International Journal of New Computer Architectures and their Applications (IJNCAA) 10(2): 23-31
The Society of Digital Information and Wireless Communications, 2020 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Network Traffic Classification Using Ensemble Learning with Time Related Features

Muhammad Anwarul Azim, Tanvir and Mohammad Khairul Islam
Department of Computer Science and Engineering
University of Chittagong, Chittagong – 4331, Bangladesh
azim@cu.ac.bd, tanvir.cse@std.cu.ac.bd and mkislam@cu.ac.bd

ABSTRACT

Network traffic classification has become important with the rapid growth of the Internet and online applications. Though, there were researches that applied different machine learning algorithms for traffic classification purposes, the continuous expansion of technologies and applications in stationary and mobile are creating a dynamic environment. Because of encryption in today's Internet, traffic classification still poses a great deal of concern for researchers and network communities. This work proposes ensemble learning including Voting, Bagging, and Boosting for traffic classification, and then compares them with their own base classifiers when used individually. Time-related features are focused which are independent of data encryption on the UNB ISCX dataset, containing flow duration, inter-arrival time, byte rate, packet rate, etc. Among different techniques, Random Forest outperforms nearly all others with respect to various evaluation matrices such as accuracy, precision, recall, and f1-score. In the case of VPN traffic and Non-VPN traffic, it gives almost 90.65% and 95.42% accuracy respectively. In the case of combined VPN and Non-VPN traffic, we achieve 90.18% accuracy for classifying traffic categories which is a significant improvement from previous works.

KEYWORDS

Network Traffic Classification, Encrypted Traffic, Time Related Features, Ensemble Learning and Machine Learning.

1 INTRODUCTION

Network traffic refers to the amount of data moving across a network at a given point in time. In networking, the term throughput means how much data was transferred from a source in a specific duration. Besides, bandwidth tells how much data could theoretically be transferred from a source at an instance in time. Due to the rapid growth of high traffic throughput on various applications and high traffic bandwidth on the Internet, classification of network traffic is essential in advanced network management system for better resource allocation and Quality of Service (QoS). As an example of the importance of network traffic classification, one can think of the asymmetric architecture of today's network access links which have been designed based on the assumption that clients download more than they upload. However, the pervasiveness of symmetric demand applications (such as peer-to-peer applications, voice over IP, video calls, etc.) has changed the clients' demand to deviate from the assumption of asymmetric nature to the symmetric. Thus, to provide satisfactory network usability to the user, various application-level knowledge is required to provide proper resources to such symmetric applications. Recently, with the widespread use of encryption techniques in network applications, traffic encryption has become the standard practice which brings new challenges to traditional traffic classification methods. Now, on the Internet, most of the traffic encrypted by applications is known as regular encryption (e.g., WhatsApp) but in some cases, this encrypted traffic can be further encrypted through protocol encapsulation (e.g., Skype traffic through VPN). There are various types of features for classifying network traffic such as payload-based, header-based but they are not encryption independent, therefore, we have utilized time-related features that are encryption independent. This paper focuses on regular, encrypted traffic classification and protocol encapsulated traffic classification for the identification of traffic type. Ensemble learning includes a wide range of research efforts that seek to find the best method to build a combination of classifiers. The main

contributions of this paper are: 1) Developing network traffic classifier model using various ensemble learning techniques for the robustness of the model. 2) Observing the effect of using various types of real-world scenarios with different flow time out durations. 3) Comparison with other classical and base machine learning approaches to evaluate the effectiveness of this proposed work.

2 BACKGROUND

The whole world is dependent on the Internet nowadays. Therefore, various types of applications are emerging day by day which belong to a wide variety of traffic types. Moreover, in today's Internet, people need security as there are possibilities of various cyber-crimes such as cyber stealing, information leakage, etc. So, there are lots of encryption in today's network traffic to provide security and privacy to the user data on the internet which other people can obtain. This continuous evolution and creation made the traffic classification process even harder for the network community as well as to the researchers.

2.1 Network Traffic Classification

A complete network module would provide both security and better QoS to its users. Security includes both intrusion detection and malware detection. An intrusion detection system is software or hardware designed to detect any malicious activity or attack against the system or network [1]. In general term malware stands for malicious software which has been designed to achieve some targets such as collecting sensitive data, accessing private computer systems even sometimes harming the systems [2]. QoS is related to Network traffic classification in which network traffic is classified into different categories of traffics according to their types such as email, chat, voice call, video call, etc. It is also a basic requirement for providing a good QoS in network management which is the ability to provide various band allocations to different traffic types. Network traffic can be classified into two basic processes: online and offline. In online, traffic is instantly classified into different categories according to their patterns. The online process is used for QoS, routing, etc. purposes. In offline, at first network traffic is collected and stored, then it is classified into different categories according to their patterns. This process is used for pricing, analyzing, anomaly detection, etc. The proposed system can work in both online and offline processes though in the online process there would be a little bit of latency.

2.2 Machine Learning

Machine learning (ML) has emerged as a useful technique for modeling problems that are otherwise difficult to formulate exactly. Traditionally, computers are manually programmed to perform a task. Using machine learning, some portion of the human contribution can be replaced by a machine learning algorithm. With the increase of computational capacity and data, machine learning is becoming more and more practically attractive over the years. Learning algorithms are widely used nowadays in network traffic classification.

For network traffic classification purposes in this work, some classical machine learning techniques are used. They are the Decision Tree (DT) and K-Nearest Neighbors (KNN). A short description of them is given below:

DT: A Decision tree is a flowchart like a tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

KNN: It depends on the metric used to calculate the distance between example points. The output of the classification is a class membership, which is determined according to the majority vote of its K nearest neighbors.

2.3 Ensemble Learning

Ensemble learning tries to improve results by combining several machine learning models. This technique allows the creation of a better predictive model compared to a single model. Ensemble models combine the decisions from various machine learning models to enhance the overall performance. Most ensemble methods use a single base learning algorithm to produce

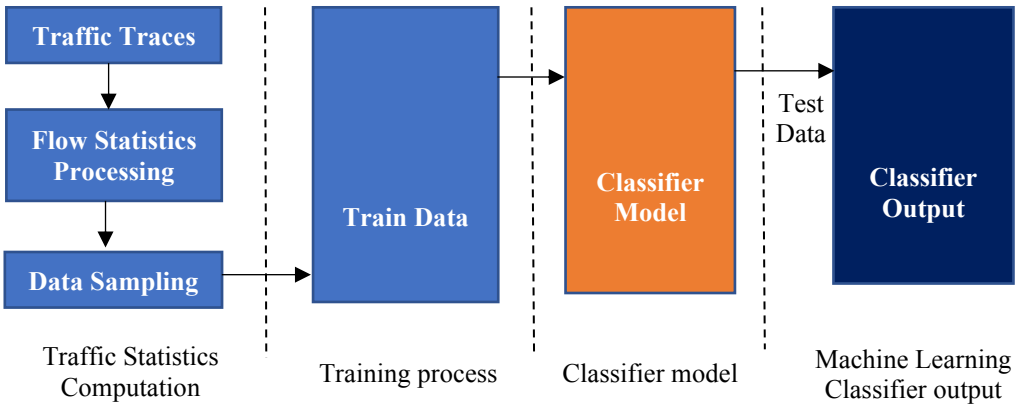

Figure 1. Classical Machine Learning Architecture for Network Traffic Classification

homogeneous base models, i.e., models of the same type, leading to Homogeneous Ensembles. Some methods use heterogeneous models, i.e., models of different types, leading to Heterogeneous Ensembles. To be more accurate than any of its individual members in ensemble methods, the base models have to be as accurate and diverse as possible.

Ensemble learning is of various types, such as Bagging, Boosting, Voting, and Stacking. The Random Forest classifier is an extension of the Bagging classifier. In this research, both Bagging (BG) and Random Forest (RF) are used for the comparison of their classification results. In the case of Boosting, Gradient Boost (GB) is used and in the case of Voting, Max Voting (MV) is used in this research. Stacking takes more time than all other ensemble learning methods. Due to the infeasibility of the running time of Stacking, we don't use it here.

A list of used ensemble learning techniques is given below:

MV: In Max Voting each base model makes a prediction and votes for each sample. Only the sample class with the highest votes is included in the final predictive class. For the base model in this work Decision Tree, K-Nearest Neighbors, and Random Forest was used.

Bagging: A Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction.

Gradient Boost: It produces a prediction model in the form of an ensemble of weak prediction models, typically a decision tree. It builds the model in a stage-wise fashion as other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Random Forest: Random Forest is also built on the Bagging technique but while Bagging considers all the features in data to build the model, Random Forest randomly selects a number of features for each sub-set to build the sub-models to combine the prediction.

3 RELATED WORK

From the early days of the Internet, network traffic classification has been a major concern for the network research community.

In Subhabarta et al. [3], there was port-based IP traffic classification. However, due to dynamic port allocation nowadays it's not efficient. Then the payload-based traffic classification method came to light. But, due to the complexity of the processes and encryption in today's IP traffic, it's also not feasible. The next traffic classification approach was based on statistical traffic properties (such as the distribution of flow duration, flow idle time, packet inter-arrival time, and packet length), Vern et al. [4]. After a statistical property-based approach, a traffic classifier needed to handle a large amount of data, thereby the machine learning approaches appeared in existence. Nguyen et al. [5] and Velan et al. [6] provided surveys of network traffic classification research done using ML.

In the area of network traffic classification, existing work exploring the use of meta-learning is limited, He et al. [7] used Stacking and Voting and proposed a new machine learning model that combines ensemble learning with co-training techniques. The work of Wang et al. [8] proposed a classification approach based on sub-flow characteristics using meta-learning. In their work, a flow truncation method was developed for real-time processing, and an aggregation machine learning system based on the accuracy of each

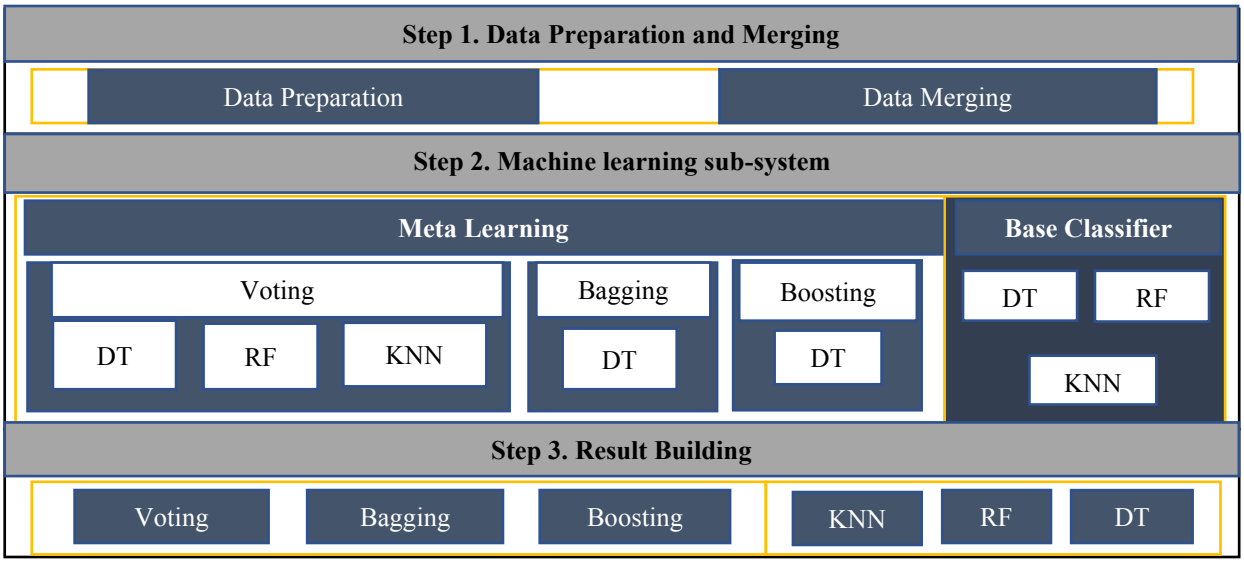

Figure 2. Ensemble Learning Architecture for Network Traffic Classification.

classifier for different applications. Gmez et al. [9] also use ensemble learning for classifying network traffic. The authors compare the performance of seven popular ensemble algorithms based on Decision Trees, focusing on model accuracy, latency, and byte accuracy. Additionally, an ensemble classifier is presented. However, they used a combination of packet size-based and Transmission control protocol window-based features and extremely large datasets. Possebon et al. [10] also used a meta-learning approach for traffic classification, however, the dataset they used was not encryption independent and was not versatile enough.

In this work, we tried to solve the traffic classification problem using a meta-learning approach in consideration of traffic encryption and achieved better accuracy to identify the traffic type.

4 METHODOLOGY

This section discusses our proposed method of network traffic classification using classical machine learning approaches and ensemble learning. The system architecture and framework of the Network Traffic Classification model are also discussed here.

4.1 Classical Machine Learning

The architecture of the Network Traffic Classifier using Classical Machine Learning is shown in Fig. 1. Here are the steps to this process:

- First, a mix of 'traffic traces' is collected that contain both instances of the application of interest and instances of other interfering applications (such as Browsing, Chat, Streaming, etc.).
- The "flow statistics processing" step involves calculating the statistical properties of these flows (such as mean forward inter-arrival time, duration, etc.) to generate time-related features.
- An optional next step is "data sampling", designed to narrow down the search space for the ML algorithms when faced with extremely large training datasets (traffic traces). The sampling step extracts statistics from a subset of instances of various application classes and passes these along to the classifier to be used in the training process.
- A Train data process is used to train the ML classifier. For this research Decision Tree, KNN, and Random Forest were used to create the classification model.
- The next part of this process is a classification model.
- Then test data is used in the classification model for evaluating the performance of the model.

4.2 Ensemble Learning

Ensemble methods combine several decision tree classifiers to produce better predictive performance than a single decision tree classifier. The main principle behind the ensemble model is that a group of individual learners come together to form a strong learner, thus, increasing the

accuracy of the model. Fig:2 represents the architecture of Ensemble Learning for Network Traffic Classification. There are three steps in Ensemble learning architecture for Network Traffic Classification, they are discussed below.

Step 1: Data Preparation and Merging: The ISCX [11] dataset from the Canadian Institute of Cyber Security has been used for our traffic

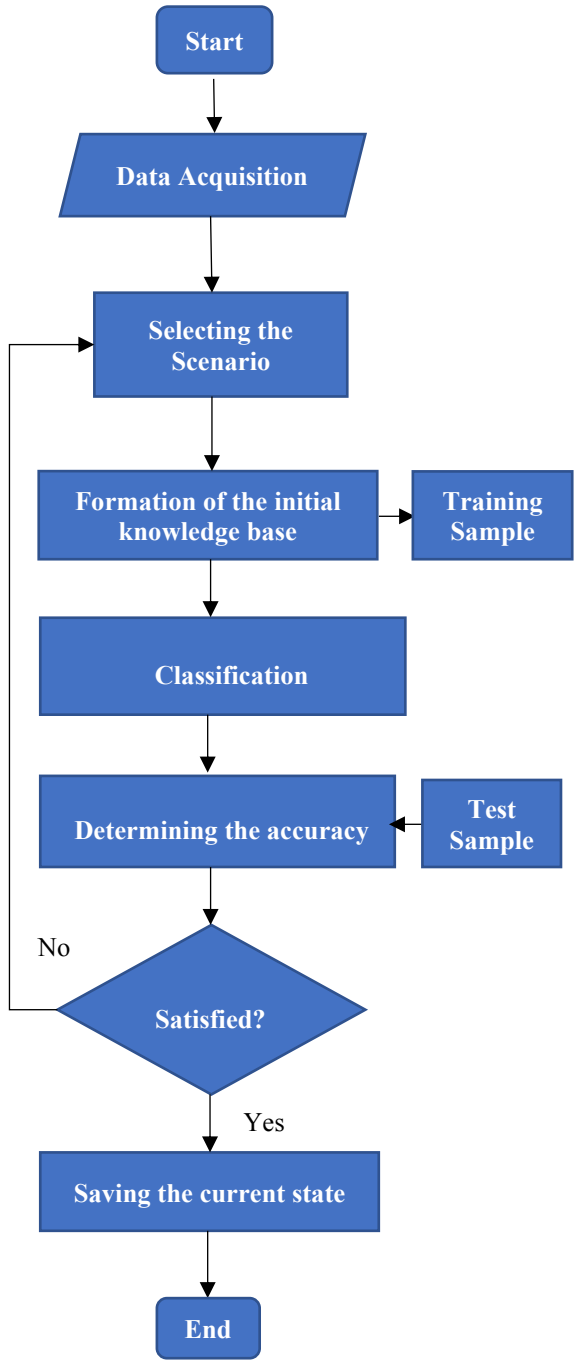

Figure 3. Work Flow of Network Traffic Classification

classification. To do our research we changed the format of data from Attribute Relation File Format (.arff) to Comma Separated Value (.csv). To increase the number of data we have merged the dataset of 15s duration and 30s duration.

Step 2: Machine Learning Sub-system: This step may be called the classification stage also. The processed data from the data collection and labeling step is used here for meta-learning and also for base classifiers.

Classifier: For base classifier Decision Tree, K – Nearest Neighbors and Random Forest were used in this work. The reason behind using these classifiers was the fact that these are unrelated algorithms with different error rates for the same dataset and widely used in the literature in the context of meta-learning.

Meta-Learning: We have used Voting, Bagging, and Boosting as meta-learning techniques. The parameters of base classifiers were kept the same in both ensemble learning and individual machine learning for a fair evaluation. The Boosting technique used here is the Gradient Boost. For Voting, Max Voting was used with Decision Tree, Random Forest, and K-Nearest Neighbors. For Bagging, Decision Tree was used as the base. Random Forest is also included in Bagging. The fundamental difference between Random Forest and Bagging is that in Random forests, only a subset of features is selected at random out of the total and the best split feature from the subset is used to split each node in a tree, unlike in bagging where all features are considered for splitting a node.

Step 3: Result Building: To evaluate each technique, different metrics were produced from the obtained results including recall, precision, F1-score and mean accuracy for a given number of experimental repetitions.

4.3 Workflow of Network Traffic Classification

There are several types of traffic available on the Internet. Based on encryption, there are both encrypted and unencrypted traffics. Encrypted traffic might be with a VPN or without a VPN and the same case for unencrypted traffic too. For better security purposes, lots of data on the Internet is encrypted. All these variations on the data pose a lot of difficulties in the traffic classification problem.

The workflow is given for classifying network traffic according to their types:

I. In the first step, the fragment of the intercepted traffic is collected.
II. The classifier scenario is selected.
III. Based on the features indicated in the scenario, a training sample is formed to build the initial knowledge base.
IV. After analyzing the given features of the test sample, the classification process is accomplished.
V. If the accuracy satisfies the requirements, then the current state is saved. Otherwise, the cycle returns to the selection of the type of scenario.

A block diagram of this whole process is shown in Figure 3.

5 EXPERIMENTAL ENVIRONMENT

The proposed method is implemented on the Google Colab which is a Jupyter notebook that run in the cloud and are highly integrated with Google Drive making them easy to set up, access and share and it is free for Machine Learning researchers. The system has 12 Gigabytes of RAM, Intel Xeon CPU clocked at 2.20GHz, and a powerful Nvidia Tesla K80 GPU. We implement the program in python programming language using the PyCharm IDE. The libraries require in this experiment are Scikit-learn, NumPy, Pandas and Matplotlib.

5.1 Data Collection and Labeling

We use the ISCX internet traffic dataset [11] provided by the Canadian Institute of Cyber Security. This is a real-world internet traffic dataset with enough diversity and quantity. The dataset is created by transferring messages between two users Alice and Bob using applications like Skype, Facebook etc. The dataset is distributed in two types such as: VPN and Non-VPN. Each type is categorized into 7 classes such as Chat, Mail, File Transfer, Browsing, P2P, VoIP and Streaming. This arrangement is known as Scenario A. The classes are reorganized to generate another scenario known as Scenario B. In scenario B, we keep both VPN and non-VPN data of every class together. Thus, we generate 7 classes from VPN and non-VPN data. The dataset includes a collection of real-world data, generating flow-based statistical time-related features, and labeling them according to their classes, for example, labeling Firefox and Google Chrome data as the Browsing category.

5.2 Dataset Description

This dataset contains a regular session and a session over VPN, therefore, containing a total of 14 traffic categories: VOIP, VPN-VOIP, P2P, VPN-P2P, etc. The dataset includes 4 flow out durations: 15s, 30s, 60s, 120s. The dataset contains traffic of various types of protocols, the list of captured protocols and applications are given in Table 1.

5.3 Scenarios and Features

The dataset has two scenarios, they are: Scenario A, Scenario B and is based on flow based statistical time related features.

Scenario A: In scenario A, at first, traffic data is divided into two basic classes: VPN and Non-VPN. Then for VPN data, all classes are combined, so there are VPN Browsing, VPN chat, VPN mail, VPN streaming, VPN P2P, VPN FT, and VPN VoIP. The same categorization is true for Non-VPN data. There are also 7 classes for data without the VPN, they are browsing, mail, chat, VoIP, streaming, P2P, and FT. Here we considered VPN class.

Table 1. List of captured protocols and applications

Traffic	Protocols and Applications
Web Browsing	Firefox and Chrome
Email	SMTPS, POP3S and IMAPS
Chat	ICQ, AIM, Skype, Facebook and Hangouts
Streaming	Vimeo and YouTube
File Transfer	Skype, FTPS and SFTP using FileZilla and an external service.
VoIP	Facebook, Skype and Hangouts voice calls (1h duration)
P2P	uTorrent and Transmission (BitTorrent)

Table 2. Accuracy comparison between our proposed method and existing methods for **VPN** data of scenario A. The values represent accuracy in percentage. The best results are given in bold-face with blue color.

Methods	Scenario A, VPN			
	15s	30s	60s	120s
Gill et al. [11]	83.71	85.29	79.42	81
Battalov et al. [12]	83	84.6	81.7	83.8
Andres et al. [13]	88.28	89.03	86.66	86.43
This Work	**90.65**	**88.36**	**95.42**	**94.14**

Table 3. Accuracy comparison between our proposed method and existing methods for **non-VPN** data of scenario A. The values represent accuracy in percentage. The best results are given in bold-face with blue color.

Methods	Scenario A, Non-VPN			
	15s	30s	60s	120s
Gill et al. [11]	88.14	85.43	85.43	84.23
Battalov et al. [12]	88.9	87.5	85.4	88.2
Andres et al. [13]	93.09	91.75	**90.63**	91.74
This Work	**95.42**	**93.03**	89.95	**94.14**

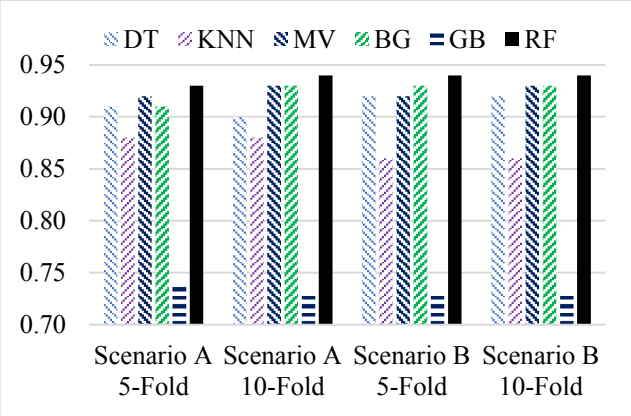

Figure 4. Accuracy of Scenario A & Scenario B using different classifiers with 5-Fold & 10-Fold Cross Validations.

the VPN and Non-VPN classes were combined which made a total of 14 classes, they are VPN browsing, browsing, VPN chat, chat, VPN mail, mail, VPN streaming, streaming, VPN P2P, P2P, VPN FT, FT, VPN VoIP, VoIP. In the second scenario, all these VPN and Non-VPN data are

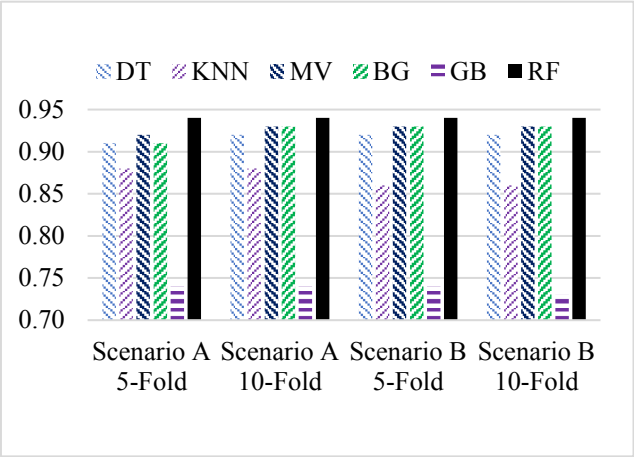

Figure 5. Precision of Scenario A & Scenario B using different classifiers with 5-Fold & 10-Fold Cross Validations.

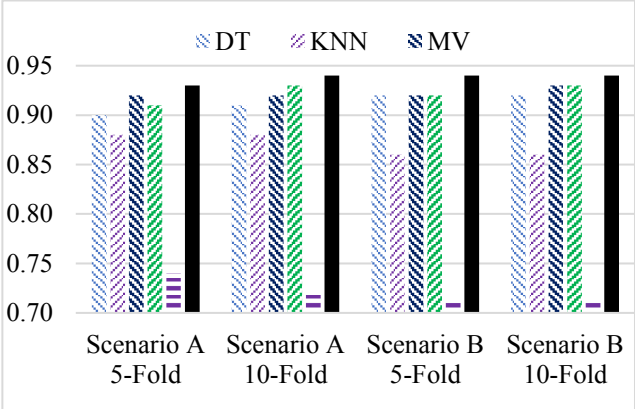

Figure 6. F1-Score of Scenario A & Scenario B using different classifiers with 5-Fold & 10-Fold Cross Validations

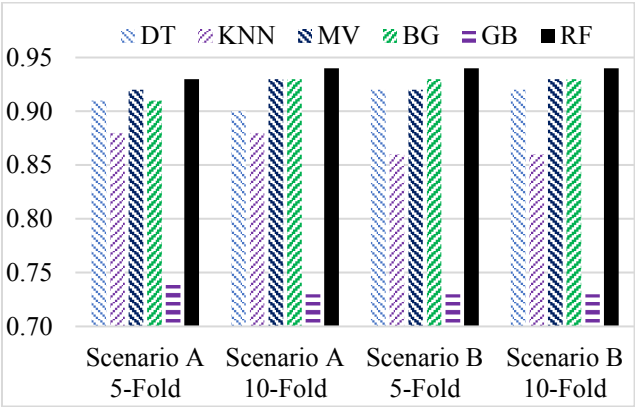

Figure 7. Recall of Scenario A & Scenario B using different classifiers with 5-Fold & 10-Fold Cross Validations

combined to form a single dataset where classes are: browsing, chat, mail, FT, VoIP, streaming, and P2P.

Features: There are various types of features for traffic classification such as port-based, payload-based, statistical-based, etc. To make this system encryption independent time related statistical-based features were used. In time-related features,

statistical features like minimum, maximum, mean, and standard deviation are calculated to generate from this dataset where the direction of the traffic flow is considered.

6 RESULTS & DISCUSSION

There are various types of evaluation metrices to measure the quality of the machine learning models. The following evaluation metrics are used to measure the quality of a machine learning model in our experiment:

(i) Recall: Recall is the ratio of correctly predicted positive observations to all observations in a class, also known as Sensitivity.

(ii) Precision: The ratio of correctly predicted positive observations to the total predicted positive observations, also known as Specificity.

(iii) F1-score: Weighted average of Recall and Precision.

(iv) Accuracy: The ratio of correctly predicted observation to the total observations.

In our experiment we use both 5-fold and 10-fold cross validations. In case of VPN classes of scenario A, the comparison is shown in Table 2. In case of No-VPN part of scenario A, the comparison with existing work is represented in Table 3.

Scenario B represents a more real-world framework, where while on the Internet some data might be encrypted through a VPN or some data might not. In case of scenario B, we have also obtained the accuracy and precision of 90.18% and 0.85 respectively via a very simplistic approach rather than a complex process where accuracy was 84% [12] [13] and precision was 78% [11]. Our research is a noteworthy improvement in traffic classification.

In addition to that, to measure the effectiveness of our proposed work we have combined the data of 15s and 30s durations to generate a new dataset where in scenario A we consider VPN part and in scenario B we consider the combination where all VPN & Non-VPN data was combined to generate 7 classes. Among the validation sets, we have got the best result for 10-fold cross validation which is almost 94% of accuracy and precision. Accuracy, precision, f1-Score and recall of this obtained result are shown in Figure 4, Figure 5, Figure 6 and Figure 7 respectively.

7 CONCLUSION & FUTURE WORK

Network traffic classification is a beneficial task in today's Internet world for a better quality of service, security and other purposes. It's also important for the Internet research community. Ensemble learning provides a more robust structure to a classifier model which combines multiple machine learning models to build a more powerful model for classification purposes.

This work explores the opportunities and methods that are required to classify network traffic into different categories according to their patterns. We have used classical machine learning methods namely, Decision Tree and K-Nearest Neighbors which were applied to classify network traffics in various scenarios such as VPN or Non-VPN and on different applications such as Facebook, Web Browser, VoIP etc. Some ensemble learning techniques were also used to resolve the traffic classification. In the case of ensemble learning, the Bagging, Random Forest, Max Voting (Decision Tree, Random Forest, K-Nearest Neighbors as the base classifier), Boosting (Gradient Boost) have been used.

Among them, RF nearly outperforms all other methods. After RF, Bagging comes to the scene as a second-best performer.

All these models have been used in the ISCX dataset which provides a wide variety of data (Chat, Mail, Browser, etc.) and different scenarios (VPN or non-VPN) of real-world network traffic. In the case of scenario A, we achieve 90.65% and 95.42% of accuracy for VPN and non-VPN data respectively. For scenario B, we have achieved 90.18% accuracy which is a significant improvement from the best previous work with accuracy of 89%.

Furthermore, we have combined the dataset of 15s & 30s durations for both scenario A & scenario B to measure the effectiveness of this work and achieved a satisfactory result.

As the Internet world is expanding rapidly, in future more diverse real-world traffic can be used to generate a better traffic classification model. Deep learning is also an emerging field in data science. We aim to use it for network traffic classification in our future work.

REFERENCES

1. Othman, Suad & Alsohybe, Nabeel & Ba-Alwi, Fadl & Zahary, Ammar.: Survey on Intrusion Detection

System Types. In: International Journal of Cyber-Security and Digital Forensics (IJCSDF) 7(4): 444-462,2018.

2. Damshenas, Mohsen, et al.: A survey on malware propagation, analysis, and detection. In: International Journal of Cyber-Security and Digital Forensics, vol. 2, no. 4, 2013.
3. Subhabrata S, Oliver S and Dongmei W.: Accurate, scalable in network identification of P2P traffic using application signatures. In: WWW2004, New York, NY, USA, May 2004.
4. Vern P.: Empirically derived analytic models of wide-area TCP connections. In: IEEE/ACM Trans. Networking, vol. 2, no. 4, pp. 316–336, 1994.
5. Thuy N, Grenville A.: A survey of techniques for internet traffic classification using machine learning. In: Commun. Surveys Tuts. 10(4), 56–76 (2008).
6. Petr V, Milan C, Pavel C, Martin D.: A survey of methods for encrypted traffic classification and analysis. In: International Journal of Network Management25(5), 355–374 (2015).
7. Haitao H, Xiaonan L, FeiTeng M, Chunhui C and Jianmin W.: Network traffic classification based on ensemble learning and co-training. In: Science in China Series F: Information Sciences, vol. 52, no. 2, pp. 338–346, Feb 2009. [Online]. Available: https://doi.org/10.1007/s11432-009-0050-8.
8. Changyu W, Xiaohong G, and Tao Q.: A traffic classification approach based on characteristics of subflows and ensemble learning. In: IFIP/IEEE Symposium on Integrated Network and Service Management (IM), May 2017, pp. 588–591.
9. Santiago G, Beln M, Antonio E, and Luis C.: Ensemble network traffic classification. In: Computer Networks, vol. 127, no. C, pp. 68–80, Nov. 2017.
10. Isadora P, Anderson S, Lisandro G, Alberto F, Angelos M.: Improved Network Traffic Classification Using Ensemble Learning. In: 2019 IEEE Symposium on Computers and Communications (ISCC) (2019): 1-6.
11. Gerard G, Arash L, Mohammad M and Ali G.: Characterization of encrypted and VPN traffic using time-related features. In: Journal, In Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP 2016) (pp. 407-414).
12. R Battalov, A Nikonov, M Gayanova, V Berkholts and R Gayanov.: Network traffic analyzing algorithms on the basis of machine learning methods. In: V International Conference on "Information Technology and Nanotechnology" (ITNT-2019).
13. Julian C, Agapito E, Juan M, Alvaro G.: QoS-Classifier for VPN and Non- VPN traffic based on time-related features. In: Computer Networks 144 (2018): 271 – 279.