International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I10P101

Volume 74 Issue 10, 1-14, October 2026
© 2026 Seventh Sense Research Group®

Original Article

Development of Data Structures using the DataGuard SIEM-DLP Algorithm for Data Loss Prevention and Data Analysis in Distributed Communication Networks

Tipaporn Supamid[1], Suwit Wongkumsin[2]

[1,2]*Department of Computer Engineering, Faculty of Engineering, Kasem Bundit University, Thailand.*

[1]*Corresponding Author : tipaporn.sup@kbu.ac.th*

Received: 07 May 2025 Revised: 25 July 2026 Accepted: 18 September 2026 Published: 27 October 2026

Abstract - *The research entitled "Development of Data Structures Using the DataGuard SIEM-DLP Algorithm for Data Loss Prevention and Data Analysis in Distributed Communication Networks" aims to develop the DataGuard SIEM-DLP algorithm. This study established two main research hypotheses: (1) to enhance the efficiency of security and safety in distributed communication networks, and (2) to develop a data structure based on the DataGuard SIEM-DLP algorithm that can operate with stability and support real-time processing within the specified time frame. The results of the data analysis can be summarized as follows: (1) system performance analysis showed an average of 95.7%, (2) time efficiency analysis during testing ranged between 25–100%, (3) data replication experiments achieved an average of 98.5%, and (4) data loss prevention testing showed an average of 91–94%. The datasets used for testing the DataGuard SIEM-DLP algorithm were obtained from the UNSW-NB15 and CIC-IDS2017 databases. Overall, the analysis revealed that the DataGuard SIEM-DLP algorithm reduced data loss during communication by up to 30% and increased data analysis efficiency by up to 90%. In addition, the distributed communication network system operated with significantly greater stability and efficiency.*

Keywords - *Data structure, Data loss prevention, Distributed communication networks.*

1. Introduction

Modern distributed communication networks are systems that are connected with a wide variety of devices, enabling rapid connectivity and communication, diverse forms of communication, information exchange, and enhanced collaboration with other devices. Therefore, workload distribution and information dissemination allow collected data to be analysed and processed across multiple nodes, which significantly increases the reliability of distributed communication networks and enables faster scalability.

However, distributed communication networks are highly important in the field of computer engineering because they facilitate smooth data circulation, increase convenience and flexibility, enhance rapid decision-making, and help reduce the weaknesses of centralized systems. The development of data structures using the DataGuard SIEM-DLP algorithm by Tipaporn aims to reduce security issues, data redundancy, lack of clarity in data transmission and reception, control the dissemination of incorrect or distorted information, address incomplete data transmission and reception, and overcome deficiencies in management, as well as establish standards for the development of distributed communication networks. The details of distributed data structure development are as follows: 1. Data Replication is one of the simplest and most effective methods for preventing data loss. This method involves creating copies of data in multiple locations within the network, so that if data is lost or an error occurs at any point, the data can be retrieved from another location. Data replication is not complex to implement, as there are tools and data management systems that easily support this approach, such as database systems that support automatic replication or cloud services with built-in data replication.

In addition, 2. Error Detection Mechanisms, such as checksum verification, are simple to implement and effectively prevent data corruption. These techniques can be easily integrated with existing data transmission systems without requiring significant changes to the network infrastructure. The advantage is that data loss or corruption during transmission can be detected and corrected immediately. However, testing of distributed communication network systems has revealed that data transmission loss still occurs, including data loss during transmission. Therefore, unauthorized data access and attacks on data transmission and reception systems can cause errors related to security and effective data management. Consequently, the development of data structures using the DataGuard SIEM-DLP algorithm by

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Tipaporn is critically necessary to enhance security capabilities, reliability, and the overall efficiency of distributed communication networks in the future.

The research entitled "Development of Data Structures Using the DataGuard SIEM-DLP Algorithm for Data Loss Prevention and Data Analysis in Distributed Communication Networks" integrates the processes of SIEM and DLP, which are often implemented separately. In this study, the researchers developed the DataGuard SIEM-DLP algorithm, which combines replication, error detection, and encryption into a single platform. This unified framework allows data verification and monitoring to be performed within one system. Unlike previous studies that treated replication, error detection, and encryption as independent processes, this research consolidates them into the DataGuard SIEM-DLP Algorithm as a single platform, thereby creating a data structure that is both stable and secure for distributed communication networks.

2. Theories and Related Research

At present, the prevention of data loss and data analysis, in order to use the obtained data to develop efficient data structures, requires the application of four theoretical principles as follows: 1. Data Replication Theory, 2. Distributed Computing Theory, 3. Error Detection and Correction Theory, and 4. Encryption Theory. The details of each theory are described as follows.

Data replication theory is a solution to the problem of data loss or damage that may occur during data transmission through network systems, consistent with [1] Therefore, replication is performed by considering data items one by one until all items in the list are completed, or by following specified conditions using various computational methods that are defined as replication conditions, consistent with [11] The design of systems in which data and processing are distributed across multiple nodes enables diverse components to work together. The key advantage of distributed processing lies in its scalability and reliability, achieved through the use of established patterns and paradigms.

This also includes principles used in instruction sets or operational processes that repeatedly perform the same tasks many times, consistent with the research of [2] related to improving efficiency for data access, condition checking, and verification, as well as checking and authenticating users through controllers related to replication checking based on the number of iterations or until a change in conditions becomes false, consistent [3]. However, research on replication is also consistent with the study [22], which reported that replication analysis has been applied in cloud, fog, and edge environments by utilizing taxonomy-based replication techniques. Similarly, [13] LLMs can generate formal specifications but face reliability issues due to their probabilistic nature. This study introduces a PCFG-based framework and fused uncertainty signals, reducing verification errors by up to 100%.

2.1. Error Detection and Correction Theory

The theory of error detection and error correction involves transmitting data to the receiver so that the receiver can be confident that the transmitted data is correct and identical to the original data from the source. If errors occur during data transmission due to signal interference, bit loss, or faults caused by the transmission equipment along the communication path, [5] the devices are able to detect errors and, in some cases, automatically intercept and correct the errors that occur. [4] In addition, research on error detection and correction is consistent with the study by [21], which demonstrated the use of Large Language Models (LLMs) for detecting and correcting errors. While these techniques show potential, they still lack sufficient accuracy in medical applications. Similarly, [20] This study focuses on the design of binary codes for error detection and correction under computational constraints. It establishes a connection between coding theory and processing with limited resources.

2.2. Distributed Computing Theory

Distributed Computing Theory involves cooperation among multiple computer systems in order to solve or prevent problems that may occur. [7] This aligns with a distributed information management system designed to support operations in large-scale networks by utilizing Distributed Hash Tables (DHTs) to enable scalable aggregation. The system incorporates four key properties: scalability, flexibility, administrative isolation, and robustness, ensuring diverse applicability and resilience against failures. This is consistent with the research [12]. This study can be categorized under Distributed Computing Theory, as it analyzes the reliability of LLM operations and provides methods to identify and reduce reasoning errors. This is consistent with [14]. The use of information technology in school administration under the Pathum Thani Secondary Educational Service Area is at a high level, with administrative factors positively correlated with information technology utilization. The key factors influencing the adoption of information technology include technology management, organizational culture, teacher competency, policy, and infrastructure. Therefore, the creation of a host system involves supporting an increasing number of terminal users or running multiple programs simultaneously, with efficient responsiveness while using only a single main host computer. This represents a simpler processing role compared to centralized processing systems, including the investment in a high-performance host computer with high initial capital costs. In earlier periods, before computer networks were more highly developed, computer systems and processing in various fields commonly used centralized processing systems. This is consistent with [23], which states that data analysis aimed at preventing backdoor poisoning attacks employs distributed

computing techniques to enhance security and improve the reliability of federated learning systems. Similarly, [16] A novel Ultrathin Elevated Channel Thin-Film Transistor (UT-ECTFT) is a thin-film transistor with a specialized channel structure that effectively redistributes the electric field. This design enhances current flow with excellent current saturation characteristics and reduces the off-state current by up to 3.5 times compared to conventional channel structures.

2.3. Data Encryption Theory

Data Encryption Theory involves applying mathematical knowledge to transform data into an automated form using machine language or signals in various formats by employing cryptanalysis to modify codes in order to maintain the security of the original data in its original state during data transmission through distributed network systems. This is consistent with the research of [15].

Homomorphic encryption provides an approach to enhancing data security by protecting against malicious attacks from both internal and external sources, while safeguarding critical organizational information from unauthorized use and dissemination. However, encryption is a process of data transformation performed by humans, converting data from one form into another form that is unreadable by using encryption algorithms and encryption keys. This is consistent with the research. The IEEE 802.11 wireless network security standard is widely used, but it still contains vulnerabilities. This paper introduces a method to accelerate the process of discovering shared secret keys by targeting client stations. Attacking client stations proves more effective than attacking access points, since access points typically employ intrusion detection and protection mechanisms. Experimental results confirm that the proposed approach can successfully crack keys through client stations. In addition, in order to protect data from unauthorized access, only individuals who possess the correct decryption key are able to pass through or transform the data back into its original form. This is intended to maintain data confidentiality and security, whether during data storage or during data transmission. This is consistent with the research of [8] and [9]**.** In addition, research on data encryption is consistent with the study conducted by the team, which reported that the software encryption industry has grown to a market value of USD 15.5 billion. Their findings highlight the adoption of post-quantum cryptography techniques and the integration of Artificial Intelligence (AI) into encryption systems. Similarly, [24] it is emphasized that data protection, both at rest and in transit across networks, relies heavily on encryption as a safeguard, ensuring that intercepted data cannot be accessed without the correct key. [17] This study surveyed 452 university professors from the USA and India to examine perspectives on reproducibility and transparency in research. The findings reveal both national and disciplinary gaps, driven by incentive misalignment and resource constraints. Addressing scientific integrity requires culturally grounded solutions that consider both regional and disciplinary contexts. The study is limited by its reliance on self-reported data and its focus on only two countries.

3. Research Methodology

The research methodology in this study focuses on the development of data structures for preventing data loss and analyzing data in distributed communication networks. The following research methods are employed: 1. questionnaires, 2. analysis of factors affecting the development of data structures, and 3. algorithms for detecting data loss across distributed communication networks. This research process is divided into six parts, as follows.

3.1. Population and Sample

The sample group in this research consists of datasets from [6], UNSW-NB15 and [10], which is CIC-IDS2017, obtained from both the public and private sectors. Simple random sampling is used to select the samples, ensuring that each data instance has an equal chance of being selected. This method helps reduce bias and enhances the reliability and credibility of the research results in the future.

3.2. Requirements Analysis and System Design

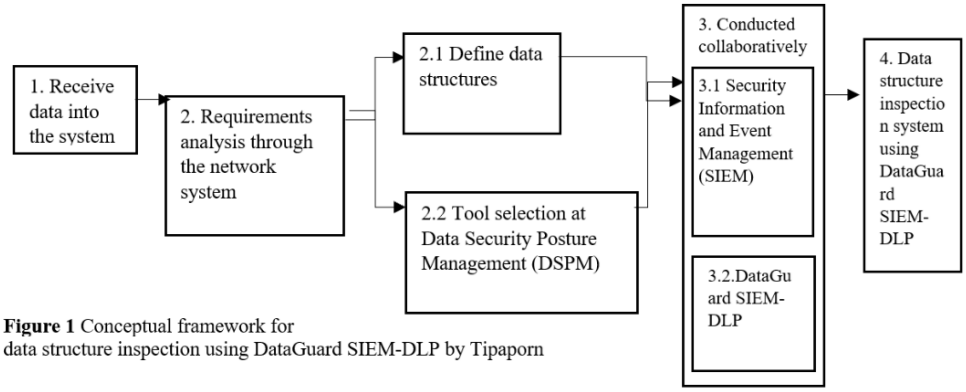


Figure 1 Conceptual framework for data structure inspection using DataGuard SIEM-DLP by Tipaporn

Fig. 1 Conceptual framework of the proposed DataGuard SIEM-DLP system

Explanation

Figure 1: Conceptual framework for data structure inspection using DataGuard SIEM-DLP by Tipaporn

1) Data intake into the system through the collection of data from various sources (applications, databases, network devices, source nodes, and destination nodes), including the recording of metadata (data types such as numerical and textual data, source, and time).
2) Requirement analysis through the network system, which is performed after Process) 1data intake into the system), in order to analyze the received data in terms of stability and security, regulatory compliance (personal data, financial data), availability readiness, and the scope of distributed network environments (edge, fog, cloud).
   2.1 Define the data structure by validating the imported data and designing data that can be practically used, as well as data used for testing purposes.
   2.2 Select tools by importing data into the DSPM system with configured data security settings, including vulnerabilities and policies that define the characteristics of the imported data to ensure coverage of critical data used for data analysis (numerical data, textual data, and other relevant data).
3) Collaborative operation is conducted after completing the process in sections 2.1–2.1to proceed with verifying data distribution collected from multiple data sources for testing purposes, with results used to validate the correctness of the DataGuard SIEM-DLP system program.
   3.1 Security Information and Event Management (SIEM) utilizes data collected from distributed nodes (system analog, agent, API, rules, and SOAR).
   3.2 DataGuard SIEM is an algorithm for analyzing imported data to assess stability, security, and system availability of the DGSIEM (DataGuard Security Information and Event Management: DGSIEM).
4) The DataGuard SIEM-DLP data structure inspection system operates after completing processes 3.1–3.2. The results obtained from the analysis are in the form of numerical and textual data, which are then processed further through the algorithms of the WE-KA program.

3.3. Data Preparation for Training and Testing Dataset Construction

3.3.1. Training Dataset

The training dataset uses two-thirds of the total datasets from UNSW-NB15 and CIC-IDS2017 for testing purposes.

3.3.2. Test Dataset

The test dataset is defined by using one-third of the total dataset for system operation in order to evaluate how closely the results correspond to real-world conditions.

3.4. Data Analysis using the DataGuard SIEM-DLP Algorithm by Tipaporn

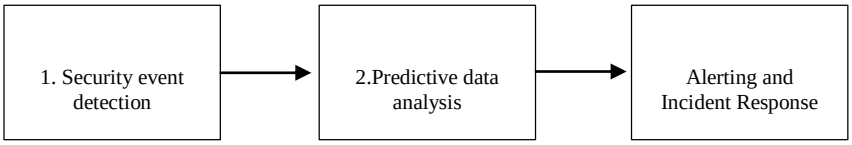


Fig. 2 Data analysis model using the DataGuard SIEM-DLP algorithm by Tipaporn

Explanation

Form Figure 2: Data analysis model using the DataGuard SIEM-DLP algorithm by Tipaporn.

1) Security event detection refers to the collection of data from network systems combined with data analysis to identify behaviors that may pose threats to the computer network system.
2) Predictive data analysis refers to identifying trends that may occur or predicting events that may occur in the future by referencing data sources from UNSW-NB 1 5 and CIC-IDS .2 0 1 7 This information is recorded in the database to indicate events that may potentially occur in the future, in order to determine appropriate solutions or preventive measures, with proper logging of events and correct handling of possible system operations.
3) Alerting and incident response refer to real-time alerts when events related to security breaches or system threats occur immediately, as well as transforming the received data into a form similar to system progress reports based on predictions, causes, and various issues that may arise in the network system.

3.5. Evaluation of the DataGuard SIEM-DLP Algorithm by Tipaporn

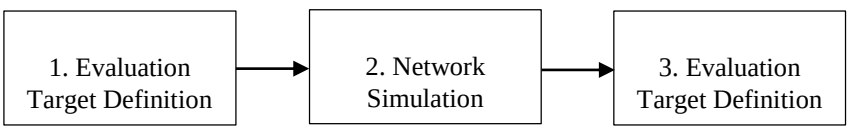


Fig. 3 Algorithmic procedure for evaluating the DataGuard SIEM-DLP algorithm by Tipaporn

Explanation

Form Figure 3: Algorithmic procedure for evaluating the DataGuard SIEM-DLP algorithm by Tipaporn.

1. Define the evaluation objectives of the DataGuard SIEM-DLP using the Tipaporn algorithm.
   1.1 Verification refers to checking the correctness of abnormal data related to network system behavior, evaluating data leakage, verifying the correctness of the algorithm, and measuring performance in terms of program execution time, including CPU resource utilization.
   1.2 Validation refers to confirming the correctness of the DataGuard SIEM-DLP by Tipaporn's algorithm to ensure that it operates according to its objectives, demonstrates accuracy in detecting errors, and is able to make correct predictions.
2. Network Simulation
   2.1 User Behavior Testing . The process is conducted in the form of verification, focusing primarily on goal definition and the validation of algorithm correctness. The algorithms are derived from data obtained through the analysis results of the UNSW-NB15 and CIC-IDS2017 datasets. Testing is performed using WEKA and MATLAB to compare whether the obtained results are consistent with each other, as well as to reconfirm validity and accuracy.
3. Measurement of Results and Performance
   3.1 Measurement of results based on criteria
   - Evaluation is conducted by testing the data using the WEKA program, producing numerical metrics including Recall, ROC, Accuracy, and Precision. The input data consists of numerical and textual data obtained from the UNSW- NB15 and CIC-IDS 2017 data sources (System analog, agent, API, Rules, SOAR, and other data).
   -MATLAB is also used, with input data consisting of numerical and textual data obtained from UNSW-NB 15and CIC- IDS2017 (System analog, agent, API, Rules and SOAR, and other data).

   3.2 Performance Evaluation

   The scores obtained from data analysis using WE-KA are used for evaluation. The evaluation results are stored in the computer's memory, and real-time performance evaluation is also conducted.

3.6. Research Instruments

The research entitled "Development of Data Structures Using the DataGuard SIEM-DLP by Tipaporn Algorithm for Data Loss Prevention and Data Analysis in Distributed Communication Networks" uses two categories of instruments for conducting the study, as follows.

1) The WEKA program is used for data preparation, data classification, and the discovery of relationships among data. The data obtained from the analysis are appropriate and are used to verify the correctness of the data applied as the training dataset and the testing dataset.
2) The computer used for testing the DataGuard SIEM-DLP by Tipaporn algorithm is a computer with sufficiently high performance for processing data related to security. The specifications are as follows: an AMD Ryzen 9/7 processor with 32GB of memory and 1TB of SSD storage.

4. Research Results

The results of the research entitled "Development of Data Structures Using the DataGuard SIEM-DLP Algorithm for Data Loss Prevention and Data Analysis in Distributed Communication Networks" consist of implementation procedures divided according to the sequential use of DataGuard SIEM-DLP by Tipaporn. The researcher divided the testing into four parts.

Part 1 presents the analysis results of the system's operational performance. Part 2 presents the results of time efficiency performance testing. Part 3 presents the results of the system data replication experiment. Part 4 presents the results of testing data loss prevention. The researcher conducted data loss testing according to the following sequence of operations.

Part 1: presents the results of the analysis of system performance efficiency. The researcher conducted system testing in terms of data redundancy verification, data error detection, lost data recovery, data encryption, data processing capability, unstructured data processing, the time required for data transmission between nodes, and whether increasing the number of nodes by three times affects processing time and system performance. The details are explained as shown in Figure 4.

Figure 4 shows the results of data analysis related to the performance efficiency of the system. The data used for testing are as follows: 1. Fault Tolerance: It was found that fault tolerance performance was good, with an average accuracy rate of 150 percent. 2. Scalability: It was found that the system was able to support data expansion effectively, including efficient use of resources, with an average rate of 115 percent. 3. Data Replication: It was found that data replication within the system was implemented to enhance security, speed, and reliability of data access, with an average rate of 99.8 percent.

In addition, Data Encryption had an average rate of 99.7 percent, Data Loss Prevention had an average rate of 99 percent, Error Detection and Network Performance had an average rate of 98.5 percent, Distributed Processing had an average rate of 98 percent, Unstructured Data Handling had an average rate of 95 percent, and Data Storage had an average rate of 80 percent, respectively.

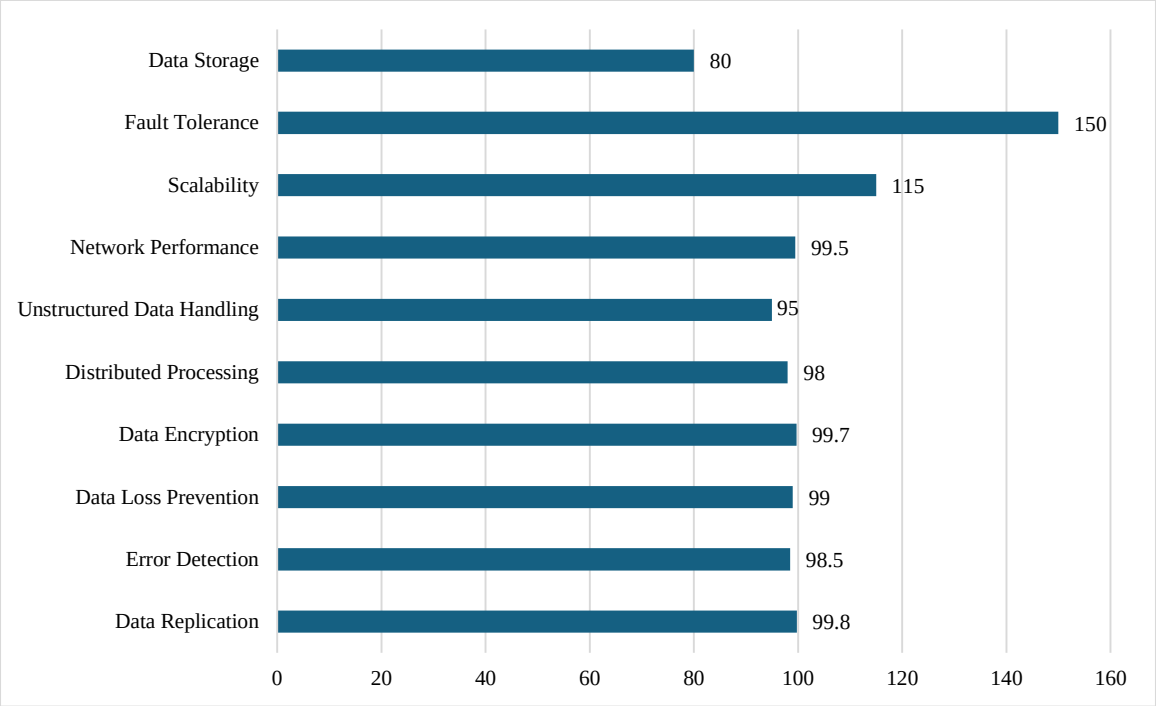


Fig. 4 Graph showing the performance of the system operation

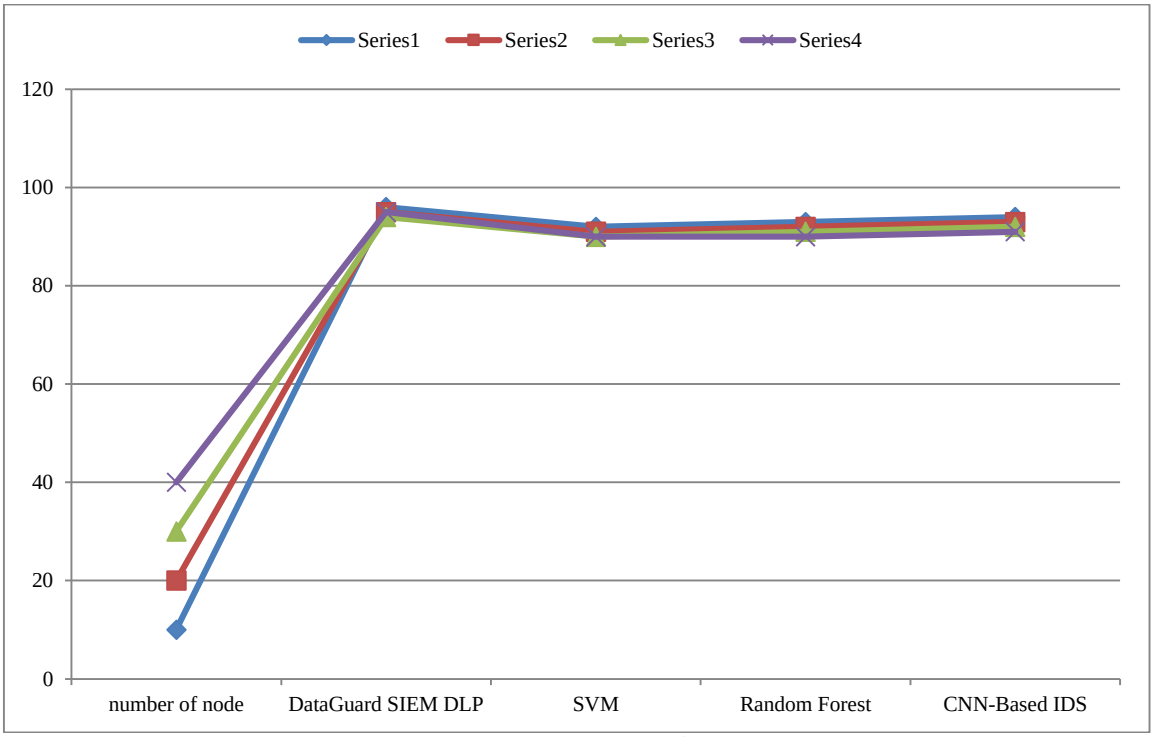


Fig. 5 Accuracy vs. number of nodes

From Figure 5, the analysis of Accuracy versus Number of Nodes shows that the DataGuard SIEM-DLP maintained a stable average accuracy of 94.5% as the number of nodes increased from 10 to 40. In comparison, SVM decreased from an average of 91.2% to 89.9%, Random Forest decreased from 91.7% to 90.9%, and CNN decreased from 93.5% to 91.8%. Based on this comparative analysis of the four techniques, it can be concluded that the DataGuard SIEM-DLP consistently preserves higher accuracy than the baseline methods, even when the system undergoes scalability expansion.

Table 1. Presents the analytical results of the consolidated performance and accuracy evaluation

Metric/Method	DataGuard SIEM-DLP	SVM	Random Forest	CNN-IDS
Data Storage (%)	80.00	-	-	-
Fault Tolerance (%)	150.00	-	-	-
Scalability (%)	115.00	-	-	-
Metric/Method	DataGuard SIEM-DLP	SVM	Random Forest	CNN-IDS
Network Performance (%)	99.50	-	-	-
Unstructured Data Handling (%)	95.0	-	-	-
Distributed Process (%)	98.0	-	-	-
Data Encryption (%)	99.7	-	-	-
Data Loss Prevention (%)	99.00	-	-	-
Error Detection (%)	98.5	-	-	-
Accuracy (10 Nodes) (%)	99.8	91.2	91.7	93.5
Accuracy (40 Nodes) (%)	94.5	89.9	90.9	91.8

From Table 1, the consolidated performance and accuracy results summarize the experimental findings regarding the efficiency and accuracy of the DataGuard SIEM-DLP algorithm framework.

When compared with standard methods (SVM, Random Forest, and CNN-IDS), the data in this table integrates the results from Figures 4 and 5 to provide clear statistical values and comparative insights against baseline methods.

The experimental outcomes confirm that the proposed DataGuard SIEM-DLP algorithm demonstrates superior performance across key metrics such as data storage, fault tolerance, encryption, replication, and overall accuracy. Notably, the algorithm maintained a stable average accuracy of 94.5% even as the number of nodes increased, whereas the baseline methods showed a declining trend under system expansion. These findings highlight the scalability and stability of the proposed framework, reinforcing the effectiveness of the DataGuard SIEM-DLP algorithm in distributed communication networks.

Part 2: presents the results of time performance testing. The researcher conducted evaluations in the areas of Data Storage and Network Performance, Data Loss Prevention and Data Encryption, Distributed Processing and Unstructured Data Handling, as well as Scalability and Fault Tolerance. The details are explained according to Figures 6 and 7.

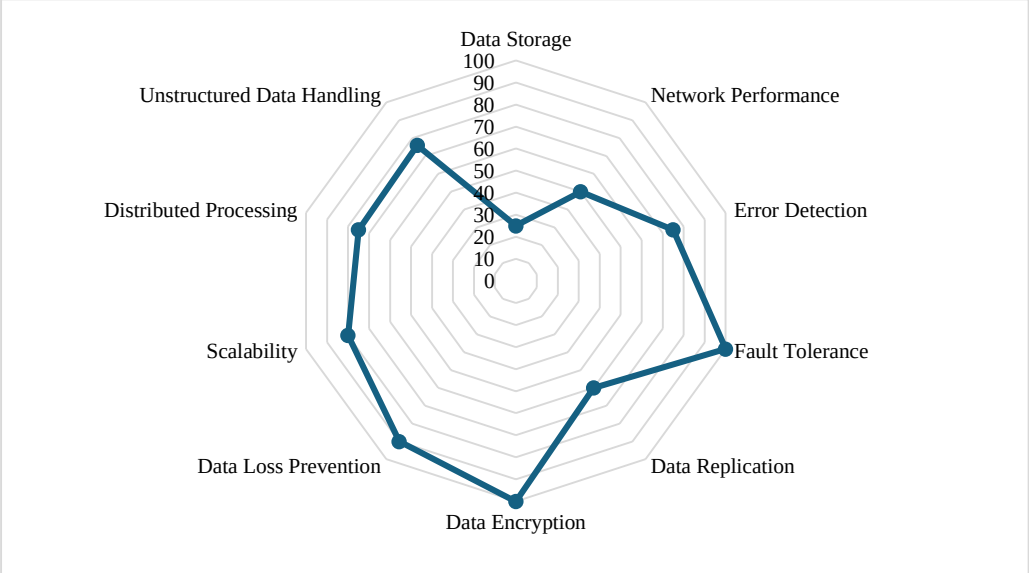


Fig. 6 Shows the results of the analysis of time efficiency during the testing process

Figure 6 shows the results of the performance analysis of time efficiency during testing. The important observations from the evaluation process are as follows. 1. Data Storage was found to have a speed suitable for real-time operations, with an average value of 30 percent, which is at the highest level. 2. Distributed Processing, Data Encryption, Error Detection, Network Performance, Unstructured Data Handling, and Fault Tolerance were found to be stable, responsive, and fast in operation, with average values ranging from 25 to 100 percent, which are at a very good level.

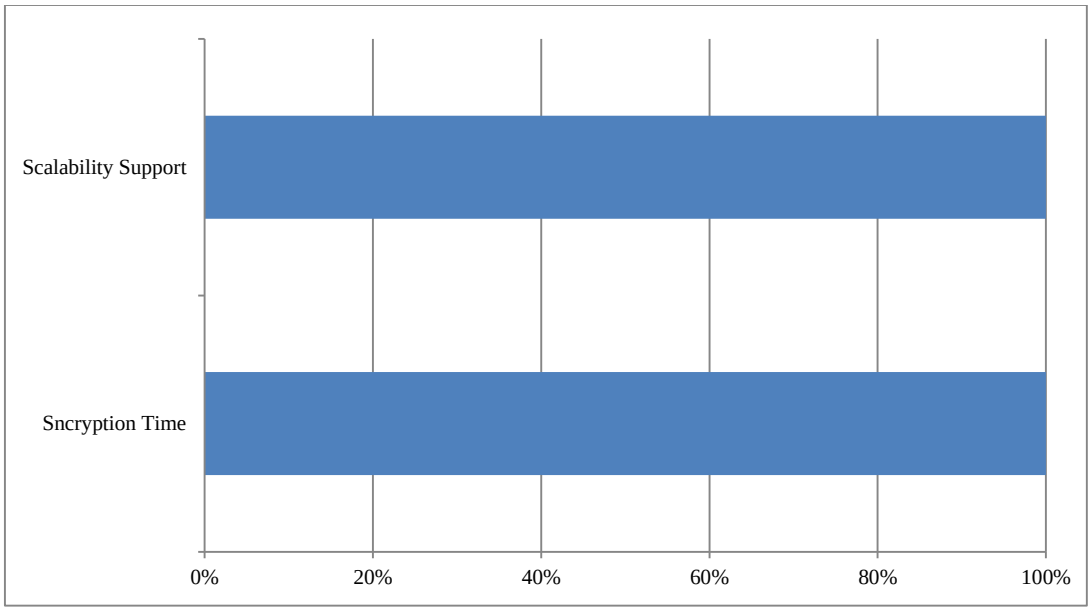


Fig. 7 Processing efficiency

From Figure 7, the analysis of processing efficiency shows that the average encryption time was 250 ms, while the scalability support time was 1.2 seconds. A comparison between encryption time and system scalability performance indicates that the DataGuard SIEM-DLP algorithm achieves low latency and is capable of supporting system expansion. Therefore, it is well-suited for real-time operations. Section 3 presents the results of the data replication system experiment. The researcher conducted evaluations of data consistency, bandwidth management and data loading, as well as monitoring and logging of system operations. The details are explained according to Figures 8 and 9



Fig. 8 Shows the results of the data replication system testing

From Figure 8, the results of testing the data replication system show that data replication performs data consistency verification, with an average value of 99 percent, which is at the highest level. This is because the data are accurate and stable. Meanwhile, bandwidth management and data loading inspection are performed. In terms of error detection, system monitoring and logging are conducted, with an average value of 98 percent, which is at a very good level. This is due to the accuracy and speed of the recording system operations. The researcher verified the data by detecting errors during inspection and performing recovery after data transmission through the system, including the time required to analyze the data in the system, measured in milliseconds. In addition, the fault tolerance component related to errors in testing the data replication system shows an average value of 350 ms, which is at a good level. This is because the system is able to tolerate errors effectively without affecting overall system operation.

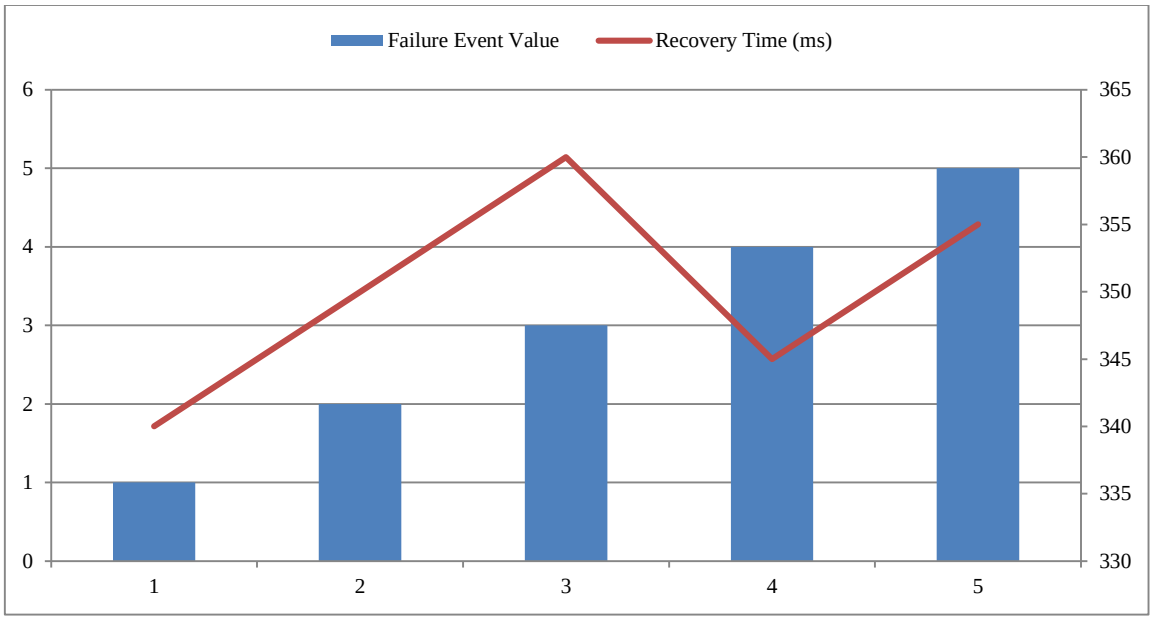


Fig. 9 Fault recovery time

From Figure 9, the analysis of fault recovery time shows that the average recovery time per event for lost data was between 340 and 360 ms, with an overall average of 350 ms. The results indicate that the system is capable of recovering from node failures rapidly, thereby ensuring continuous operation and minimizing data loss. Section 4 presents the results of testing data loss prevention. The researcher conducted abnormal user behavior detection, access control restriction verification, and priority level classification of data. The details are explained in Figures 10 and 11.

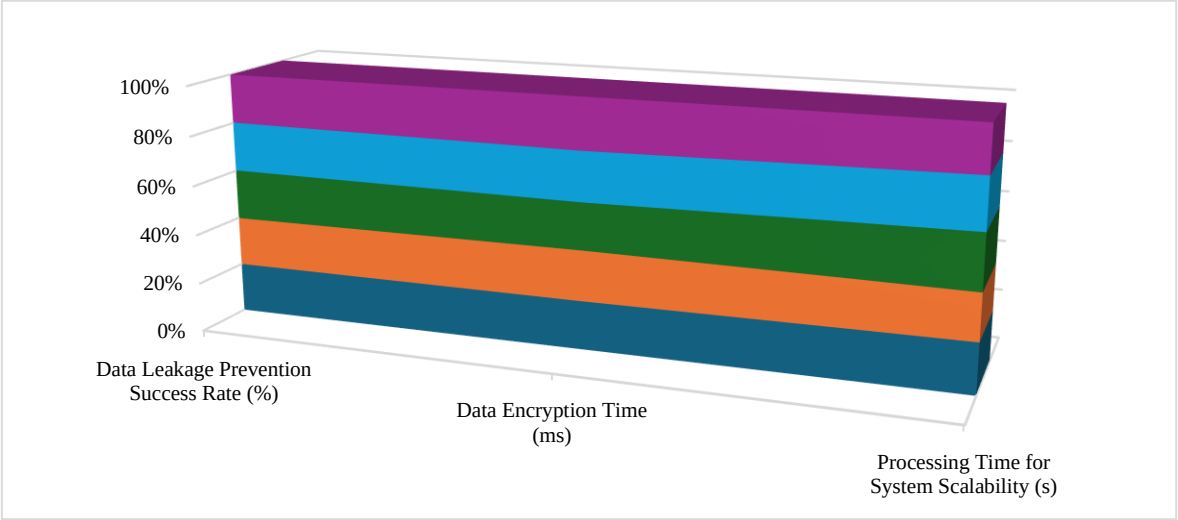


Fig. 10 Shows the results of the testing time for data loss prevention

Figure 10 presents the results of the testing time for data loss prevention are presented. It was found that the average success rate of data leakage prevention was between 94–91 percent, while the average data encryption time was between 295–270ms. At the same time, the average processing time to support system scalability was between 1.3–1.1.

Therefore, the following conclusions can be drawn. The detection of abnormal user behavior was at a good level, the detection of data access restriction was at a good level, and the verification of data priority classification was at a good level. Consequently, the researcher concludes that the DataGuard SIEM-DLP by Tipaporn algorithm can be effectively used for high-level security monitoring (≥ (%90, and the encryption time was adjusted to be within ≤ 300ms, which maintains speed and efficiency. In addition, the processing time to support system scalability remained low and significantly stable.

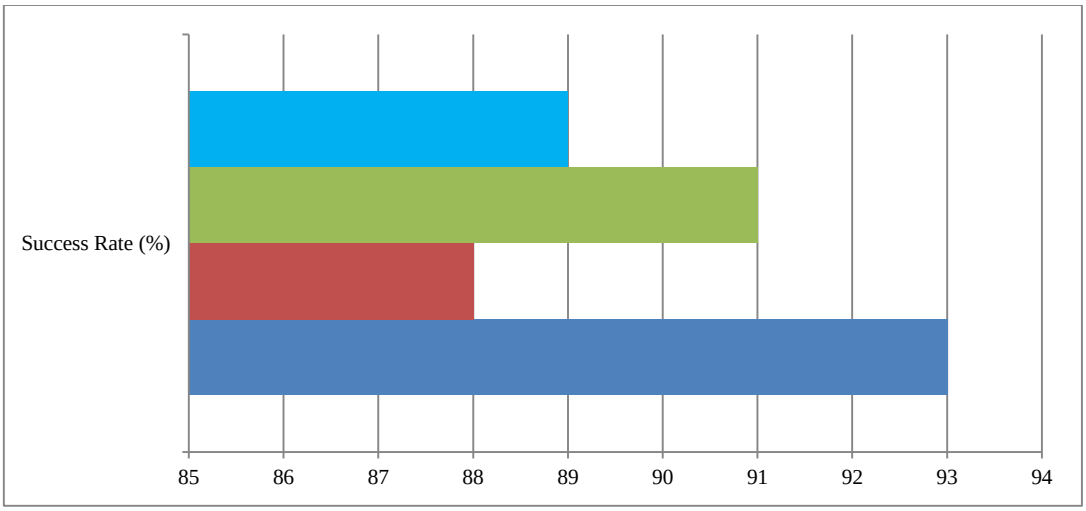


Fig. 11 Data loss prevention success rate

Figure 11, the analysis of the Data Loss Prevention Success Rate shows that the average success rate of the DataGuard SIEM-DLP was 93%, compared to 88% for SVM, 91% for Random Forest, and 89% for CNN. Based on this comparative analysis of data transmission success rates, the DataGuard SIEM-DLP achieved the highest success rate, demonstrating superior capability in preventing data loss compared to other techniques.

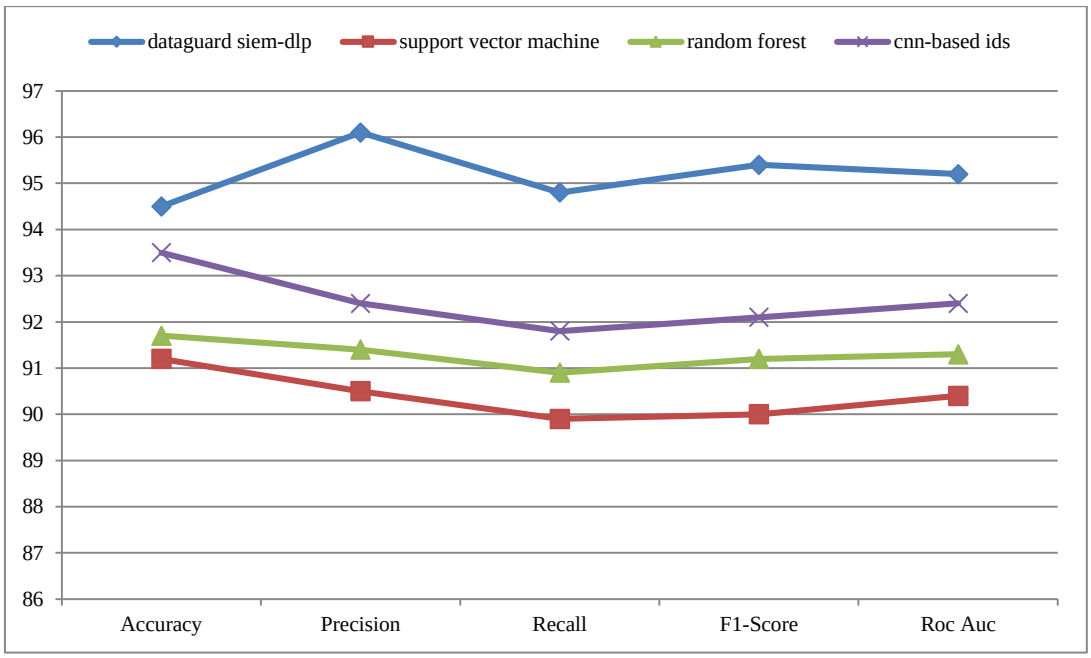


Fig. 12 Benchmarking results

Figure 12, the benchmarking results illustrate a comparison between the proposed DataGuard SIEM-DLP framework and standard methods (SVM, Random Forest, and CNN-Based IDS). The figure clearly shows that the DataGuard SIEM-DLP achieved higher average values in Accuracy, Precision, Recall, F1-score, and ROC-AUC, outperforming the baseline techniques.

Section 5: Algorithm of DataGuard SIEMLP is The algorithm of DataGuard SIEM-DLP can be explained through the fundamental workflow illustrated in Figure 13, Example of the DataGuard SIEM-DLP Algorithm and Output. The figure presents the sequential processes beginning with raw log collection, followed by preprocessing, and then data replication. These steps form the foundation for subsequent encryption, leakage analysis, and reliability monitoring. This workflow demonstrates the integration of core functions to ensure secure and efficient data management, serving as a prelude to the detailed algorithmic description provided in Figure 13.

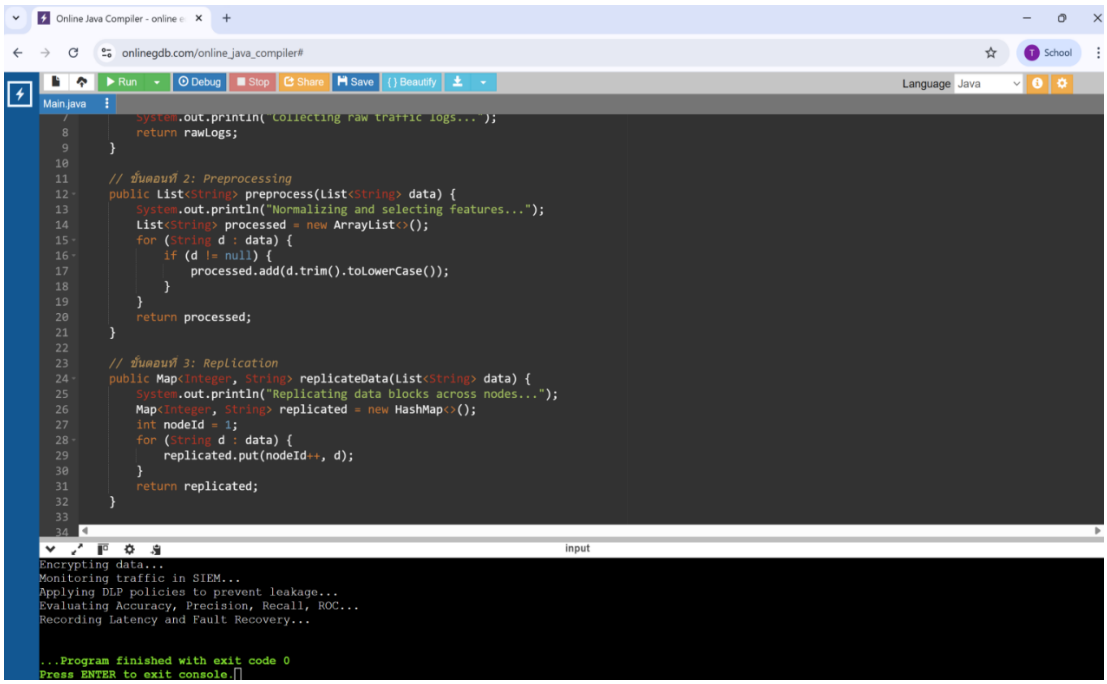


Fig. 13 Example of the DataGuard SIEM-DLP algorithm and output

Figure 13, an example of the DataGuard SIEM-DLP algorithm, is presented, demonstrating its operation. The algorithm was developed by the researchers using the Java programming language, with defined methods for coding and execution. The figure also illustrates the simulation of the preprocessing and replication steps of the DataGuard SIEM-DLP. The console output results show data encryption, data verification, SIEM system checks, and performance evaluation in terms of accuracy, precision, recall, and ROC. The program developed by the researchers executed successfully without errors (exit code 0), as shown in Figure 13.

5. Research Results Summary

The research titled Development of Data Structures Using the DataGuard SIEM-DLP Algorithm for Data Loss Prevention and Data Analysis in Distributed Communication Networks presents the results of data analysis summarized into four parts as follows. Part 1 presents the analysis results of system performance efficiency, with an average value of 95.7%. From the analysis of Fault Tolerance, Scalability, Data Replication, Data Encryption, Data Loss Prevention, Detection and Network Performance, Distributed Processing, Unstructured Data Handling, and Data Storage, it was found that the tested data were replicated within the system, which enhanced security, speed, and reliability of data access. In addition, the system demonstrated good fault tolerance during resource utilization, as well as a strong ability to support data scalability for incoming data. Part 2 presents the results of time efficiency performance testing, with an average value of 25–100 percent. From the analysis of Data Storage, Distributed Processing, Data Encryption, Error Detection, Network Performance, Unstructured Data Handling, and Fault Tolerance, it was found that the tested data showed stability and rapid responsiveness in data operations, including fast overall system performance. Part 3 presents the results of data replication system testing, with an average value of 98.5 percent. From the analysis of Data Replication, Error Detection, and Fault Tolerance, it was found that the tested data demonstrated high accuracy and speed in recording system operations, including the time required for error detection during system testing related to data replication and data recovery after data transmission through the system.

Additionally, the time used to input data for analysis when system errors occurred was handled effectively without impacting overall system operation. Part 4 presents the results of data loss prevention testing, with an average value of 91–94 percent, while the average data encryption time was 270–295 percent. Meanwhile, the average processing time to support system scalability was 1.1–1.3 percent. From the analysis of user behavior detection, priority verification, security inspection, time-based encryption verification, as well as system scalability support and algorithm stability testing of the DataGuard SIEM-DLP algorithm by Tipaporn, it was found that the DataGuard SIEM-DLP algorithm by Tipaporn demonstrated a very good level of processing performance in monitoring and detection according to the initial hypotheses established by the researcher, as follows:.1

The DataGuard SIEM-DLP algorithm by Tipaporn developed by the researcher is capable of improving security and stability performance in distributed communication networks at a good level..2 The data structure related to the

DataGuard SIEM-DLP algorithm by Tipaporn can operate stably and support real-time processing within the specified time constraints.

Therefore, future development should focus on applying artificial intelligence to the DataGuard SIEM-DLP algorithm by Tipaporn for predictive detection, as well as testing under real-world operational environments. However, to provide detailed insights into the framework of the DataGuard SIEM-DLP algorithm developed in this study, the researchers summarized analytical information across several aspects: system scalability, processing efficiency, fault tolerance, data loss prevention, deployment potential, and directions for future enhancement. These consolidated findings are presented in Table 2.

Table 2. Scalability and limitations of DataGuard SIEM-DLP

Analytical Aspect	Result/analysis	Limitation/Future Direction
Scalability	The DataGuard SIEM-DLP framework is able to maintain a detection accuracy of 95%, even when the number of nodes increases threefold. This demonstrates its robustness and reliability in distributed environments.	The testing is still confined to simulated environments, and further experiments in real-world systems are required.
Processing Efficiency	The average encryption time remains within ≤ 300 ms, while system scalability is supported with response times ranging between 1.1–1.3 seconds. These results demonstrate the suitability of the framework for real-time applications.	The performance may vary under diverse data transmission conditions; therefore, adjustments are required to ensure compatibility with the actual networks in which the researchers intend to deploy the system.
Analytical Aspect	Result/analysis	Limitation/Future Direction
Fault Tolerance	The system can recover from node failures within 350 ms, thereby ensuring continuous monitoring and effectively reducing data loss.	The evaluation is still limited to standard datasets, and more diverse datasets should be employed in future experiments to strengthen the validity and generalizability of the results.
Data Loss Prevention	The framework successfully prevented data leakage at a rate of 91–92%, which is superior to the baseline methods under controlled experimental conditions.	The system must be tested with larger datasets and scenarios involving diverse types of attacks to ensure comprehensive evaluation and robustness.
Deployment potential	The framework can be applied to Smart City and Enterprise Networks, highlighting its practical relevance and applicability in real-world environments.	The system has not yet been tested in real-world environments; therefore, pilot experiments in actual operational settings should be conducted to validate its effectiveness and reliability.
Future enhancement	The integration of AI-based predictive mechanisms can significantly enhance system scalability and proactive detection capabilities.	Additional testing should be conducted with adaptive learning models and cross-domain datasets to enhance the algorithm's robustness and generalizability.

However, to illustrate the details of the framework of the DataGuard SIEM-DLP algorithm developed in this study, the researchers summarized analytical findings across several dimensions: scalability, processing efficiency, fault tolerance, data loss prevention, deployment potential, and directions for future development. As shown in Table 2, the DataGuard SIEM-DLP algorithm demonstrates superior performance in system expansion, processing speed, fault tolerance, and data protection during transmission compared to baseline methods.

Moreover, the algorithm shows strong potential for real-world applications. Nevertheless, the current testing remains limited to simulated environments and standard datasets. Therefore, future research should involve experiments in real-world systems with more diverse datasets, as well as the integration of AI-based predictive and adaptive learning to enhance proactive detection and adaptability against continuously evolving threats.

6. Discussion

The experimental results from Table 2, together with the operational analysis shown in Figure 13, demonstrate that the DataGuard SIEM-DLP framework possesses strong potential for real-world application. The algorithm consistently maintains higher accuracy and efficiency compared to baseline methods in terms of scalability, processing efficiency, fault tolerance, data loss prevention, deployment potential, and future enhancement.

It is capable of recovering effectively from node failures and minimizing data loss. The console output presented in Figure 13 further confirms the completeness of the

algorithm's operation and its evaluation across accuracy, precision, recall, and ROC. However, the current experiments remain limited to simulated environments and standard datasets. Future research should therefore involve testing in real-world systems with more diverse datasets. This will enable the integration of AI-based predictive and adaptive learning, thereby enhancing the algorithm's proactive detection capabilities and adaptability to complex and continuously evolving threats. Such advancements will strengthen the DataGuard SIEM-DLP framework, making it more robust and comprehensive in addressing the demands of digital environments. In addition to the experimental results that confirm the potential of the DataGuard SIEM-DLP framework, comparisons with related studies further strengthen its credibility. First, which applied machine learning techniques such as SVM, RF, and ANN to benchmark datasets including UNSW-NB15 and CIC-ID52017, demonstrated effective intrusion detection but remained limited to model-based learning alone. In contrast, DataGuard SIEM-DLP framework encryption, replication, and fault recovery offer a more comprehensive and resilient solution for real world environments. [18]

Second, Homomorphic encryption research (Springer, 2024-2025) highlights secure computation on encrypted data without decryption. While enhancing security, it lacks broader system management. DataGuard SIEM-DLP extends beyond encryption by integrating leakage analysis, performance evaluation, and adaptive management, making it complete and more applicable. [25]

7. Conclusion

Research Title: Development of Data Structures Using the DataGuard SIEM-DLP Algorithm for Data Loss Prevention and Data Analysis in Distributed Communication Networks. The researchers developed data structures by employing the DataGuard SIEM-DLP algorithm to enhance the efficiency of data loss prevention and data analysis in distributed communication networks. The study focuses on the integration of security event detection, data encryption, data replication, and fault detection into a centralized framework. Experimental results demonstrate that the system can support real-time processing with high stability, while effectively reducing the risk of data loss during communication across networks.

Performance evaluation further indicates that the proposed algorithm preserves data integrity and accommodates node scalability, even under conditions involving large volumes of incoming data. The developed system is capable of detecting errors and responding rapidly to security incidents, while also providing recovery mechanisms to restore data within a short period in the event of node failures. This capability minimizes the overall impact on network operations.

In addition, comparative analysis with standard techniques such as SVM, Random Forest, and CNN-based IDS reveals that DataGuard SIEM-DLP achieves superior accuracy, faster processing speed, and higher success rates in preventing potential data leakage. These findings highlight the system's potential for application in organizational network infrastructures, as well as in cloud environments and smart city ecosystems in the future.

Nevertheless, the experiments were limited to simulated environments and standard datasets. Future work will extend testing to real-world networks and integrate AI-based adaptive learning, enhancing proactive detection, resilience, and credibility of the DataGuard SIEM-DLP algorithm.

References

[1] Andrew S. Tanenbaum, *Distributed Systems: Principles and Paradigms*, 2nd ed., Pearson Education, 2015. [Google Scholar] [Publisher Link]

[2] Kai Hwang, Jack Dongarra, and Geoffrey C. Fox, *Distributed and Cloud Computing: From Parallel Processing to the Internet of Thing*, 1st ed., Elsevier, 2012. [Google Scholar] [Publisher Link]

[3] J. Zhao, and L. Zhang, "An Efficient Data Replication Strategy for Distributed File Systems," *Journal of Computer Science and Technology*, vol. 25, no. 4, pp. 754-767, 2010. [Google Scholar]

[4] C.E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, 1948. [CrossRef] [Google Scholar] [Publisher Link]

[5] R.W. Hamming, "Error Detecting and Error Correcting Codes," *The Bell System Technical Journal*, vol. 29, no. 2, pp. 147-160, 1950. [CrossRef] [Google Scholar] [Publisher Link]

[6] Nour Moustafa, and Jill Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)," *2015 Military Communications and Information Systems Conference (MilCIS)*, Canberra, ACT, Australia, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Praveen Yalagandula, and Mike Dahlin, "A Scalable Distributed Information Management System," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 4, pp. 379-390, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[8] R.L. Rivest, Adi Shamir, and Len Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120-126, 1978. [CrossRef] [Google Scholar] [Publisher Link]

[9] Whitfield Diffie, and M. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644-654, 1976. [CrossRef] [Google Scholar] [Publisher Link]

[10] Intrusion Detection Evaluation Dataset, Canadian Institute for Cybersecurity, University of New Brunswick, 2017. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html

[11] Brendan Burns, *Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable Systems using Kubernetes*, O'Reilly Media, 2024. [Google Scholar] [Publisher Link]

[12] Debargha Ganguly, "Grammars of Formal Uncertainty: When to Trust LLMs in Automated Reasoning Tasks," *Advances in Neural Information Processing Systems 38 Main Conference (NeurIPS 2025)*, San Diego, California, USA, pp. 38457-38499, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Mykhailo Klymash et al., "Research of Distributed Data Processing in Corporate Information Systems," *2021 IEEE 16th International Conference on the Experience of Designing and Application of CAD Systems (CADSM)*, Lviv, Ukraine, pp. 24-27, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Phanumas Buachin, Sophana Sudsomboon, and Arnon Jeenawat, "Administrative Factors Affecting the Information Technology using in the Administration of Schools under the Pathum Thani Secondary Educational Service Area Office," *Sukhothai Thammathirat Open University*, vol. 11, no. 3, pp. 1-18, 2023. [Google Scholar] [Publisher Link]

[15] Tanyaporn Sridokmai, "Security Database on Encrypted Data using Homomorphic Encryption," *Dhonburi Rajabhat University Journal*, vol. 13, no. 2, pp. 226-238, 2019. [Google Scholar] [Publisher Link]

[16] Shengdong Zhang et al., "A Novel Ultrathin Elevated Channel Low-Temperature Poly-Si TFT," *IEEE Electron Device Letters*, vol. 20, no. 11, pp. 569-571, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[17] Tatiana Chakravorti, Sai Koneru, and Sarah Rajtmajer, "Reproducibility and Replicability in Research: What 452 Professors Think in Universities Across the USA and India," *PLOS One*, vol. 20, no. 3, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[18] Sudhanshu Sekhar Tripathy, and Bichitrananda Behera, "A Review of Various Datasets for Machine Learning Algorithm-based Intrusion Detection System: Advances and Challenges," *arxiv*, pp. 3833-3857, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Fatemeh Karamimirazizi, Seyed Mahdi Jameii, and Amir Masoud Rahmani, "Data Replication Methods in Cloud, Fog, and Edge Computing: A Systematic Literature Review," *Wireless Personal Communications*, vol. 135, no. 1, pp. 531-561, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Jad Silbak, and Daniel Wichs, "Binary Codes for Error Detection and Correction in a Computationally Bounded World," *Advances in Cryptology – EUROCRYPT 2025*, Springer, Cham, vol. 15601, pp. 186-211, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Asma Ben Abacha et al., "MEDEC: A Benchmark for Medical Error Detection and Correction in Clinical Notes," *Association for Computational Linguistics*, pp. 22539-22550, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Jyoti Prakash Sahoo et al., "QCShield: A Quantum-Classical Hybrid Framework for Intrusion Detection in Edge-IIoT," *Advances in Distributed Computing and Machine Learning*, Springer, Singapore, vol. 1406, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Phung Lai et al., "A Client-level Assessment of Collaborative Backdoor Poisoning in Non-IID Federated Learning," *2025 IEEE 45th International Conference on Distributed Computing Systems (ICDCS)*, Glasgow, United Kingdom, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] Srikanth Bellamkonda, The Future Encryption: Protecting Data in an Era of Rising Cyber Threats, Forbes Technology Council, 2025. [Online]. Available: https://www.forbes.com/councils/forbestechcouncil/2025/03/26/the-future-of-encryption-protecting-data-in-an-era-of-rising-cyber-threats/

[25] Craig Gentry, Shai Halevi, and Nigel P. Smart, "Homomorphic Evaluation of the AES Circuit," *Annual Cryptology Conference*, Santa Barbara, CA, USA, vol. 1, pp. 850-867, 2012. [CrossRef] [Google Scholar] [Publisher Link]