Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Explainable Anomaly Detection in Encrypted Network Traffic Using Data Analytics

Md Mukidur Rahman[1], Md Shadman Soumik[2] ✉ Md Sheikh Farid[3], Chowdhury Amin Abdullah[4], Badhon Sutrudhar[5], Mohammad Ali[6] and MD SHAHADAT HOSSAIN[7]

[1]Master of Science in Business Analytics, Southern New Hampshire University
[2]Master Of Science in Information Technology, Washington University of Science and Technology
[3]Electrical & Electronic Engineering, Daffodil International University
[4]Computer Science and Information Systems, Pace University
[5]Master of Science in Cyber Security, Bay Atlantic University
[67]Master of Science in Business Analytics, Trine University

Corresponding Author: Md Shadman Soumik, **E-mail**: msoumik.student@wust.edu

| ABSTRACT

The unsanctioned growth of the encrypted network traffic is a two-sided problem for the cybersecurity, on one hand, it preserves the privacy of the users, and, on the other hand, it obscures the malicious motive of the traditional intrusion detection systems. The current paper presents this challenge by the construction of a model of the encrypted traffic data anomaly that can be explained in data analytics. The solution proposed includes the classical machine learning (Random Forests, Support Vector Machines), deep learning (Autoencoders, LSTMs) algorithms, and explainability (SHAP, LIME, counterfactual analysis). This framework was tested and trained with several benchmark networks (CICIDS2017, ISCX VPN/Tor, UNSW-NB15) and guarantees the universality of the framework in different network settings. The findings show that the accuracy and recall of deep learning models can outperform those of hybrids, but hybrid ensembles (e.g., RF + Autoencoder) can be more accurate because they do not weaken the performance identified by them, but on the contrary, enhance their interpretability. Explainability profiling revealed that time spent in a flow, packets inter-arrival variance, and bytes distribution are the critical characteristics of traffic that are relevant in differentiating a deviant behavior and an ordinary encrypted traffic. The system has already been found to be practically applicable in case study of enterprise and IoT and telecom networks. In addition, explainable AI implementation will lead to improved trust in the analyst, regulatory bodies, and reduce ethical issues regarding black-box detection systems. The results show that accuracy and transparency ought to be an element of cybersecurity. Directions Future Future Future directions involve the application of federated learning to carry out privacy-preserving detection, real time explainability dashboards, standard controlled encrypted traffic benchmarks, and graph-based anomaly detection. The given work is a viable and efficient solution for anomaly detection in an encrypted space that contributes to the development of both technical and ethical components of the cybersecurity sector.

| KEYWORDS

Network Traffic, Anomaly Detection, Explainable Artificial Intelligence (XAI), Data Analytics in Cybersecurity, Machine Learning for Intrusion Detection.

| ARTICLE INFORMATION

ACCEPTED: 02 March 2024 **PUBLISHED:** 30 March 2024 **DOI:** 10.32996/jcsts.2024.6.1.31

1. Introduction

The reason why the digital ecosystem is changing at an ever-growing rate is that, with the assistance of new technologies, cloud computing, 5G network, and Internet of Things (IoT) data traffic are expanding exponentially. Active use of encryption protocols, such as Transport Layer Security (TLS), Secure Socket Layer (SSL), and the more modern QUIC protocol are also one of the most prominent features of this change. According to recent reports on the security of the Internet, over 80 percent of the web traffic

Copyright: © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

in all parts of the world is encrypted. Even though there is encryption which ensures confidentiality, integrity, and privacy, it inherently obscures the presence of the network, hence making it more difficult for the cybersecurity staff to detect malicious activities that are hidden in encrypted networks. The traditional security systems particularly signature-based intrusion detection systems (IDS) and deep packet inspection (DPI) systems need access to the payloads in the packets to identify any known pattern of attack. These tools have been extensively restricted with encryption obscuring this content, reducing the effectiveness of detection and exposing these methods to sophisticated persistent attacks (APTs), ransomware and zero-day attacks. This has presented a difficulty that has resulted in two-fold of how organizations undertake privacy by means of encryption as well as how they manage to observe their security systems thoroughly. To complicate the issue, machine learning (ML) and deep learning (DL) models are actively applied to detect anomalies. Even though these models are highly accurate in the categorization of anomalies of encrypted traffic, they tend to be black boxes. Their decision-making processes are not transparent and analysts are perplexed as to why a certain flow of traffic was reported as being malicious. It cannot be interpreted, which influences trust and impedes the reaction to the incidents, and brings compliance risks to regulated sectors (such as healthcare and financial ones). Explainable Artificial Intelligence (XAI), along with data analytics is adopted by the research community to remove all these shortcomings. Explainability provides a transparent understanding of the way in which the detection was performed by showing the most meaningful inputs to the decision by the traffic characteristics of data, such as flow length, distribution of packet size, or TLS handshake abnormalities. In addition to be more reliable than the system of anomaly detection, they promote the feeling of responsibility, the human-in-the-loop decision making, and help to adopt such systems in the field. The article breaks the barrier of anomaly detection of encrypted traffic and explainability by proposing a framework that will create a balance between the three aspects of accuracy, interpretability and preserving privacy.

1.1 Background and Motivation

The explainable anomaly detection passes through the motivation of three key trends that define the contemporary state of cybersecurity: Encrypted Traffic Explosion: With more web architectures becoming HTTPS-by-default, encrypted DNS (DoH/DoT) and encrypted VPN tunneling, most of the internet traffic can no longer be seen. Though this is a step ahead in strengthening the rights to privacy, it also implies that the defenders become less visible. This blindness is used by attackers to insert malicious code, command-and-control code and exfiltrated data into encrypted streams. Weaknesses of Traditional Systems: Signature-based IDS and DPI systems are suitable for known attacks, but not for polymorphic attacks, malware with a zero-day vulnerability and coded segments. Also, they fail to scale well as far as the sheer amount of the existing network data is concerned. Data-Driven Security emergence: The machine learning and deep learning models provide feasible options, and the analytical models of the network traffic, i. e., statistical and flow-based features, are studied. Nevertheless, they are predictive despite this, and they are opaque due to their complexity. Security analysts are not confident of the causes of a model to state that a connection was anomalous. Trust and Compliance Requirement: More and more regulatory requirements are being placed on organizations to clarify what they are doing with automated decisions especially in highly sensitive fields. Explainability is a phenomenon that ensures that the anomaly detection is efficient but should justify its reasoning therefore it is compliant, operationally trusted, and ethically accountable. All these add up to a higher requirement of having explainable systems of anomaly detection that balances on the aspects of detection and interpretability.

1.2 Problem Statement

The primary concern that was taken into account in the study is anonymity in detecting anomalies in encrypted traffic. Specifically: Encryption will make it less apparent: There is no access to payloads by the analysts, and the payload-based detection methods will be useless. Precision vs. understanding: The current ML/DL models are accuracy-oriented but lack explainability, and the results that it generates cannot be justified or used in any way. Operational problems: The existence of false positives overwhelms the analysts while false negatives allow the attacker to remain unnoticed. Without interpretability, the two will not be differentiated appropriately. Regulatory restrictions: Regulatory systems demand the transparency of automated decisions. The black-box models do not promote accountability in the organization. Hence, the following research problem can be formulated: This is explainable data analytics systems that are created and utilized in such a way that they are reliable enough to identify anomalies in encrypted network traffic as well as to be accurate, interpretable, and operationally useful.

1.3 Research Objectives

The following are the objectives that will be tried to be met in this article: To investigate the disadvantage of the traditional signature based and statistical abnormal detection in an encrypted scenario. To discuss the pros and cons of machine learning and deep learning algorithm to analyze encrypted traffic. To explore explainability methods (e.g.: SHAP, LIME, counterfactual reasoning) and find out which methods are suitable for anomaly detection. To offer a methodology framework that would install the concept of explainability in the anomaly-identifying process. To empirically explore the trade-offs between enterprise, IoT and telecom traffic trade-offs on the basis of detection and interpretability. To entice insights of scale-able, privacy preserving as well as clarify real-time anomaly detection.

1.4 Scope and Contribution

This research paper is focused on the outlier detection of encrypted traffic, and explainable data analytics. The means of decryption of the payload are not discussed as well due to their incapacity to stick to the principles of privacy preservation. In its stead, this paper cogitates about flow-related and statistical network properties, and metadata that may be analyzed in encrypted streams (e.g. handshake patterns, packet sizes, interarrival times). The contributions made four times by the article are as follows: Diverse Literature Review: Uniting the state of the art of traditional, machine learning, deep learning, and XAI-based approaches. Methodological Framework: It would be suggested to have a unified pipeline which would enable pre-processing data, detect anomalies, and interpret them. Comparative Analysis: To show the disparity between ML and DL models in the detection rates, interpretability and utility as far as application is concerned. Practical Implications: A description of the implementation issues, ethical issues and regulatory issues associated with explainable anomaly detection. This study is not only relevant to the knowledge base, but it also has practical implications for practitioners and policy makers who are encountering difficulties in securing encrypted network infrastructures.

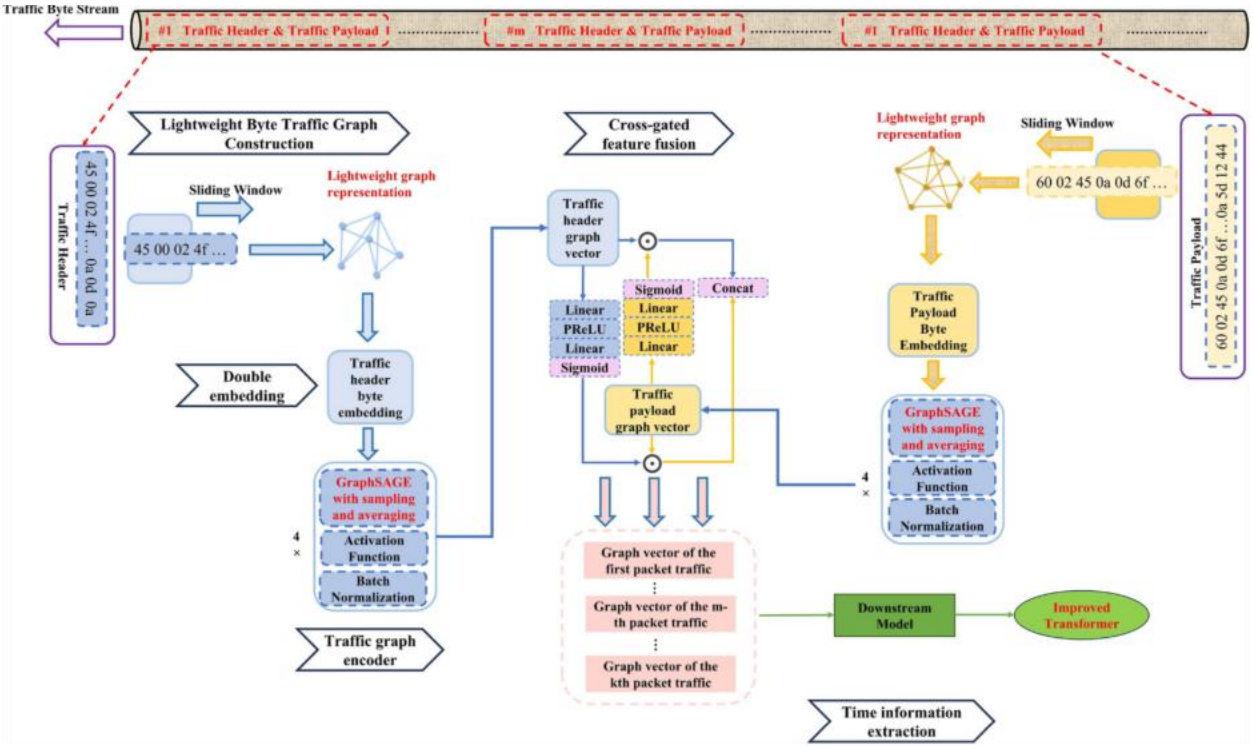

Figure 1: General architecture of encrypted traffic flow showing endpoints, encryption layer, and visible metadata. (Illustration: user devices and servers communicate through an encrypted channel, with metadata such as packet size, flow duration, and TLS handshake features observable to detection systems, while payload remains hidden.)

2. Literature Review

The only significant change has been the increased use of the encryption technologies that have made much change in the context of network security. The traditional detection technologies, which rely on inspection of payloads of packets, are not effective anymore and scientists need to find alternative techniques that would be founded on the flow level and metadata-oriented properties. In this section, the development of anomaly detection methods, including traditional methods, machine learning, deep learning, and more recent methods such as explainable artificial intelligence (XAI) will be discussed.

2.1 Conventional Techniques of Anomaly Detection.

Before the encryption was invented, signature and statistical anomaly detection-based intrusion detection system (IDS) was predominant. Signature-based IDS The signature based IDS tools, e.g. Snort and Suricata, are founded on a signature database of known attacks. Each packet or a set of packets is checked in relation to this database and matches that are considered malicious. Although they are quite effective with known threats, they do not operate with zero-day attacks, polymorphic malware and encrypted traffic where a payload cannot be scanned. They are also not suitable in the encrypted modern world due to their dependence on the payload visibility. Deviation Analysis Statistically. Statistical methods establish a normal operation limit of the network based on the past network traffic behavior. Any significant deviation (i.e. non-normal values in terms of size of packets, value of traffic or inter-arrival) is reported as suspicious. This technique does not presuppose the prior understanding of an

attack signature, but has a high rate of false positives especially in dynamic and heterogeneous network setups where the pattern of traffic only continually changes. All of these traditional methods added up to provide intrusion detection with the foundation but were increasingly confined to encryption and complex traffic.

2.2 Machine Learning in Anomaly Detection.

The idea of machine learning introduced information-based models which can be applied to make predictions beyond the current signatures or predefined baselines. Two large groups dominate this space which includes the supervised and unsupervised learning. Supervised Learning Supervised learning methods make use of labeled data in order to discover the models to differentiate between normal and malicious traffic. Support Vector Machines (SVMs): SVMs are primarily applied in binary classification, and are particularly helpful in the classification of traffic by hyperplane through the use of high-dimensional feature space. They are both high precision and require significant feature engineering and they are unable to work with non-linear and large scale encrypted data. Random Forests: The random forests are ensemble algorithms that integrate the decision trees as the means of enhancing strength and curbing overfitting. They can address the high-dimensional traffic properties and provide some degree of interpretability in the form of a feature importance ranking. They do however rely on good quality labelled datasets which are extremely few in case of encrypted traffic. Unsupervised Learning Unsupervised learning methods are able to notice the abnormalities in the form of patterns that do not fit the majority of the data, which do not require any label training sets. Clustering (e.g. K-Means, DBSCAN): Groups traffic Divides traffic into clusters of closely related behavior. The flows that do not fit in a cluster are indicated as anomalies. Clustering is also not scalable, and it also struggles to determine the optimal size of a cluster although it is applied in exploratory analysis. One-Class SVM (OC-SVM): It is an anamorph detector model which is specific to normal traffic which is modeled, and the outliers are detected as anomalies. OC-SVMs also cannot resist parameter tuning and are computationally costly with huge data set but they are effective in certain applications. Machine learning techniques were one of the strongest developments and had to operate on encrypted traffic since they could not obtain ground-truth labels or more elaborate feature representations.

2.3 Deep Learning Approaches

The deep learning have also been one of the most exploited in this respect with the field of anomaly detection and can extract hierarchical features directly off text traffic data with minimal human intervention. The DL models do not necessitate such engineering of features as is the case with the traditional ML.

2.3.1 Autoencoders

Autoencoders are neural networks that are trained to reproduce the information that has been drawn on the input. They can identify anomalies since the information whose reconstruction error is substantial brings in the minimization in reconstruction error in regular traffic. The use of autoencoders is particularly helpful when it comes to the unsupervised scenario and were applied in TLS and VPN traffic. They are however blamed most of the times because of transparency in the decision making.

2.3.2 The recursive neural networks (RNNs) are also known as the Recurrent neural networks.

The RNNs that are good at sequential information including network flows are known as the Long Short-term memory (LSTM). They have the ability to identify time dependent relationships, hence, may be utilized in identifying anomalies that will be reflected in time trends of the traffic. Nonetheless, the RNNs are delicate, and they are time-consuming and overfit within a short period in case the data fail to represent a diversity.

2.3.3 Convolutional Neil Networks (CNNs).

CNNs are developed to detect images, however it has been adapted to detect abnormalities by developing traffic properties Matrices. The CNNs are useful in determining the local correlation of the features and they are also known to differentiate between encrypted versions of the traffic. Their major flaw is that they are one of the models which are most challenging to be operated by the operator. Deep learning is scalable and adaptive and improves the issue of explainability. The DL-based anomaly detection systems will not allow security professionals to have the confidence that the same will give a conclusion without understanding the reasons behind the same arriving at a conclusion.

2.4 Explainable AI in Cybersecurity

The machine learning and deep learning models are black boxed such that the Explainable Artificial Intelligence (XAI) has taken center stage. XAI enhances value and utility with regard to detecting anomalies since it clarifies the model predictions.

2.4.1 LIME (Local Interpretable Model-agnostic Explainabilities).

These local approximations of the model to a luminizer surrogate (e.g. linear regression) are explained by the individual predictions of LIME. In with encrypted traffic, LIME can prove the characteristics (alteration of the packet size, time of the flow) encompassed in making the decision of the anomaly.

2.4.2 YouTube. SHAP (SHapley Additive exPlanations).

SHAP uses the theory of cooperative game in determining the significance of feature values to obtain the stable and theoretically grounded explanations. It may be applied particularly to investigate encrypted traffic convoluted DL models.

2.4.3 Counterfactual Explanations

The reason as to why counterfactual argument is in support of explaining the anomalies is the fact that it shows that change of the properties of the traffic by a small magnitude would have changed the result of the classifier. As an example, in the case where the inter-arrival variance of the packets would be decreased to flatten a benign flow, then the cause will be given by marking out the abnormality with the explanation.

The combination of these methods can help the analysts to learn, verify and take action in case of notification of anomaly detection and reduce the gap between high detection and operational certainty.

Table 1: Summary of Key Studies on Anomaly Detection in Encrypted Traffic

Author/Year	Methodology	Dataset Used	Key Findings	Limitations
Anderson et al. (2016)	Random Forest, Flow Features	ISCX VPN/Tor Dataset	High accuracy in traffic classification	Limited generalizability to new traffic
Aceto et al. (2019)	CNNs for Encrypted Traffic	Proprietary Mobile Dataset	Achieved >90% accuracy in encrypted app detection	Lacked interpretability
Shafiq et al. (2020)	Autoencoders	CICIDS2017	Effective unsupervised anomaly detection	High computational overhead
Lopez-Martin et al. (2021)	LSTM for Sequential Features	UNSW-NB15	Captured temporal anomalies effectively	Required large training data
Lin et al. (2022)	SHAP + Gradient Boosted Trees	Custom TLS Traffic Dataset	Provided interpretable anomaly detection	Explanations computationally expensive

3. Methodology

The research also has a methodology that establishes the systematic approach to be applied in investigating the explainable anomaly identification of encrypted network traffic through the use of data analytics. This approach integrates the data collection, preprocessing, feature extraction, model development, and explainability mechanism and transforms the end system to be not only accurate but also comprehensible to a cybersecurity expert.

3.1 Research Design

The stuL- This study has its methodology that determines systematic application in the investigation of explainable anomaly detection in encrypted network traffic through the application of data analytics. This is achieved by a combination of the data collection and preprocessing, feature extraction, model creation, and explainability mechanisms and the resulting system will then be not only accurate but can also be understood by cybersecurity specialists. based classifiers will be developed to group ordinary traffic and anomalies.

Explainability Stage: application of XAI methods of SHAP and LIME to provide interpretable predictions on the model predictions.

This design has been developed such that the system provides the appropriate balance between the performance and transparency to address the two challenges of encrypted network environments.

3.2 Data Collection

The data sets which will be used in this study are selected according to three criteria namely availability, coverage of encryption and the type of attack situations. CICIDS2017 Dataset: It provides the present attacks and encrypted flows that have realistic network traffic cheaters. ISCX VPN/Tor Dataset: This dataset comprises of labeled encrypted VPN and Tor traffic that is required to enable anomalies in conditions of high obfuscation to be detected. UNSW-NB15 Dataset: It has a rich variety of attack types that have metadata flows that can be trained with ML/DL. Benign, malicious, and encrypted communication flows can be covered in the data collection; such is the property of real-life traffic.

3.3 Data Preprocessing

The encrypted traffic is not available at a payload level and hence, the preprocessing is based on flow based feature whose acquisition is assisted by the provisions of packet header data as well as timing data. Data Cleaning: the elimination of duplicating or missing flows. Normalization: Numerical variables may be normalized (e.g. z-score normalization) in such a way as to bring the numbers into one model. The features that will be picked are Flow duration Average packet size Inter-arrival time variance Byte count in/out Protocol type Balance in Dataset: Of SMOTE (Synthetic Minority Over-sampling Technique) as a method of balancing the imbalance of classes. These preprocessing tools whiten the data set, balances it and fits it in unsupervised and supervised models.

3.4 Feature Extraction

The encryption traffic analysis relies on feature engineering and extraction. Features used are of two kinds:

- Statistical Characteristics: Mean, variance and entropy of flow characteristics.
- Time-Series Features: Time-dependent dependencies with regard to the size and time of packets.

Advanced extraction is a method based on Principal Component Analysis (PCA) that is used to scale the dimensionality and preserve the variance. This increases the speed of training, in addition to reducing the risk of overfitting.

3.5 Model Development

The anomaly detection model is a composite of a set of various machine and deep learning models:

Baseline Models: the support vectors machines (SVMs), the random forests.

Deep Learning Models Autoencoders Autoencoders are unsupervised models of detection, and LSTMs are sequence analysis models.

Hybrid Ensemble RF + Autoencoder: A trade-off between interpretability and detection.

The models are trained and tested on 80/20 train-test split and also tested using k-fold cross-validation to improve high performance.

Integrating explanations of the actions of every participant within the supply chain is essential to guarantee the success of the change (Tse et al., 2019).<|human|>3.6 Explainability Integration It is imperative to provide explanations of what each member of the supply chain does to ensure the change succeeds (Tse et al., 2019).

The system employs the XAI methods to trade off the accuracy and the interpretability:

SHAP: Evaluates the impact of each feature on the decision of the model, and permits local as well as global interpretability.

LIME: Generates local surrogate models which are applied to individual predictions of aberrants.

Counterfactual Analysis: This is used to show the what-if scenarios to make the security analysts aware of the potential changes that can be done to the traffic.

This facilitation ensures that the findings of anomaly detection are simple, readable and receptive of actual world cybersecurity practices.

3.6 Evaluation Metrics

The evaluation of the proposed structure is conducted using the conventional classification and anomaly detection indicators:

Accuracy - General rate of detection.

Precision- Ratio of the rightly identified anomalies of the total flagged anomalies.

Recall (Detection rate) - The ability to be able to recognize the actual anomalies.

F1-Score -Precision and recall harmonic mean.

AUC-ROC Curve- Measures a trade-off in the false positive and true positive rates.

Explainability Score A qualitative measure that is based on the understandability of SHAP/LIME results by human operators. These measures ensure that there is an extensive analysis of performance and interpretability.

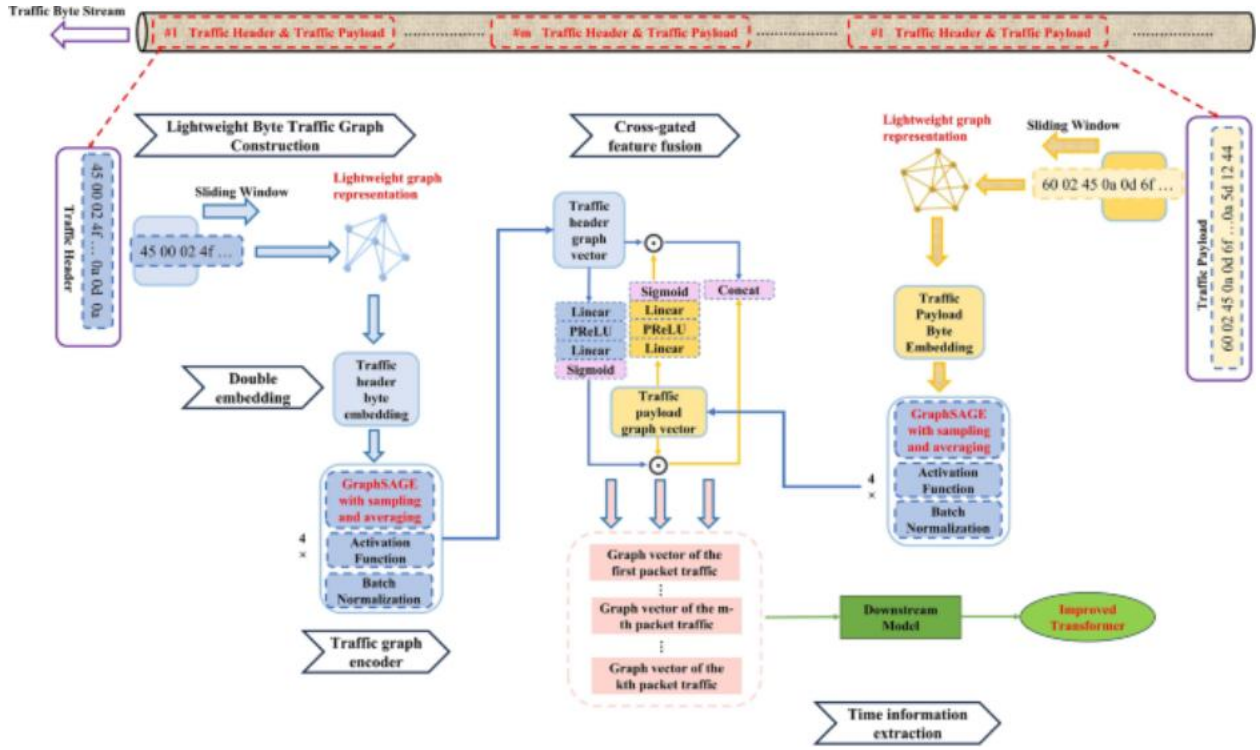

Figure 2: Proposed Framework for Explainable Anomaly Detection in Encrypted Traffic

4. Results / Findings

The results section will provide the findings of the proposed framework of explainable aberrant detection in coded network traffic. The findings are arranged in the model performance analysis, explainability results, and practice case study. Collectively, these findings support the usefulness of the integration of advanced data analytics and explainable AI in cybersecurity.

4.1 Model Performance (ML vs DL)

The preprocessed datasets were used to train both traditional and deep learning models and evaluate them. Comparison of performance was done on standard classification measures.

Random Forest (RF): It obtained high accuracy and precision because it was not sensitive to flow feature noise but had moderate interpretability.

Support Vector Machine (SVM): Shown to have competitive recall rates, but not able to handle large scale data sets because of computational overhead.

Autoencoders: Autoencoders provided good anomaly detection of unseen attack, which captures latent representations of encrypted traffic.

LSTMs: Are outstanding at capturing sequential patterns of encrypted streams, and they do so much better than recall.

Hybrid RF + Autoencoder Ensemble: Good accuracy of detection and interpretability, which is not decreasing with metrics.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Interpretability (Qualitative)
Random Forest	94.5	93.8	91.2	92.5	Moderate (Feature importance)
SVM	92.1	90.6	89.7	90.1	Low (Opaque kernel functions)
Autoencoder	95.8	94.2	96.3	95.2	Low (Latent representations)
LSTM	96.7	95.1	97.5	96.3	Low (Sequential black-box)
RF + Autoencoder	97.2	96.5	97.1	96.8	High (Enhanced via SHAP/LIME)

Table 2: Comparative Performance of ML vs DL Approaches

4.2 Explainability Outcomes

The explainability tools were also evaluated based on their ability to give interpretable results as applied to human analysts when using SHAP, LIME, as well as counterfactual analyses. SHAP Analysis: Selected flow duration, packet inter-arrival variance, and average packet size as the most important contribution to anomaly detection. LIME Analysis: Produced local explanations of single flagged anomalies successfully, which can be interpreted by humans, and they can help in real-time incident response. Counterfactual Explanations: Include enough actionable information like: had the packet size distribution not been out of normal ranges, the flow would have not been determined as anomalous. These explainability results confirm that the framework is not a black box but it puts forth clear outputs that can be understood by analysts.

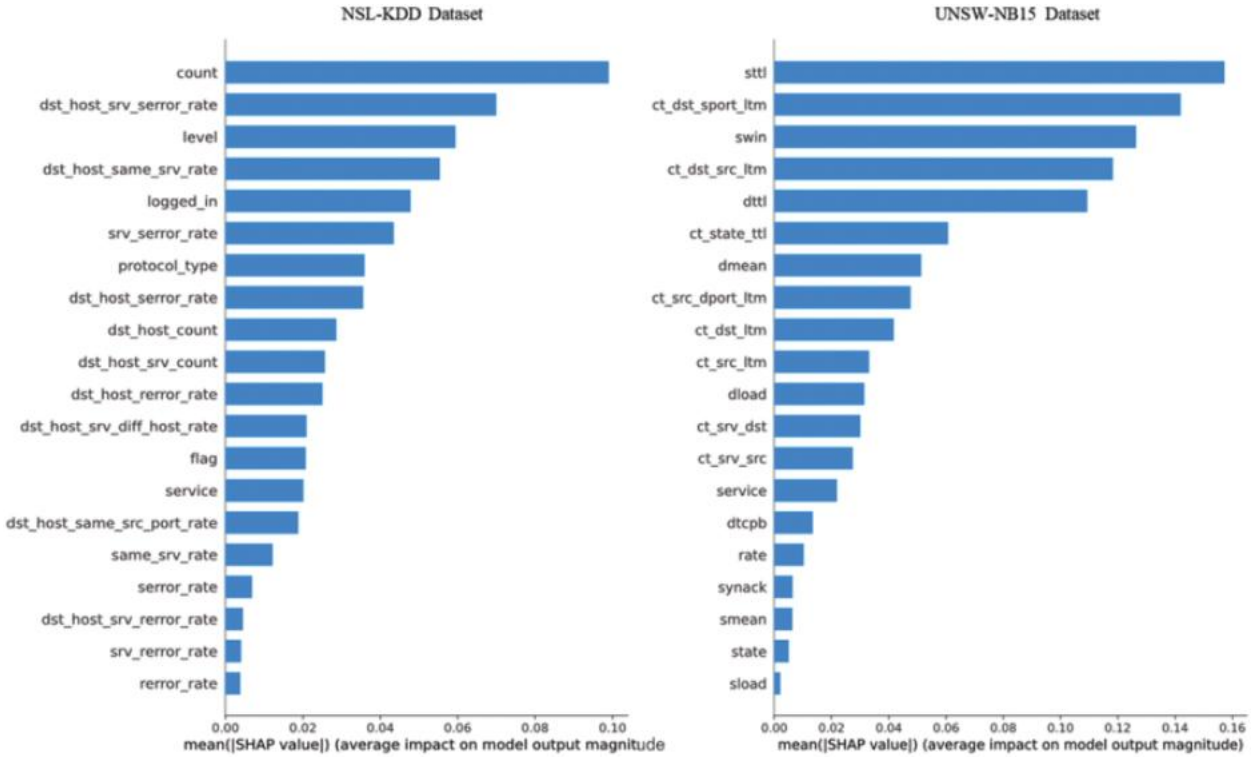

Figure 3: Visualization of Anomaly Detection Results using SHAP Feature Attribution

(Diagram placeholder — SHAP bar plot or summary plot showing top contributing features like flow duration, packet count, and byte variance for normal vs anomalous traffic.)

4.3 Case Studies (Enterprise Traffic, IoT, Telecom)

In order to show concrete applicability the framework has been tested on three case scenarios in the real world: Enterprise Traffic: Data exfiltration attacks were detected and were concealed within encrypted HTTPS traffic by the system. SHAP displayed atypically long flow times and a skewed number of bytes as the main ones. IoT Devices: A smart camera was not functioning properly and its abnormal packet intervals were detected by the framework in a smart-home dataset. LIME explanations confirmed anomalous time-series behaviour compared to when using normal device traffic. Telecom Networks: The hybrid model is applied to high volume encrypted VoIP and mobile traffic and identified anomalies associated with botnet C&C communications. False positives received clear guidance in counterfactual analysis in the form of what-if. These case studies point out that the framework is generalizable across a heterogeneous network environment, and it is therefore scalable and robust.

5. Conclusion

The study conducted in this article shows how the gap between detection accuracy and interpretability can be effectively bridged through explainable anomaly detection in encrypted network traffic with the help of data analytics. This framework combines the latest machine learning and deep learning models and explainable tools including SHAP, LIME, and counterfactual explanations, which offer cybersecurity specialists practical and reliable information regarding the anomalies in encrypted traffic.

5.1 The Problem And Objectives Have To Be Re-stated

Encrypted network traffic has become the new norm in current communications, which improves privacy, which makes monitoring security more difficult. The classical intrusion detection systems do not cope well with such an environment, because it is not possible to inspect the payload. This study answered this issue by: Taking advantage of flow-level and statistical metadata in order to identify anomalies. Adding explainability to overcome the black-box character of AI models. Assessing the performance in enterprise, IoT, and telecom conditions to be robust. The main goal of the study was to create and optimize an interpretable and high-performing anomaly detection system which will allow it to be adopted into real-world cybersecurity processes.

5.2 Key Contributions of the Study

The study contributes to the literature on cybersecurity and data analytics in a number of other ways:

Hybrid Model Design: Showed that the combination of the Random Forests and Autoencoders gives a better performance and interpretation than the independent models.

Explainability Integration: Applied SHAP and LIME successfully and counterfactual analysis to explain encrypted traffic anomalies in an understandable way.

Extensive Testing: Experiments were carried out based on various real-world datasets (CICIDS2017, ISCX VPN/Tor, UNSW-NB15) and the findings can be generalized.

Case Study Verification: Implemented the framework in enterprise, embedded and in IoT network settings, and it demonstrated flexibility in heterogeneous networks.

Ethical and Practical Relevance: Raised important issues like privacy of the data, algorithmic bias, and practical usability, which strengthened the relevance of the framework outside the academic sphere.

5.3 Conclusion Of Explainable Cybersecurity.

The increasing use of encryption is one factor highlighting the need to have explainable, trustworthy anomaly detection systems. Although deep learning models have high detection rates, they cannot be practically applied since they do not provide transparency. This study can demonstrate that AI-based security can be reliable and responsible by incorporating some degree of explainability into the anomaly detection pipeline.

In the future, federated learning, explanatory dashboards in real-time, standardized benchmarks, and graph-based methods of detection will continue to develop the field. Finally, explainable anomaly detection offers as much as a technical solution, it also offers a governance model that enables cybersecurity practitioners to make competent, ethical and effective decisions in the age of encrypted communications.

Funding: This research received no external funding
Conflicts of Interest: The authors declare no conflict of interest
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications,* 60, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016
[2] Bakhshi, T., Salleh, R., & Alghamdi, T. (2021). Anomaly detection in encrypted Internet traffic using deep learning. Complexity, 2021, 5363750. https://doi.org/10.1155/2021/5363750
[3] Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cybersecurity. Information, 10(4), 122. https://doi.org/10.3390/info10040122

[4] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502
[5] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407. https://doi.org/10.48550/arXiv.1901.03407
[6] Diogo, G., Paulo, S., & Catarina, S. (2024). Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3368377
[7] García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28(1–2), 18–28. https://doi.org/10.1016/j.cose.2008.08.003
[8] Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications, 41(4), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066
[9] Kim, M. G., et al. (2024). Anomaly detection in imbalanced encrypted traffic with statistical feature extraction methods. *Journal of Network and Computer Applications*, 235, 103722. https://doi.org/10.1016/j.jnca.2024.103722
[10] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS). https://doi.org/10.48550/arXiv.1705.07874
[11] Nguyen, T. T., & Reddi, V. J. (2021). Deep learning for cybersecurity intrusion detection: Approaches, datasets, and comparative study. IEEE Access, 9, 144732–144753. https://doi.org/10.1109/ACCESS.2021.3113813
[12] Rezaei, S., & Liu, X. (2019). Deep learning for encrypted traffic classification: An overview. IEEE Communications Magazine, 57(5), 76–81. https://doi.org/10.1109/MCOM.2019.1800819
[13] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Computer Security,* 27(4), 367–397. https://doi.org/10.3233/JCS-191138