scenarios with the same tool, it is seen that the complexity affects the measurement greatly. The SIEM application prototype allows organizations and researchers to benchmark detection capabilities, as well as factors that affect these capabilities.

The two times for TTR were 11.71 seconds and 30.80 seconds, leading to an MTTR of 21.255 seconds. This, while it is used as a meantime, can be flawed due to the reasons above, where, instead, there were two simple scenarios with both times of 11.71 seconds and 15.2 seconds, it would lead to a lower MTTR. This lower MTTR does not necessarily indicate that the response was strictly better/handled better, due to the fact that the complexity of the scenario was smaller, leading to an easier time responding. Due to this, it is difficult to compare which scenario the analyst responded to /handled the attack faster/better, being a great flaw of this metric as well as the additional reasons stated earlier in this study. This could be a path for future work, to improve on this issue in order to have more meaningful metrics.

5.5 Addressing a Challenge

Although the primary aim of this study is to highlight performance metrics, their measurement, and associated challenges, some suggestions arise from the observed limitations of time-based metrics. Given that time-based metrics can differ depending on the context, they may be disaggregated into sub-metrics based on the type of attack. For instance, MTTR could be broken down into a sub-metric that is measured specifically for simple attacks such as brute force. In such cases, attack complexity may become a negligible factor in the overall measurement. This approach could assist in establishing more objective metrics—one of the key challenges previously discussed—although it may not offer a complete solution. However, disaggregating time-based metrics into sub-metrics does not address all issues, such as the challenge posed by high volumes of attacks. For example, five simple attacks occurring simultaneously can skew the MTTR results, as later incidents could take longer to address due to analyst overload. One possible solution is to implement automation or distribute alerts across multiple analysts. This would allow parallel analysis of incidents, thereby improving the accuracy of scenario-specific MTTR measurements. However, there is future work in defining and establishing these changes to the metrics to investigate whether this suggestion aids in answering these challenges that have been discussed.

6 CONCLUSION

In conclusion, this study has examined some cybersecurity performance metrics that are well-defined and accompanied by established methods of measurement. Metrics such as MTTD, MTTR, MBS, and detection accuracy-as well as others briefly referenced-are widely used and prevalent in the literature that has been covered. Despite their adoption, many of these metrics face challenges which have been discussed, such as the complexity of attacks which was shown in the simple SIEM demonstration, reliance on third parties, volumes, as well as many other challenges faced by analysts.

Future work could expand on these challenges by proposing new metrics or refining existing ones to apply more specifically to certain scenarios, thereby improving clarity and reducing ambiguity. There

remains a notable research gap in the development of objective performance metrics, which continues to pose a challenge for SOC analysts [1]. This gap highlights an important area for future research. Furthermore, expanding on multiple metrics in detail as covered in this study, and addressing their challenges and solutions could form the basis for future work. Additionally, expanding the SIEM prototype to include a broader range of metrics and simulating more complex scenarios—such as varying attack volumes, analyst tools, or operational constraints—could further demonstrate the limitations and variability of these metrics can be an adequate improvement in future work.

REFERENCES

[1] Agyepong, E., Cherdantseva, Y., Reinecke, P., and and, P. B. Challenges and performance metrics for security operations center analysts: a systematic review. Journal of Cyber Security Technology 4, 3 (2020), 125–152. https://doi.org/10.1080/23742917.2019.1698178.

[2] Agyepong, E., Cherdantseva, Y., Reinecke, P., and Burnap, P. A systematic method for measuring the performance of a cyber security operations centre analyst. Computers & Security 124 (2023), 102959. https://www.sciencedirect.com/science/article/pii/S0167404822003510.

[3] AON. Intangible versus tangible risks comparison report: De-risking ai, ip, and cyber. Tech. rep., AON, (2024). https://www.aon.com/en/insights/reports/2024-intangible-versus-tangible-risks-comparison-report.

[4] Baron, H. The state of non-human identity security. Tech. rep., *Astrix, (2024). https://astrix.security/learn/whitepapers/the-state-of-non-human-identity-security/.

[5] BLACKDUCK. Global state of devsecops. Tech. rep., BlackDuck, (2024). https://www.blackduck.com/resources/analyst-reports/state-of-devsecops.html.

[6] Burke, J. The best incident response metrics and how to use them. Tech. rep., Tech Target, (2024). https://www.techtarget.com/searchsecurity/tip/The-best-incident-response-metrics-and-how-to-use-them.

[7] Dorofee, A. J., Killcrece, G., Ruefle, R., and Zajicek, M. T. Incident management capability metrics version 0.1. (2007), https://api.semanticscholar.org/CorpusID:1229898.

[8] Forsberg, J., and Frantti, T. Technical performance metrics of a security operations center. Computers & Security 135 (2023), 103529. https://www.sciencedirect.com/science/article/pii/S0167404

[9] Hammond, A., (2024). 12 incident response metrics your business should be tracking, https://www.intigriti.com/blog/business-insights/12-incident-response-metrics-your-business-should-be-tracking.

[10] Hodge, S. Key metrics to measure the effectiveness of your incident response. (2025), https://www.cyberriskinsight.com/cyber-incident/key-metrics-measure-effectiveness-incident/.

[11] Jürgen Großmann, Michael Felderer, F. S. Risk Assessment and Risk-Driven Quality Assurance. Springer Cham, (2017. https://doi.org/10.1007/978-3-319-57858-3.

[12] KUMAR, H. R. Cybersecurity incident response and forensics: Comparative analysis and proposals for improvement. 197–201. (2024), https://ijirt.org/publishedpaper/IJIRT167700_PAPER.pdf.

[13] NIST. Computer security incident handling guide. Tech. Rep. SP 800-61 Rev. 2, National Institute of Standards and Technology, (2012). https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.8
61r2.pdf.

[14] Olajide, A. O., (2024). Measuring Incident Response Effectiveness Key Metrics and KPIs for Business Analysts, https://modernanalyst.com/Resources/Articles/tabid/115/ID/66
Incident-Response-Effectiveness-Key-Metrics-and-KPIs-for-Business-Analysts.aspx.

[15] Pendleton, M., Garcia-Lebron, R., and Xu, S. A survey on security metrics. CoRR abs/1601.05792 (2016). http://arxiv.org/abs/1601.05792.

[16] Security Onion. Is security onion for me? https://securityonionsolutions.com/, Accessed last: (2025-06-12).

[17] SIRP, (2024). incident response metrics that make a difference, https://sirp.io/blog/incident-response-metrics-that-make-a-difference/.

[18] Soulé, J., Wilford, G., Broyles, P., LaPlant, M., and Maier, G. Defining standardized performance capability metrics for incident management teams based on resource typing levels, 10 (2021). https://doi.org/10.13140/RG.2.2.13480.70409.

[19] Sritapan, Vincent; Stewart, W. Z. J., and Rohm, C. T. J. Developing a metrics framework for the federal government in computer security incident response. https://doi.org/10.58729/1941-6687.1170,(2011).

[20] Timothy I Alatise, O. E. N. Threat detection and response with siem system. int j commun inf technol. (2024), https://doi.org/10.33545/2707661X.2024.v5.i1a.78.

[21] Wazuh. The open source security platform. https://wazuh.com/, Accessed: (2025-06-12).

[22] Álvaro Rocha, Carlos Hernan Fajardo-Toro, J. M. R. Developments and Advances in Defense and Security. Springer Singapore, (2023). https://doi.org/10.1007/978-981-19-7689-6.