RESEARCH ARTICLE OPEN ACCESS

Self-Attention Mechanisms for Detecting Anomalies in Encrypted Network Traffic: A Systematic Review

Sabuhee, Pratik Buchke
Department of Computer Science & Engineering
Oriental Institute of Science & Technology, Bhopal

ABSTRACT

The growing use of encryption in network communications has made it a critical cybersecurity problem to identify abnormalities in encrypted information. Conventional anomaly detection techniques often have trouble decrypting encrypted communication, which results in inefficiencies and privacy issues. Self-attention mechanisms have shown impressive ability in recent years to capture intricate patterns and relationships within sequential data, especially when used in transformer-based models. The use of self-attention techniques to identify abnormalities in encrypted network data is examined in this systematic study. We investigate many deep learning architectures that use self-attention for traffic analysis, such as Transformer, BERT, and its variants. We also highlight current developments in model optimisation and performance assessment, and we address important issues including scalability, real-time detection, and feature extraction from encrypted data. This review sheds light on the efficacy, constraints, and potential avenues for future study of self-attention-based anomaly detection in encrypted network settings by combining the results of current investigations.

Keywords: Self-Attention Mechanism, Anomaly Detection, Encrypted Network Traffic, Deep Learning, Traffic Analysis, Feature Extraction, Real-Time Detection, Machine Learning

I. INTRODUCTION

Cybersecurity has grown to be a top issue for both people and organisations due to the quick development of digital communication and the growing dependence on networked technologies. A series of unauthorised operations carried out on a computer or network with the intent to compromise stability, steal confidential information, or alter data is called an intrusion [1]. old security systems like firewalls and rule-based detection mechanisms often find it difficult to detect complex assaults as cyber threats change, particularly in encrypted settings where old approaches are unable to adequately analyse network data. Using software or hardware-based technologies, intrusion detection is essential for spotting malicious activity occurring inside a system or network. Intrusion Detection Systems (IDS) are made to identify many types of assaults, such as malware penetration, host-based intrusions including illegal logins, and denial-of-service (DoS) attacks. IDSs dynamically monitor network traffic and system records to spot suspicious activity, in contrast to firewalls, which serve as static defence measures by preventing unwanted access. In order to identify possible threats, IDS uses audit logs from host systems and network packet capture in promiscuous mode. The IDS creates alerts when it detects malicious behaviour, alerting stakeholders and system administrators to take remedial action.

Although encryption improves data security and privacy, intrusion detection becomes more difficult as its usage in network communications grows. Conventional techniques depend on looking for irregularities in packet payloads, but payload visibility is limited in encrypted data, making it difficult to distinguish between malicious and legal activities. Modern intrusion detection systems use sophisticated machine learning methods, including self-attention processes, to examine encrypted network traffic in order to overcome this difficulty. Self-attention mechanisms have shown remarkable efficacy in processing sequential data and are often used in transformer-based models such as BERT. They are appropriate for detecting irregularities in encrypted settings because they are able to capture contextual linkages and long-range interdependence in network data. Without depending on deep packet inspection, the system may learn to identify departures from typical traffic patterns by integrating self-attention into IDS frameworks. Contextual comprehension of network flow sequences, scalability in handling high traffic volumes, and decreased false positive rates by concentrating on the most relevant aspects of network behaviour are the main advantages of self-attention techniques.

Traditional intrusion detection methods have limits as cyber threats become more complex, especially in encrypted contexts where direct inspection is impractical. One interesting method for enhancing

anomaly detection capabilities in encrypted communication is to include self-attention techniques into IDS systems. Without sacrificing data privacy, these deep learning models improve security threat identification's precision, scalability, and effectiveness. In order to build a strong, AI-driven cybersecurity ecosystem, future research should concentrate on refining self-attention-based models for real-time performance, enhancing their capacity to adapt to new attack methods, and combining them with other security measures.

II. BACKGROUND AND MOTIVATION

As services become more digitally connected and encrypted conversations grow more common, maintaining cybersecurity has become a major concern. Particularly in encrypted settings, traditional network security measures like firewalls & rule-based intrusion detection systems (IDS) have not been able to identify complex cyberthreats. By prohibiting unwanted access to sent data, encryption improves data privacy. However, it also diffuses network traffic, making it more difficult for traditional intrusion detection systems to discern between benign and malevolent activity. By examining system logs and network traffic, intrusion detection systems are essential for spotting security risks. However, deep packet inspection (DPI) and signature-based detection are key components of classic IDS techniques, and they lose their efficacy when traffic is encrypted. Because of this restriction, sophisticated methods that can identify irregularities in encrypted network traffic without compromising privacy are now required.

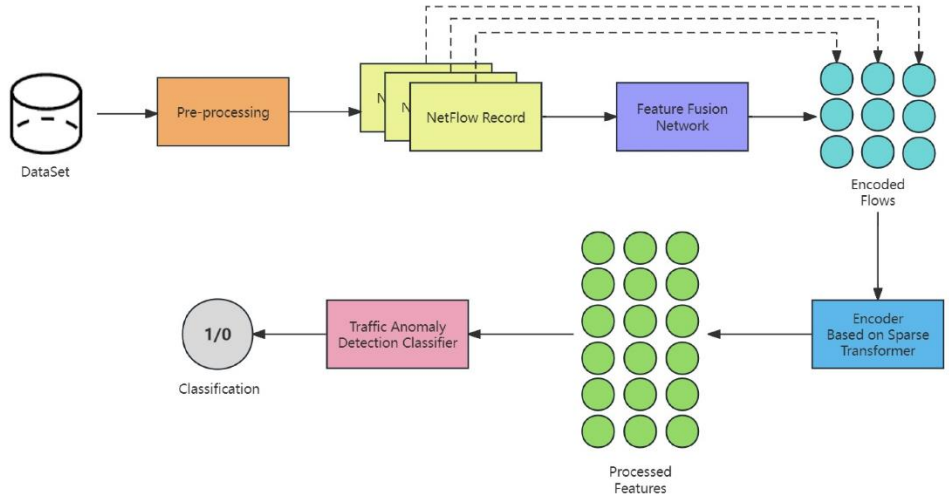

Figure 1 Anomaly Detection

Figure 1 illustrates how anomaly detection works. Strong techniques for network traffic analysis have been made possible by recent developments in deep learning and artificial intelligence (AI). Among them, self-attention mechanisms have shown exceptional ability in identifying intricate patterns in sequential data, especially those used in transformer-based models such as BERT. Self-attention models are well suited for examining encrypted traffic patterns because, in contrast to conventional machine learning methods, they are capable of efficiently understanding contextual relationships. The purpose of this research is to investigate how well self-attention processes perform in encrypted network contexts for anomaly detection. This strategy seeks to solve important issues including feature extraction from encrypted data, scalability, and real-time detection by using deep learning architectures like Transformers. A comprehensive review of current approaches will provide light on their advantages, disadvantages, and room for development, directing future studies to create strong cybersecurity frameworks powered by AI.

III. EVOLUTION OF ANOMALY DETECTION IN NETWORK SECURITY

- **Conventional Intrusion Detection Systems (IDS)**

To find cyberthreats, early intrusion detection systems used rule-based and signature-based techniques. These systems functioned by comparing prepared attack signatures with patterns of network traffic. Even while

they worked well for established threats, they had trouble spotting fresh or developing assaults, particularly in encrypted settings.

- **Methods of Statistical and Machine Learning**

Researchers developed statistical & machine learning-based intrusion detection systems (IDS) to overcome the drawbacks of rule-based detection. IDS was able to identify abnormalities based on departures from typical network behaviour by using techniques including support vector machines (SVM), decision trees, and clustering. Nevertheless, these techniques often needed a lot of feature engineering and had trouble processing high-dimensional traffic data.

- **Analysing Network Traffic using Deep Learning**

Neural networks such as Convolutional Neural Networks (CNNs) & Recurrent Neural Networks (RNNs) were used for anomaly detection as deep learning gained popularity. Although feature extraction & pattern identification were enhanced by these models, their ability to effectively handle sequential data and long-range interdependence was limited.

- **Transformer Models and Self-Attention Mechanisms**

Self-attention mechanisms were introduced by recent developments, especially Transformer-based models like Vision Transformers (ViTs) and BERT. Natural language processing (NLP) was transformed by these models, which have subsequently been modified for network traffic monitoring. Transformers are quite good at detecting anomalies in encrypted communication because, in contrast to CNNs and RNNs, they are very good at capturing global dependencies.

- **Detection and Optimisation in Real Time**

Transformer-based IDS model optimisation for real-time performance is the main focus of current research. The speed and accuracy of detection are improved by innovations like sparse attention, effective network flow tokenisation, and hybrid AI-driven security frameworks. These advancements open the door to cybersecurity solutions that are both scalable and private.

IV. SELF-ATTENTION-BASED DETECTION ALGORITHM

Anomaly detection in encrypted network data has been transformed by self-attention methods, especially those found in Transformer-based systems. A synopsis of the main models that use self-attention to improve network security is provided below.

1. Models Based on Transformers

Self-attention, which enables deep learning architectures to capture long-range relationships between features, was first presented by transformer-based models. Transformers, as opposed to conventional convolutional neural networks (CNNs), analyse all of the information at once, allowing for global context awareness. The fundamental process, referred to as multi-head self-attention, helps the model comprehend connections across a picture by giving distinct input characteristics varying degrees of priority. Originally intended for natural language processing, the Transformer model was subsequently modified for computer vision applications by [2]. Modern self-attention-based object identification frameworks were made possible by this change. Transformers make systems more adaptable and scalable by doing away with the locality restrictions that convolutional layers impose. Transformers are very good at recognising things in crowded and complicated situations because they can dynamically reweight characteristics depending on context. Because of their adaptability, several sophisticated detection algorithms have been developed, making self-attention a crucial part of contemporary computer vision research.

2. Transformer of Vision (ViT)

One innovative approach for image processing is the Vision Transformer (ViT), which uses self-attention mechanisms in lieu of convolutional layers. ViT splits pictures into fixed-size patches and handles each patch as an input token rather than processing whole images as separate entities. By using this technique, the transformer may better represent features by modelling long-range relationships between various picture areas. ViT employs multi-head self-attention to examine the links between patches and positional encodings to preserve spatial information. ViT improves accuracy, particularly in large-scale datasets, by capturing global picture context in a single forward pass, in contrast to typical CNNs that depend on local receptive fields. For best results, ViT needs a substantial amount of pretraining on large datasets like ImageNet. Self-attention-based architectures may successfully replace traditional convolutional networks in computer vision, as shown by ViT's improved performance in picture classification and object recognition tasks, despite its high computational cost [3].

3. DETR (Detection Transformer)

A novel object detection model called the Detection Transformer (DETR) does away with the necessity for conventional anchor boxes and post-processing methods like non-maximum suppression (NMS). Rather, DETR directly predicts item positions using bipartite matching and a set-based global loss function. The model's transformer encoder-decoder architecture comes after a convolutional backbone for feature extraction. DETR is quite good at identifying overlapping things because of its self-attention mechanism, which aids in the analysis of connections between objects. DETR eliminates the requirement for heuristic-based region recommendations, which are employed in conventional object detectors like Faster R-CNN, by processing pictures as a whole. DETR streamlines the object detection pipeline, but its main flaw is that it requires a lot of processing power and has delayed convergence during training. Notwithstanding this drawback, DETR represents a major breakthrough in transformer-based detection due to its capacity to manage intricate object interactions at different sizes, which has influenced the creation of more effective self-attention-based models [4].

4. The Swing Transformer

By adding shifted windows and hierarchical feature representation, the Swin Transformer improves the Vision Transformer (ViT). Swin Transformer preserves multi-scale feature representation by dividing pictures into increasingly smaller windows, in contrast to ViT, which analyses full images with a set patch size. Swin Transformer can effectively capture fine-grained information while preserving computing efficiency because to its hierarchical method. By improving information flow between adjacent areas, the shifted window mechanism lessens the loss of spatial context. Swin Transformer gets around ViT's drawbacks, which include its inability to handle high-resolution visuals, by including self-attention at many scales. Swin Transformer is also more memory and compute efficient, which makes it appropriate for situations involving large-scale object identification and segmentation. It has shown to be a formidable substitute for conventional CNN-based backbones because to its exceptional performance in benchmarks like as COCO and ImageNet, underscoring the promise of hierarchical self-attention mechanisms in visual identification tasks. [5]

5. R-CNN Sparse

An effective object identification technique called Sparse R-CNN eliminates the need for laborious candidate selection by optimising region recommendations via self-attention processes. Sparse R-CNN learns to dynamically provide a fixed set of high-quality suggestions, in contrast to typical R-CNN models that produce hundreds of region proposals. Multiple self-attention iterations are used to improve these suggestions, allowing the model to concentrate on the most relevant item locations. Sparse R-CNN maintains good detection accuracy while drastically lowering computational cost by doing away with the need for dense region proposal networks. This method is perfect for real-time applications as it increases inference speed as well. Sparse R-CNN may recognise objects more effectively in complicated and congested situations because it can adaptively modify recommendations utilising self-attention. Sparse R-CNN is one of the most effective self-attention-based object detection frameworks on the market because it provides a simplified pipeline with less redundant computations than standard detectors [6].

6. Pyramid Vision Transformer

The Pyramid Vision Transformer (PVT) introduces a hierarchical feature extraction method to improve self-attention-based detection. PVT allows multi-scale feature learning by gradually shrinking the feature maps, in contrast to ViT, which analyses all picture patches at a single scale. PVT is more suited for dense prediction applications like object identification and segmentation because of its pyramid structure, which resembles conventional CNN backbones. By including spatial-reduction attention, the model lowers computational complexity without sacrificing its capacity to represent global interdependence. PVT is a great option for large-scale object identification jobs because it effectively strikes a balance between efficiency and accuracy. It has been used in many vision applications because to its superiority over traditional transformers in handling high-resolution pictures and detecting minute objects. PVT provides an ideal solution for real-world situations where computing economy and performance are crucial by combining progressive feature reduction with self-attention [7].

7. Deformable DETR

By addressing the original DETR model's delayed convergence and computational inefficiencies, Deformable DETR enhances it. Deformable DETR proposes deformable attention modules that adaptively concentrate on the most significant characteristics, rather than evenly attentive to all parts of the picture. Particularly for tiny and closely spaced objects, this focused attention method greatly increases object identification accuracy and speeds up training. With sparse spatial sampling, Deformable DETR improves performance while preserving DETR's benefits,

including set-based prediction and end-to-end training. Additionally, the model incorporates multi-scale feature fusion, which improves its ability to handle objects of different sizes. Deformable DETR uses less computing power than DETR while achieving better detection results and quicker convergence. In real-world applications, where recognising tiny and obstructed objects is critical, its dynamic attention zone adjustment makes it very successful [8].

V. DATASETS AND FEATURES

Both private and publicly accessible datasets are essential for self-attention-based model training and validation. Because of the labelled network traffic data in these datasets, supervised learning techniques for anomaly detection are made possible. Some of the most popular datasets in this field are listed below:

1. CICIDS 2017 (Canadian Institute for Cybersecurity Intrusion Detection System 2017)

Due to its ability to replicate real-world network traffic situations, including both typical and attack patterns, the CICIDS 2017 dataset is often utilised in cybersecurity research. Numerous cyberattack methods, including Distributed Denial-of-Service (DDoS), Brute Force, Botnet, and SQL Injection assaults, are included in the dataset. It is ideal for training anomaly detection models since it offers both packet-based and flow-based information. CICIDS 2017 is a useful dataset for assessing how well self-attention systems handle encrypted data since it contains both encrypted and non-encrypted traffic.

2. UNSW-NB15

A contemporary network dataset called UNSW-NB15 was created to overcome the shortcomings of earlier datasets like as KDD99 and NSL-KDD. It is appropriate for study on intrusion detection as it records modern cyberattacks such as shellcode, exploits, fuzzers, and backdoors. There are 49 characteristics in the collection, which include time-related statistics, fundamental connection metrics, and other statistical qualities. It allows both supervised and semi-supervised learning techniques for anomaly detection since it has labelled data. UNSW-NB15 is a popular dataset for deep learning models because of its realistic network traffic distribution and range of attack methods.

3. CTU-13 Botnet Dataset

The purpose of the CTU-13 Botnet Dataset is to identify botnet activity in network traffic. In addition to regular and background network data, it includes traffic that was recorded from other botnet families, including Zeus, Virut, and Neris. The dataset is helpful for time-series analysis and anomaly identification since it offers information at both the packet and flow levels. Since botnets often employ encryption to avoid detection, CTU-13 is a useful dataset for evaluating how well self-attention-based algorithms detect abnormalities in encrypted traffic that are associated with botnets.

4. ISCX VPN-nonVPN Dataset

Distinguishing between Virtual Private Network (VPN) and non-VPN traffic is the main goal of the ISCX VPN-nonVPN Dataset. Given that VPNs encrypt data, rendering conventional signature-based techniques useless, this dataset is crucial for studies in privacy-aware intrusion detection. Researchers may create models that can categorise encrypted traffic using statistical and flow-based characteristics instead of payload inspection because to the dataset's inclusion of a variety of application-layer protocols operating across both VPN and non-VPN connections.

5. MAWI Dataset

An ISP-level network's actual internet traffic traces make up the MAWI (Measurement and Analysis on the WIDE Internet) Dataset. In contrast to simulated datasets, MAWI offers real-world traffic patterns that include both typical and unusual network activity. It is especially helpful for analysing time-series anomaly detection models, identifying zero-day assaults, and researching long-term traffic patterns. It aids in testing the generalisability of deep learning models in practical situations since it incorporates encrypted network traffic.

6. TON_IoT Dataset

With its integration of network traffic, system logs, and telemetry data from IoT devices, the TON_IoT Dataset is specifically designed for Internet of Things (IoT) networks. It includes a variety of IoT hacks, including efforts at data exfiltration, Mirai botnet infections, and misuse of MQTT. The dataset is useful for researching cross-domain anomaly detection, which requires analysing both conventional network and Internet of Things traffic simultaneously. TON_IoT is very useful for creating self-attention-based models that can manage contemporary cybersecurity concerns since encrypted IoT traffic is present.

VI. RELATED WORK

Finding abnormalities has become a crucial cybersecurity problem as network information is increasingly encrypted. Because payload information is not readily available, encrypted data often presents challenges for traditional anomaly detection

techniques. Patterns and irregularities in encrypted communication may be detected without decryption thanks to recent developments in deep learning, especially self-attention methods. This systematic review examines the literature on self-attention-based anomaly detection models, evaluates their efficacy, and identifies present issues and potential avenues for further study in this developing area.

Wanting Hu et. al. [9] To detect and reduce security risks, network traffic anomaly detection is crucial. The intricacy and encryption of network communication provide difficulties for conventional detection techniques. In order to increase the accuracy of anomaly detection, recent research has concentrated on deep learning approaches, including feature engineering and self-attention processes. The significance of feature extraction in creating high-quality datasets has been underlined by several research. Through the integration of feature selection criteria from many sources, researchers have improved the representation of network traffic in datasets such as UNSW-NB15. Using machine learning models like XGBoost, the DNTAD dataset, which was created using sophisticated feature-processing methods, has shown enhanced anomaly detection performance. The capacity of deep learning models, especially LSTM-based techniques, to identify time-series relationships in network data has made them popular. By giving traffic variables varying degrees of priority, the self-attention mechanism, when combined with LSTM, improves detection accuracy and enables models to discover more significant patterns. Self-attention-enhanced LSTM models perform better than conventional techniques in identifying abnormalities, according to ablation experiments. This study looks at previous studies on self-attention methods for encrypted traffic anomaly detection, emphasising developments in deep learning models and dataset generation as well as how well they handle intricate traffic patterns.

MS Alshehri et. al [10] The Industrial Internet of Things (IIoT) is particularly susceptible to security issues and privacy abuses because of the integration of several smart devices and communication protocols. Several studies have looked at intrusion detection systems (IDS) that use machine learning (ML) and deep learning (DL) to identify hostile behaviour in IIoT networks. Unfortunately, since repetitive data artificially inflates model performance and leads to poor generalisation, imbalanced datasets often present issues for existing models. Moreover, the presence of duplicated data from other classes reduces detection accuracy. To get around these limitations, researchers have examined self-attention techniques and deep convolutional neural networks (DCNNs) for enhanced anomaly identification. Recent studies propose models that integrate self-attention with DCNNs to improve the feature extraction process and capture complex patterns in network traffic data. To ensure dependable training outcomes and remove redundancy, advanced data-cleaning techniques have also been created. Performance evaluations on datasets like IoTID20 and Edge-IoTset demonstrate that self-attention-based models perform better than traditional ML and DL methods. Comparative evaluations with prior studies demonstrate the advantages of integrating self-attention into deep learning systems for IoT security. This article provides an overview of recent advancements and discusses how they have enhanced IoT network monitoring and intrusion detection.

Table 1 Related Work on Intrusion Detection

Study	Focus Area	Methodology	Key Findings	Limitations
Anderson et.al [10]	Intrusion Detection Concept	Introduced the concept of anomaly detection	Laid the foundation for intrusion detection systems (IDS)	Lacked modern ML/DL advancements
Chowdhury et al. [11]	Machine Learning for Intrusion Detection	Hybrid ML model using simulated annealing & SVM	Improved classification of network incursions	Limited generalization to advanced cyber threats
Yang Min et al. [12]	Semi-Supervised Intrusion Detection	Multi-level feature extraction & semi-	Enhanced detection with limited labeled data	Still vulnerable to advanced evasion techniques

		supervised learning		
Tuor et al. [13]	Deep Learning for Anomaly Detection	Deep autoencoders for unsupervised learning	Demonstrated efficiency in detecting temporal patterns	Struggled with real-time application
Yan et.al. [14]	CNN-based Intrusion Detection	CNN + Generative Adversarial Network (GAN)	Improved detection of synthetic attack patterns	High computational cost
Wu et al. [15]	Transformer-Based IDS	RTIDS with position-embedding technology	Improved sequence-based anomaly detection	Requires large datasets for training
Alkhatib et al. [16]	Transformer for Network Security	BERT-based sequence learning for CAN	Effective learning of sequential patterns in network traffic	Limited application beyond CAN systems
Vaswani et al. [17]	Transformer Architecture	Self-attention mechanism in deep learning	Revolutionized sequential data processing	Lacks direct application to network security
Dosovitskiy et al. [18]	BERT Model	Pre-trained transformer for text-based learning	Achieved state-of-the-art NLP performance	Not directly applicable to network anomaly detection
S. Ullah et. al [19]	IoT Network Monitoring & Malicious Activity Detection	Deep Convolutional Neural Network (DCNN)	Achieved **77**.55% accuracy using IoTID20 dataset	Lower accuracy compared to hybrid models
A. Khacha et. al [20]	IoT Intrusion Detection	CNN + LSTM Hybrid Model	98.69% accuracy using Edge-IIoTset dataset	Focuses mainly on IoT sub-categories, lacks generalization
P. Dini et. al. [21]	IIoT Intrusion Detection	Extreme Learning Machine, SVM, & Rule-Based IDS	High accuracy (97.83% KDD99, 96.59% UNSW-NB15, 92.54% CSE-CIC-IDS-2018, 97.27% Edge-IIoTset)	Requires multiple datasets for evaluation
A. A. Alashhab et. al. [22]	SDN-Enabled IoT Cyberattack Detection	LSTM Model	98.88% accuracy in LDDoS detection	Focuses only on SDN-enabled networks
D. Javeed et. al. [23]	Smart Agriculture	B-GRU + LSTM Hybrid Model	**98.32% accuracy** in DDoS detection	Limited application beyond smart

	Cyberattack Detection		in Edge-IIoTset dataset	agriculture networks

VII. EVALUATION METRICS

Several common performance criteria are used to assess binary classifier efficacy. These metrics assess how well the classifier can distinguish between attack and typical cases. The following are the main terms used in categorisation evaluation:

- True Positive (TP): The number of attack flows correctly classified as attacks.
- False Positive (FP): The number of benign flows incorrectly classified as attacks.
- True Negative (TN): The number of benign flows correctly classified as normal.
- False Negative (FN): The number of attack flows incorrectly classified as normal.

These numbers are used to compute several performance measures that evaluate the accuracy, precision, recall, and general efficacy of classifiers.

Accuracy

By calculating the percentage of properly categorised examples (attack and normal) compared to the total number of instances, accuracy (ACC) indicates the classifier's overall correctness. It is described as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

When attack and normal classes are balanced, accuracy is helpful. However, accuracy by itself may not be enough to assess the classifier's efficacy in situations when the dataset is unbalanced.

Precision

Precision quantifies the proportion of correctly classified attacks out of all instances predicted as attacks. It is calculated as:

$$Precision = \frac{TP}{TP + FP}$$

A high precision value indicates that the classifier minimizes false positives, ensuring that normal traffic is not mistakenly classified as an attack. This is particularly important in security applications where false alarms can lead to unnecessary investigations.

Recall

Recall, also known as Sensitivity or True Positive Rate (TPR), measures the classifier's ability to correctly identify actual attack instances. It is given by

$$Recall = \frac{TP}{TP + FN}$$

A higher recall value indicates that the classifier is effective in detecting attacks, minimizing the risk of missed threats. However, improving recall may sometimes lead to an increase in false positives.

F1-Score

The F1-score provides a balanced measure of classifier performance by considering both precision and recall. It is the harmonic mean of precision and recall and is defined as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1-score is particularly useful when dealing with imbalanced datasets, as it ensures that both false positives and false negatives are accounted for in performance evaluation.

VIII. CHALLENGES AND LIMITATIONS

Self-attention-based detection algorithms have a number of issues and restrictions that affect their effectiveness, scalability, and practicality despite their remarkable performance. These difficulties are brought on by practical deployment limitations, data reliance, and computational complexity.

1. Expensive computation

The high computational demand of self-attention mechanisms is one of their main problems, particularly in models such as Transformers and Vision Transformers (ViT). Self-attention affects every input token at once, in contrast to conventional convolutional neural networks (CNNs), resulting in quadratic complexity with respect to input size. Because of this, these models are resource-intensive and need strong GPUs or TPUs for both training and inference, which restricts their use in real-time applications and low-power devices.

2. High Requirements for Data

For self-attention-based models to be trained effectively, large datasets are usually needed. For instance, since Vision Transformers (ViTs) lack the inductive biases that CNNs are naturally equipped with, such as locality and translation invariance, they perform poorly when trained on limited datasets. In order to get around this, models often use pretraining with proprietary datasets or large-scale datasets like ImageNet, which adds cost and time to the training process.

3. Training Time and Slow Convergence

Self-attention-based models often exhibit slower rates of convergence when compared to conventional object identification techniques. Models like as DETR (Detection Transformer), for instance, need a lot of training time to perform competitively. The necessity for significant global feature interactions is the cause of this poor training speed, which might make it difficult to employ these models practically in applications that need to be deployed quickly.

4. Having Trouble Managing Small Items

Self-attention mechanisms often have trouble identifying little items in vast pictures, even when they are good at capturing long-range relationships. Fine-grained local characteristics may be diluted by the global nature of self-attention, making it challenging to recognise tiny or closely spaced things. By implementing hierarchical attention and adaptable feature focussing, certain architectures—like Swin Transformers and Deformable DETR—try to solve this problem, although difficulties still exist.

5. Limited Capability to Interpret

Like other deep learning techniques, self-attention models have interpretability issues. It is difficult to comprehend how self-attention mechanisms make decisions, in contrast to rule-based systems or more straightforward machine learning methods. In high-stakes applications like cybersecurity, healthcare, and finance, where explainability is essential for accountability and trust, this black-box aspect presents issues.

6. Adversarial Attack Sensitivity

According to recent studies, self-attention-based models are susceptible to adversarial assaults, in which little changes in the input data may cause the model to predict things incorrectly. By altering input patterns in a manner that is invisible to humans but has a big influence on model outputs, attackers may take advantage of this flaw. Strong defence mechanisms, such adversarial training or model regularisation, are needed to address this problem.

7. Problems with Scalability and Deployment

Scalability issues arise when self-attention-based models are used in practical settings. It is challenging to incorporate these models into edge devices, Internet of Things applications, and mobile systems because to their high memory usage and computational demands. Furthermore, their effectiveness in large-scale production settings is still a work in progress, necessitating improvements like hardware acceleration strategies and model compression.

8. Limitations of Current Approaches

Although self-attention-based detection algorithms are effective in many applications, they have a number of drawbacks.

- **High Computational Complexity**: Due to their quadratic temporal complexity, transformers are resource-intensive and challenging to implement in real-time applications or on edge devices.
- **High Data and Training Requirements:** These models perform poorly when there is a lack of data since they need a lot of labelled data and pretraining.
- **Slow Training and Optimisation Problems**: Training is computationally costly and fraught with difficulties including unstable gradients and hyperparameter sensitivity, which prolongs optimisation.
- **Difficulty with tiny and Overlapping items:** In complicated situations, these models' accuracy is diminished due to their inability to recognise tiny items or differentiate between overlapping ones.
- **Adversarial Attack Vulnerability:** Security-critical applications may be impacted by self-attention models' susceptibility to subtle, undetectable changes in input data.
- **Lack of Interpretability:** In high-stakes applications like healthcare and finance, the "black box" nature of these models restricts comprehension and confidence.
- **Scalability and implementation Issues:** Although methods like pruning and distillation are being investigated to address this, implementation on real-world devices is challenging due to high memory and processing power requirements.

IX. RECENT TRENDS AND FUTURE DIRECTIONS

As encryption has been more widely used and cyber threats have become more sophisticated, current research has concentrated on enhancing anomaly detection using cutting-edge deep learning architectures. BERT and Vision Transformers (ViTs) are two examples of transformer-based models that have become popular because of their capacity to identify long-range relationships in sequential data. Additionally promising for improving detection accuracy and resilience are hybrid deep learning techniques that include CNNs, RNNs, and self-attention processes. Federated learning, which enables models to be trained over dispersed networks without disclosing raw data, is also being investigated as a privacy-preserving technique. Researchers are creating lightweight and effective designs, such sparse transformers and attention-based pruning approaches, to allow real-time detection because of the high computational cost of transformers. Additionally, dataset restrictions and class imbalances in training models are being addressed with the use of data augmentation approaches and synthetic data synthesis utilising generative adversarial networks (GANs).

Notwithstanding these developments, a number of issues still need to be resolved in order to increase the effectiveness and suitability of self-attention-based anomaly detection in encrypted network data. Lightweight transformer systems that can manage high-speed network traffic with little computing cost are necessary for scalability and real-time detection. In order to sustain detection accuracy over time, adaptive learning models that dynamically adapt to changing assault patterns will be essential. Enhancing self-attention mechanisms' interpretability and explainability may also help cybersecurity experts make better decisions by offering more insights into how abnormalities are identified. For practical implementation, these models must be smoothly integrated with current security frameworks, such as Intrusion Detection Systems (IDS) and Security Information and Event Management (SIEM) programs. Furthermore, models' resilience may be increased by guaranteeing cross-dataset generalisation by training and testing in various network contexts. Finally, to maximise resource utilisation, energy-efficient designs must be created, especially for deployment in edge and Internet of Things devices. Future studies may greatly enhance the efficacy and practicality of self-attention mechanisms for encrypted traffic anomaly detection by tackling these issues, which will help create a more secure and private cybersecurity environment.

X. CONCLUSION

As network communications become increasingly dependent on encryption, more advanced anomaly detection methods that don't sacrifice efficiency or privacy have to be developed. Conventional methods often have trouble deciphering encrypted communication, which causes issues with scalability and accuracy. The use of self-attention techniques, namely in transformer-based designs like BERT, for identifying irregularities in encrypted network data has been investigated in this systematic research. When compared to traditional machine learning models, our data shows that self-attention mechanisms perform better in detecting intricate sequential patterns. Real-time detection capabilities, processing cost, and feature extraction from encrypted data are still significant challenges, however. These issues have been addressed in part by recent developments in model optimisation, hybrid architectures, and dataset augmentation techniques nevertheless, further study is required to improve these models for real-time and large-scale deployment. To bring it up, self-attention-based methods provide improved accuracy and flexibility, suggesting a potential path for encrypted traffic anomaly detection. To enable realistic application in real-world cybersecurity scenarios, future research should concentrate on designing lightweight structures, combining multimodal data sources, and increasing model efficiency.

REFERENCES

[1]. Enrolf Carl, Schultz Eugene & Mellander Jim 2004, Intrusion Detection and Prevention, McGraw Hill.

[2]. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. *European Conference on Computer Vision (ECCV)*, 213–229.

[3]. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*.

[4]. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *International Conference on Computer Vision (ICCV)*, 10012–10022.

[5]. Sun, P., Zhang, R., Jiang, Y., Kong, T., Li, F., Chen, H., ... & Yuan, Z. (2021). Sparse R-

CNN: End-to-end object detection with learnable proposals. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14454–14463.

[6]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 5998–6008.

[7]. Wang, W., Xie, E., Li, X., Fan, D. P., Song, K., Liang, D., ... & Luo, P. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. *International Conference on Computer Vision (ICCV)*, 568–578.

[8]. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2021). Deformable DETR: Deformable transformers for end-to-end object detection. *International Conference on Learning Representations (ICLR)*.

[9]. Hu, W., Cao, L., Ruan, Q., & Wu, Q. (2023). Research on anomaly network detection based on self-attention mechanism. *Sensors*, *23*(11), 5059.

[10]. dos Santos, F.P.; Ribeiro, L.S.; Ponti, M.A. Generalization of feature embeddings transferred from different video anomaly detection domains. *J. Vis. Commun. Image Represent.* **2019**, *60*, 407–416.

[11]. Nam, M.; Park, S.; Kim, D.S. Intrusion detection method using bidirectional GPT for in-vehicle controller area networks. *IEEE Access* **2021**, *9*, 124931–124944.

[12]. Chowdhury; Nasimuzzaman, M.; Ferens, K.; Ferens, M. Network intrusion detection using machine learning. In Proceedings of the International Conference on Security and Management (SAM), Las Vegas, NV, USA, 25–28 July 2016; The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp): Las Vegas, NV, USA, 2016.

[13]. Tuor, A.; Kaplan, S.; Hutchinson, B.; Nichols, N.; Robinson, S. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. In Proceedings of the Workshops at the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.

[14]. Yan, Q.; Wang, M.; Huang, W.; Luo, X.; Yu, F.R. Automatically synthesizing DoS attack traces using generative adversarial networks. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 3387–3396.

[15]. Wu, Z.; Zhang, H.; Wang, P.; Sun, Z. RTIDS: A robust transformerbased approach for intrusion detection system. *IEEE Access* **2022**, *10*, 64375–64387

[16]. Alkhatib, N.; Mushtaq, M.; Ghauch, H.; Danger, J.-L. Can-bert do it? controller area network intrusion detection system based on bert language model. In Proceedings of the 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 5–8 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–8.

[17]. Tuor, A.; Kaplan, S.; Hutchinson, B.; Nichols, N.; Robinson, S. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. In Proceedings of the Workshops at the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.

[18]. Yan, Q.; Wang, M.; Huang, W.; Luo, X.; Yu, F.R. Automatically synthesizing DoS attack traces using generative adversarial networks. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 3387–3396.

[19]. S. Ullah, J. Ahmad, M. A. Khan, E. H. Alkhammash, M. Hadjouni, Y. Y. Ghadi, F. Saeed, and N. Pitropakis, ‘‘A new intrusion detection system for the internet of things via deep convolutional neural network and feature engineering,’’ Sensors, vol. 22, no. 10, p. 3607, 2022.

[20]. A. Khacha, R. Saadouni, Y. Harbi, and Z. Aliouat, ‘‘Hybrid deep learning- based intrusion detection system for industrial internet of things,’’ in 2022 5th International Symposium on Informatics and its Applications (ISIA). IEEE, 2022, pp. 1–6.

[21]. P. Dini, A. Begni, S. Ciavarella, E. De Paoli, G. Fiorelli, C. Silvestro, and S. Saponara, ‘‘Design and testing novel one-class classifier based on polynomial interpolation with application to networking security,’’ IEEE Access, vol. 10, pp. 67 910–67 924, 2022.

[22]. A. A. Alashhab, M. S. M. Zahid, A. Muneer, and M. Abdullahi, ‘‘Low-rate ddos attack detection using deep learning for sdn-enabled iot networks,’’ International Journal of Advanced Computer Science and Applications, vol. 13, no. 11, 2022.

[23]. D. Javeed, T. Gao, M. S. Saeed, and P. Kumar, ‘‘An intrusion detection system for edge-envisioned smart agriculture in extreme environment,’’ IEEE Internet of Things Journal, 2023