e-FlowPrint: Enhanced FlowPrint for Robust Unknown Traffic Detection Using Uncertainty Measures Inspired by Active Learning

Marziyeh Bayat , Zhino Naghshbandi , Mehdi Teimouri *

Posted Date: 14 April 2025

doi: 10.20944/preprints202504.0946.v1

Keywords: Network Traffic Analysis; Traffic Classification; Mobile Application Identification; Open-Set Recognition; Unknown Traffic Detection; Robustness; Encrypted Traffic

 doi:10.20944/preprints202504.0946.v1

Article

e-FlowPrint: Enhanced FlowPrint for Robust Unknown Traffic Detection Using Uncertainty Measures Inspired by Active Learning

Marziyeh Bayat, Zhino Naghshbandi and Mehdi Teimouri *

University of Tehran, Iran

* Correspondence: mehditeimouri@ut.ac.ir

Abstract: The increasing prevalence of encryption enhances network traffic confidentiality and integrity but complicates network management and security by obscuring traffic flows. This challenge makes detecting cyberattacks and enforcing policies increasingly difficult. Encrypted Traffic Intelligence (ETI), particularly network traffic classification (NTC), offers solutions using machine learning techniques. However, practical implementation remains challenging due to the inherent complexity of network environments, where traffic feature distributions vary because of factors such as network topology and delay. This variability undermines the robustness of classifiers trained on static datasets. Moreover, dynamic environments increase the likelihood of encountering unknown traffic, where inaccurate identification can lead to high false positive rates, unacceptable in critical applications like billing and cybersecurity. To address these challenges, we propose e-FlowPrint, an enhanced FlowPrint-based open-set recognition (OSR) classifier designed for robust unknown traffic detection. Inspired by uncertainty sampling techniques in active learning, we introduce two novel methods: Probability Anomaly Recognition (PAR) and Entropy-Based Uniformity Analysis (EnUniA). PAR utilizes the disparity between the highest and second-highest classification probabilities; a small disparity indicates uncertainty, suggesting that the sample is likely to be unknown. EnUniA calculates entropy values across all classes, where high entropy indicates a uniform distribution of probabilities, further increasing the likelihood of the sample being unknown. We evaluate the proposed model using the ITC-Net-blend-60 dataset across diverse real-world network environments and conduct long-term performance assessments through three-year network condition simulations. Our results demonstrate that e-FlowPrint improves FlowPrint's unknown traffic detection performance by approximately 30%. Additionally, it enhances overall classifier performance by 2% in varying network environments.

Keywords: network traffic analysis; traffic classification; mobile application identification; open-set recognition; unknown traffic detection; robustness; Encrypted Traffic

1. Introduction

The advent of encryption has greatly improved the confidentiality and integrity of network traffic [1], protecting sensitive data from unauthorized access and interception. This advancement ensures secure communication for both organizations and individuals. However, the widespread adoption of encryption also poses significant challenges for network management and security. By concealing the contents of data packets, encryption restricts network administrators' visibility into the underlying traffic flows. As a result, they can no longer identify cyberattacks or implement traffic management policies [2].

This is where Encrypted Traffic Intelligence (ETI) comes into play. ETI is a technology that provides valuable information about network traffic without decryption and compromising users' privacy.

 doi:10.20944/preprints202504.0946.v1

One of the key functions of ETI is network traffic classification (NTC) [3]. NTC employs advanced techniques such as machine learning (ML), deep learning (DL), and graph science to categorize network traffic. It can be performed at different levels of granularity, including protocol, application, and service levels, with each level having its own distinct methods and uses [4].

For example, in video applications like Netflix and YouTube, encryption complicates the distinction between downloading video and streaming on demand. With ETI, operators can identify the underlying service. They can prioritize streaming over downloads or compress on-demand streams to minimize buffering during network congestion. Similarly, frequently accessed video content from platforms like Instagram can be cached, while packets from email applications such as Gmail or Outlook can be run through extra filtering to detect security threats hidden in file attachments.

Additionally, application awareness is essential for access management in corporate environments, especially those supporting remote work or BYOD policies. Shadow IT activity can be managed by whitelisting/blacklisting specific applications — a critical security measure in sensitive sectors like data centers, where restricting access to vulnerable applications helps safeguard infrastructure [5].

Therefore, insights delivered by NTC suite support a wide range of networking solutions, such as routers, network packet brokers, policy control engines, and security tools, including next-gen firewalls and cloud access security brokers [6].

Network environments are highly complex [7]. Despite extensive research has been conducted on the subject, practical implementation faces two key challenges: model robustness and detecting unknown traffic [8]. This complexity directly impacts model robustness. In real-world scenarios, the behavior of network traffic is highly affected by a range of factors such as network topology, bandwidth, and latency, causing significant variations in traffic feature distributions. These factors may vary across different network segments or even within the same network segment over time [9]. Recent studies [9,10] have demonstrated that while existing classifiers perform well when trained and tested on conventional machine learning datasets—where the data is divided into training and testing sets—they often suffer significant performance degradation when applied to different datasets. Given the substantial impact of network environments on the distribution of network traffic features and, consequently, on model performance, it is crucial to develop models that demonstrate robustness and compatibility across variant network environments [11].

Beyond model robustness, this dynamic environment also necessitates robust unknown traffic detection, which is crucial for real-world deployment. The ability to identify whether or not a test sample belongs to one of the semantic classes in a classifier's training set that is called open-set recognition (OSR) or out-of-distribution (OOD) detection, has received significant attention in recent years [12]. Nonetheless, most network traffic classification is created with a closed-world assumption, where all potential classes are known during training [13]. This approach limits classifiers to predefined label sets, hindering their ability to recognize new traffic data. Accordingly, unknown traffic data is misclassified as a known class, leading to a high false positive rate. In many use cases, false positives in the classification results are unacceptable. For example, operators have to rule out false positives in the area of billing, as this can have a negative revenue impact and potentially damage the brand image. In cybersecurity use cases, false positives can make a huge difference and enable attacks and data breaches.

When the traffic network environment is constantly changed and updated, then the likelihood of appearing unknown traffic flow is higher [8]. Therefore, it is essential to identify and classify unknown traffic accurately.

The majority of web traffic comes from mobile devices. In 2025, mobile devices (excluding tablets) accounted for over 62% of global web traffic [14]. The affordability of mobile devices, coupled with the rapid expansion of network infrastructure such as 5G and the popularity of mobile applications and services, has driven a significant increase in mobile traffic volumes [15]. For this reason, we focused on classifying mobile applications. However, identifying mobile applications

 doi:10.20944/preprints202504.0946.v1

poses more greater challenges than classifying other types of network traffic, such as Internet of Things (IoT) traffic. This is primarily due to the homogeneous yet highly dynamic and continuously evolving nature of mobile apps [5]. Notably, new updates to an application can change its behavior, which complicates the classification process. Furthermore, mobile applications can behave differently depending on the device and operating system, they are running on [16]. As a result, two aforementioned challenges—model robustness and unknown traffic detection—are particularly pronounced in mobile traffic classification. These observations become the motivation for our work.

In this study, we present a novel and robust model for mobile traffic classification, with a particular focus on detecting unknown traffic in dynamic and complex network environments. Unlike existing methods, our approach integrates uncertainty measures inspired by active learning techniques, which significantly improve the detection of unknown traffic. We introduce two key components: Probability Anomaly Recognition (PAR), which analyzes the disparity between the highest and second-highest classification probabilities, and Entropy-Based Uniformity Analysis (EnUniA), which measures the uniformity of classification probabilities across all classes. These innovations allow our model to more accurately identify unknown traffic, which is crucial for real-world applications such as cybersecurity and billing, where misclassification can lead to severe consequences.

The model's performance is evaluated using the ITC-Net-blend-60 dataset, an in-house dataset specifically collected to address the challenges of mobile network traffic classification. To further validate the model, we conduct extensive long-term simulations replicating real-world network conditions, enabling us to assess its robustness and adaptability in dynamic environments. Our results show a significant improvement in unknown traffic detection, with a 30% enhancement over traditional methods, as well as a 2% overall improvement in classification performance across varying network conditions.

The rest of the paper is organized as follows. Section 2 reviews some of the most significant and recent studies on mobile network traffic classification. In Section 3, we introduce our proposed method. Section 4 presents the results of applying this method in both invariant and variant network environments. Section 5 discusses the experimental findings in greater detail. Finally, we conclude the paper in Section 6.

2. Literature Review

In this section we provide an overview of key network traffic classification methods. Due to the prevalence of encryption, feature extraction has increasingly focused on time-based characteristics, packet lengths, unencrypted layer header values, and TLS handshake information [5]. These features can be divided into three levels: "packet", "flow", and bag of Flow (BoF).

Over the past two decades, a multitude of traffic classification methods have been developed and deployed [17]. Historically, port-based approaches were dominant. However, the contemporary use of dynamic ports by many applications has significantly diminished the accuracy of this method. Studies by Moore and Papajianaki [18] and Madhukar and Williamson [19], indicate that port-based classification now accurately classifies only 30% to 70% of Internet traffic. Consequently, more sophisticated traffic classification techniques are necessary to address the complexities of modern network traffic.

Behavior-based approaches offer an alternative perspective by classifying traffic based on host (endpoint) behavior. Empirical studies have demonstrated that inter-host interactions exhibit distinct communication patterns across different applications. These patterns can be discerned by analyzing factors such as the number of communicating hosts, port usage, transport layer protocols, and the IP addresses of communicating hosts. This methodology has shown promising classification accuracy. For instance, Van Ede et al. [5] proposed a semi-supervised approach for mobile app fingerprinting based on endpoint temporal correlations while Li et al. [10] proposed a method based on graph propagation for traffic pattern detection.

 doi:10.20944/preprints202504.0946.v1

With the advancement of machine learning, it has become a central approach in encrypted network traffic classification. Machine learning-based methods can be broadly classified into three categories: statistical feature-based, temporal feature-based, and deep learning-based. Statistical feature-based methods involve extracting statistical features from traffic objects, such as packet counts, average packet lengths, and average packet arrival times per flow. These features are then used to train machine learning models for classification, as demonstrated in [20,21].

Temporal feature-based methods utilize features, such as packet length sequences or packet arrival interval sequences, as input for classifiers. Time series models, including Hidden Markov Models [22–25] and Recurrent Neural Networks[24,25], are commonly employed to learn temporal features for traffic classification.

Deep-Learning based take raw traffic data as input and combine feature extraction, learning, and classification seamlessly, such as the method presented in [6,26].

However, these traditional approaches may struggle to detect novel, unseen applications that emerge in real-world scenarios. Consequently, recent investigations have increasingly addressed the challenge of identifying unknown or out-of-distribution (OOD) data. For instance, [1,13] employed methodologies combining Convolutional Neural Networks (CNNs) and Maximum SoftMax Probability (MSP) to detect OOD samples. Additionally, S. Cruz et al. [27] investigated the detection of unknown attacks and traffic using an open-set approach, utilizing a Weighted Support Vector Machine (W-SVM) model, and Y. Zhang et al. [28]proposed a novel open-set detection method, leveraging Extreme Value Theory (EVT) and an Open-CNN architecture, to estimate the probability of unknown attacks.

These studies collectively highlight the growing importance of open-set recognition (OSR) in enhancing intrusion detection capabilities. By enabling the identification of novel applications and unknown attacks, OSR offers a promising approach to improve the security of networked systems, particularly in dynamic and evolving environments

3. Materials and Methods

3.1. Dataset

To evaluate our proposed methods, we use in-house ITC-Net-Blend-60 dataset [29]. This dataset comprises encrypted mobile traffic labeled per app. It was gathered from 60 popular Android applications in five different network scenarios. These scenarios were consistent across all the apps but varied in terms of the Internet service provider (ISP), geographic location, device, app version, and individual users. Detailed specifications of the dataset and each scenario are provided in Table 1. As indicated in FlowPrint [5] browsers behave like a platform for accessing web content rather than a dedicated application. Consequently, similar to the FlowPrint, we have entirely eliminated this category of applications from the dataset.

Table 1. ITC-Net-blend-60 dataset specifications.

Scenario ID	No. Apps	No. Bi-Flows[1]	User	Device			Location[2]	ISP[3]
				Vendor	Model	Android version		
A	59	108,370	U_1	Xiaomi	Note10 Pro	11	L_1, L_2	N_1, N_2
B	60	72,279	U_2	Samsung	A50	11	L_1, L_3	N_1, N_3
C	59	141,957	U_3	Samsung	A31	11	L_4	N_2, N_4
					Tab A7 Lite	11	L_4	N_2, N_4
D	59	106,652	U_4	Samsung	J7 Prime 2	9	L_1, L_2, L_5	N_1, N_2, N_5
E	52	47,044	U_5	Samsung	J7	6.0.1	L_6	N_2, N_6
					A12	11	L_6	N_2, N_6

 doi:10.20944/preprints202504.0946.v1

[1] The threshold of flows is set to one second. [2] L_1 = ITC Lab; L_2 = District 5, Tehran; L_3 = District 11, Tehran; L_4 =Qom; L_5 = Karaj; L_6 = District 8, Tehran. [3] N_1 = University of Tehran; N_2 = TCI; N_3 =AsiaTech; N_4 =NTC; N_5 = Shatel; N_6 = MCI.

3.2. Experimental Setup

To ensure a comprehensive evaluation of the proposed approach, we examining its performance under two network conditions: an invariant network environment and a variant network environment. The following sections provide detailed descriptions of each environment, demonstrating the thoroughness of our approach.

3.2.1. Invariant Network Environment

To simulate the invariant network environment, we combined scenarios A to E into a single dataset. We then used the cross-validation method to split this dataset into training, validation, and test sets. The sets were divided into the following ratios: 80% for training, 10% for validation, and 10% for testing, based on the number of samples in each class.

3.2.2. Variant Network Environment

To evaluate model robustness and simulate a variant network environment, we used a cross-dataset validation method. For the validation set, we selected scenario E that was kept separate and not used in the rest of the evaluation process. This approach was taken to eliminate any bias during the optimization step. Although this may result in optimal parameters being tailored to this specific scenario, our findings in the subsequent sections demonstrate that these parameters also generalize effectively to other scenarios. In each experiment, we randomly selected three scenarios for training and the remaining one for testing. This process was repeated across all four possible training and testing set combinations, and the average model performance was reported.

3.3. Model Architecture

Open-set classifiers comprise two components: a closed-set classifier and an unknown class detector, and a positive correlation exists between closed-set accuracy and open-set performance. Specifically, studies in [30] demonstrate a high Pearson Product-Moment correlation, $\rho = 0.95$, between accuracy and AUROC, indicating a near-linear relationship. While a strong closed-set classifier is crucial, it alone may be insufficient for effective open-set recognition. Given that closed-set classifiers are well-established in the traffic identification field. Nevertheless, the detection of unknown traffic patterns remains largely unexplored. In this research, we select a strong classifier from existing works and focus on developing a novel method to recognize unknown traffic patterns. The proposed Framework for network traffic classification with unknown detector is illustrated in Figure 1.

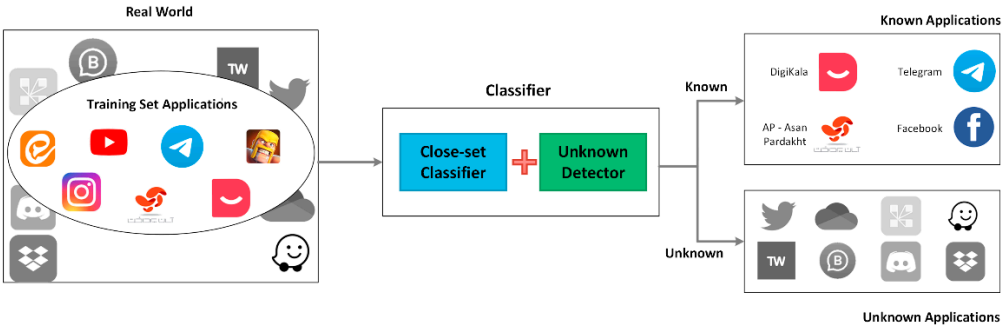


Figure 1. Framework for network traffic classification with unknown detector.

 doi:10.20944/preprints202504.0946.v1

3.3.1. Closed-Set Classifier

Given the vast range of existing closed-set classifiers, we selected a representative classifier for each of the three primary traffic classification approaches. This selection yielded the following classifiers: Flowprint [5] or behavior-based classification, App Scanner [20] for machine learning-based classification, and DT-ensemble [21]) for deep learning-based classification. We then assessed their performance in variable and fixed network environments to identify the most effective model.

As shown in Figure 1, the FlowPrint classifier exhibits greater robustness and experiences less performance degradation than the other three classifiers. This advantage is due to the distinct features used by FlowPrint compared to the other classifiers. Specifically, FlowPrint utilizes IP addresses, port numbers, and certificates as its features.

In comparison, the other two classifiers use statistical features such as packet length and packet arrival time interval in the flow. In a short time, statistical features may be affected by various network conditions, like latency and network congestion. Another key distinction is that FlowPrint employs the Bag of Flows (BoF) as its traffic object in decision-making. Although this led to a longer timeline, it helps the classifier make decisions based on more information and thus predict the label of the samples more accurately.

Given FlowPrint's robust performance across both stable and variable environments, we selected it as the closed-set classifier. Notably, FlowPrint incorporates an unknown data detection method, similar to the Maximum Softmax Probability (MSP) approach in OOD detection. This method classifies traffic as unknown when the maximum similarity between a test sample's fingerprint and the fingerprints of training applications falls below a predefined threshold. In subsequent sections, we will refine this approach to enhance unknown traffic detection in both stable and dynamic network environments.

3.3.2. Unknown Detector

Drawing inspiration from the methods presented in active learning, we propose our method to detect unknown traffic. Active learning is a technique wherein machine learning algorithms strategically select the most informative training data rather than relying on passive data acquisition. This approach is particularly advantageous when data labeling is expensive or time-intensive [31]. In active learning, the focus is on identifying low-confidence data to enhance model training. Conversely, in out-of-distribution or unknown data detection, we aim to identify the same low-confidence data but with the objective of preventing the model from making predictions on these instances. Building on the concept of 'uncertainty sampling' in active learning, we propose two novel methods as follows to detect unknown traffic.

- **Probability Anomaly Recognition (PAR):** The PAR method, inspired by the Best versus Second Best (BvSB) approach in active learning, utilizes the disparity between the highest and second-highest classification probabilities to make decisions. When the disparity is small, it indicates uncertainty in the model's prediction, with the highest probability being only slightly greater than the second-highest. This reduced disparity suggests that the sample is more likely to be unknown, as the model struggles to confidently distinguish between classes. The operation function of this detector is shown in equation (1) that $\hat{y}1$ and $\hat{y}2$ represent the classes with the highest and second highest estimated probabilities according to the θ model.

$$Label = \begin{cases} Known & if\ P_\theta(\hat{y}_1|x) - P_\theta(\hat{y}_2|x) > th; \\ Unknown & Otherwise; \end{cases} \quad (1)$$

- **Entropy-Based Uniformity Analysis (EnUniA):** The EnUniA method, inspired by entropy-based techniques in active learning, calculates the classification probabilities for all classes. Higher entropy values indicate a uniform distribution over the classes, suggesting that the model is uncertain about the correct classification. This uniformity in probability distribution is a strong indicator of the likelihood that the sample is unknown. Entropy values range from zero

 doi:10.20944/preprints202504.0946.v1

(indicating a clear, confident classification) to the logarithm of the number of classes (indicating maximum uncertainty). The operation function of this detector is shown in equation (2).

$$Label = \begin{cases} Known & if \ -\sum_i P_\theta(\hat{y}_i|x) \log P_\theta(\hat{y}_i|x) < th; \\ Unknown & Otherwise; \end{cases} \quad (2)$$

To compute these two metrics in the FlowPrint method, we first determine the similarity between each sample and all fingerprints obtained during the training phase. Since each application may have multiple fingerprints, we consider the maximum similarity value for each class. Subsequently, we apply the proposed methods to perform the final evaluation.

3.4. Optimal Threshold Determination

In the real world, a decision-making model requires a threshold for operation. However, determining the optimal threshold presents a challenge due to inherent trade-offs. Typically, threshold values are empirically derived and depend on specific use cases [32]. Furthermore, theoretical analysis demands the determining optimal thresholds to evaluate models beyond the scope of ROC/AUC metrics.

In this study, we employed the validation dataset to ascertain the optimal threshold. The validation set excludes samples from unknown classes to prevent potential bias toward treating a specific application as an unknown class during the threshold optimization process. We defined the optimal threshold as the point where the F1 score is maximized.

Specifically for each network condition, we partitioned the validation set into two equal halves for training and testing and evaluated FlowPrint's ability to classify known applications using our two proposed methods. The results are presented in Table 2.

It is important to note that since FlowPrint is based on the temporal correlation of destination features, we did not shuffle the data during the partitioning process.

Table 2. Summary of optimal threshold values found for each method and condition.

Method	Condition	Optimal value	F1-score (%)
MSP	Invariant Environment	0.089	81.41
	Variant Environment	0.143	91.54
PAR	Invariant Environment	0.005	82.25
	Variant Environment	0.001	92.53
EnUniA	Invariant Environment	5.601	81.37
	Variant Environment	5.580	91.51

4. Simulation Results

To comprehensively evaluate and compare our method against the baseline, we conducted a series of experiments, performing each under both variant and invariant network conditions, as outlined in section 3.2.

4.1. Expriment of Appearing New Applications (Unknown Classs)

An overall evaluation, independent of optimal threshold considerations, was conducted using the ROC-AUC metric. As illustrated in Figure 2, our proposed methods achieved superior performance over the baseline in both invariant and variant network environments.

 doi:10.20944/preprints202504.0946.v1

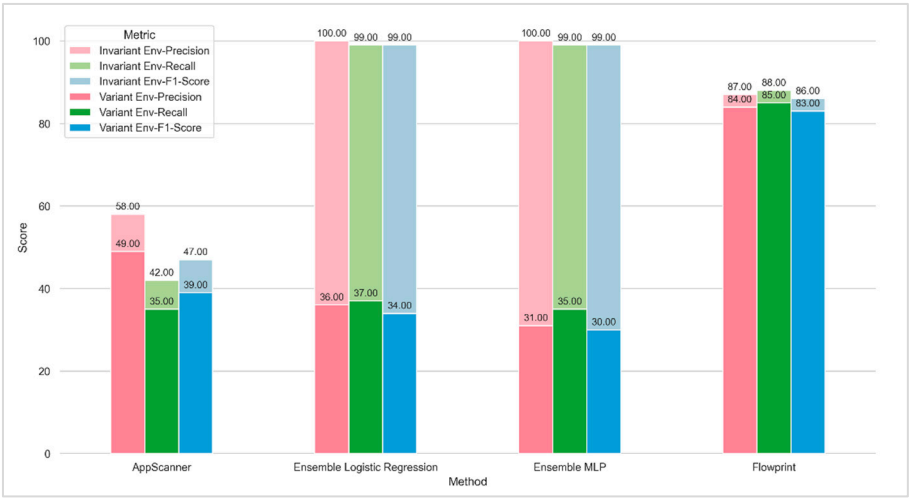


Figure 2. Robustness comparison of the Appscanner, FlowPrint, and DT-Ensemble models in the absence of unknown traffic.

In Figure 3 is illustrated the distribution of known and unknown samples from the FlowPrint classifier. As depicted, PAR and EnUniA concentrate unknown data within a specific region, facilitating identification via appropriate threshold selection.

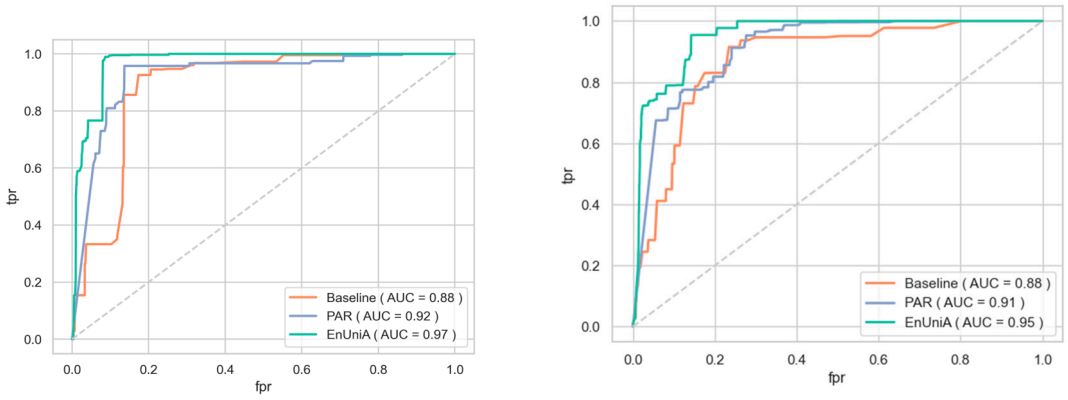


Figure 3. Comparison of ROC curves and AUC values for unknown traffic detection under (**a**) invariant and (**b**) variant network conditions. .

To enable a direct performance comparison with the authors' approach, our proposed methods were evaluated using the obtained thresholds. The results, presented in Table 3, detail FlowPrint's accuracy with various detectors in invariant and variant network environments. For concise analysis, each table entry reports the performance on the unknown class and the macro-average across all 50 classes (49 known, 1 unknown).

 doi:10.20944/preprints202504.0946.v1

Table 3. FlowPrint performance with different unknown traffic detectors under invariant and variant network environments.

Results of the Method and threshold proposed by the FlowPrint authors for an invariant network environment				Results of the Method and threshold proposed by the FlowPrint authors for a variant network environment			
	Precision	**Recall**	**F1-Score**		**Precision**	**Recall**	**F1-Score**
Unknown Traffic	0.59	0.21	0.29	Unknown Traffic	0.44	0.23	0.30
Macro average	0.85	0.87	0.84	Macro average	0.84	0.84	0.81
Results of the PAR method using the optimal threshold in an invariant network environment.				Results of the PAR method using the optimal threshold in a variant network environment.			
	Precision	**Recall**	**F1-Score**		**Precision**	**Recall**	**F1-Score**
Unknown Traffic	0.46	0.62	0.52	Unknown Traffic	0.46	0.67	0.54
Macro average	0.89	0.87	0.86	Macro average	0.87	0.84	0.83
Results of the EnUniA method using the optimal threshold in an invariant network environment.				Results of the EnUniA method using the optimal threshold in a variant network environment.			
	Precision	**Recall**	**F1-Score**		**Precision**	**Recall**	**F1-Score**
Unknown Traffic	0.82	0.49	0.60	Unknown Traffic	0.76	0.80	0.59
Macro average	0.87	0.88	0.85	Macro average	0.85	0.85	0.83

As demonstrated, our proposed methods are more efficient in both invariant and variant environments, improving unknown data detection by up to 50% and overall recognition accuracy by up to 2%. To address potential concerns regarding threshold optimization, we calculated the optimal threshold for the conventional method using the approach described in Section 3.4. As shown in **Table 4**, these thresholds yielded lower performance in both environments compared to the authors' proposed thresholds. Consequently, they were excluded from the comparative analysis.

Table 4. FlowPrint unknown traffic detection performance at optimal thresholds derived from: (a) invariant network environment, (b) variant network environment.

	Precision	**Recall**	**F1-Score**
Unknown Traffic	0.19	0.11	0.13
Macro average	0.84	0.87	0.83

(**a**)

	Precision	**Recall**	**F1-Score**
Unknown Traffic	0.51	0.05	0.08
Macro average	0.83	0.84	0.81

(**b**)

4.2. Simulation of Long-Term Performance

Destination features (IP address, port) and TLS certificates can change over time due to server replication/migration or certificate renewals. To assess the proposed model's long-term performance,

 doi:10.20944/preprints202504.0946.v1

we simulated these feature changes and measured their impact. Given that application IP addresses change on average once every three months, we modeled this using a Poisson random variable with rate = 1/3. For 49 known applications in the training dataset, assuming independence, the combined rate follows a Poisson random variable with rate = 49/3. We assumed the same change rate for TLS certificates, resulting in a second Poisson random variable with rate = 49/3. We conducted 500 simulations over a 36-month period. Based on Monte Carlo methodology, with a variance error of 0.1 and a 95% confidence level, the minimum simulation count required is 384; we opted for 500 simulations. During each simulation, we trained the model on the training data and randomly modified IP addresses and TLS certificates in the testing data according to the two Poisson random variables. Given the domain-based nature of TLS certificates, this random selection provides a suitable approximation of the model's performance.

Simulation results in both network environments are depicted **Figure 4**. The two proposed methods demonstrated superior performance compared to the authors' baseline.

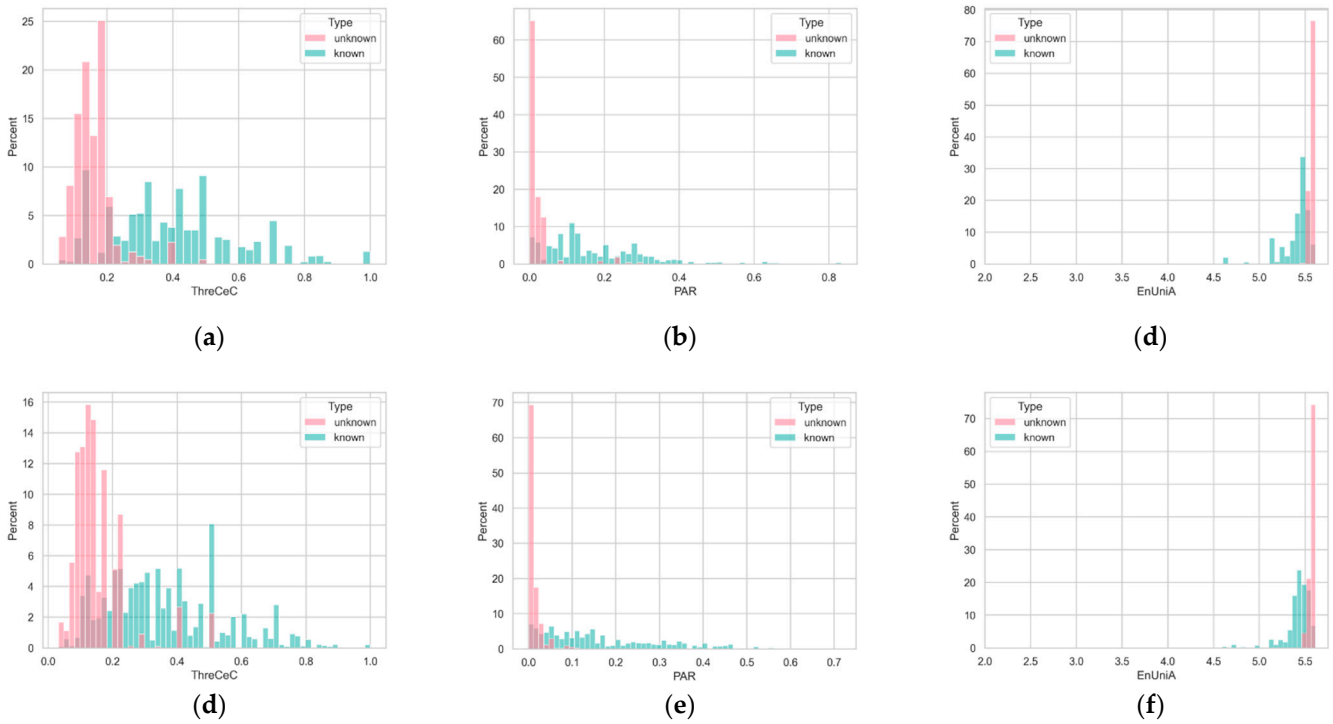
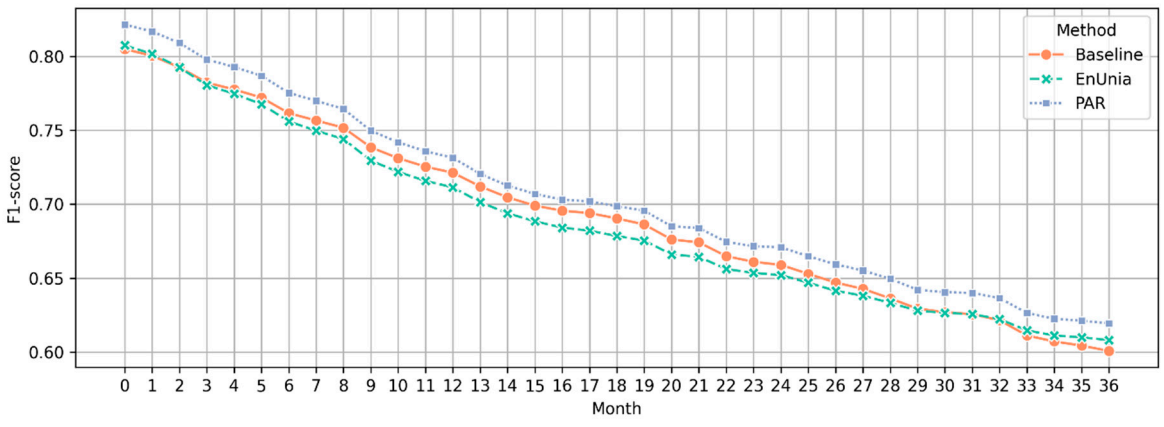
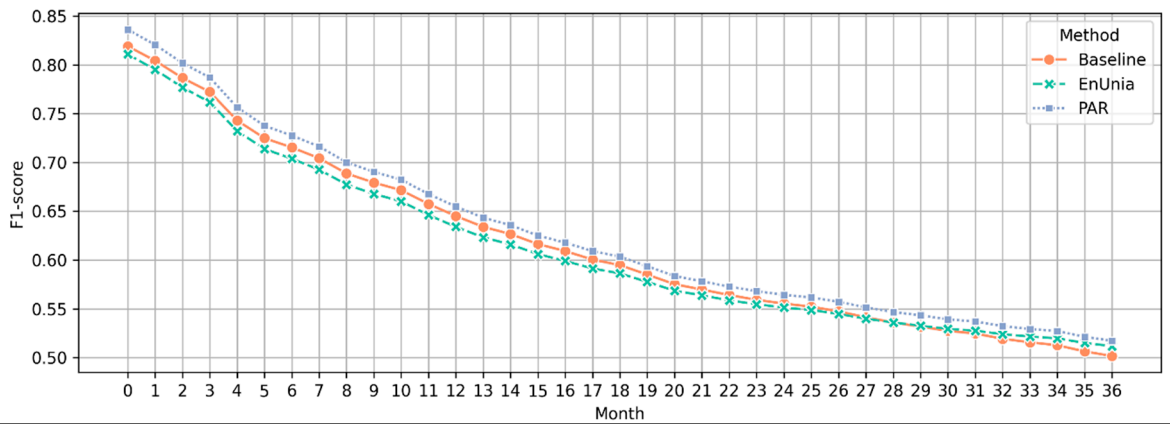


Figure 4. Distribution diagrams comparing known and unknown data in invariant (**a-c**) and variant (**d-f**) network environments, using Baseline (**a, d**), PAR (**b, e**), and EnUnia (**c, f**).



 doi:10.20944/preprints202504.0946.v1



(**b**)

Figure 5. Results of long-term performance simulation under (**a**) invariant network environment, and (**b**) variant network environment.

5. Discussion

Our proposed unknown traffic detection methods, PAR and EnUniA, outperform the Baseline. This superiority is attributed to the inherent tendency of classifiers to identify similarities between entirely novel data and trained classes, a consequence of traffic homogeneity. The assumption that a model can assign unknown data to a single class with negligible probability is, therefore, unrealistic. Leveraging this observation and drawing inspiration from active learning sampling techniques, we developed PAR and EnUniA. These methods exhibit a significant advantage over Baseline for classifiers utilizing traffic objects larger than flows, while performing comparably for those using sub-flow objects. This distinction arises from the binary nature of flows: they are either homogeneous or heterogeneous. When analyzing multi-flow objects, such as bag flows, homogeneous flows demonstrate similarity with known classes, enhancing the performance of PAR and EnUniA. However, with sub-flow objects, only homogeneous sub-flows benefit from our proposed methods.

Overall, PAR and EnUniA demonstrate superior performance compared to Baseline. Furthermore, their minimal computational overhead and straightforward post-processing make them broadly applicable across various classifier types.

6. Conclusions

Network traffic identification and classification are essential for network management and security, supporting applications such as capacity planning and anomaly detection. However, dynamic real-world networks present challenges due to changing traffic feature distributions. This study addresses the need for robust classifiers and the detection of potentially hazardous unknown traffic. Inspired by active learning and exploiting application traffic homogeneity, we present a robust open-set recognition (OSR) classifier, utilizing FlowPrint, a state-of-the-art method for closed-set classifier and PAR and EnUniA methods for unknown traffic detection. Compared to the base classifier, our methods demonstrate superior performance. In particular, EnUniA improves FlowPrint's unknown traffic detection in variant environments by approximately 30% (76% precision, 50% recall, and 59% F1-score) and enhances overall classifier performance by 2%.

Author Contributions: Conceptualization, M.B.; methodology, M.B.; software, M.B. validation, M.B.; data curation, M.B.; writing—original draft preparation, M.B., Z.N.; writing—review and editing, M.B., Z.N., M.T.; visualization, M.B., Z.N.; supervision, M.T.; project administration, M.T.;

Funding: This research received no external funding.

 doi:10.20944/preprints202504.0946.v1

Data Availability Statement: The data utilized in this study can be freely and openly accessed on Mendeley Data under the name ITC-Net-Blend-60.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ETI	Encrypted Traffic Intelligence
NTC	Network Traffic Classification
ML	Machine Learning
DL	Deep Learning
OSR	Open-set Recognition
OOD	Out-of-Distribution
IoT	Internet of Things
MSP	Maximum SoftMax Probability
PAR	Proximity Ambiguity Resolver
EnUnia	Entropy-based Uncertainty Analyzer
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
BoF	Bag of Flow

References

1. M. H. Pathmaperuma, Y. Rahulamathavan, S. Dogan, A. M. Kondoz, and R. Lu, "Deep Learning for Encrypted Traffic Classification and Unknown Data Detection," *arXiv preprint arXiv:2203.15501*, 2022.
2. M. Shen *et al.*, "Machine Learning-Powered Encrypted Network Traffic Analysis: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials*, 2022.
3. J. Liu, J. Wang, T. Yan, F. Qi, and G. Chen, "Unknown Traffic Recognition Based on Multi-Feature Fusion and Incremental Learning," *Applied Sciences*, vol. 13, no. 13, p. 7649, 2023.
4. A. Azab, M. Khasawneh, S. Alrabaee, K.-K. R. Choo, and M. Sarsour, "Network traffic classification: Techniques, datasets, and challenges," *Digital Communications and Networks*, vol. 10, no. 3, pp. 676-692, 2024, doi: 10.1016/j.dcan.2022.09.009.
5. T. van Ede *et al.*, "FlowPrint: Semi-supervised mobile-app fingerprinting on encrypted network traffic," in *Network and Distributed System Security Symposium (NDSS)*, 2020, vol. 27.
6. M. Lotfollahi, M. J. Siavoshani, R. S. H. Zade, and M. Saberian, "Deep packet: A novel approach for encrypted traffic classification using deep learning," *Soft Computing*, vol. 24, no. 3, pp. 1999-2012, 2020.
7. Y. Feng *et al.*, "Unmasking the Internet: A Survey of Fine-Grained Network Traffic Analysis," *IEEE Communications Surveys & Tutorials*, 2025.
8. M. S. Sheikh and Y. Peng, "Procedures, Criteria, and Machine Learning Techniques for Network Traffic Classification: A Survey," *IEEE Access*, 2022.
9. W. Li, X.-Y. Zhang, H. Bao, Q. Wang, and Z. Li, "Robust network traffic identification with graph matching," *Computer Networks*, vol. 218, p. 109368, 2022.
10. W. Li, X.-Y. Zhang, H. Bao, H. Shi, and Q. Wang, "ProGraph: Robust Network Traffic Identification With Graph Propagation," *IEEE/ACM Transactions on Networking*, 2022.
11. N. Malekghaini *et al.*, "Deep learning for encrypted traffic classification in the face of data drift: An empirical study," *Computer Networks*, vol. 225, p. 109648, 2023.
12. J. Yang, K. Zhou, Y. Li, and Z. Liu, "Generalized out-of-distribution detection: A survey," *International Journal of Computer Vision*, vol. 132, no. 12, pp. 5635-5662, 2024.
13. J. Zhang, F. Li, F. Ye, and H. Wu, "Autonomous Unknown-Application Filtering and Labeling for DL-based Traffic Classifier Update," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, 2020: IEEE, pp. 397-405.

 doi:10.20944/preprints202504.0946.v1

14. "Mobile internet traffic as percentage of total web traffic in January 2025, by region." statista. https://www.statista.com/statistics/306528/share-of-mobile-internet-traffic-in-global-regions/ (accessed 25 March, 2025).
15. C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications surveys & tutorials,* vol. 21, no. 3, pp. 2224-2287, 2019.
16. V. F. Taylor, R. Spolaor, M. Conti, and I. Martinovic, "Robust smartphone app identification via encrypted network traffic analysis," *IEEE Transactions on Information Forensics and Security,* vol. 13, no. 1, pp. 63-78, 2017.
17. A. Biernacki, "Identification of adaptive video streams based on traffic correlation," *Multimed. Tools Appl.,* pp. 1-21, 2019.
18. A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques," in *Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, 2005, pp. 50-60.
19. A. Madhukar and C. Williamson, "A longitudinal study of P2P traffic classification," in *14th IEEE International Symposium on Modeling, Analysis, and Simulation*, 2006: IEEE, pp. 179-188.
20. V. F. Taylor, R. Spolaor, M. Conti, and I. Martinovic, "Appscanner: Automatic fingerprinting of smartphone apps from encrypted network traffic," in *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2016: IEEE, pp. 439-454.
21. O. Aouedi, K. Piamrat, and B. Parrein, "Ensemble-based Deep Learning model for network traffic classification," *IEEE Transactions on Network and Service Management,* 2022.
22. H. Ozkan *et al.*, "Multimedia traffic classification with mixture of Markov components," *Ad Hoc Networks,* vol. 121, p. 102608, 2021.
23. M. Shen, M. Wei, L. Zhu, and M. Wang, "Classification of encrypted traffic with second-order markov chains and application attribute bigrams," *IEEE Transactions on Information Forensics and Security,* vol. 12, no. 8, pp. 1830-1843, 2017.
24. G. D'Angelo and F. Palmieri, "Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction," *Journal of Network and Computer Applications,* vol. 173, p. 102890, 2021.
25. X. Ren, H. Gu, and W. Wei, "Tree-RNN: Tree structural recurrent neural network for network traffic classification," *Expert Syst. Appl.,* p. 114363, 2020.
26. C. Xu, J. Shen, and X. Du, "A method of few-shot network intrusion detection based on meta-learning framework," *IEEE Transactions on Information Forensics and Security,* vol. 15, pp. 3540-3552, 2020.
27. S. Cruz, C. Coleman, E. M. Rudd, and T. E. Boult, "Open set intrusion recognition for fine-grained attack categorization," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, 2017: IEEE, pp. 1-6.
28. Y. Zhang, J. Niu, D. Guo, Y. Teng, and X. Bao, "Unknown network attack detection based on open set recognition," *Procedia Computer Science,* vol. 174, pp. 387-392, 2020.
29. M. Bayat *et al.*, "ITC-Net-blend-60: a comprehensive dataset for robust network traffic classification in diverse environments," *BMC Research Notes,* vol. 17, no. 1, p. 165, 2024.
30. S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Open-Set Recognition: a Good Closed-Set Classifier is All You Need?," in *International Conference on Learning Representations*, 2022.
31. B. Settles, "Active learning literature survey," 2009.
32. L. Du, Z. Gu, Y. Wang, and C. Gao, "Open world intrusion detection: An open set recognition method for can bus in intelligent connected vehicles," *IEEE Network,* 2024.