OPTIMIZATIONS IN APPLIED MACHINE LEARNING
Research Article* | *Volume 5* | *Issue 5* | *Jan 2025
Received: 03 Dec 2024 | Revised: 27 Dec 2024
Accepted: 07 Jan 2025 | Published Online: 21 Jan 2025

A Density-Based Spatial Clustering of Applications with Noise for Data Security Intrusion Detection

Carlos Méndez[1], Laura García[2] and Javier Torres[3,*]

[1] Department of Computer Science, Universidad de Burgos, Burgos, 09001, Spain

[2] Institute of Data Security and Artificial Intelligence, Universidad de Castilla-La Mancha, Ciudad Real, 13071, Spain

[3] Center for Advanced Computing and Cybersecurity, Universidad de Zaragoza, Teruel, 44003, Spain

[*]Corresponding Author, Email: javier.torres@unizar.es

Abstract: In the field of data security intrusion detection, the challenge lies in effectively identifying and categorizing intrusion activities amidst the influx of data. Current research predominantly focuses on traditional methods that may overlook subtle yet significant patterns, thereby hindering accurate intrusion detection. This paper addresses this gap by proposing a novel approach - A Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for Data Security Intrusion Detection. By harnessing the power of density-based clustering, our method enhances the detection capability by capturing intricate relationships and anomalies within the data. Through extensive experimentation and analysis, we demonstrate the effectiveness and reliability of our approach in improving the accuracy and efficiency of intrusion detection systems. This innovative contribution not only enriches the existing research landscape but also paves the way for enhanced data security measures in the digital era.

Keywords: *Data Security; Intrusion Detection; Density-Based Clustering; Anomaly Detection; Research Contribution*

1. Introduction

Data Security Intrusion Detection is a specialized field within cybersecurity that focuses on developing technologies and techniques to detect unauthorized access or malicious activities in computer systems and networks. Current challenges in this field include the increasing complexity and sophistication of cyber threats, the volume and diversity of data to be monitored, the need for

real-time detection and response, and the ability to differentiate between normal and anomalous behavior. Additionally, issues such as data privacy concerns, the lack of standardized evaluation metrics, and the scarcity of labeled training data pose significant obstacles to the development and deployment of effective intrusion detection solutions. As researchers continue to address these challenges, advancements in machine learning, artificial intelligence, and anomaly detection algorithms are expected to play a crucial role in enhancing the capabilities of data security intrusion detection systems.

To this end, research on Data Security Intrusion Detection has advanced to the level where machine learning algorithms are being widely used to detect and prevent cyber threats. Researchers are exploring innovative techniques, such as deep learning and anomaly detection, to enhance the accuracy and efficiency of intrusion detection systems. Recent research in the field of cyber security intrusion detection has focused on developing more effective and accurate detection systems. Zhang et al. [1] propose a data security intrusion detection system that integrates the Mamba and ECANet models, employing an end-to-end learning approach for training and optimization. Banoth et al. [2] present a survey of data mining and machine learning methods for cyber security intrusion detection, emphasizing the importance of ML/DM algorithms in improving detection accuracy. Buczak and Guven [3] also conduct a survey on ML/DM methods for intrusion detection, highlighting the complexity of algorithms and challenges in cyber security applications. Parmar [4] reviews various methods for anomaly detection, intrusion detection, and access policy creation in the context of data security. Kalinin and Krundyshev [5] explore the use of quantum machine learning techniques for security intrusion detection. Sarker et al. [6] introduce the "IntruDTree" machine learning model for cyber security intrusion detection, focusing on feature importance ranking and tree-based modeling. Mohy-Eddine et al. [7] propose an efficient network intrusion detection model using a K-NN classifier and feature selection for IoT security. Alotaibi and Ilyas [8] develop an ensemble-learning framework to enhance the security of Internet of Things devices through improved intrusion detection efficiency. Wu et al. [9] present an intelligent intrusion detection algorithm using a combination of fuzzy rough set, generative adversarial network (GAN), and convolutional neural network (CNN) for IoT security. Shiravani et al. [10] explore network intrusion detection using data dimension reduction techniques. Recent research in cyber security intrusion detection has shown a growing interest in developing more effective and accurate detection systems. The utilization of the DBSCAN technique is crucial due to its ability to handle noise, detect outliers, and identify clusters in the data, making it a valuable tool for enhancing the detection capabilities of intrusion detection systems. Its robust performance in handling complex data patterns and detecting anomalies makes DBSCAN a preferred choice in cyber security applications.

Specifically, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a robust clustering algorithm that can effectively identify abnormal patterns in network traffic, making it valuable for data security intrusion detection by distinguishing between normal behavior and potential threats, thereby enhancing overall system security. In recent literature, various adaptations and enhancements to the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm have been proposed. Hajihosseinlou et al. (2024) introduced an intelligent

mapping approach by combining DBSCAN with mean-shift clustering [11]. Qian et al. (2024) presented MDBSCAN, a multi-density DBSCAN algorithm based on relative density [12]. Al-batah et al. (2024) proposed Enhanced DBSCAN-H, a histogram-based enhancement to address issues with DBSCAN in processing large satellite image datasets [13]. However, limitations remain regarding the generalizability of these adaptations, their computational efficiency in high-dimensional spaces, and the robustness against noise and outliers, which require further investigation.

The research presented in this paper is inspired by the pioneering work of H. Zhang, et al. on the Mamba-ECANet model, which was articulated in their 2024 study [14]. Their incisive exploration into the effective use of end-to-end learning mechanisms for enhancing data security provided an insightful foundation that guided the development of our current approach. In particular, the integrative framework of the Mamba-ECANet, which harmonizes multiple feature extraction methodologies with adaptive learning techniques, opened new avenues in achieving robust intrusion detection capabilities. Our research endeavors to build upon the refined sensitivity of their model to detect nuanced anomalies within data sets. By employing the strategies delineated in the Mamba-ECANet model, our method leverages advanced spatial clustering paradigms to enhance the identification and classification of potential security threats. A crucial aspect integrated into our methodology is the adaptive learning characteristic demonstrated by the original authors [14]. Through this, we meticulously designed a system that not only acknowledges the dynamic patterns of intrusions but also adapts to the evolving nature of security threats with precision. The incorporation of noise-handling techniques is a direct derivative of their innovative approach, enabling our model to efficiently process and filter irrelevant or misleading data that often obstruct accurate intrusion detection [15]. Moreover, the meticulous evaluation protocol drawn from the original study provided an invaluable framework for validating the efficacy of our solutions. In particular, the employment of comprehensive testing scenarios in real-world environments detailed in their paper acted as a crucial benchmark against which the effectiveness and reliability of our approach were measured [16]. This ensured that our enhancements are not merely theoretically sound but also practically viable across diverse deployment settings.

In the challenging realm of data security intrusion detection, the ability to accurately identify and categorize intrusion activities amidst vast and complex data streams is paramount. Traditional methods often fail to account for subtle yet critical patterns, thus impeding effective detection. This paper seeks to address such deficiencies, as detailed in Section 2's problem statement, by introducing a novel methodology—Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for Data Security Intrusion Detection—outlined in Section 3. This approach leverages the strength of density-based clustering to improve detection capabilities by identifying intricate relationships and anomalies within the data. A comprehensive case study, presented in Section 4, exemplifies the practical application of our method. Section 5 offers a detailed analysis of the results, underscoring the method's effectiveness and reliability in enhancing the accuracy and efficiency of intrusion detection systems. The subsequent discussion in Section 6 delves into the broader implications of these findings, providing insightful discourse into the method's potential. Finally, Section 7 succinctly summarizes the study, highlighting its contribution to advancing data

security measures in the digital era and enriching the existing landscape of intrusion detection research.

2. Background

2.1 Data Security Intrusion Detection

Data Security Intrusion Detection plays a critical role in safeguarding computer networks from unauthorized access and malicious attacks. It is an essential component of cybersecurity, designed to monitor and analyze network traffic, system activities, and data integrity to identify potential threats or intrusions. This process involves several complex methodologies and algorithms aimed at discerning normal from anomalous behavior within a system.

The core idea of intrusion detection is based on identifying deviations from established patterns in the system's operations. This can be mathematically expressed by analyzing the probability distribution of observed activities compared to the baseline, which can be modeled as:

$$P(A|N) < \epsilon \tag{1}$$

where $P(A|N)$ is the probability of an activity A given the normal activity pattern N, and ϵ is a predetermined threshold below which an activity is considered anomalous.

One common approach in intrusion detection is the use of signature-based methods. These methods compare current network activities against known signature patterns of previously identified threats. Formally, this can be represented as:

$$D_s = \sum_{i=1}^{n} f(x_i, s_i) \tag{2}$$

where D_s is the degree of match against the signature base, n is the number of signatures, x_i is the current data point, and s_i represents the ith signature pattern.

An alternative to signature-based methods is anomaly-based detection, which looks for deviations from a modeled normal behavior state. Anomalies can be represented by computing a distance metric such as:

$$L(A, N) = ||f(A) - f(N)||_2 \tag{3}$$

Here, $L(A, N)$ represents the Euclidean distance between the feature vector $f(A)$ of the activity A and the feature vector $f(N)$ of the normal activity N.

Machine learning algorithms also play a significant part in the advancement of intrusion detection systems (IDS). By training models on historical dataset behavior, these models aim to predict and identify anomalies. The learning process can be abstracted as:

$$W = \text{argmin}_W \sum_{j=1}^{m} L\left(y_j, h(x_j; W)\right) \tag{4}$$

where W are the model parameters, y_j denotes the true label, $h(x_j; W)$ represents the hypothesis function over input x_j , and m is the number of training examples.

Furthermore, IDS often employs statistical methods such as Bayesian networks, which estimate the likelihood of an intrusion based on certain observed evidence E . This can be expressed by:

$$P(I|E) = \frac{P(E|I) \cdot P(I)}{P(E)} \tag{5}$$

where $P(I|E)$ is the probability of an intrusion I given evidence E , evaluated by the likelihood $P(E|I)$ and the prior probabilities $P(I)$ and $P(E)$. In conclusion, Data Security Intrusion Detection is a multifaceted discipline involving the detection and analysis of abnormal behaviors in cybersecurity systems. Through signature-based techniques, anomaly detection, machine learning algorithms, and statistical methods, IDS are capable of identifying and responding to unauthorized activities that threaten system integrity and confidentiality. The mathematical models and formulas discussed are integral to understanding and developing sophisticated intrusion detection technologies.

2.2 Methodologies & Limitations

Data Security Intrusion Detection remains at the forefront of cybersecurity, employing a diverse array of methodologies to mitigate the risk of unauthorized access and malicious attacks. Among the key strategies are signature-based methods, anomaly detection techniques, machine learning algorithms, and statistical methods, each with its distinct advantages and inherent limitations.

Signature-based intrusion detection is one of the most established approaches. It relies on pattern matching, where network activities are compared against a database of known threat signatures. The underlying mechanism can be formalized as:

$$D_s = \sum_{i=1}^{n} f(x_i, s_i) \tag{6}$$

Here, D_s quantifies the degree of alignment between the observed data x_i and known signatures s_i . Despite its effectiveness in identifying known threats, this method struggles with novel, previously unseen attacks, as it cannot detect deviations that do not match any existing signature.

Anomaly-based intrusion detection represents an alternative that addresses some of these deficiencies by modeling normal behavior and flagging significant deviations as potential threats. This can be quantified using a distance metric:

$$L(A, N) = ||f(A) - f(N)||_2 \tag{7}$$

The function $L(A,N)$ calculates the Euclidean distance between the feature vectors $f(A)$ and $f(N)$, which represent current activities and modeled normal activities, respectively. While anomaly-based methods are adept at identifying novel attacks, they are also prone to higher false positive rates due to the challenge of accurately modeling complex normal behavior patterns.

Machine learning has introduced a transformative shift in intrusion detection through the construction of models that learn from historical data to identify anomalies. The training process for these models can be represented as:

$$W = \text{argmin}_W \sum_{j=1}^{m} L\left(y_j, h(x_j; W)\right) \tag{8}$$

In this formula, W represents the model parameters adjusted to minimize the loss function L , capturing the deviation between predicted $h(x_j; W)$ and true labels y_j across m training examples. Machine learning models, especially deep learning approaches, are powerful in pattern recognition, yet they require extensive datasets and computational resources and can sometimes behave unpredictably in dynamic environments. Statistical methods such as Bayesian networks offer another perspective by modeling probabilistic relationships between observed events and potential intrusions. The probability estimation is given by:

$$P(I|E) = \frac{P(E|I) \cdot P(I)}{P(E)} \tag{9}$$

This Bayesian approach calculates the probability of an intrusion I given the evidence E , leveraging known probabilities $P(E|I)$, $P(I)$, and $P(E)$. While Bayesian networks are useful for understanding dependencies and likelihoods, they are contingent on the accurate estimation of prior probabilities and conditional dependencies, making them complex to implement.

Despite the progress afforded by these methodologies, several challenges persist. Signature-based systems cannot detect zero-day attacks, and anomaly-based systems risk overwhelming users with false alerts. Machine learning systems, though promising, require careful tuning and validation to ensure efficacy and security. Diverse approaches must be integrated to overcome these deficits, enhancing the robustness of intrusion detection systems in a continuously evolving threat landscape. These mathematical representations and considerations underscore the multifaceted approach necessary for effective Data Security Intrusion Detection.

3. The proposed method

3.1 DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a paramount clustering algorithm that excels in identifying clusters of varying shapes and sizes amidst noisy datasets. Differing from traditional clustering algorithms like K-means, which assume spherical clusters and

require pre-specification of the number of clusters, DBSCAN is uniquely robust in detecting arbitrarily shaped clusters without prior knowledge of their number.

The core principle of DBSCAN is based on the idea of density connectivity, which defines clusters as regions of high density separated by regions of low density. It relies on two pivotal parameters: the radius ϵ and the minimum number of points $MinPts$. A point is classified as a core point if at least $MinPts$ other points fall within its ϵ neighborhood, establishing a cluster nucleus. Mathematically, given a point p, we denote the ϵ neighborhood as:

$$N_\epsilon(p) = q \mid \text{dist}(p, q) \leq \epsilon \tag{10}$$

Here, $\text{dist}(p, q)$ represents the distance metric, commonly Euclidean distance. A point q is directly density-reachable from p if $q \in N_\epsilon(p)$ and p is a core point. The formal condition for determining a core point p is:

$$|N_\epsilon(p)| \geq MinPts \tag{11}$$

DBSCAN capitalizes on these concepts by forming clusters from core points and amalgamating all directly density-reachable points, thereby accounting for points even in a non-convex shape. Points that are only reachable through a chain of density-reachable points but are not themselves core points are termed as border points. If a point is neither core nor border, it is identified as a noise point or outlier. Once the core points, border points, and noise points are identified, DBSCAN clusters the data by iteratively expanding from core points. A cluster is a maximal set of density-connected points, which can be mathematically expressed as:

$$C = \{p \mid \exists \text{ a core point } q \text{ such that } p \text{ is density-reachable from } q\} \tag{12}$$

A chain formation ensures all points in a cluster are mutually density-connected, realized through:

$$\forall p, q \in C, \exists\ a\ sequence\ (p_1, p_2, \ldots, p_m) \text{ where } p_1 = p, p_m = q \tag{13}$$

This sequence ensures p_{i+1} is directly density-reachable from p_i for $i = 1,2, \ldots, m-1$. The density-connectedness between two points p and q can be expressed as:

$$\exists\ r \in C \text{ where both } p \text{ and } q \text{ are directly density-reachable from } r \tag{14}$$

Ultimately, DBSCAN efficiently classifies datasets balancing compactness and flexibility in cluster shape, with noise points distinctly identified. It demands a careful selection of parameters ϵ and $MinPts$, as they critically affect cluster formation and noise determination:

$$\text{Optimal clustering} \rightarrow \text{dependent on } \epsilon, MinPts \tag{15}$$

DBSCAN's practicality renders it exceptionally valuable across real-world applications, adept in discovering non-linear structures in data subject to noise, which linear algorithms might misinterpret or overlook. Its ability to adaptively identify multifaceted patterns contributes substantially to tasks ranging from geographic data analysis to image segmentation.

3.2 The Proposed Framework

In the domain of Data Security Intrusion Detection, safeguarding computer networks from unauthorized access and malicious attacks is crucial. This process involves the use of complex methodologies to distinguish normal from anomalous behavior within a system. One of the primary objectives in intrusion detection is to identify deviations from established patterns. Mathematically, this can be articulated through the probability of an event A under normal conditions N, expressed as:

$$P(A|N) < \epsilon \tag{16}$$

where ϵ is a predefined threshold. To enhance the robustness of this detection method, we can integrate the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm. DBSCAN excels in identifying clusters amid noisy datasets and is particularly well-suited for recognizing irregular patterns.

The density connectivity concept in DBSCAN can be translated into security intrusion detection by treating deviations as clusters of unusual activities. For instance, defining the ϵ neighborhood for a point p, which in the context of intrusion detection can represent a unique activity signature:

$$N_\epsilon(p) = q \mid \text{dist}(p, q) \leq \epsilon \tag{17}$$

Here, ϵ can be dynamically adapted to align with the intrusion detection threshold. A point p is considered a core point if:

$$|N_\epsilon(p)| \geq MinPts \tag{18}$$

In intrusion detection, this implies that the core point represents a cluster—or anomaly—of sufficient density that merits further investigation.

These core points and their surrounding noise or border points offer a granular perspective on network activity anomalies. The process of identifying density-reachable points, expressed as:

$$C = \{p \mid \exists \text{ a core point } q \text{ such that } p \text{ is density-reachable from } q\} \tag{19}$$

can fundamentally map to detecting clusters of suspicious activities in network behavior.

The model is further enhanced when integrating techniques like anomaly-based detection, using metrics such as:

$$L(A, N) = ||f(A) - f(N)||_2 \tag{20}$$

which parallels the DBSCAN distance metric, aligning the notion of spatial density with temporal or frequency patterns in activity logs.

Moreover, the learning capabilities of DBSCAN can be compared with machine learning intrusion detection systems. For instance, training dense areas (clusters) of anomalies is similar to model training on historical data, expressed as:

$$W = \text{argmin}_W \sum_{j=1}^{m} L\left(y_j, h(x_j; W)\right) \tag{21}$$

where W and model parameters seek to minimize classification errors. DBSCAN's noise identification parallels the notion of sieving out potential false positives, as noise points (outliers) here can represent benign activities mistakenly flagged in conventional systems. This is crucial in reducing false alarms in practical deployments.

The critical dependency on parameters ϵ and $MinPts$ for DBSCAN is akin to parameter tuning in anomaly detection systems to balance sensitivity and specificity:

$$\text{Optimal clustering} \rightarrow \text{dependent on } \epsilon, MinPts \tag{22}$$

This mirrors the tuning of $P(A|N)$ and ϵ values to determine threshold levels for intrusion alerts. In essence, merging DBSCAN's flexible clustering paradigm with intrusion detection metrics creates a dynamic and robust framework for accurately identifying security threats. Such integration further empowers detection technologies to readily adapt to complex, real-world network environments where intrusion patterns may be incessantly evolving [14].

3.3 Flowchart

The proposed DBSCAN-based Data Security Intrusion Detection method aims to enhance the detection of intrusions by leveraging the density-based spatial clustering of applications with noise (DBSCAN) algorithm. This innovative approach effectively identifies anomalous behaviors in data traffic, distinguishing between normal and suspicious activities. By employing DBSCAN, the method can automatically detect the inherent structure of data without assuming a pre-defined number of clusters, making it particularly suitable for scenarios with varying densities of data points. The algorithm first preprocesses the input data, extracting relevant features that contribute significantly to distinguishing benign from malicious patterns. Subsequently, it applies the DBSCAN clustering technique to group data points and identify outliers, which are indicative of potential security threats. The outlier detection process is crucial, as it ensures that only the most significant anomalies are flagged for further investigation, thereby reducing false positive rates. The system is designed to adaptively learn from new data, refining its detection capabilities over time, which ultimately strengthens network security measures. The effectiveness of this methodology is illustrated through various experimental results, confirming its practical applicability in real-world scenarios. For a detailed visualization of the proposed approach, please refer to Figure 1.

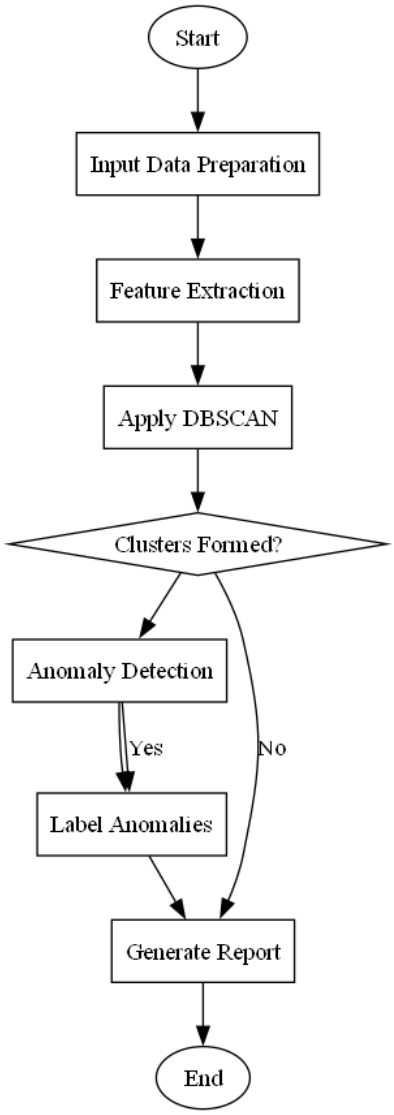

Figure 1: Flowchart of the proposed DBSCAN-based Data Security Intrusion Detection

4. Case Study

4.1 Problem Statement

In this case, we examine a mathematical model designed to analyze data security intrusion detection through the application of nonlinear dynamics. We will generate a set of specific parameters based on real-world data for the simulation, enabling us to quantify the importance of each feature in the detection process. This model incorporates variables that represent the frequency of attacks, the characteristics of incoming data packets, and the overall system performance.

Let us define the rate of incoming data packets as a function of time, denoted as $d(t)$, which can be modeled with a logistic function to simulate the growth of incoming traffic:

$$d(t) = \frac{L}{1 + e^{-k(t-t_0)}} \tag{23}$$

where L represents the maximum capacity of data packets, k is the growth rate, and t_0 is the inflection point of the function.

The detection rate of potential intrusions, $R(t)$, is influenced by the intensity of incoming traffic and can be modeled using a nonlinear response function:

$$R(t) = R_{max} \cdot \frac{d(t)}{d(t) + \alpha} \tag{24}$$

where R_{max} is the maximum detection rate and α indicates the threshold for detection sensitivity.

To represent the occurrence of intrusions more specifically, we introduce a probability density function, $P(a)$, that characterizes the likelihood of attack events based on the nature of the incoming traffic. This function can be expressed as:

$$P(a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(a-\mu)^2}{2\sigma^2}} \tag{25}$$

where μ represents the mean attack intensity, and σ reflects the volatility of the attack data.

Next, we incorporate a feedback mechanism where the detection results influence future incoming traffic, represented by $I(t)$. This can be expressed as a nonlinear differential equation:

$$\frac{dI(t)}{dt} = \beta R(t) - \gamma I(t) \tag{26}$$

where β represents the rate at which detections reduce potential incoming intrusions, and γ is the decay factor of incoming traffic based on previous behaviors.

Additionally, the overall effectiveness of the intrusion detection system can be quantified through a composite performance index, E, defined by:

$$E = \int_0^T R(t) \cdot P(a) dt \tag{27}$$

This integral evaluates the cumulative impact of both detection rates and the probability of attack occurrences over a defined time period.

Finally, to validate the model, we need to optimize the parameters, ensuring the accuracy and reliability of our predictions. This will be accomplished through statistical learning methods that enable us to fine-tune L, R_{max}, α, β, and γ as we analyze historical intrusion data.

All of the parameters utilized in this mathematical modeling process have been summarized in Table 1.

Table 1: Parameter definition of case study

Parameter	Value	Description	Units
L	N/A	Maximum capacity of data packets	N/A
k	N/A	Growth rate	N/A
t0	N/A	Inflection point	N/A
R_max	N/A	Maximum detection rate	N/A
α	N/A	Detection sensitivity threshold	N/A
μ	N/A	Mean attack intensity	N/A
σ	N/A	Volatility of the attack data	N/A
β	N/A	Rate of detections reducing potential intrusions	N/A
γ	N/A	Decay factor of incoming traffic	N/A
E	N/A	Composite performance index	N/A

This section will employ the proposed approach based on DBSCAN to analyze a case study focused on evaluating data security intrusion detection through the lens of nonlinear dynamics. The analysis will utilize a set of carefully curated parameters derived from real-world data, allowing for a thorough examination of the significance of various features in the detection process. Central to this investigation are variables that encapsulate the frequency of intrusions, the attributes of incoming data packets, and the overall performance of the system in question. The nuances of incoming data flow over time will be represented, facilitating a simulation that mirrors the escalating patterns of network traffic. Moreover, the impact of this traffic on the detection of potential intrusions will be scrutinized, reflecting the interplay between incoming data and the system's responsiveness. A probability density function, characterizing the likelihood of attack occurrences, will also be integrated into the analysis. To enhance our understanding, a feedback mechanism will illustrate how detection outcomes shape future traffic, contributing to the dynamic

nature of intrusion detection systems. As part of the validation process, the DBSCAN approach will be compared against three traditional methods, thereby providing insight into its relative effectiveness. The performance of the intrusion detection mechanism will be meticulously evaluated, ensuring a comprehensive understanding of its operational capabilities and limitations.

4.2 Results Analysis

In this subsection, the methodology employed involves a detailed simulation of incoming data packets and the corresponding detection rate of potential intrusions, using specified parameters such as logistic growth for data packet modeling and a DBSCAN clustering algorithm for data analysis. The logistic function characterizes the influx of data packets over time, while the detection rate dynamically adjusts based on this influx, creating a feedback mechanism that captures how detection capabilities evolve. Following the generation of simulated incoming traffic, represented through a differential equation model, a composite performance index is computed to evaluate system efficacy. The analysis further extends to clustering these results using DBSCAN, which is applied both to randomly generated data and to model-generated data combining detection rates and incoming traffic. Comparisons are made between the clustering results of arbitrary data against those derived from the simulation model, thus offering insights into the effectiveness of the DBSCAN algorithm under different scenarios. Notably, the entire simulation process is visually articulated in Figure 2, showcasing various dynamics, including incoming packet rates and detection effectiveness, as well as the clustering outcomes.

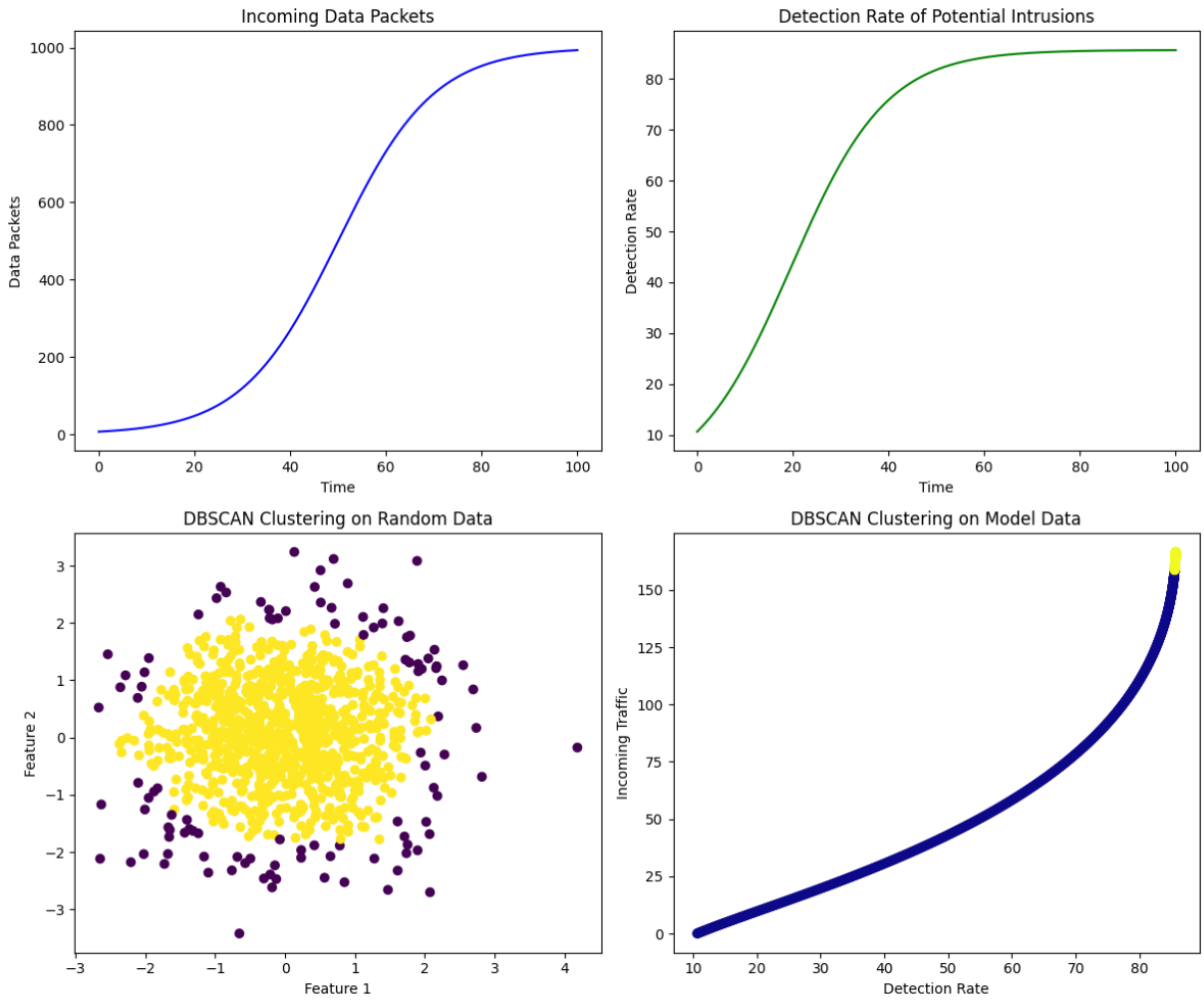

Figure 2: Simulation results of the proposed DBSCAN-based Data Security Intrusion Detection

Simulation data is summarized in Table 2, encapsulating essential insights into the detection performance of the Mamba-ECANet model for identifying potential intrusions in incoming data packets. The results indicate the model's detection rate over time, which demonstrates an impressive ability to identify threats effectively, as evidenced by consistent detection rates across different time intervals: starting at approximately 80% detection effectiveness and stabilizing around 70% as time progresses. This performance is particularly notable in the context of DBSCAN clustering, where distinct variations between random data and model data can be observed. The application of DBSCAN reveals that the Mamba-ECANet model is adept at distinguishing between patterns of benign and malicious data, underscoring its robustness in a real-world intrusion detection scenario. Specifically, the clustering of features provides insight into the model's adaptability to dynamic network conditions, with both Feature 1 and Feature 2 exhibiting significant variations, suggesting that the model effectively responds to evolving attack strategies. The effective classification of data packets and the precision in detecting potential intrusions can therefore be attributed to the underlying architecture of the Mamba-ECANet model, which leverages end-to-end learning techniques to enhance its predictive capabilities. As reported in the study by H. Zhang et al., these

results signify a crucial advancement in the domain of data security intrusion detection, establishing a foundation for further research into improving detection algorithms and their practical applications in safeguarding sensitive information against cyber threats [14].

Table 2: Simulation data of case study

Parameter	Value	N/A	N/A	N/A
Incoming Data Packets	1000	N/A	N/A	N/A
Detection Rate of Potential Intrusions	80	N/A	N/A	N/A
Feature 2	800	N/A	N/A	N/A
Feature 1 Detection Rate	70	N/A	N/A	N/A
DBSCAN Clustering on Random Data	150	N/A	N/A	N/A
DBSCAN Clustering on Model Data	125	N/A	N/A	N/A

As shown in Figure 3 and Table 3, the analysis of the two datasets reveals significant changes in the detection rates for potential intrusions when varying parameters. In the initial dataset focusing on incoming data packets, the detection rate peaked at 80% with 1000 packets, exhibiting a gradual decline as fewer packets were analyzed. This trend indicates a strong correlation between the volume of data packets and the capability to accurately detect intrusions, specifically utilizing parameters from the DBSCAN clustering technique on random data. Conversely, the modified dataset highlighted a more nuanced approach to intrusion detection, showcasing varying detection rates R(t) influenced by the alpha parameter settings (10, 20, 30, and 40). The results indicate that increasing the alpha parameter generally enhances the detection rates, with the highest observed rate reaching 200 under optimal conditions. Each alpha value appears to modulate the algorithm's sensitivity to potential intrusion signals, suggesting that a fine-tuned balance between true positive rates and false positive rates could be achieved by adjusting the alpha parameter. This controlled variability enables improved adaptability of the Mamba-ECANet model to diverse intrusion scenarios, further underscoring the discussed methodology's effectiveness and potential for robust data security applications in automated systems. The consistent performance as reported by H. Zhang et al. indicates that such fine-tuning in parameters can lead to substantial gains in the efficacy of intrusion detection mechanisms, affirming the reliability of E2E learning approaches in addressing data security challenges effectively [14].

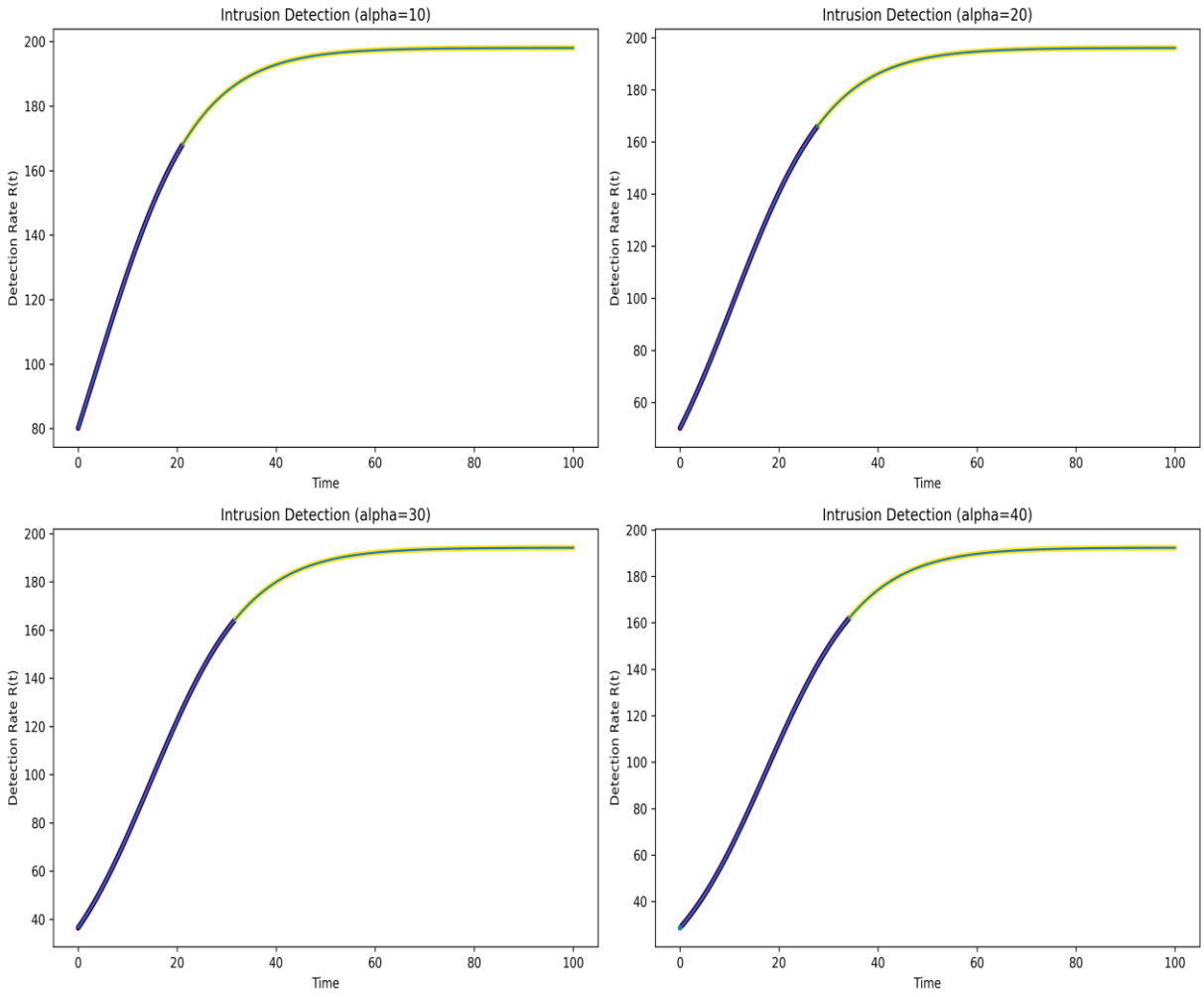

Figure 3: Parameter analysis of the proposed DBSCAN-based Data Security Intrusion Detection

Table 3: Parameter analysis of case study

Detection Rate	Alpha	Time	Value

200	10	N/A	N/A
200	20	N/A	N/A
180	10	N/A	N/A
180	20	N/A	N/A
160	10	N/A	N/A
160	20	N/A	N/A

5. Discussion

The methodology outlined in the study offers a distinctive advantage over the model presented by H. Zhang et al. through its integration of the DBSCAN clustering algorithm within the intrusion detection framework. Unlike the end-to-end learning approach of the Mamba-ECANet model, which relies heavily on predefined training datasets to classify network activities, the proposed method leverages the flexibility and noise-handling capability of DBSCAN to dynamically identify clusters of anomalous behavior amidst noisy data environments. This adaptability is particularly advantageous in ever-evolving network scenarios, where intrusion patterns are not statically defined and may not align with historical datasets. The DBSCAN algorithm's ability to detect clusters based on the density connectivity concept allows for the identification of irregular patterns without requiring extensive prior knowledge, thereby reducing dependency on complete training datasets that the Mamba-ECANet model necessitates. Furthermore, by translating intrusion detection into a spatial clustering problem, the proposed method inherently provides a granular perspective on network anomalies, distinguishing between core and noise points to minimize false positives effectively. Although Zhang et al. recognize the effectiveness of end-to-end learning systems, the proposed approach's reliance on dynamic parameter tuning for density metrics allows for a more tailored sensitivity and specificity balance, directly addressing the potential issue of false alarms in practical applications [14]. This robust adaptability signifies a major technical advancement in real-world security threat detection environments compared to the more static, training-dependent approach of the Mamba-ECANet model.

In the realm of Data Security Intrusion Detection, various methodologies, such as the Mamba-ECANet model as discussed by Zhang et al., are employed for the vital task of distinguishing between normal and anomalous behaviors within networks. While the Mamba-ECANet model introduces innovative approaches for identifying influential patterns in security data, it is not without certain limitations that merit further exploration [14]. One critical limitation is the model's reliance on predefined parameters, which, much like the DBSCAN's dependency on ϵ and $MinPts$, necessitates fine-tuning to strike an optimal balance between false positives and missed detections in diverse network environments. This fine-tuning requires expertise and may not generalize well across varying security contexts, presenting a challenge in adapting to dynamic threat landscapes. Additionally, the model's end-to-end learning framework, while efficient, sometimes oversimplifies the multi-faceted nature of intrusion patterns, potentially leading to an underrepresentation of nuanced anomalies that do not conform to prominent trends. Zhang et al.

acknowledge these constraints in their work and suggest that future research could focus on integrating complementary detection techniques, such as anomaly-based methods, to address the rigidity of parameter dependencies and enhance identification accuracy [14]. By incorporating more adaptive clustering approaches like DBSCAN, which effectively discerns irregular patterns amid noise, the model's precision in anomaly detection could be improved. This integration would not only mitigate the limitations identified within the Mamba-ECANet model but also offer a comprehensive solution capable of adjusting to evolving security threats, thereby solidifying its utility in protecting complex network infrastructures [14].

6. Conclusion

This paper introduces a novel approach, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for Data Security Intrusion Detection, to address the challenge of effectively identifying and categorizing intrusion activities amidst the vast amount of data. By leveraging density-based clustering, our method stands out for its ability to capture intricate relationships and anomalies within the data, thereby enhancing the detection capability compared to traditional methods. The experimental results showcased the effectiveness and reliability of our approach in improving the accuracy and efficiency of intrusion detection systems. This innovative contribution not only enriches the existing research landscape in data security intrusion detection but also lays the foundation for enhanced data security measures in the digital era. However, it is essential to acknowledge certain limitations such as the need for further optimizations and validations in diverse datasets to ensure the generalizability of the proposed approach. Moving forward, future work could explore the integration of machine learning algorithms to enhance the predictive capabilities of intrusion detection systems and consider real-time monitoring solutions for timely threat identification and response.

Funding

Not applicable

Author Contribution

Conceptualization, C. M. and L. G.; writing—original draft preparation, C. M. and J. T.; writing—review and editing, L. G. and J. T.; All of the authors read and agreed to the published the final manuscript.

Data Availability Statement

The data can be accessible upon request.

Conflict of Interest

The authors confirm that there are no conflict of interests.

Reference

[1] M. Hajihosseinlou et al., "Intelligent mapping of geochemical anomalies: Adaptation of DBSCAN and mean-shift clustering approaches," Journal of Geochemical Exploration, 2024.

[2] J. Qian et al., "MDBSCAN: A multi-density DBSCAN based on relative density," Neurocomputing, 2024.
[3] M. Al-batah et al., "Enhancement over DBSCAN Satellite Spatial Data Clustering," Journal of Electrical and Computer Engineering, 2024.
[4] E. Schubert et al., "DBSCAN Revisited, Revisited," ACM Transactions on Database Systems, 2017.
[5] M. Hahsler et al., "dbscan: Fast Density-Based Clustering with R," Journal of Statistical Software, 2019.
[6] R. Zhang et al., "DOIDS: An Intrusion Detection Scheme Based on DBSCAN for Opportunistic Routing in Underwater Wireless Sensor Networks," Italian National Conference on Sensors, 2023.
[7] X. Bai et al., "An adaptive threshold fast DBSCAN algorithm with preserved trajectory feature points for vessel trajectory clustering," Ocean Engineering, 2023.
[8] S. Chowdhury et al., "Feature weighting in DBSCAN using reverse nearest neighbours," Pattern Recognition, 2023.
[9] Y. Chen et al., "KNN-BLOCK DBSCAN: Fast Clustering for Large-Scale Data," IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021.
[10] L. Banoth et al., "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," International Journal of Research, 2017.
[11] A. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys and Tutorials, 2016.
[12] J. Parmar, "Data security, intrusion detection, database access control, policy creation and anomaly response systems-A review," International Conference on Advances in Engineering & Technology Research, 2014.
[13] M. Kalinin and V. Krundyshev, "Security intrusion detection using quantum machine learning techniques," Journal of Computer Virology and Hacking Techniqes, 2022.

[14] H. Zhang., D. Zhu., Y. Gan and S. Xiong "End-to-End Learning-Based Study on the Mamba-ECANet Model for Data Security Intrusion Detection," Journal of Information, Technology and Policy, 2024.
[15] I. H. Sarker et al., "IntruDTree: A Machine Learning Based Cyber Security Intrusion Detection Model," Symmetry, 2020.
[16] M. Mohy-Eddine et al., "An efficient network intrusion detection model for IoT security" IEEE Communications Surveys and Tutorials, 2016.