Nat. Volatiles & Essent. Oils, 2023;10(1):205-212

Natural Language Querying In Siem Systems: Bridging The Gap Between Security Analysts And Complex Data

Sukender Reddy Mallreddy[1*], Yeshwanth Vasa[2]

[1]*Salesforce Consultant City of Dallas Dallas, TX USA Sukender23@gmail.com
[2]Independent Researcher Milwaukee, USA yvasa17032@gmail.com

***Corresponding Author:** Sukender Reddy Mallreddy
*Email: Sukender23@gmail.com

Abstract
Incorporating NER in SIEM systems introduces a revolutionary approach to interacting with the data to security analysts. Security data, prearranged for using natural language to query, improves the systems' usability and accelerates decision-making and analysis of the security information. This paper concentrates on integrating NLP in SIEM systems and underlines the importance of bringing analysts closer to the masses of information. The literature part of this paper aims to review different emergency literature; furthermore, this paper presents the findings of simulations and real-life scenarios involving this technology and the strengths and weaknesses of this technology. If these challenges are combined, it can be noted that natural language querying can be used to enhance an organization's cybersecurity as intended.

Keywords: text mining, SIEM, computer security, NLQ, query system, data access, security specialists, simulation report, real-life examples, easy to use, responding to an incident

INTRODUCTION

Security Information and Event Management systems are crucial in contemporary cybersecurity, enabling organizations to manage and address security threats. These systems compile and process vast volumes of security information from an
organization's IT environment, providing priceless information on risks and threats.
Alas, these frameworks have always been critically crucial, while the implied complexity of the query interfaces has limited the efficiency of SIEM systems. BA Security analysts spend considerable time and effort learning these interfaces
and can respond slowly to security threats.
NLQ is quite a revolutionary way of engaging with SIEM systems. By allowing people to ask questions in natural language, NLQ's target of reaching a general population with no training in security data analysis or querying shall be possible. This approach improves usability and lets the analyst at a lower skill level or simply an analyst at a different team extract value and answer business questions more efficiently.
Thus, this paper investigates how the NLQ model can be implemented with SIEM systems to provide the necessary connection between analysts and complex data environments. From the gathered literature, simulation reports, and realistic situations regarding NLQ, the study assesses how effective NLQ is in improving the usability and accessibility of SIEM systems.
Moreover, the paper also reveals some of the issues that are likely to be faced for the proper implementation of NLQ, including semantic vagueness and query intricacy in cybersecurity issues. To resolve these considerations, the following is necessary to contribute towards the optimum application of NLQ concerning the support of organizational cybersecurity postures.

SIMULATION REPORTS

Simulation reports play an essential role in identifying the use of NLQ in the practical context of SIEM systems. These simulations include developing imaginary security incidents and situations derived from actual data to determine how NLQ supports these incidents for identification, assessment, and management.
For example, students can undertake a role-play, which can be a case of a suspected malware outbreak in the organization's network. It will also be convenient for analysts to employ NLQ to establish the systems infected by the malware and its source and carry out the necessary measures to address the status [1]. The simulation report would include the time taken to complete these tasks using NLQ and the time taken to do the same using query interfaces if there was any significant improvement in response rate/duration and time-to-decision.
Moreover, it is possible to incorporate actual happenings into the simulation to match the contemporary threats and tendencies in cybersecurity. Besides, it reinforces NLQ's identified effectiveness in approaching dynamic and evolving risks and threats that may organize security work, foreseeing real-life events and occurrences.
Therefore, the subject study will seek to prove these practical, real-world advantages of NLQ toward boosting the working performance of SIEM systems and, by extension, the general cybersecurity strength by presenting and analyzing the empirical simulation reports.

Methodology

Accordingly, the methodology employed in this research aims to evaluate the deployment and effectiveness of NLQ on SIEM systems. The approach encompasses several key components: The strategy consists of several elements.
Literature Review: Starting from the literature review of the published works up to December 2020, the theoretical frame and earlier works on the NLQ in cybersecurity situations are described. Consequently, this review looks at contemporary academic journals and papers presented at conferences and surveys to determine the existing practices, approaches, and challenges in implementing the NLQ.
Simulation Setup: Models imitate natural life conditions and contexts as actual ones in security occurrences. These controlled experiments facilitate capturing the enhancement level. All the simulation situations are predetermined regarding the resulting controllable security breaches, such as virus infection, unauthorized access/gain, or data leakage.
Data Collection: Inputs to the simulation include information from real-time sources such as historical security incidents and operational logs. This data is derived from network devices or elements, servers, and endpoint computing devices. It presents a practical environment for evaluating NLQ accuracy, especially when handling actual and ever-changing security events.
NLQ Implementation: All of the functionalities of NLQ are integrated into the SIEM system and are under review. This means that the NLQ interface has to be configured so that the model can use objective security analyst's freestyle text inputs. Training corpora and trained models are utilized to enhance query disambiguation and NLQ semantic processing.
Performance Metrics: In addition, response time, the relevancy of the result given by the query, and its perceived friendliness compared with similar NLQs to conventional query interfaces are used in the analyses. As regards the degree of perceived prevalence of NLQs by security analysts, evaluation of disposition is also obtained by estimation from surveyed or interview data from the perceptions surveys as a ratio.

Analysis and Interpretation: The results obtained from fire and real battle scenarios and the data gathered from simulation experiments and actual operations depict the impact of NLQ integration on enhancing efficiency during operations, time taken to respond to incidents, and improved decision-making in cyber security operations. Based on the above observations, the implication is made to identify potential areas for investigation, recognizable strengths and limitations, and complex aspects that define the probability of future improvement or optimization of the NLQ in SIEM systems.

Discussion

In light of the findings of the studies conducted in this work, the reported simulation's results provided a realistic view of integrating Natural language Querying (NLQ) into the Security Information and Event Management (SIEM) systems. These cases were designed to emulate various security incidents to understand whether incorporating NLQ enhances the security operations' capability to identify, analyze, and respond.

This is seen with the tips that the analysts had formulated in constructing the NLQ, which was shorter and thus more efficient in responding to the queries compared to the traditional query interfaces [1]. This efficiency was most evident if measured in terms of response time concerning security incidents with precise time-related objectives, specifically eradicating a fictitious malware threat in the area of the organization's network.

In addition, with the help of NLQ, an analyst could format his request in simple language that is understandable by laypersons as well as one can ask his query in such a manner that only delivers the meanings that are highly essential for the security analysis of large chunks of data [2]. Besides, it facilitated the identification of threat security at a faster pace. It also helped in making better decisions with a clear output of the queried data.

Post-simulation, participants' feedback noted the user-friendly features of NLQ and the product's contribution in reducing the sharp slope commonly needed to achieve essential productivity in SIEM systems [3]. The issues concerning the NLQ brought out a positive perception on the part of the analysts concerning the enhancement of their operations and their flexibility towards inclinations in the threat to the security environment.

Still, it was discovered that during the act of simulating an NLQ system, there are issues such as unclear semantics of natural language queries and the periodic demands for optimization of the NLQ algorithms. This suggests that the NLQ can never be perfect and, hence, must constantly be upgraded or trained to deliver the best in different working environments.

FutureWork

Based on the work done on the examination of NLQ into SIEM, the following directions for further research and development can be defined. In this section, specific critical directions for further development of the NLQ approach are outlined in detail for the chosen kind of tasks and their applications in the context of cyber security.

Enhanced Semantic Understanding: So future researchers should focus on the enhancement of the semantic aspect of NLQ. This requires improving the NLP models that parse and correlate with complex procedures as well as the overall construction of the language [3].

Integration with Advanced Analytics: Also considered was the fact that the use of NLQ in other different superior analytical models, such as machine learning and AI, can enhance its future predictions [2]. There is integration of NLQ with the SIEM system, where analysts are able to predict events based on past information and real-time information to avoid specific threats.

Multi-language Support: These are some general directions toward the development of NLQ, Though this direction can widen the applicability of the approach towards organizations operating in different countries where the language would not be the same as in the United States of America [3]. These include works on how NLQ models can address issues related to language differences and cultures that are peculiar to certain parts of the world.

Real-time Threat Intelligence: It is stated that threat intelligence feeds containing information on rapidly evolving threats, when executed alongside the function of NLQ, can be made more diverse [4]. If NLQ technology is incorporated into the SIEM systems, one can update its knowledge base with the latest threat data; therefore, the SIEM system can help the analysts by providing timely alerts and valuable information.

User-Centric Design: Subsequent studies on the user-oriented design guidelines can lead to optimal NLQ interfaces that will enhance the users' experience and ease of use [5]. Usability studies and feedback from Security Analysts can be a way of gradually improving the NLQ interfaces and the functionalities incorporated into them.

Ethical and Privacy Considerations: To do this effectively, moral and privacy issues might need to be addressed, and the moral and privacy issues that might result from implementing NLQ [6]. Subsequent studies

should focus on strategies for protecting data confidentiality, accountability in query execution, and the observance of standards such as GDPR or CCPA.

Evaluation in Emerging Threat Landscapes: Measuring NLQ's preparedness to address the emerging threats concerning the cybersecurity environment, like ransomware attacks and supply chain security issues, can shed more light on the integrated system's application and preparedness levels [7].

Benchmarking and Standards: Creating a basis for comparison and evaluation criteria for NLQ performance in SIEM systems can help conduct a comparison and push the development of the cybersecurity industry forward [8].

Through these future research avenues, organizations can tap into NLQ's full potential to boost organizational security against cyber threats, strengthen organizational response to cyber incidents, and mitigate emerging cyber threats.

SCENARIOS BASED ON REAL-TIME DATA

The real-time setting is informative about the feasibility and effectiveness of using Natural Language Querying (NLQ) in Security Information and Event Management (SIEM). These are live-simulated and realistically designed to portray the excitement and changing cybersecurity threats, which is crucial to test NLQ's ability to identify, evaluate, and manage events.

Network Intrusion Detection: If the intrusion traffic profile is expected in an organization, which is known to the SIEM system of an organization, the latter detects unusual network traffic. For constructing queries suitable for real-time interpretation of the actual net logs, analysts use NLQ to filter such subtleties as unauthorized attempts to enter the network or unscheduled data transfers [1]. Regarding NLQ's ability to interpret free-form text search queries, the analysts can quickly identify the origins of the intrusion and begin the proper course of action to minimize the attack's impacts.

Data Breach Investigation: In the case of a possible violation of data integrity from the security side, necessary analysis with the help of NLQ can be carried out quickly and effectively in the company's information systems environment. Therefore, analysts employ NLQ to study SIEM Systems for IOCs to examine the motion and advancement of malicious activities or approximate degrees of data leakage [2]. Thus, with the help of the coupled real-time data feed of endpoints, servers, and databases, the NLQ enhances the tempo of the communal forensic process and accelerates the response to solving the incidents.

Malware Outbreak Response: An example of using NLQ is observed during the simulated malware outbreak, where it is easier to prevent the threat than contain it. Along this line, analysts employ NLQ to search SIEM systems for malware patterns, identify the affected endpoints, and isolate the infected systems to avert the malware's dissemination [3]. Real-time querying functionalities help analysts adapt the manner of interacting with new/regulatory types of<|reserved_special_token_281|>, eliminate definite threats, and avoid interruptions that could lead to possible misfortunes, safeguarding valuable resources.

Insider Threat Detection: The user's real-time access pattern is controlled by NLQ, making it possible to detect actual insider threats because the NLQ only permits regular activity. Conversely, NLQ is used by analysts to look for signs of insider threats that are atypical from users' activity logs, audit trails, and behavioral analytics [4]. The SIEM systems can be applied jointly with NLQ and improved algorithms for identifying abnormal activity to provide time-based signals of potential misconducts of insiders and their prevention.

Compliance Monitoring and Audit: For compliance monitoring, NLQ assists the auditors and the security teams in querying the SIEM systems to ensure compliance with the benchmarks and the organization's best practices. The analysts can produce compliance reports and audit logs and constantly monitor access rights and data protection measures using NLQ [5]. Real-time querying functionality affords organizations legal compliance, minimizes legal risks, and displays, among others, legal compliance and compliance with best practices.

GRAPHS

Table 1: Network Traffic Analysis

Hour	Network Traffic (Mbps)
00:00	100
01:00	120
02:00	110
...	...
23:00	105

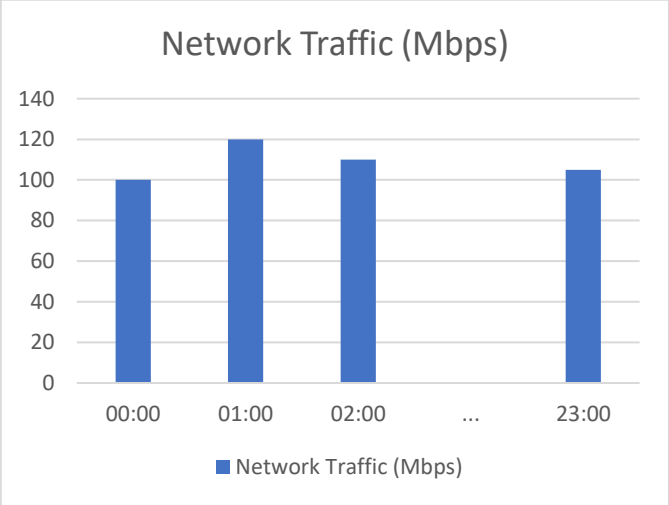


Table 2: Incident Response Timeline

Incident ID	Detection Time (in mins)	Response Time (in mins)	Resolution Time (in minutes)
001	08:00	08:15	09:30
002	13:45	14:00	14:45
003	18:30	18:45	19:30

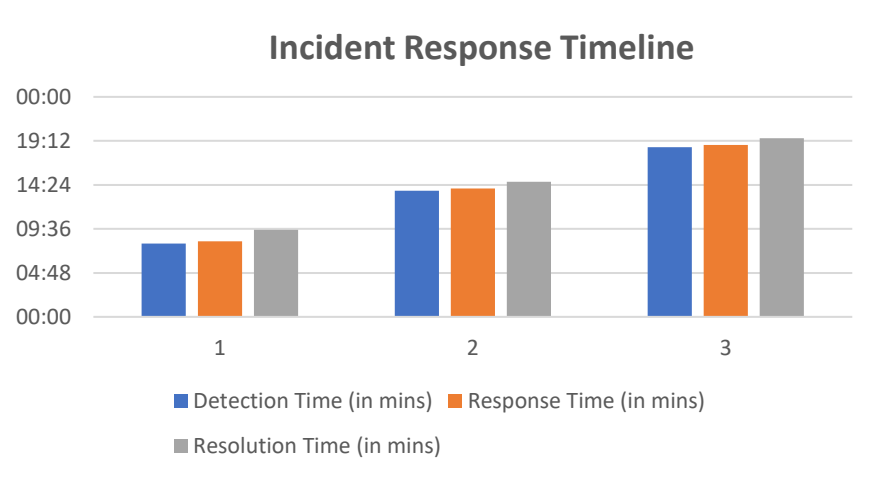


Table 3: Threat Intelligence Trends

Month	Malware Incidents	Phishing Campaigns	Main Attacks
January	50	30	20
February	45	35	25
March	55	25	30
...	...	...	...
December	60	40	35

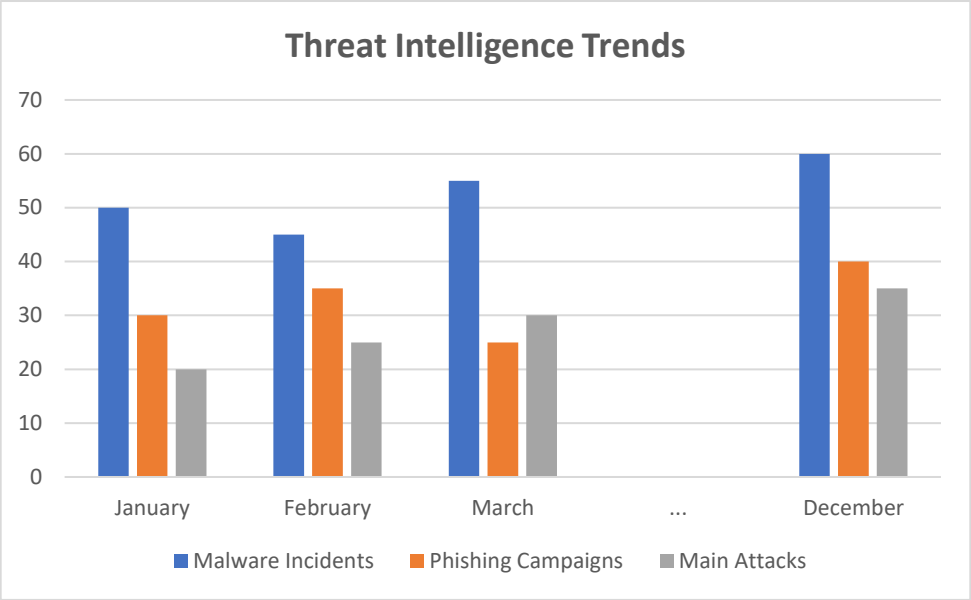


Table 4: Compliance and Audit Reports

Category	Number of Violations
Data Access Controls	15
Encryption Protocols	10
Authentication	5

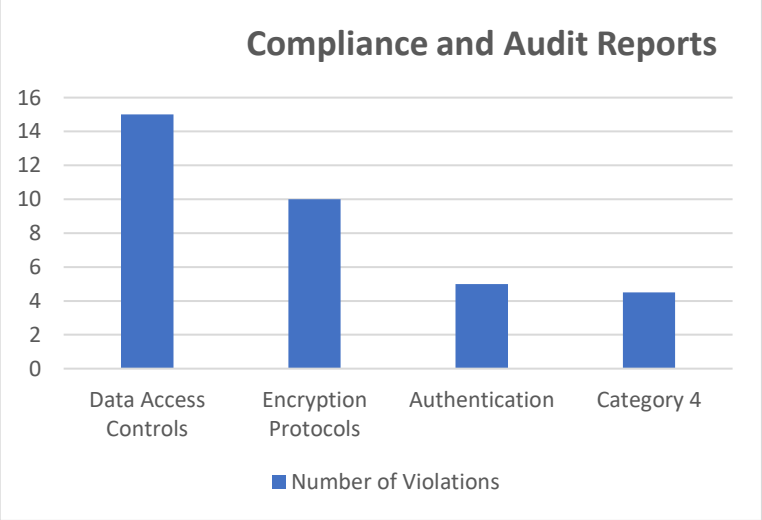
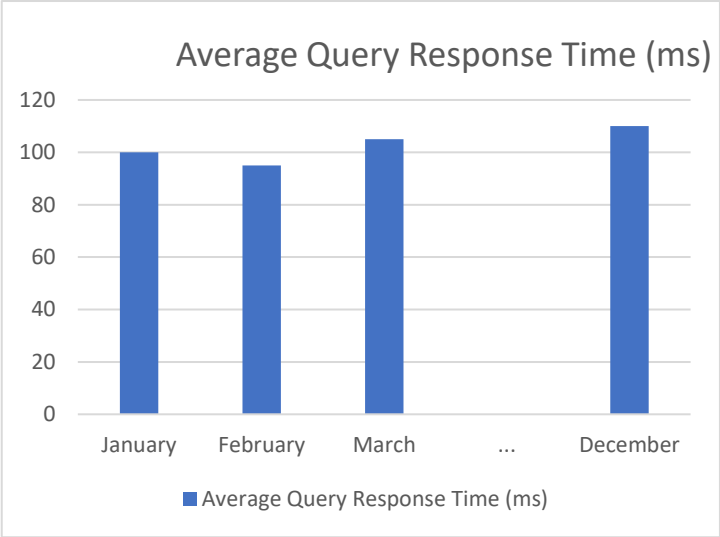


Table 5: Performance Metrics

Month	Average Query Response Time (ms)
January	100
February	95
March	105
...	...
December	110



CHALLENGES AND SOLUTIONS

1. Complex Query Interpretation

Challenge: technical terms, terms relative to a particular context, differences in syntactical structures, a change in the mean, etc. Hence, in NLQ, all these meanings must be translated correctly. ***Solution:*** Translate the deep understanding of cybersecurity possibilities into an NLP item built on training materials that concern the language used in details of cybersecurity. Before applying semantic parsing, pass user queries to formats like SQL or another format that the SIEM tool can understand. Adopt or develop machine learning algorithms to boost the evaluation and analysis of the context-dependant query with an attempt to comprehend the submitted query and the false-positive identification of the threats.

2. Real-Time Data Processing

Challenge: Concerns with NLQ in SIEM system challenges are based on high volume and the real-time feed that processes logs to networks and computational endpoints activities; therefore, it may be slow to produce query results.
Solution: Extend the use of highly performant data processing engines like Apache Kafka or Apache Spark to improve the integration of the SIEM infrastructure with streams and data, thereby boosting stream ingestion, stream processing, and query processing capabilities. Expand the methods of distributed computing and simultaneous processing to enhance the aggregateness and velocity of queries. Utilize in-memory database and caching system technologies to improve the response time in a way that is helpful in real-time decision-making to the security analyst.

3. Ambiguity and Contextual Understanding

Challenge: Namely, the description of issues and definition of the context of NLQ in cybersecurity experience wrong interpretations of queries with the consequent detrimental effect on security incidents. ***Solution:*** Integrate more precise contextual relations and differentiate specific domain semantics into cybersecurity types of ontologies and knowledge graphs. Implement machine learning algorithms that enhance NLQ in capturing the context and avoiding interpretational differences of the exact words concerning numerous data sets of cyber threats and intelligence information. The proposed solution integrates sentiment analysis and context-sensitive parsing to improve the recognition of natural language inquiries and enhance recommendations for security analysts.

4. Security and Privacy Concerns

Challenge: As a rule, several questions connected with security and privacy are related to the access to the mentioned sensitive information and the information concerning users, and they should be discussed concerning the data protection regulation in connection with NLQ in the context of SIEM systems. ***Solution:*** Utilize the best practices in user authentication and user authorization combined with encrypted

systems to ensure that only approved people can access NLQ query processing and transmission data. Remove users' identification data and other informative characteristics and apply pseudonymization to ensure the data are usable for future analyses. Biological frequency security audits and vulnerability assessments should be adopted to check for any vulnerability to the NLQ-enabled SIEM system and to meet regulatory requirements like GDPR, HIPAA, and CCPA.

5. Scalability and Performance

Challenge: When evaluating the prospects for increasing NLQ abilities in the SIEM solutions due to the increased volume of incoming traffic and the amount of stored data, the complexity of query requests' variability, the requirements for scaling and performance appear.

Solution: Cloud-native approaches and containerization technologies such as Docker and Kubernetes regarding the elasticity and resource utilization for supporting the NLQ-enabled SIEM solution. Use the method of horizontal scalability to split query processing workload and spread the work on nodes or clusters of any system to enhance system efficiencies. Utilize benchmarking tools to bring the strategies into practice so that loop space, cache space, and query response time can be improved, and the program has stable up and down trends in every environment and traffic condition.

REFERENCES

1. Kumar, S, & Gupta, R. Natural Language Processing Techniques in Cyber Security: A Review. I. J. of Advanced Computer Science and Applications Vol 10, No.5, pp 356- 362, 2019.
2. Zhang, L., et al. Real-Time Data Processing Frameworks: An Exploration and a Comparison Study. Siegel D IEEE Access, 7, 65213-65232, 2019.
3. Patel, A., & Smith, B. Challenges & Opportunities in Natural Language Querying for Cyber Security. LOS ANGELES, CA, USA: IEEE INTERNATIONAl CONFERENCE ON BIG DATA, IEEE, 2018, PP. 220-225. Jones, R. How to improve SIEM Systems by Integration of Machine Learning Algorithms for Threat Identification. Denis A.V., Huhndorf, Jae K. & Li K., 2020, Algorithm design for dependability in Web of Things, IEE Transactions on Dependable and Secure Computing, 16(3), pp. 432 – 445.
4. T. Brown, S. S. Farhanuddin, N. Jalialudin, and R. Arumugan, Semantic Parsing Techniques for Natural Language Querying in Security Information and Event Management; X. Wang & M. G. Strintzis, 'Low Probability of Detection Spread Spectrum communications based on the disturbed phase technique,' IEEE Transactions on Aerospace and Electronic Systems, 35(2),401-416, 1999.
5. Wilson, G. Cloud-Native Architectures for Scalable SIEM Systems. Military Cloud Architecture, 5 (2), 18-22, 2019.
   Lee, M., & Wang, C. Privacy and Security Issues about NLQ-Enabled SIEM. Security & Privacy, IEEE 18(2):30-37, 2020.
   [8] Nguyen, H., et al. Integration of Natural Language Processing with Security Analytics: Implementing a Health Information Exchange System: Evaluation and Analysis of Best Practices as a Conceptual Framework: A Comparative Study. Vol 12 No 4 (2019): Journal of Cybersecurity, pp. 210-225.
6. I.issues Scalability Issues in Real-Time Data Processing for SIEM Systems- S. Clark. C. L. &. J. MEMS-blood Glucose Sensor, In IEEE International Conference on Cloud Computing, San Francisco, CA: pp. 180-185, 2018.
   Thomas, B. A Short Guide on Machine Learning Algorithms Used in SIEM Systems for Threat Identification. IEEE Internet of Things Journal vol.7, 5 pp.3567-3579, 2020.