Journal of Network and Information Security
11 (2) 2023, 30-39
http://www.publishingindia.com/jnis

The Math of SIEM Analysis: Evaluation of Key Next-Gen SIEM Features using Validation

Ertuğrul AKBAŞ
Computer Enginering Departmant, Istanbul Esenyurt University,
Istanbul, Turkey. Email: eakbas@gmail.com

Abstract: In the contemporary landscape of cybersecurity, Security Information and Event Management (SIEM) systems stand as pivotal guardians, entrusted with the critical task of collecting, correlating, and scrutinizing vast volumes of security data. As cyber threats continue to evolve at an unprecedented pace, the demand for Next-Generation SIEMs (NG-SIEMs) has witnessed a notable surge. These advanced systems come armed with heightened functionalities and sophisticated attributes to tackle the ever-changing threat landscape effectively. This manuscript presents a systematic proposal for systematically contrasting pivotal aspects of NG-SIEMs through rigorous validation processes. By instituting a well-structured approach for these comparisons, organizations can make judicious decisions, selecting the most fitting NG-SIEM solution meticulously tailored to their specific security needs. Our methodology involves not only the identification of these advanced features but also their detailed analysis in real-world scenarios. Through meticulous validation and real-world simulations, this research aims to shed light on the practical effectiveness of NG-SIEMs in diverse cybersecurity environments. By bridging the gap between theoretical attributes and practical applicability, this study contributes significantly to the understanding of NG-SIEMs' capabilities. Furthermore, the insights derived from these analyses serve as valuable guides for organizations aiming to fortify their cybersecurity postures against the ever-evolving and increasingly sophisticated cyber threats of the modern digital era.

Keywords: SIEM, Correlation engine, Log retention.

I. Introduction

Security Information and Event Management (SIEM) solutions have become essential instruments in the constantly changing field of network activities. They provide insights into anomalies in the network and guarantee improved security, control, and efficiency. This study explores the crucial field of Next-Generation SIEMs (NG-SIEMs) and emphasizes the need of a systematic strategy for assessing their basic characteristics.

With the increase in security warnings, it is now more important than ever to have cutting-edge technologies to counter new cyberthreats. With features like sophisticated threat detection techniques, machine learning capabilities, behavioral analytics, real-time monitoring, scalability, and integration with cloud services, NG-SIEMs are the progression of classic SIEMs. In particular, we concentrate on elements like correlation, threat detection, suspicious activity identification, and live log management in order to thoroughly evaluate and contrast various contemporary SIEM solutions.

Because of the wide range of features and functionalities available, choosing among the various NG-SIEM products can be difficult. This paper emphasizes that performing comparisons requires a

methodical and data-driven approach. Our suggested approach consists of analyzing correlation and log retention capabilities.

The subsequent sections of the paper are structured as follows:

- Section II: Presents related work encompassing both proprietary and open-source SIEM systems currently prevalent in the market.
- Section III: Systematic analysis of correlation features of commercially available SIEM products.
- Section IV: Systematic analysis of log retention features of commercially available SIEM products.
- Section V: Draws conclusive insights and outlines recommendations for future research directions.

In conclusion, this paper illuminates the importance of NG-SIEMs in addressing contemporary cybersecurity challenges. By offering a rigorous methodology for systematic comparison, it aids organizations in selecting the optimal NG-SIEM solution that aligns with their security needs and aspirations.

II. Related Work

Organizations have recognized the paramount importance of securing their digital assets. Consequently, the demand for comprehensive security solutions is steadily increasing, prompting numerous players and vendors to introduce their own Security Information and Event Management (SIEM) systems. In this context, we explore some of the prominent SIEM systems that are classified within the top-tier category of the latest Gartner report [15]. Additionally, we will delve into various other SIEM systems, presenting a comprehensive analysis of their distinct features.

Before embarking on a systematic analysis and comparison of these products, we will provide an overview of prior research. Muhammad Sheeraz *et al.* [1] introduced a comparative table in their work, while Gustavo Gonzalez Granadillo *et al.* [2] conducted another comparative study. Notably, esteemed institutions such as Gartner [3] offer commercial analyses of SIEM systems based on market trends and key vendors. Their annual reports position SIEM vendors as market leaders, challengers, niche players, or visionaries. Despite periodic evaluations of SIEM capabilities by entities like Gartner, there remains a noticeable absence of a comprehensive survey addressing these systems, their functionalities, and existing gaps.

In addition, other security-focused institutions such as TechTarget [4] and Info-Tech Research Group [5] have extensively reported on the capabilities of SIEM solutions and the methodologies for comparing and evaluating SIEM vendors. TechTarget frequently releases informative electronic guides on securing SIEM systems and delineating strategies for their management and success within enterprises [6]. Meanwhile, Info-Tech offers technical assessments of the SIEM vendor landscape [7], spotlighting the advantages and limitations of major commercial SIEM solutions. Both entities use the Gartner Magic Quadrant as a foundational framework for their evaluations, though they often overlook more intricate technical aspects that warrant consideration in future SIEM developments.

Likewise, organizations like Solutions Review [8] periodically release reports to assist prospective SIEM buyers in selecting the most suitable solution for their specific business needs.

Unlike the prevailing trend of comparative analyses, our approach deviates significantly by adopting a systematic methodology that involves a hands-on evaluation of the SIEM software under consideration. Rather than relying solely on indirect assessments, we will engage directly with the SIEM solutions, meticulously scrutinizing their functionality and performance. This approach allows us to offer a more immersive and insightful exploration of the systems in question.

Our methodology goes beyond superficial comparisons. We are committed to providing an in-depth comprehension of each SIEM software's capabilities by leveraging detailed screen captures and meticulously articulating the specific features we are scrutinizing. This involves a systematic

breakdown of the analysis process, shedding light on the rationale behind our selections and highlighting the intended purpose of each assessment.

By adopting this holistic and interactive approach, our work contributes to a more comprehensive understanding of the SIEM landscape. It bridges the gap between theoretical comparisons and practical implications, offering a richer perspective for organizations seeking effective and tailored security solutions. Through hands-on exploration and methodical analysis, we aim to empower decision-makers with the insights necessary to make informed choices in an increasingly complex cybersecurity landscape.

III. Deep Insight on Correlation

The value extracted from a SIEM product is intricately linked with the concept of correlation. It is through correlation that the true potential of a SIEM solution comes to fruition. If you are spending 80 percent of your time within a SIEM tool doing alert review and analysis, then you are on the right track [10]. A pivotal gauge in this journey lies in the distribution of your efforts and time. If a significant chunk, approximately 80 percent, of your operational endeavors within a SIEM tool is dedicated to the crucial tasks of reviewing and analyzing alerts, then it's a clear indicator that you are steering your cybersecurity strategy in the optimal direction.

This perspective underscores the pivotal role of fully harnessing the capabilities of a SIEM tool. The SIEM system serves as more than just a tool; it acts as a gateway to insights and actionable intelligence that bolsters your security posture. The substantial investment of time in alert review and analysis showcases a proactive approach to identifying potential threats and anomalies. It signifies that your SIEM tool is effectively fulfilling its purpose as a vigilant sentinel, sifting through the digital landscape to highlight the meaningful signals requiring your attention.

In essence, the true value proposition of a SIEM product materializes when it integrates seamlessly into your cybersecurity workflow, aiding in the identification and comprehension of security incidents. The 80 percent benchmark acts as a tangible metric, representing an organization's dedication to harnessing the full potential of a SIEM tool. It signifies a harmonious alignment of your efforts with the core essence of SIEM—to elevate your security stance by transforming raw data into actionable insights.

There are correlation engine comparison researches in the literature [11], but they do not delve deeply into satisfying the requirements of the current advanced threats. A detailed comparison of the correlation capacity of SIEM products technically will be given. The comparison based on the most critical correlation and detection capabilities:

- Rule Chain (Multi-Sage Rules)
- Correlation Logic
- List/Watchlist Management
- Real-Time Correlation
- Cross-Correlation
- Correlation Operators
- Correlation Field Operators
- Correlation Field Restrictions
- Machine Learning

Rule Chain (Multi-Sage Rules)

Rule chain is the ability to combine multiple steps (rules) of a use case without any restrictions. This type of rule detects a sequence of events occurs.

SureLog multi-stage rule sample: "if a firewall admin login has occurred and after this login action there is no configuration change immediately (wait for 15 minutes) but if there is a change in the firewall after this 15 minute within 12 hours, notify".

Most of the SIEM tools like ArcSight, Logrthym, Qradar, Securonix, and SureLog support multi-stage rules. AlienVault, Trelllix, FireEye, FortiSIEM, Solarwinds LEM, ManageEngine SIEM are the other SIEM tools that support multi-stage rules with the producer notifications.

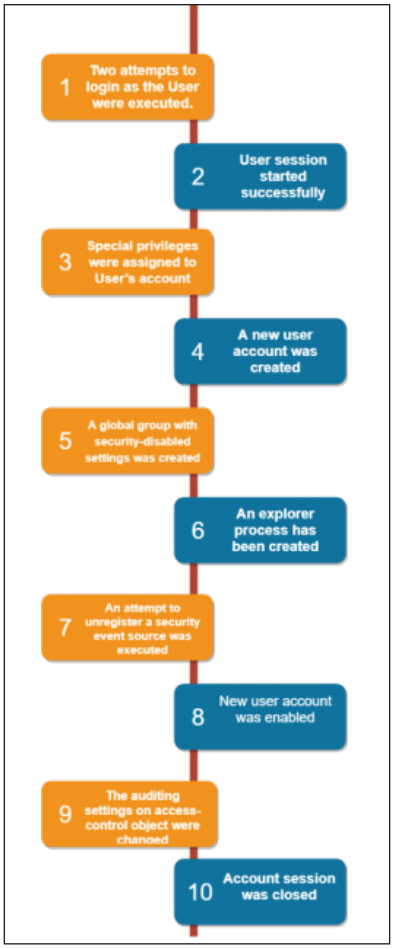

Fig. 1: SureLog Rule Chain Sample

If you wish to create a use case with Trelllix such as: "If there has been a firewall admin login and no immediate configuration change follows this login action (waiting for 15 minutes), but if a change occurs within 12 hours after the 15-minute interval, generate a notification," caution is warranted. The logic of "waiting for 15 minutes" and subsequently checking for events occurring "later within 12 hours" is not supported.

As there are two or more actions necessitating time windows, the allocated 10 minutes must be divided among them. In this particular instance, a period of five minutes is designated for each action. Following the occurrence of unsuccessful attempts within five minutes, the system initiates monitoring for a successful logon from the same IP source within the subsequent five minutes. Consequently, the implementation of wait logic between actions (rules) becomes unfeasible in this context.

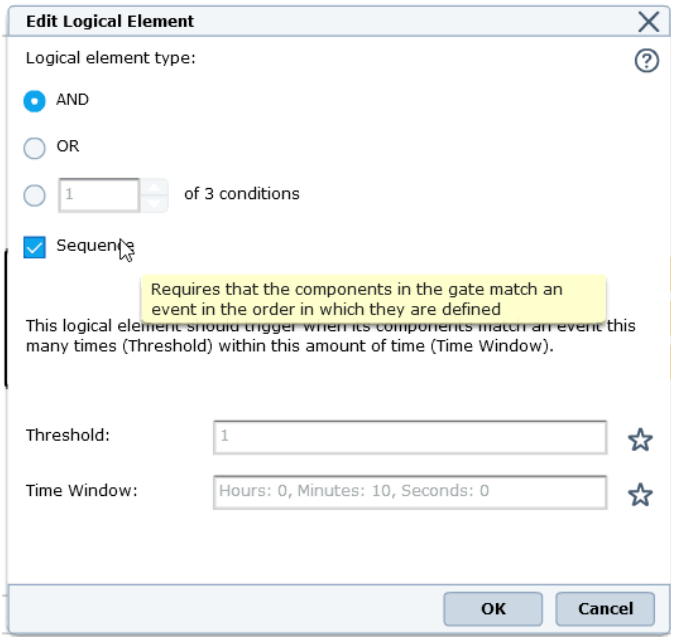

Fig. 2: Trelllix Rule Chain Sample

You should also pay the same attention to FireEye, Solarwinds LEM and ManageEngine. In ManageEngine SIEM, there is no chance to define a new rule type to chain. Also, there are schema fields restrictions to link rule chains.

Logpoint and Rapid7 do not support this kind of correlation capability.

One another requirement when chaining rules, is cross-linking of rule fields. As an example: If a device is the destination of a brute force attack and then this destination device is the source of the port scan, detect this device. While in some products this context is established by using only the fields allowed in that SIEM product's schema, in products like SureLog, ArcSight, such limitations do not exist.

Correlation Logic

Rules are discriminators used to find a certain behavior. If their designer knows what it's searching for, they will be invaluable tools. To design a rule without any limits or barriers, the correlation logic of the rule engine must be very powerful and flexible. It is hard to test the correlation logic of the SIEM tools. One of the simplest ways is to try to implement a discriminator use case (correlation rule). For example:

- "Detects more than three authentication failures from the same user within five minutes without any successful login in-between."
- This logic seems simple but "without any successful login in-between" is different. SureLog correlation engine can detect this use case.
- Arcsight also can detect similar use cases.
- If you want to detect this use case with Splunk, it might be possible to do with "transaction" events. But those searches are very taxing in the search head.
- Rapid7 and Logpoint have the same issues with Splunk.
- FortiSIEM, ManageEngine SIEM, Trelllix, Solarwinds LEM does not support this feature.
- Another test use case is detecting changes.

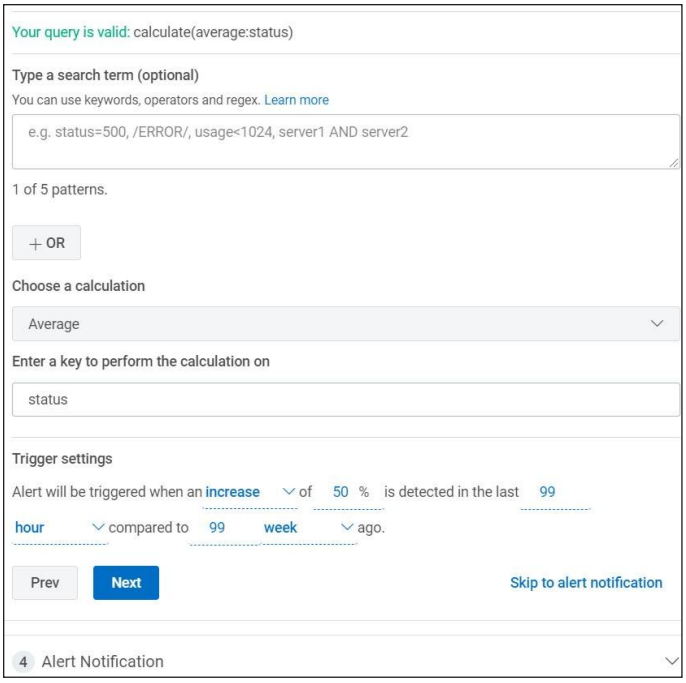

Fig. 3: Rapid7 Change Detection Wizard

There are many SIEM solutions that detects changes successfully like Rapid7, Qradar and SureLog.

Another example is "Never Seen Before Type of Rules" [12, 13]. While Arcsight, Exabeam, Qradar, Rapid7, Securonix, and SureLog have this capability, some of them do not support this type of use cases.

If we continue with the decisive scenarios that are not available in every SIEM product, we can use below rules:

- If a machine bandwidth usage is > 200Mbytes within 5 minutes or If a user to DSTIP bandwidth usage is > 500Mbytes within 10 minutes.
- If an account has not been used at least in the last 30 days, notify/lock/delete this account. (This use case is mandatory for FedRAMP Moderate; Control AC-2(3) nd NIST 800-53; Control AC-2(3)).

List/Watchlist Management

Arcsight, Logrthym, Qradar, Securonix, and SureLog have a strong list management feature. Both products support simple lists, multi-dimensional lists, complex

lists, lists with 20 columns. Also, those products add, delete, modify, list items dynamically, or manually. ANET SureLog has additional list operators like count, sum, compare, check case sensitivity in lists, tables, and cells [14].

Fig. 4: SureLog List Management Wizard

In some SIEM solutions Dynamic list usage in correlation rules is limited to one dimension like FortiSIEM, Logpoint, Trellix, RSA.

Without an advanced list/watchlist management, it is not possible to detect advanced attacks.

Real Time Correlation

Arcsight, Fireye, FortiSIEM, Logrthym, ManageEngine SIEM, Trelllix, Qradar, RSA NetWitness, Solarwinds LEM, SureLog has a real-time correlation capability. If you use Splunk ES for real time detection, you have to consider "Each real-time search unpreemptively locks 1 core on EVERY INDEXER and on your Search Head".

Elastic also has no real-time correlation feature.

Correlation Operators

Arcsight, Logrthym, Qradar, SureLog have a strong correlation operator support like:

- And
- Or
- Fallowed by Within
- Not Fallowed by Within
- At the Same Time
- Before X time
- After Y time

Not all SIEMs support all of the above operators. Trelllix has some missing operators like "At the Same Time", "Before", "Not Fallowed by Within".

Solarwinds LEM documents mention some other correlation limits. For example, you cannot create a rule using "NOT FALLOWED BY" operator. "At the Same Time", "Before" are an example of other missing correlation operators. Only "AND", "OR" Operator supported. "NOT" Operator is not supported. Also, other operators listed above are not supported.

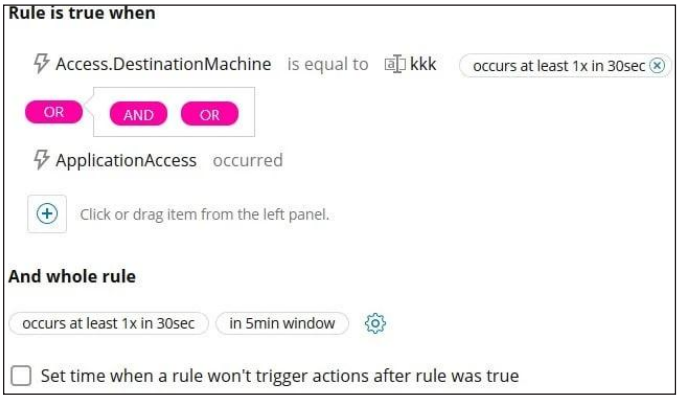

Fig. 5: Solarwinds LEM Operators

ManageEngine EventLog Analyzer correlation has only one operator "Fallowed by Within". Many operators are missing like" Not Fallowed by Within". Also, other operators listed above are not supported like "At the Same Time", "Before".

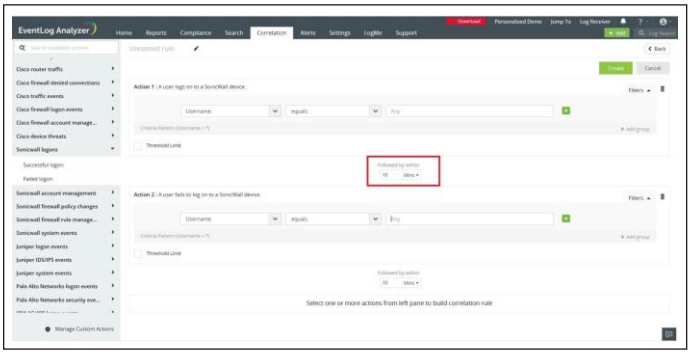

Fig. 6: ManageEngine EventLog Analyzer Operators

FortiSIEM, RSA NetWitness does not support all of the above operators like: "At the Same Time", "Before X Time".

Logpoint does not have this kind of correlation capability. Mainly it is a search-based tool.

Rapid7 does not have this kind of correlation capability. Mainly it is a search-based tool.

Machine Learning

SureLog, IBM QRadar, Arcsight, LogRhythm, Exabeam, Securonix, NetWitness Platform has NLP/ML/AI features like DGA detection, outlier detections, rarity detection, similarity detection. LogPoint uses 3rd party UEBA tool Fortscale (RSA Now).

Rare and Abnormal events are common use cases for UEBA. Exabeam and Securonix support that kind of event. SureLog also detects rare and abnormal events.

Spike, suspicious (Outliers) events are common use cases for UEBA. Exabeam and Securonix support that kind of event. SureLog, IBM QRadar, Arcsight, LogRhythm, Exabeam, Securonix, NetWitness Platform also detects outliers.

IV. Log Retention Capabilities Analysis

The techniques and tools used by SIEM products to keep logs up to date at the backend vary greatly. For example, log compression is used by IBM QRadar and SureLog, however the compression ratios differ significantly. Additionally, Elasticsearch uses Apache Lucene's mathematical formula to extend logs. As a result, comparable steps are followed by Elastic technology-based solutions like ManageEngine for raw logs, Logrhythm, and FortiSIEM (if configured with it after installation).

IBM Qradar

How much space is used per day in bytes can be calculated with the following formula:

[eps rate] * ([AveragePayloadSize in bytes] + [AverageRecordsSize in bytes]) * 86400 [15]

Splunk

You can estimate how much index disk space you will need for a given amount of incoming data. Typically, the compressed raw data file is 10% the size of the incoming, pre-indexed raw data. The associated index files range in size from approximately 10% to 110% of the raw data file. The number of unique terms in the data affect this value [16].

Trellix SIEM

Due to the number of enabled standard indexes on McAfee ESM, you can add only 5 indexes to an accumulator field. If you need more than 5, you can disable up to 42 unused standard indexes (such as sessionid, src/dst mac, src/dst port, src/dst zone, src/dst geolocation).

Trellix ESM uses standard indexes to generate queries, reports, alarms, and views. If you disable an index, McAfee ESM notifies you when it can't generate a query, report, alarm, or view due to a disabled index, but it does not identify which index is disabled. Due to this limitation, do not disable standard indexes unless needed [17].

ElasticSearch (Lucene Based Solutions)

You can estimate how much index disk space you will need for a given amount of incoming data. Disk space used (original) = 1/3 original for each indexed field + 1 * original for stored + 2 * original per field with term vectors [18].

SureLog

SureLog compresses indexes. Compressing indexes give SureLog the advantage of real-time search capability for years (Hot storage). An example of a SureLog disk capacity requirement of a hot storage for max 5000 EPS for one year is about 5 GB. When SureLog disk usage for hot storage compares to Elasticsearch and Lucene based systems, the result depicted in the below graph.

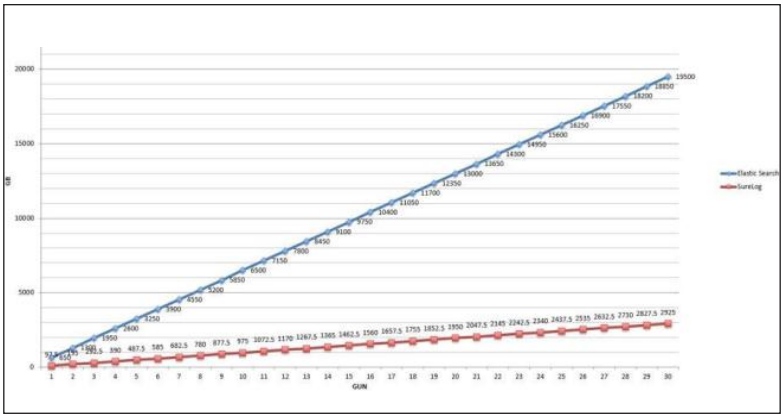

Fig. 7: SureLog vs Elasticsearch Disk Usage

Hence, the situation remains consistent in Elastic-based solutions. Additionally, a crucial point to note is that traditional databases inherently struggle to sustain logs in real-time for extended periods. Users must carefully calculate the disk costs associated with these technologies. In attempting to evade these costs, users might find themselves entangled in the laws, standards, and regulations outlined above, or worse, experience security vulnerabilities.

V. Conclusion

In conclusion, this article has shed light on the pivotal role that Security Information and Event Management (SIEM) systems play in modern cybersecurity. By aggregating, correlating, and analyzing vast amounts of security data, SIEM systems provide crucial insights to organizations in their ongoing battle against evolving cyber threats. With the ever-changing threat landscape, the demand for Next-Generation SIEMs (NG-SIEMs) has surged, offering advanced features and capabilities.

The proposed systematic approach to comparing NG-SIEM features through validation is essential in enabling organizations to make well-informed decisions about their security infrastructure. This approach assists in selecting the most suitable NG-SIEM solution tailored to specific security requirements. Key features such as correlation enhanced by machine learning and real-time hot log management have been identified as paramount in this assessment.

Throughout the paper, an exploration of the intricate details of SIEM features has been conducted. Notably, the paper has addressed the significance of correlation in optimizing the utilization of SIEM tools. The observation that spending approximately 80 percent of operational time on alert review and analysis within a SIEM tool indicates an effective security approach is particularly noteworthy. This metric underscores the vital role of a SIEM tool in transforming raw data into actionable insights, contributing to a proactive security stance.

The paper has further delved into technical aspects of correlation capabilities, highlighting specific features such as multi-stage rules, correlation logic, list/watchlist management, real-time correlation, correlation operators, and machine learning. These facets showcase the diversity in capabilities among various SIEM solutions.

Additionally, the importance of immediately available (hot, online) log management in incident response has been emphasized. Live logs hold immense value in providing real-time insights into ongoing activities, aiding in timely incident detection and response. However, challenges arise in managing the voluminous log data efficiently, necessitating careful planning, resource allocation, and adherence to regulations.

The significance of adhering to laws and standards surrounding log retention has been stressed, with MITRE, government recommendations, and RFP requirements guiding organizations in establishing

appropriate retention periods. Live logs compression methods have been discussed, revealing the diversity in approaches employed by different SIEM solutions.

In sum, this article provides a comprehensive exploration of SIEM systems, their critical features, and the methodology for comparing and assessing NG-SIEM solutions. By offering a deep understanding of these aspects, the article equips organizations with the knowledge to make informed decisions in an ever-evolving cybersecurity landscape.

References

[1] Muhd. Sheeraz, Muhd. A. Paracha, M. Ul Haque, Muhd. H. Durad, S. Muhd. Mohsin, S. S. Band, and A. Mosavi, "Effective security monitoring using efficient SIEM architecture," *Human-Centric Computing and Information Sciences*, vol. 8, 2018, doi: https://doi.org/10.22967/HCIS.2023.13.017.

[2] G. G. Granadillo, S. González-Zarzosa, and R. Diaz, "Security information and event management (SIEM): Analysis, trends, and usage in critical infrastructures," *Sensors*, vol. 21, 2021, doi: http://doi.org/10.3390/s21144759.

[3] Gartner, "Macro factors that will shape the 2020s." Accessed: May 31, 2023. [Online]. Available: https://www.gartner.com/en

[4] TechTarget, "TechTarget SearchSecurity." Accessed: May 31, 2023. [Online]. Available: http://searchsecurity.techtarget.com

[5] InfoTech, "Info-Tech Research Group." Accessed: May 31, 2023. [Online]. Available: http://www.infotech.com/

[6] TechTarget, "SearchSecurity. How to define SIEM strategy, management and success in the enterprise." Accessed: Jan. 13, 2023. [Online]. Available: https://searchsecurity.techtarget.com/essentialguide/How-to-define-SIEM-strategy-management-and-success-in-the-enterprise

[7] Info-Tech Research Group, "Vendor landscape: Security information & event management," in Optimize IT Security Management and Simplify Compliance with SIEM Tools; Technical Report; Info-Tech Research Group: London, ON, Canada, 2015.

[8] Solutions Review, "Security information and event management vendor map." Accessed: Dec. 14, 2022. [Online]. Available: https://solutionsreview.com/security-information-event-management/security-information-event-management-vendor-map

[9] O. Podzins, and A. Romanovs, "Why SIEM is irreplaceable in a secure IT environment?," in *Proceedings of the 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, Vilnius, Lithuania, Apr. 25-25, 2019, doi: https://doi.org/10.1109/eStream.2019.8732173.

[10] I. Valenzuela, "Your SIEM questions answered." Accessed: Aug. 23, 2023. [Online]. Available: https://www.sans.org/blog/your-siem-questions-answered/

[11] L. Rosa, P. Alves, T. Cruz, P. Simões, and E. Monteiro, "A comparative study of correlation engines for security event management," in *Proc. of 10th Int. Conf. on Cyber Warfare and Security (ICCWS-2015),* ISBN: 978-1-910309-98-8, ISSN: 2048-9897, Mar. 2015.

[12] E. Akbaş, "Never seen before type of rules with surelog SIEM." Accessed: Aug. 2023. [Online]. Available: https://medium.com/@eakbas/never-seen-before-type-of-rules-with-surelog-siem-cb3c0a7dc0c3

[13] E. Akbaş, "At the same time SIEM operator." Accessed: Aug. 2023. [Online]. Available: https://drertugrulakbas.medium.com/at-the-same-time-siem-operator-be8d6598b7b8

[14] SureLog, "SureLog lists." Accessed: Sep. 2023. [Online]. Available: https://medium.com/@surelog/surelog-lists-b952aca0a047

[15] IBM Qradar. Accessed: Aug. 2023. [Online]. Available: https://www.ibm.com/support/pages/qradar-how-determine-average-event-payload-and-record-size-bytes-updated

[16] Splunk. Accessed: Aug. 2023. [Online]. Available: https://docs.splunk.com/Documentation/Splunk/8.0.0/Capacity/Estimateyourstoragerequirements

[17] Trellix (McAfee). Accessed: Aug. 2023. [Online]. Available: https://docs.trellix.com/bundle/enterprise-security-manager-11.5.x-installation-guide/page/GUID-2F189D5A-AC92-4965-80A4-03EE2272F37C.html

[18] Apache Lucene. Accessed: Aug. 2023. [Online]. Available: https://lucidworks.com/post/estimating-memory-and-storage-for-lucenesolr/