


Enhancing Cloud Network Security with Innovative Time Series Analysis

Amer Al-Mazrawe   [Alkafeel University | aamiersame@alkafeel.edu.iq]

Bahaa Al-Musawi  [University of Kufa | bahaa.almusawi@uokufa.edu.iq]

 College of Technical Engineering, Alkafeel University, Column 23, Al-Nidaa Neighborhood, next to Al-Amirat Residential Complex, Najaf Al-Ashraf, Iraq.

Received: 31 July 2024 2022 • **Accepted:** 07 December 2024 2024 • **Published:** 03 February 2025

Abstract Cloud computing has revolutionized computing infrastructure abstraction and utilization, distinguished by its cost-effective and high-quality services. However, the challenge of securing cloud networks persists, mainly due to the broad exchange of data and the inherent complexity of these techniques. Anomaly detection emerges as a promising solution to improve cloud network safeness, presenting perception into system behavior and alerting operators for further actions. This paper offers a novel time series analysis method for detecting anomalies in cloud networks. Our technique employs innovative time series analysis techniques based on a matrix profile, and the Kneedle algorithm to identify multi-dimensional anomalous patterns within multiple features extracted from network traffic streams. To evaluate the efficacy of our approach, we implemented timestamp-based and index-based methods to two distinct datasets: the most widely used UNSW-NB15 and the recently introduced CICIoT2023 datasets. The results highlight the efficacy of our proposed method in identifying cloud network anomalies. It achieved an impressive accuracy of 99.6% and an F1-score of 99.8% using the timestamp-based analysis method. For the index-based analysis method, accuracy reached 98%, accompanied by an outstanding F1-score of 99.9%.

Keywords: Cloud networks, Anomaly Detection, Matrix profile, Time series analysis, Cloud security, IoT, Cyber-attacks

1 Introduction

The remarkable advancements in cloud computing, driven by its flexibility, cost-effectiveness, and benefits, have led to widespread adoption across numerous institutions and enterprises. However, a persistent concern among companies revolves around the protection of privacy and ensuring cloud security. This challenge remains a focal point for cloud service providers. A recent survey by Baron [2022] highlighted that data loss from the cloud ranks as the top concern for cloud networks.

Multiple studies and research attempts have made significant contributions to the field of cloud security, covering various aspects, with data monitoring within cloud networks being a particularly significant focus. The motivation behind implementing cloud monitoring is to maintain control, allocate resources efficiently, and enhance security by storing and analysing all measurements gathered from various cloud components Syed *et al.* [2017]. Cloud network monitoring plays a crucial role in upholding the efficiency, accessibility, and protection of cloud-centric applications. It involves proactive management of infrastructure, issue resolution, and the guarantee of a smooth user experience Nzanzu *et al.* [2022].

Several techniques and algorithms have been introduced to monitor cloud networks by capturing traffic transfer between cloud components to alert network operators. While several approaches rely on machine learning techniques, network operators have been hesitant to adopt them for several reasons. Firstly, these solutions were initially introduced as 'black boxes' making it challenging To discuss the steps

and rationale behind these models' decision-making, which makes them a harder sell compared to simpler but often less effective rule-based approaches. Secondly, training machine learning models with imbalanced datasets can lead to biases in favour of the majority classes Jacobs *et al.* [2022]. Finally, anomaly detection techniques based on machine learning may become less effective over time due to changes in the underlying network traffic caused by changes in topology or settings. Consequently, a new round of training becomes necessary to maintain optimal performance.

Although time series analysis has been around for a long time, it remains a powerful tool for predicting trends and identifying anomalies. Time series analysis serves the purpose of detecting anomalous occurrences within the behaviour of cloud networks, potentially signalling system failures or cyberattacks. These anomalies can significantly disrupt service reliability and result in substantial financial repercussions. Due to the varied and diverse characteristics of cloud environments, conventional methods may not be suitable for addressing the challenges arising from cloud security, performance issues, and various application workloads. To tackle these challenges, new anomaly detection methods in cloud environments should consider the analysis of both historical and real-time data streams Garg *et al.* [2019].

Based on this motivation, we propose a novel approach for identifying anomalies in cloud networks using a matrix profile framework and the Kneedle algorithm. The matrix profile is an innovative data structure designed for time series analysis, enabling the discovery of repeated patterns, correlations, and outliers within the time series. It is a robust and

scalable tool that operates effectively with minimal parameter requirements Yeh *et al.* [2016b]. This capability allows for the detection of anomalies that may indicate system failures, performance degradation, security breaches, or other unusual behaviours in the cloud environment De Paepe *et al.* [2020].

The evaluation of our proposed approach yields promising results in the detection of anomalies in cloud networks, whether using timestamp-based or index-based. Our new approach achieved an accuracy of 99.6% and an F1-score of 99.8% using timestamp-based analysis for the UNSW-NB15 dataset. Furthermore, for index-based analysis with the CIIoT2023 dataset, our approach attained an accuracy of 98% and an F1-score of 99.9%.

In summary, this paper's contributions can be outlined as follows:

- We conducted feature selection investigations on two datasets, UNSW-NB15 and CIIoT2023, and proposed subsets of 9 and 10 features, respectively, for detecting various attacks supported by these datasets ii.
- We present a new method for identifying abnormalities in cloud networks, utilizing a multidimensional matrix profile and the Kneedle algorithm without the need to establish a profile for normal behaviour.
- We assess the efficacy of our proposed technique using both timestamp-based and index-based methods for two different datasets, demonstrating the effectiveness of our approach in detecting anomalies in cloud networks.

The subsequent sections of this article are structured in the following manner: Section 2 explores related works. Section 3 provides a background of the matrix profile and multidimensional matrix profile. Section 4 describes our proposed system design. In Section 5, we discuss and outline our results, and finally, conclude our work in Section 6.

2 Related Work

This section summarizes previous works that employ statistical, machine learning, deep learning, and hybrid methods to identify abnormalities in cloud networks.

Agrawal *et al.* in Agrawal *et al.* [2016] proposed an automatic anomaly detection system based on robust Principal Component Analysis (PCA) for cloud networks. Robust PCA employs a recursive Singular Value Decomposition (SVD) and a threshold to convert the initial data into a reduced-rank representation. Their work focused on identifying abnormal behaviours in servers within cloud networks. The suggested system achieved an accuracy rate of 87% and an F1-score of 86% when evaluated on Yahoo benchmark datasets. However, the proposed method was not evaluated with well-known datasets.

Huang *et al.* in Huang *et al.* [2017] Proposed a solution called Relaxed Linear Programming SVDD (RLPSVDD) to address the issue of a high false-positive rate in Support Vector Data Description (SVDD). The RLPSVDD algorithm aims to generate a flexible data representation and then apply linear programming to detect anomalies in time series data. While the RLPSVDD approach has demonstrated efficacy in

detecting anomalies in cloud networks, it involves more effort for parameter tuning.

Saljoughi *et al.* in Saljoughi *et al.* [2017] Suggested an innovative approach utilizing artificial intelligence methods for identifying irregularities in cloud networks. This method uses multilayer perceptron neural networks to classify attacks and utilizes the particle swarm algorithm to enhance the accuracy of the classifier. The suggested technique's performance was evaluated using the NSL-KDD and KDD Cup 1999 datasets, resulting in 99.4% accuracy for KDD Cup 1999 and 98.08% for NSL-KDD. Nevertheless, the optimization of parameters for the suggested approach demands a significant level of effort.

Ding *et al.* in Ding *et al.* [2018] Presented is a novel real-time detection technique named RADM, which utilizes hierarchical temporal memory (HTM) and Bayesian networks (BN) to identify anomalies in multivariate time series data. The HTM is employed for anomaly detection, whereas the BN is utilized for validation purposes. The evaluation of the suggested approach using the NAB dataset yielded a 77% accuracy rate. Nevertheless, it exhibits a significant incidence of false alarms.

Schmidt *et al.* in Schmidt *et al.* [2018] introduced a technique that utilizes the online Autoregressive Integrated Moving Average (ARIMA) model for anomaly event detection in cloud monitoring. OpenStack cloud computing systems generated a dataset for the evaluation of the methodology. The results exhibited exceptional detection rates and few false positives. The proposed method failed to identify cloud anomalies on virtual cloud services.

Zhang *et al.* in Zhang *et al.* [2019] presented Active Transfer Anomaly Detection (ATAD) is a method of identifying anomalies in an unlabelled dataset by comparing it to other datasets. To address the difficulty of acquiring an adequate amount of tagged data for cloud monitoring in the face of large volumes and dispersed locations. This approach integrates active learning and transfer learning methodologies to intelligently identify a limited number of samples from the target dataset. The evaluation of ATAD using the NAB and Yahoo datasets resulted in an F1-score of 99.2%. Nevertheless, the proposed method needed to be evaluated with larger-scale datasets.

Lou *et al.* in Lou *et al.* [2019] suggested a Max-Min method that employs distance and support vector data to identify anomalies in cloud systems. It detects anomalies by capturing several metrics derived from cloud components, including network traffic, CPU use, and memory usage. The evaluation of the proposed approach achieved a 95.0% accuracy rate by monitoring ten servers. However, the presented approach requires that it be evaluated with more than ten servers or larger datasets.

Yasarathna and Munasinghe in Yasarathna and Munasinghe [2020] Utilized an Autoencoder and One-class Support Vector Machine (OCSVM) to accurately identify anomalies in cloud network data. The effectiveness of the OCSVM and Autoencoder was demonstrated by evaluating the proposed technique utilizing two datasets, YAHOO and UNSW-NB15. The results demonstrated that the Autoencoder performed better than the OCSVM. Nevertheless, fitting a neural network model of the proposed method takes significant

time.

Al-Bakaa et al. in Al-Bakaa and Al-Musawi [2022] introduced a new method based on Recurrence Quantification Analysis (RQA), a non-linear statistical analysis technique, to identify anomalous deviations in network traffic. The proposed approach was evaluated using the UNSW-NB15 dataset, yielding a 96.71% accuracy and a 92.33% F1-score. The method outperforms previous works using one feature and effectively identifies hidden attributes in the underlying series of an individual feature. However, it is worth noting that the suggested method necessitates estimating RQA parameters before deployment.

Parameswarappa et al. in Parameswarappa et al. [2023] Utilized machine learning techniques to propose an innovative firewall approach for improving the security of cloud-based computing. The proposed methodologies forecast the ultimate classification of attacks by combining past node assessments with the present determination made by the machine learning algorithm, a method known as the 'most frequent decision'. The effectiveness of the suggested method was evaluated by employing the UNSW-NB15 dataset, resulting in an accuracy and F1-score of 97.68%. Nevertheless, it incurs significant costs in terms of computational resources.

Cheema et al. in Cheema et al. [2023] came up with a new way to do structural health monitoring (SHM) in sensor array networks that combine matrix profiling from time series data mining with optimal transport theory. The methodology has several advantages: it performs well in small data scenarios without requiring extensive training, supports unsupervised and semi-supervised learning, adapts to online learning environments, and is resilient to sensor network issues. Notably, it requires no predefined thresholds and offers high interpretability. The method was validated through diverse case studies, including aerospace simulation datasets, experimental building data from Los Alamos National Laboratories, and field data from a cable-stayed bridge in Sydney, Australia. Results demonstrated their effectiveness in accurately tracking progressive damage levels, showcasing its applicability in real-world SHM scenarios.

This study presents a novel method for detecting anomalies in cloud networks. Our approach is founded on the utilization of Matrix Profile, a time series analysis technique introduced to yield valuable insights into the underlying patterns and structures within the data. The MP algorithm has been characterized as robust, scalable, and parameter-free. It is simple and capable of handling missing data effectively. The evaluation of our proposed approach demonstrates a promising performance in terms of accuracy and low rate of false alarms.

3 Multi-Dimensional Matrix Profile

In this section, we introduce the fundamentals of matrix profile and highlight its characteristics. We also explore the fundamentals of a multi-dimensional matrix profile.

3.1 Background of Matrix profile

The matrix profile (MP) was developed by Yeh, Zhu, et al. Yeh et al. [2016b] to offer numerous advantages in time series data mining tasks such as providing precise results with dismissals in motif, discord, or time series joins. It is simple, parameter-free, scalable, and highly space-efficient, enabling the processing of massive datasets Madrid et al. [2019]. The MP involves of two fundamental elements: a distance profile and a profile index. The distance profile refers to a vector that consists of the minimal Euclidean distances, while the profile index includes the index of its nearest neighbouring element Scott et al. [2024]. In other words, it is the position of its most identical sub-sequence. Fast computation and detection times are two of the matrix profile's most important features Alzahrani et al. [2022] and Zhu et al. [2016]. The MP streamlines the process by eliminating the need for users to set similarity or distance thresholds for time series joins. Its construction is easily parallelizable in distributed systems, leveraging hardware efficiently. Moreover, it maintains a constant time complexity even as the sequence length increases, a rarity among alternative methods. The Matrix profile's deterministic construction time allows for accurate computation duration predictions. Lastly, it excels at handling missing data, guaranteeing accurate answers with no false negatives Yeh et al. [2017a].

Below, we will outline some of the key terminology used in time series and matrix profiles:

A subsequence: is a contiguous set of data points from a time series. subsequences of a fixed length are used for pattern recognition, anomaly detection, and similarity search.

The Euclidean Distance: is a fundamental measure of distance between two points in Euclidean space Liberti et al. [2014]. In essence, it is the straight-line distance between two points, calculated using the Pythagorean theorem. **Figure 1** illustrates an example of subsequences and Euclidean distance Yeh et al. [2016b].

Euclidean Distance

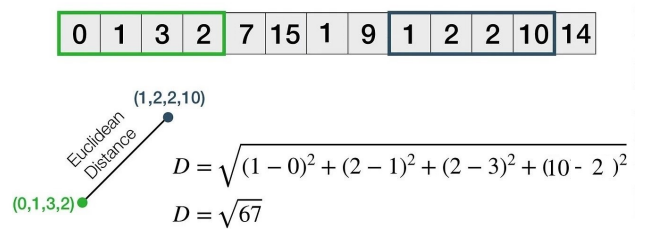


Figure 1. An example of calculating the Euclidean distance.

The distance matrix refers to the similarity between subsequences in a time series. It uses Euclidean distance to calculate the similarity degree Madrid et al. [2019]. **Figure 2** Provides a demonstration of a distance matrix.

A time series discord refers to a subsequence with the greatest distance from its closest neighbor.

The MP can be calculated using several algorithms, such as STAMP, STAMPI Yeh et al. [2016a], STOMP Zhu et al. [2016], SCRIMP++ Zhu et al. [2018a], and GPU-STOMP Zhu et al. [2018b] which use both the available compu-

	0.37	0.47	0.78	0.65	0.99	0.81	0.51	0.68	0.57	0.993
0.62		0.65	0.51	0.56	0.75	0.88	0.9	0.8	0.64	0.897
0.29	0.49		0.42	0.29	0.18	0.64	0.86	0.78	0.36	0.863
0.72	0.69	0.59		0.62	0.91	0.18	0.36	0.56	0.63	0.913
0.94	0.94	0.83	0.99		0.72	0.91	0.92	0.37	0.39	0.989
0.9	0.5	0.59	0.47	0.35		0.39	0.67	0.61	0.65	0.899
0.7	0.56	0.62	0.61	0.18	0.53		0.78	0.37	0.47	0.91
0.47	0.78	0.47	0.46	0.56	0.67	0.37		0.39	0.53	0.779
0.84	0.46	0.89	0.97	0.57	0.18	0.35	0.73		0.61	0.97
0.56	0.6	0.73	0.86	0.76	0.59	0.42	0.64	0.94		0.936
A										B

Figure 2. Example of the distance matrix (A), and its corresponding MP (B).

tational resources and domain limitations for best effectiveness. The mSTAMP algorithm is quick without high-performance hardware, demonstrating the algorithm's efficiency and effectiveness Yeh *et al.* [2017a].

MP analysis relies on the window size, which determines the length of subsequence examined in time series data. The choice of window size affects the analysis's sensitivity and granularity. Smaller window sizes detect short-term patterns and anomalies, while larger ones capture long-term trends but may miss short-lived anomalies Scott *et al.* [2024]. Factors like signal periodicity, and time series length are considered when selecting the best window size. Experimenting with different window sizes and comparing results can help find the best comparison between accuracy and computational efficiency Alzahrani *et al.* [2022]. Nilsson *et al.* in Cheema *et al.* [2023] indicated that certain knowledge of the way to capture data can assist in generating rough estimates that are typically effective for identifying optimal window sizes.

Calculating a MP involves several steps. The first step is to calculate the distance matrix between each subsequence using the Euclidean distance. After that, an exclusion zone is established to skip insignificant correspondences. Next, the maximum values are extracted from the distance matrix to identify potential anomalies in the matrix profile vector. Finally, the timestamps or indices of the potential anomalies are determined. **Figure 3** illustrates these steps.

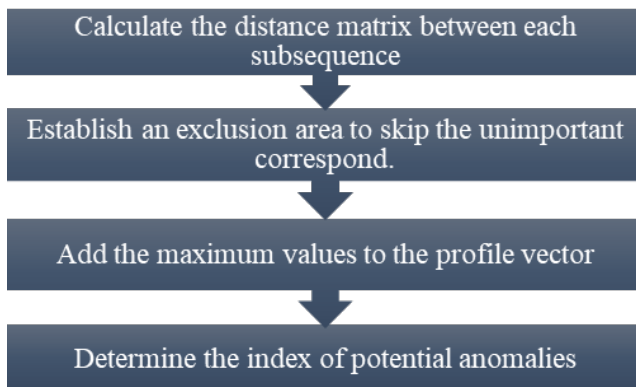


Figure 3. Steps involved in the MP procedure.

3.2 Multi-Dimensional Matrix Profile

A matrix profile multidimensional time series refers to the application of the matrix profile technique to analyse and extract patterns from time series data that have multiple dimensions Yeh *et al.* [2017b]. Multi-dimensional MP exceeds the limitations of one-dimensional matrix profiles by holistically computing Z-normalized Euclidean distances between subsequences within multi-dimensional time series data. This approach discloses complex patterns and relationships that reside across multiple dimensions, empowering the discovery of recurrent motifs, and anomalies. Multi-dimensional matrix profiles are different from 1-dimensional matrix profiles since it's not just a stack of one-dimensional profiles.

Yeh *et al.* in Yeh *et al.* [2017b] introduced "mSTAMP" as an extension of the original MP framework to efficiently find similar motifs and discord in time series data. The mSTAMP algorithm uses the MP as a fundamental to simultaneously identify discords or motifs across multiple time series samples. mSTAMP can find recurring patterns across many samples quickly and easily by using the Matrix Profile's ability to quickly calculate distances between subsequences. This approach allows mSTAMP to handle large datasets efficiently and discover meaningful patterns that might not be evident by analyzing each time series independently.

Satopää *et al.* in Satopää *et al.* [2011] proposed the Kneedle algorithm. The Kneedle algorithm is a method used for automatically determining the "knee" point in a curve or elbow. In time series data analysis, a knee point represents a significant change or inflection point in the curve's slope, indicating a transition between two distinct regions or behaviours. The Kneedle algorithm is often applied in various fields, such as anomaly detection, clustering, and optimization. The main goal of the Kneedle algorithm is to find the knee point in each curve without requiring prior knowledge of the data's underlying distribution or characteristics. It works by iteratively examining the distances between consecutive points in the curve and calculating the change in slope at each step. When the algorithm identifies a point where the slope change is significant, it marks that point as the knee point. This indicates the transition between two segments of the curve, offering valuable insights into the data's structure or optimal parameter selection.

By integrating the strengths of the matrix profile algorithm mSTAMP and the Kneedle algorithm, we establish an effective approach for detecting anomalies in multidimensional time series data. Consequently, our proposed system employs mSTAMP and Kneedle algorithms to detect anomalies in cloud networks, utilizing multiple features extracted from cloud networks as inputs.

4 System Design

The suggested system comprises five components, as illustrated in **Figure 4** data source, preprocessing, feature selection, Multidimensional Anomaly Detection (MAD), and alarm. In the preprocessing stage, missing instances are replaced, and non-numerical attributes are encoded. Feature selection involves dimensional reduction and the selection

of the most effective features. The MAD stage identifies anomalies using a multi-dimensional matrix profile.

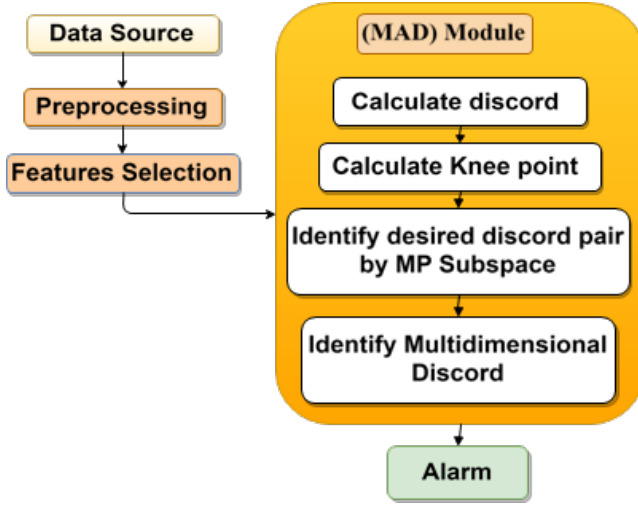


Figure 4. System design.

4.1 Data Source

The process of evaluating an anomaly detection system represents a challenge for researchers due to the challenges of injecting various types of attacks into a real network without impacting users. Additionally, network operators are unwilling to allow researchers to validate their approaches on real networks due to privacy concerns. Consequently, researchers have resorted to using their testbed and injected synthesized attacks to evaluate their approaches, and they have publicly published the generated datasets. Each of these datasets possesses unique characteristics concerning format, the frequency of multiple attacks, and the duration of these attacks. To assess the efficacy of the proposed system, we utilized two datasets. These are the UNSW-NB15 and CICIOT2023 datasets. The UNSW-NB15 dataset is the most widely used dataset that has been used for evaluating anomaly detection methods in the Internet of Things (IoT) Al-Bakaa and Al-Musawi [2021] and cloud computing networks Manimurugan [2021]. The CICIOT2023 is a more contemporary dataset with more than 32 recent attacks. This extensive test underscores the effectiveness of our methodology in detecting both traditional and novel attack vectors, offering a detailed analysis of its flexibility and precision across many contexts. We chose to utilize the UNSW-NB15 dataset, despite its age of eight years, for several reasons. Firstly, our paper introduces a new approach, necessitating a comparison of its performance with similar works that have utilized a well-established dataset for benchmarking. Secondly, this dataset addresses numerous shortcomings observed in older datasets, including the presence of multiple missing values. Thirdly, it encompasses nine of the most critical attack types that pose significant threats to networks Moustafa and Slay [2015]. Finally, the UNSW-NB15 dataset, comprising 49 features, provides the means to detect various attacks it supports Zoghi [2020]. Furthermore, we evaluated our approach using the recent CICIOT2023 dataset Neto *et al.* [2023] for several reasons. Firstly, there are architectural similarities

between IoT and cloud networks Firouzi *et al.* [2022] and Liu *et al.* [2020]. Secondly, our approach introduces novelty, making a comprehensive evaluation essential to determine its effectiveness in addressing modern issues observed in recent datasets, including emerging types of attacks. Thirdly, to provide a comprehensive examination of network activities, the CICIOT2023 dataset integrates real-time attack data with typical network behaviors. Comprising 47 features, this dataset is specifically designed to detect various types of attacks, some of which can pose significant threats to cloud networks. Moreover, it includes the simulation of 33 attacks to promote the development of security analytics technology. However, it is important to note that, unlike the UNSW-NB15 dataset, the CICIOT2023 dataset does not include timestamps in its CSV files. Consequently, we opted to use an index-based approach for this dataset rather than a time-based one. This decision allows us to assess the efficacy of our technique using an index-based dataset as well.

4.2 Preprocessing

A preprocessing covers the procedures of encoding features by mapping non-numeric ones to numeric values; for instance, the service feature's categorical discrete values "Normal" and "Attack" are converted to "0" and "1," respectively. Another example is encoding protocol from text to a number, such as "HTTP" to be "1." Next, we detect missing values like "#DIV/0!," "?," and "-" and replace those missing occurrences with the value that occurs most frequently. During the preprocessing of the UNSW-NB15 dataset, we utilize a timestamp to group the data into time windows of (5, 10, 15, and 20) time intervals and select the most frequent value for each column. Grouping aims to capture patterns and trends within specific periods, potentially revealing attack signatures or trends that might be obscured in individual data points. We found that grouping by a 10seconds window produced the best results. This step improves the efficiency of the execution time. In contrast, we utilize the CICIOT2023 dataset as an index based.

4.3 Features Selection

Feature selection involves selecting effective features from a larger set of available attributes to improve model performance, reduce overfitting, enhance interpretability, and reduce computational complexity Zebari *et al.* [2020]. Numerous algorithms have been introduced to find out the most effective feature. Numerous algorithms have been introduced to find out the most effective feature. We use Forward Selection Ranking (FSR) and Backward Elimination Ranking (BER) methods to identify effective features in feature selection tasks as the most used methods, and their effectiveness is due to their iterative approach, computational efficiency, and ability to capture feature interactions Al-Bakaa and Al-Musawi [2021]. FSR builds optimal feature subsets by adding impactful features, while BER eliminates less relevant ones, enhancing model interpretability and reducing computational costs. We achieved 10 features in the CICIOT2023 dataset and 9 features in UNSW-NB15. The

selected feature for the UNSW-NB15 and the CICIoT2023 datasets can be found in **Table 1** and **Table 2** respectively.

4.4 Multidimensional Anomaly Detection

Multidimensional Anomaly Detection (MAD) employs two algorithms to identify multidimensional discords. First, the core algorithm "mSTAMP" computes discords in a multidimensional time series. It is important to observe that this algorithm may encounter irrelevant dimensions. Failing to disregard these irrelevant dimensions can complicate the identification of abnormalities linked to relevant dimensions. One approach to address this challenge is by identifying the elbow or knee point. The Kneedle algorithm chooses the "knee" by plotting the maximum matrix profile value for each dimension found in the previous step against the dimensions and picking the dimension that has a "turning point" with concave curvature. As a result, the matrix profile subspace uses selected dimensions to identify dimensions with real synchronous anomalies, indicating multidimensional discords. This process is crucial in identifying dimensions with synchronous anomalies. In the Kneedle algorithm, the online parameter influences the detection and adjustment of the knee point. Selection of Online Mode, the algorithm operated in online=True mode, enabling it to iteratively refine and adjust the knee point with each data point processed. In this form, the algorithm initially recognized index 6 as a preliminary knee point, but later substituted it with index 9 due to its superior curvature and status as the peak. This tendency is a deliberate characteristic of the online mode, enabling the knee point to adjust to the most significant curvature as the algorithm advances through the data. Figures **Figure 5** and **Figure 6** illustrate the online parameter effect.

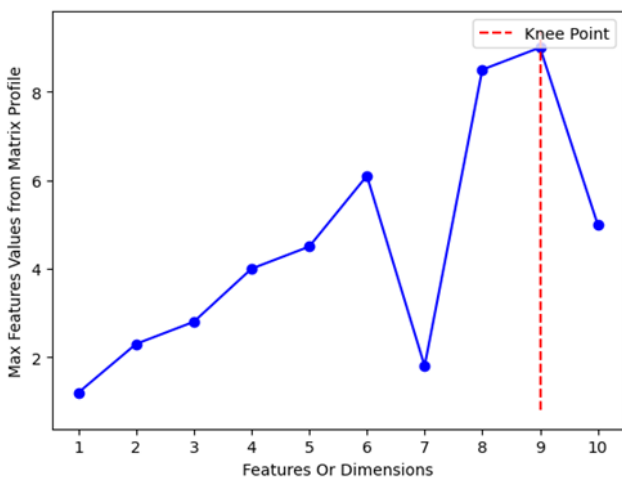


Figure 5. An example of identifying Knee using the Kneedle Algorithm with online parameter = True.

When analysing a dataset using the matrix profile, it is crucial to consider window and subset sizes as significant parameters. For large datasets, there are two options: investigating the entire database or dividing it into subsets and iterating through each to cover the entire database. The former approach is time-consuming and computationally expensive, while the latter is more efficient in terms of accuracy and de-

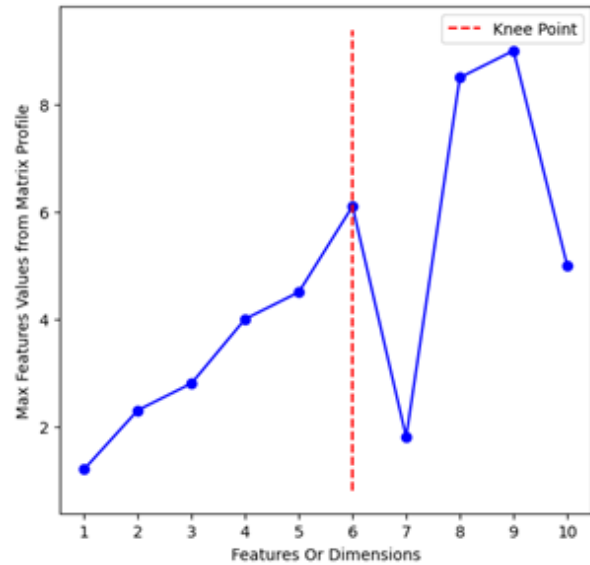


Figure 6. An example of identifying Knee using the Kneedle Algorithm with online parameter = False.

tected attacks. The matrix profile method effectively handles concept drift and adjusts to changing network behaviours by persistently examining recent data patterns using a sliding window technique. This enables the identification of shifts without dependence on static baselines. This method recalibrates analogous profiles instantaneously, allowing the system to identify and adapt to emerging, prevailing patterns as network behaviours evolve. Concurrently, this method and the Kneedle algorithm, which finds big changes in profile values, can tell the difference between normal and abnormal situations. This makes it a good choice for finding problems in dynamic cloud-based systems. In general, smaller subsets and window sizes can detect short-term patterns and anomalies more accurately, while larger ones capture long-term trends but may miss short-lived anomalies. It is crucial to observe that the best subset size and window sizes depend on various factors. Factors such as signal periodicity and time series length should be considered when selecting the best subsets and window sizes. To get the ideal balance among accuracy and computational efficiency, it is advisable to conduct experiments with various subsets and window sizes, subsequently comparing the outcomes Alzahrani *et al.* [2022]. Section 5 will offer additional analysis and explanation.

5 Results and Evaluation

The experiments were conducted on a simple PC with a 2.90 GHz Intel Core i7 processor and 8 GB of memory. All experiments were executed using Python and the STUMPY open-source library was utilized for efficient time series data analysis. The STUMPY library is primarily used for computing the matrix profile and is compatible with additional libraries like Pandas, NumPy, and Scikit-Learn. In terms of time execution, our approach based on the matrix profile showed a promising result. For example, it took 1018 seconds to analyse 80,000 instances of the CICIoT2023 dataset. The system's performance was evaluated using accuracy and

Table 1. Selected features from the UNSW-NB15 dataset.

No.	Feature Name	Description
1	Sloss	Source packets lost or retransmitted
2	State	A protocol that depends on the state, such as ACC, CLO, else (-)
3	Dur	Log full period.
4	Sbytes	Bytes from source to destination
5	Dbytes	Bytes from destination to source
6	Sttl	Time to live from source to destination
7	Dttl	Time to live from destination to source
8	Dloss	Destination packets that are deleted or transmitted
9	Service	FTP, SSH, DNS, HTTP, else (-)

Table 2. Selected features from the CICIoT2023 dataset.

No.	Feature Name	Description
1	IAT	The difference in time from the previous packet.
2	flow_duration	Time spent flowing the packet.
3	urg_count	Log full period.
4	Header Length	Length of header.
5	Rate	Transmission rate of packets inside a flow.
6	HTTPS	Verifies whether the application layer protocol is HTTPS.
7	HTTP	Verifies whether the application layer protocol is HTTP.
8	TCP	Indicates if the transport layer protocol is TCP.
9	rst_count	In a single flow, how many packets contain the RST flag.
10	syn_count	Number of packets in a flow with a SYN flag set

an F1-score. Accuracy represents the proportion of correctly classified records and can be calculated as follows:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

TP represents the number of correctly identified abnormal records, while TN represents the number of correctly classified normal occurrences. FP refers to regular records mistakenly labeled as attacks, while FN refers to total attack records classified as normal by the system. The F1-score is a crucial metric that assesses the system's ability to distinguish abnormal occurrences from normal occurrences and is computed as follows:

$$F1 - score = (2TP) / (2TP + FP + FN) \quad (2)$$

We evaluate the suggested approach to identifying attacks based on accuracy and F1-score. Since window size can impact the accuracy of the matrix profile, we used a range of window and subset sizes. Furthermore, to compute and interpret the matrix profile, we employ time series-based and index-based datasets as inputs to the suggested methods. Figures **Figure 7**, **Figure 8**, **Figure 9** and **Figure 10** demonstrate the outcomes of anomaly detection when evaluating our method with two chosen datasets, UNSW-NB15 and CICIoT2023, by varying windows and subsets. In Fig. 6 and 7, the values of windows are beginning with 3, followed by 10, and increases steadily by increments of 10 up to 160. The pattern then shifts to larger increments, beginning with 1000 and increasing by 200 up to 5200. Finally, the sequence continues with an increment of 800, starting from 6000 and culminating at 10000. The subset sizes begin at 100 and increase in increments of 100 up to 1000. After that, it progresses with larger increments of 1000, starting at 1000 and

increasing by 1000 until it reaches 10000. On the other side, in Fig. 8 and 9 the values of windows are starts with 3, then 10, and continues with increments of 10 up to 160. The pattern then progresses with varying increments: starting from 200, increasing by 50 up to 600, then by 100 up to 800, followed by 10 to 810. The sequence then jumps to larger increments, beginning with 1000, increasing by 200 up to 1200, then by 50 to 1250, followed by 220 to 1470, 220 again to 1690, and concluding with an increment of 220 to 1910. The subsets are beginning with 100, followed by 200, 300, 400, and 500, each increasing steadily by 100. The pattern then shifts to increments of 250, starting from 750 and progressing to 1000, 1250, 1500, 1750, and 2000. Finally, the sequence continues with an increment of 250, starting from 2500 and culminating at 3000. Our approach can identify anomalies by getting the best accuracy and F1-scores in the CICIoT2023 dataset, with 98.0% and 99.9%, respectively, with a window size of 1200 instances and a subset of 10000 instances. In the UNSW-NB15 dataset, we achieve the highest accuracy and F1-scores of 99.6% and 99.8%, respectively, with a window size of 1200 instances and a subset of 3000 instances. We started by using the smallest feasible window, which was size 3. We conducted incremental tests on larger sizes, assessing each option according to factors including accuracy, resource consumption, speed of anomaly detection, and suitability for the data capture intervals. Noteworthy is the fact that a larger window size makes the anomaly more visible. This highlights the significance of selecting an optimal window size that improves detection accuracy while reducing costs. The time delay to detect attacks is a critical factor for network operators. It helps to mitigate the spread of attacks and enables network operators to take timely action against various attacks. The detection time depends on the

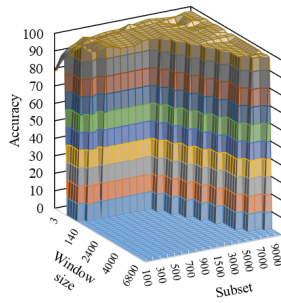


Figure 7. Effect of changing window and subset sizes on the accuracy of the CICIoT2023 dataset.

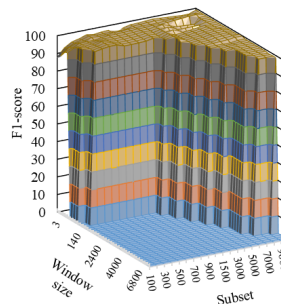


Figure 8. Effect of changing window and subset sizes on the F1-score for the CICIoT2023 dataset.

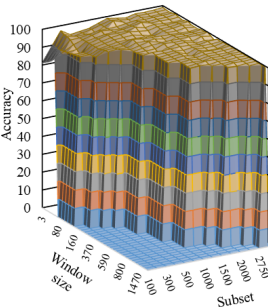


Figure 9. Effect of changing window and subset sizes on accuracy for the UNSW-NB15 dataset.

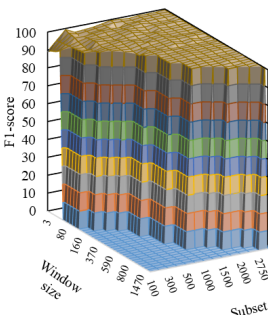


Figure 10. Effect of changing window and subset sizes on the F1-score for the UNSW-NB15 dataset.

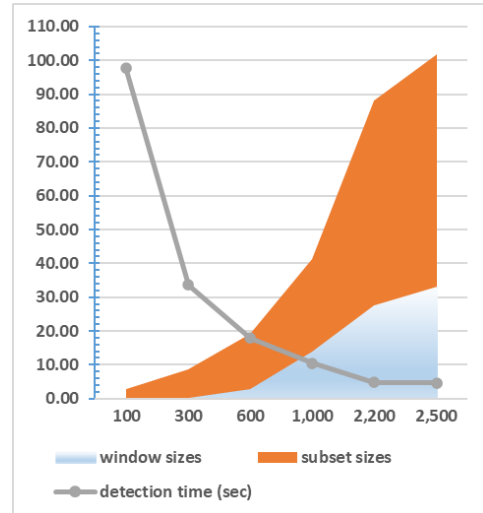


Figure 11. Correlations between detection time and window size

window size, the smaller the window, the longer the delay in the detection process. For instance, the proposed method takes approximately 4.57 seconds for the analysis of 1000 instances as a subset, and a window size of 100, this duration is to classify instances as normal or anomalies. The accuracy of the MP algorithm can be impacted by window size. Our results demonstrate that accuracy tends to decrease as the window size decreases. **Figure 11** Fig. 11 illustrates the relationship between subsets, window size and the detection time delay. To that end, we evaluate our approach using the selected features, window size and subset size in two cases where two different datasets were used. Next, we conducted a comparative analysis with other recent studies through two scenarios. Firstly, we compared our approach with studies that employed various strategies for anomaly detection and used the UNSW-NB15 dataset to evaluate their works. The evaluation process considered accuracy, F1-score, and feature count. Secondly, we compared the performance of our proposed approach with various machine-learning techniques using the CICIoT2023 dataset. The results demonstrate that our approach yields superior results.

Case 1: We conducted a comparative analysis with other recent studies that employed various strategies for anomaly detection using the UNSW-NB15 dataset. Despite not being the most recent dataset available, the UNSW-NB15 remains a trustworthy option for assessing anomaly detection models. The accuracy, F1 score, and feature count were considered during the evaluation process. The UNSW-NB15 was used to evaluate the performance of two methods, OCSVM and the autoencoder, for identifying anomalies in cloud networks Yasarathna and Munasinghe [2020]. The results showed that the OCSVM achieved an accuracy of 60.89%, while the autoencoder achieved an accuracy of 99.10% using only ten features.

Similarly, Manimurugan [2021] used the UNW-NB15 model to evaluate the performance of the proposed method based on improved naïve Bayes and PCA. The results showed the ability of the proposed system to detect anomalies with an accuracy of 92.48% and an F1 score of 91.64% using 12 features. Several machine learning techniques, including logistic regression (LR), K-nearest neighbors (KNN), decision tree (DT), extra tree classifier (ETC),

random forest (RF), gradient boosting (GB), and multilayer perceptron (MLP), have been applied to detect anomalies in cloud computing environments Parameswarappa *et al.* [2023]. The results showed that the RF model outperformed other machine learning techniques, achieving an accuracy of 97.68% and an F1-score of 97.68% using 20 features.

In contrast to the studies that employed various machine learning techniques for anomaly identification, our approach utilizes a multidimensional matrix profile—an innovative time-series analysis technique distinguished by its scalability and parameter-free characteristics. This approach is employed to identify anomalies in the underlying behavior of extracted features from cloud networks. An inherent advantage of the matrix profile is its ability to operate without training, effectively addressing the challenge of bias often associated with machine learning techniques. For instance, compared with the approach used by Al-Bakaa and Al-Musawi in Al-Bakaa and Al-Musawi [2021], which utilizes 19 characteristics to achieve a remarkable 99.96% accuracy and a 99.3% F1-score, our approach reached 99.6% accuracy and 99.8% F1-score using only nine features. Additional details of this comparison can be found in **Table 3**

Case 2: We evaluate further by applying our approach to the CICIOT2023 dataset. Unlike the UNSW-NB15 dataset, the CICIOT2023 dataset lacks support for the timestamp feature in its CSV file. Consequently, our analysis opted to utilize an index series rather than a time series. Moreover, we conducted a comparative analysis of the performance of our proposed approach with that of various machine learning techniques since the CICIOT2023 dataset has yet to be evaluated with other systems. The results demonstrate that our approach competes effectively with most machine learning methods. Importantly, our model requires no training. A detailed comparison between our proposed approach and machine learning tools applied to the CICIOT2023 dataset using ten features is highlighted in **Table 4**

6 Conclusion And Future Work

In this study, we introduced a novel approach that harnesses the power of a multidimensional matrix profile and the K-needle algorithm to detect anomalies in cloud networks. Our suggested approach demonstrated its ability to uncover anomalous patterns within multiple features extracted from network traffic streams without establishing a normal behavior profile. Our comprehensive evaluation, which was conducted using two diverse datasets, UNSW-NB15 and CICIOT2023, demonstrated the effectiveness of our approach in identifying anomalies in cloud networks. Our findings revealed impressive results, with an accuracy of 99.6% and an F1-score of 99.8% achieved using the timestamp-based analysis method for the UNSW-NB15 dataset. Furthermore, the index-based analysis method applied to the CICIOT2023 dataset yielded an accuracy of 98% and an outstanding F1-score of 99.9%. These results underscore the robustness and applicability of our proposed technique in enhancing cloud network security. Further investigations to extend our approach to real-time anomaly detection in cloud networks would be invaluable.

The study conclusively demonstrates SwinT's effectiveness in classifying COVID-19 from radiographic images, emphasizing its superior generalization capabilities across diverse datasets, a critical factor for real-world clinical applications.

Integration into Clinical Practice The integration of the SwinT into clinical settings should be judicious, with the model serving as an augmentation to, rather than a replacement for, human expertise. This ensures that diagnostic processes benefit from both advanced AI capabilities and professional medical judgment.

Challenges and Future Directions Addressing the challenges of data leakage and dataset diversity is critical for advancing the practical application of these models. Future research should aim to broaden the Swin Transformer's application scope to include other medical conditions and imaging modalities and integrate XAI techniques to enhance the transparency of its diagnostic processes.

Advancing Medical AI Research Building partnerships with medical institutions for access to comprehensive and varied datasets will be essential. These collaborations, along with ongoing model refinement based on real-world clinical feedback, will be crucial for the successful implementation of AI technologies in healthcare settings.

Concluding Reflections The Swin Transformer stands out as a transformative tool in medical diagnostics, capable of significantly enhancing the accuracy and efficiency of radiological assessments. As the medical AI field evolves, the thoughtful integration of such technologies into clinical practice is imperative, ensuring they align with ethical standards and contribute positively to patient care.

Declaration

Acknowledgements

We extend our profound gratitude to Mr. Euclides Carlos Pinto Neto, a faculty member at the University of New Brunswick (UnB), for his invaluable support in completing this work successfully

Funding

This manuscript did not receive any external funding.

Authors' Contributions

Amer Al-Mazrawe: Contributed to designing the methodology and the writing of the manuscript, conducted the analysis, and implemented the time series analysis techniques. Bahaa Al-Musawi: Conceptualized the study and contributed to editing the manuscript.

Competing interests

The authors declare that they have no competing interests.

Table 3. A comparison between the proposed approach and previous works that used the UNSW-NB15 dataset.

No.	Work	Method	Accuracy	F1-score	No. of Features
1.	Yasarathna and Munasinghe [2020]	Autoencoder	99.1%	–	10
2.	Manimurugan [2021]	Naïve Bayes and PCA	92.48%	91.64%	12
3.	Al-Bakaa and Al-Musawi [2021]	DT and RF classifiers	99.96%	99.3%	19
4.	Al-Bakaa and Al-Musawi [2022]	Quantification Analysis	96.71%	92.33%	1
5	Parameswarappa <i>et al.</i> [2023]	LR	92.85%	92.86%	20
		KNN	95.04%	95.05%	
		DT	96.33%	96.33%	
		ETC	97.53%	97.53%	
		RF	97.68%	97.68%	
		GB	95.85%	95.85%	
		MLP	96.39%	96.39%	
6.	Proposed Approach	matrix profile	99.6%	99.8%	9

Table 4. Using ten features, comparing the proposed method and machine learning techniques on the CICIoT2023 dataset.

No.	Methods	Accuracy	F1-score
1.	LR	98.0%	99.0%
2.	KNN	99.0%	99.0%
3.	DT	99.0%	99.0%
4.	ETC	99.0%	99.0%
5.	RF	99.0%	99.0%
6.	GB	99.0%	99.0%
7.	MLP	96.0%	98.0%
8.	Proposed Approach	98.0%	99.9%

Availability of data and materials

Appendices A and B present our efforts to evaluate the proposed method using the UNSW-NB15 and CICIoT2023 datasets to achieve optimal accuracy and F1 scores.

References

- Agrawal, B., Wiktorski, T., and Rong, C. (2016). Adaptive anomaly detection in cloud using robust and scalable principal component analysis. In *2016 15th international symposium on parallel and distributed computing (ISPDC)*, pages 100–106. IEEE. DOI: 10.1109/ISPDC.2016.22.
- Al-Bakaa, A. and Al-Musawi, B. (2021). Improving the performance of intrusion detection system through finding the most effective features. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, pages 1–9. IEEE. DOI: 10.1109/ICOTEN52080.2021.9493564.
- Al-Bakaa, A. and Al-Musawi, B. (2022). A new intrusion detection system based on using nonlinear statistical analysis and features selection techniques. *Computers & Security*, 122:102906. DOI: 10.1016/j.cose.2022.102906.
- Alzahrani, M. A., Alzahrani, A. M., and Siddiqui, M. S. (2022). Detecting ddos attacks in iot-based networks using matrix profile. *Applied Sciences*, 12(16):8294. DOI: 10.3390/app12168294.
- Baron, H. (2022). Report by cloud security alliance. Available at: <https://cloudsecurityalliance.org/>.
- Cheema, P., Alamdari, M. M., Vio, G., Azizi, L., and Luo, S. (2023). On the use of matrix profiles and optimal transport theory for multivariate time series anomaly detection within structural health monitoring. *Mechanical Systems and Signal Processing*, 204:110797. DOI: 10.1016/j.ymssp.2023.110797.
- De Paepe, D., Haute, S. V., Steenwinckel, B., De Turck, F., Ongena, F., Janssens, O., and Van Hoecke, S. (2020). A generalized matrix profile framework with support for contextual series analysis. *Engineering Applications of Artificial Intelligence*, 90:103487. DOI: 10.1016/j.engappai.2020.103487.
- Ding, N., Gao, H., Bu, H., Ma, H., and Si, H. (2018). Multivariate-time-series-driven real-time anomaly detection based on bayesian network. *Sensors*, 18(10):3367. DOI: 10.3390/s18103367.
- Firouzi, F., Farahani, B., and Marinšek, A. (2022). The convergence and interplay of edge, fog, and cloud in the ai-driven internet of things (iot). *Information Systems*, 107:101840. DOI: 10.1016/j.is.2021.101840.
- Garg, S., Kaur, K., Kumar, N., Kaddoum, G., Zomaya, A. Y., and Ranjan, R. (2019). A hybrid deep learning-based model for anomaly detection in cloud datacenter networks. *IEEE Transactions on Network and Service Management*, 16(3):924–935. DOI: 10.1109/TNSM.2019.2927886.
- Huang, C., Min, G., Wu, Y., Ying, Y., Pei, K., and Xiang, Z. (2017). Time series anomaly detection for trustworthy services in cloud computing systems. *IEEE Transactions on Big Data*, 8(1):60–72. DOI: 10.1109/TB-DATA.2017.2711039.
- Jacobs, A. S., Beltiukov, R., Willinger, W., Ferreira, R. A., Gupta, A., and Granville, L. Z. (2022). Ai/ml for network security: The emperor has no clothes. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer*

- and Communications Security, pages 1537–1551. DOI: 10.1145/3548606.3560609.
- Liberti, L., Lavor, C., Maculan, N., and Mucherino, A. (2014). Euclidean distance geometry and applications. *SIAM review*, 56(1):3–69. DOI: 10.1137/120875909.
- Liu, H., Li, J., and Gu, D. (2020). Understanding the security of app-in-the-middle iot. *Computers & Security*, 97:102000. DOI: 10.1016/j.cose.2020.102000.
- Lou, P., Yang, Y., and Yan, J. (2019). An anomaly detection method for cloud service platform. In *Proceedings of the 2019 4th International Conference on Machine Learning Technologies*, pages 70–75. DOI: 10.1145/3340997.3341005.
- Madrid, F., Imani, S., Mercer, R., Zimmerman, Z., Shakibay, N., and Keogh, E. (2019). Matrix profile xx: Finding and visualizing time series motifs of all lengths using the matrix profile. In *2019 IEEE International Conference on Big Knowledge (ICBK)*, pages 175–182. IEEE. DOI: 10.1109/ICBK.2019.00031.
- Manimurugan, S. (2021). Iot-fog-cloud model for anomaly detection using improved naïve bayes and principal component analysis. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–10. DOI: 10.1007/s12652-020-02723-3.
- Moustafa, N. and Slay, J. (2015). UNSW-NB15: A comprehensive dataset for network intrusion detection systems (unsw-nb15 network dataset). In *2015 Military Communications and Information Systems Conference (MilCIS)*, pages 1–6. IEEE. DOI: 10.1109/MilCIS.2015.7348942.
- Neto, E. C. P., Dadkhah, S., Ferreira, R., Zohourian, A., Lu, R., and Ghorbani, A. A. (2023). CICIOT2023: A real-time dataset and benchmark for large-scale attacks in iot environment. *Sensors*. DOI: 10.3390/s23135941.
- Nzanzu, V. P., Adetiba, E., Badejo, J. A., Molo, M. J., Takenga, C., Noma-Osaghae, E., and Suraju, S. (2022). Monitoring and resource management taxonomy in interconnected cloud infrastructures: a survey. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 20(2):279–295. DOI: 10.12928/telkomnika.v20i2.20503.
- Parameswarappa, P., Shah, T., and Lanke, G. R. (2023). A machine learning-based approach for anomaly detection for secure cloud computing environments. In *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pages 931–940. IEEE. DOI: 10.1109/IDCIoT56793.2023.10053518.
- Saljoughi, A. S., Mehrvarz, M., and Mirvaziri, H. (2017). Attacks and intrusion detection in cloud computing using neural networks and particle swarm optimization algorithms. *Emerging Science Journal*, 1(4):179–191. DOI: 10.28991/ijse-01120.
- Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). Finding a “kneedle” in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, pages 166–171. IEEE. DOI: 10.1109/ICDCSW.2011.20.
- Schmidt, F., Suri-Payer, F., Gulenko, A., Wallschläger, M., Acker, A., and Kao, O. (2018). Unsupervised anomaly event detection for cloud monitoring using on-line arima. In *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*, pages 71–76. IEEE. DOI: 10.1109/UCC-Companion.2018.00037.
- Scott, B. A., Johnstone, M. N., Szewczyk, P., and Richardson, S. (2024). Matrix profile data mining for bgp anomaly detection. *Computer Networks*, 242:110257. DOI: 10.1016/j.comnet.2024.110257.
- Syed, H. J., Gani, A., Ahmad, R. W., Khan, M. K., and Ahmed, A. I. A. (2017). Cloud monitoring: A review, taxonomy, and open research issues. *Journal of Network and Computer Applications*, 98:11–26. DOI: 10.1016/j.jnca.2017.08.021.
- Yasarathna, T. L. and Munasinghe, L. (2020). Anomaly detection in cloud network data. In *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, pages 62–67. IEEE. DOI: 10.1109/SCSE49731.2020.9313014.
- Yeh, C.-C. M., Kavantzaz, N., and Keogh, E. (2017a). Matrix profile vi: Meaningful multidimensional motif discovery. In *2017 IEEE international conference on data mining (ICDM)*, pages 565–574. IEEE. DOI: 10.1109/ICDM.2017.66.
- Yeh, C. C. M., Kavantzaz, N., and Keogh, E. (2017b). Matrix profile vi: Meaningful multidimensional motif discovery. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 565–574. IEEE. DOI: 10.1109/ICDM.2017.66.
- Yeh, C. C. M., Van Herle, H., and Keogh, E. (2016a). Matrix profile iii: the matrix profile allows visualization of salient subsequences in massive time series. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 579–588. IEEE. DOI: 10.1109/ICDM.2016.0069.
- Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., and Keogh, E. (2016b). Matrix profile i: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 1317–1322. DOI: 10.1109/ICDM.2016.0179.
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., and Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2):56–70. DOI: 10.38094/jastt1224.
- Zhang, X., Kim, J., Lin, Q., Lim, K., Kanaujia, S. O., Xu, Y., and Mishra, P. (2019). Cross-dataset time series anomaly detection for cloud systems. In *2019 USENIX Annual Technical Conference (USENIX ATC 19)*, pages 1063–1076. Available at: <https://www.usenix.org/conference/atc19/technical-sessions>.
- Zhu, Y., Yeh, C. C. M., Zimmerman, Z., Kamgar, K., and Keogh, E. (2018a). Matrix profile xi: Scrimp++: time series motif discovery at interactive speeds. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 837–846. IEEE. DOI: 10.1109/ICDM.2018.00099.
- Zhu, Y., Zimmerman, Z., Senobari, N. S., Yeh, C. C. M., Funning, G., Mueen, A., and Keogh, E. (2016). Matrix profile ii: Exploiting a novel algorithm and gpus to break the one hundred million barrier for time series motifs and joins. In *2016 IEEE 16th International Confer-*

ence on Data Mining (ICDM), pages 739–748. IEEE. DOI: 10.1109/ICDM.2016.0085.

- Zhu, Y., Zimmerman, Z., Shakibay Senobari, N., Yeh, C.-C. M., Funning, G., Mueen, A., Brisk, P., and Keogh, E. (2018b). Exploiting a novel algorithm and gpus to break the ten quadrillion pairwise comparisons barrier for time series motifs and joins. *Knowledge and Information Systems*, 54:203–236. DOI: 10.1007/s10115-017-1138-x.
- Zoghi, Z. (2020). *Ensemble Classifier Design and Performance Evaluation for Intrusion Detection Using UNSW-NB15 Dataset*. PhD thesis, The University of Toledo. Available at: <https://www.proquest.com/openview/7c41b65580bde34353f4a2d61523ab3f/1?pq-origsite=gscholar&cbl=18750&diss=y>.