Coarse-Grained Classification Of P2p Network Traffic Using Filter/Wrapper Features Selection

[1]Haitam A. Jamil, [2]Bushra M. Ali, [2]Ahmed E. Osman, [2]Hind G. Abdelrahim

[1]*Faculty of Computer Science and IT, University of Elimam Elmahdi , Kosti, Sudan*
[2] *Faculty of Engineering, University Technology Malaysia , Johor, Malaysia*

Abstract *— Classifying network traffic applications is needed for network security and controlling. The emergence of new Internet applications with the use of encryption techniques, gains significant attention in the last period of time. However, the problem of using huge features requires longer processing time as well as low classification accuracy. Therefore, feature selections have a significant impact on classification performance. In this paper, we propose Filter/Wrapper feature selection methods for flow-based Internet traffic Classification using Machine Learning techniques. The evaluation has been carried out through experiments on the traffic traces downloaded from different shared resources. The experiments demonstrate our approach can greatly improve the computational performance.*

Keywords *— Coarse-grained classification; features selection; wrapper approach; filter method.*

I. INTRODUCTION

Simple classification assumes that most applications use well-known port numbers, and the classifier uses this port number to identify the application type. However, most Internet applications use unknown port numbers, or more than one application uses the same port number, which indicates the failure of port base classification [1]. Another classification method is payload based (deep packet inspection), which is individual packet inspection, looking for unique signatures. However, using this technique faces two problems; first, it is difficult to detect non-standard ports by using packet inspection because these packets are encrypted. Second, deep packet inspection touches on users' privacy. In order to solve the problem of past classification methods (base port and payload inspection), machine learning (ML) technique was developed [2]. ML uses artificial intelligence to classify IP traffic, which provides a powerful solution by extracting the right information from application features. Moreover, some of the ML algorithms are suitable for Internet traffic flow classification at a high speed. Most of the proposed ML classification methods are limited to offline traffic classification and cannot support online classification. Online classification, means the decision of which packet belongs to which flow, assuming to be on the traffic speed. Such, like any hardware classifier (Packet Shaper, SANGFOR), is installed on the network path to classify with the passage of the traffic [3].

II. THE FEATURE SELECTION

Traffic features are statistical characteristics found or determined from a given data sample. In traffic classification, features are normally numeric. There is no doubt that features selection perform an vital role in flow-based traffic detection designs [4, 5]. Traffic detection methods use these features to indicate an instant to a class.

Many researchers are defined Feature selection by looking at it from different perspectives. The contents of their definitions are mostly similar. Kira in [6] defined feature selection as "*Find the minimally sized feature subset that is necessary and sufficient to the target*" and in [7] Narendra noted it as "*Select a subset of (M) features from a set of (N) features, M < N, such that the value of a criterion function is optimized over all subsets of size M*". On the other hand, Koller in [8] defined feature selection as "*The aim of feature selection is to choose a subset of features for improving prediction accuracy or decreasing the size of the structure without significantly decreasing prediction accuracy of the classifier built using only the selected features*".

Several research works have been done in the field of feature selection and traffic classification. Works in [9, 10] review the state-of-the-art approaches, techniques and available application for the classification of network data. Authors in [10] stated that feature selection methods search through each subset of features to obtain the suitable one among the opposing candidate subsets according to some evaluation function. However, this procedure is exhaustive as it attempts to find only the best one. It may be too costly and practically restrictive, even for a medium size of feature set. Other methods based on heuristic or random search methods try to decrease computational complexity by compromising performance. These methods need a stopping criterion to prevent an exhaustive search of subset.

Figure 1 illustrates the four basic steps in a typical feature selection method, this as the points of viewed of the authors in [10]. The first step is the procedure generation. It uses to generate the candidate subset of

feature. The evaluation function is the second step which evaluates the subset under examination. The third step is the stopping criterion that decides when to stop. Last, the validation procedure, checks whether the subset is valid.

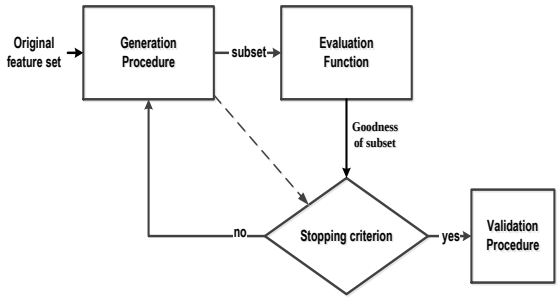

Fig 1: Feature selection process with validation

Work in [11] proposed and evaluated feature selection algorithms (FSAs) in order to understand their general behaviour on the particularities of relevance, irrelevance, redundancy and sample size of the datasets. To realize this aim, these authors designed and carried out a set of experiments using generated data sets. The result of the experiments can be viewed as a track towards obtaining useful knowledge that allows determining which algorithm to be used. The result is shown the different behaviour of the algorithms to different data particularities. For this, these authors showed the risk in relying in an individual algorithm and advised to use combinations of algorithms for more reliable assessment of feature significance.

Many algorithms have been proposed for feature selection. In particular, these algorithms help to solve one or more problems with the following characteristics. Great numbers of features, many irrelevant features, many redundant features, noisy data, continuous data and small training sets [12]. All these algorithms can be represented in a space of characteristics according to the criteria of search organization, generation of successor and evaluation measures. Search organization and generation of successor are grouped as generation procedure. These three characteristics are described briefly as follows.

First, search algorithm is used to drive the feature selection process using one of these strategies: exponential, sequential or random strategy. Second, generation of Successor is a mechanism that proposes a successor of the current hypothesis. Different operators can be considered to generate a successor: Forward, Backward, Compound, Weighting, and Random. Last, evaluation measure is a function used to evaluate the generated successor.

Most algorithms for supervised learning can be classified as a filter or a wrapper approaches [13-15]. Filter methods select a subset of features by only using intrinsic properties of the data. The subset is selected by evaluating some predefined criterion. Therefore, this concept usually considers a faster speed importantly. Moreover, the filter method is computationally less expensive and more general. Wrapper algorithms assess the quality of a given feature subset. The wrapper method can produce high classification accuracy and high computational complexity.

Moore in [16] used Fast Correlation Based Filter (FCBF) for feature reduction and Naïve Bayes algorithm to assess the feature reduction effect. The result of the overall classification accuracy based on the reduced sets is 84.06%, which is much better than using all features. On the other hand, the work in [17] also uses the Naïve Bayes algorithm with only the first five packets of the flow. The resultant effect is that only two feature, the size and direction of the initial data packet with TCP connection provides the distinction for all applications.

Jun et al. [18] applied two optimal features subsets to provide a good traffic classification accuracy. The accuracy of using the flow features subsets on Support Vector Machine (SVM) classifier is 70% while the training time was reported at 40 seconds.

Yang et al. [19] used random search algorithm for features reduction to identify P2P traffic by using SVM. However, this work did not include UDP traffic although P2P traffic consists of both TCP and UDP packets.

Auld et al. in [20] used 249 features derived from packet streams consisting of one or more packet headers. A full description of Moore features is available in [21]. One of the features selection algorithms available in WEKA tool is a correlation-based feature selection (CFS) which is a subset heuristic evaluation that takes into account the usefulness of separate features for predicting the class along with the degree of inter-correlation among them. It assigns high scores to subsets containing features that are greatly correlated with the class and have low inter-correlation with each other. The other algorithm is the Consistency-based feature selection (CON) which evaluates all of the subset of features concurrently and selects the optimal subset.

Feature selection algorithms were used to choose the best feature subsets in [21] but this process consumes much time. Moreover, most of these features are hard to be extracted from on-line traffic for on-line traffic classification.

III. METHODOLOGY

Datasets used in this work were downloaded from specific shared resources. Also, datasets were captured from the academic network in University of Technology Malaysia.

A method for selecting feature's subsets is needed in order to improve the classification accuracy, and to avoid incomprehensibility due to the large number of features investigated. In this work, we proposed filter and wrapper method for feature selection. A filter method was used to select important feature subsets

from all features in the data sets, and a wrapper method was employed for actual FS. Figure 2 shows the general framework for filter/wrapper FS approach.

Benefit and cost are used to evaluate the effectiveness of the proposed approach. These metrics depend on true positive, false positive, true negative and false negative. TP is the number of P2P class that are correctly classified, FP is the number of nP2P class that are classified as P2P class, TN is the number of non P2P (nP2P) class that are correctly classified, and FN is the number of P2P class that are classified as nP2P class. Training and testing times are used to illustrate the efficiency improvement.

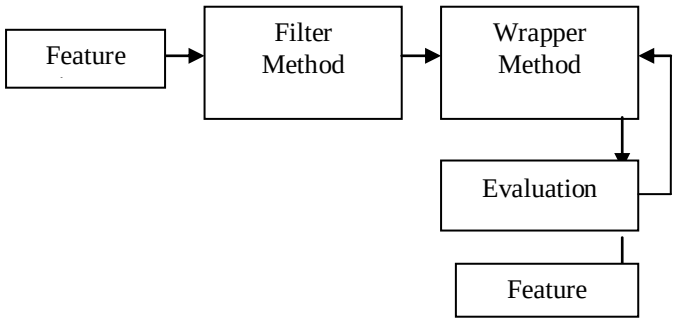

Fig 2: General framework for filter/wrapper Feature Selection approach

IV. RESULTS

This section explains the experimental results of using our proposed approach to classifying Internet traffic effectively and efficiently using feature selection methods based on filter method and wrapper techniques for content-based detection using ML techniques.

Table 1 defines the classification performance of the proposed topology. The performance accuracy of the training part using artificial neural network model is 98.00% and 97.90% for testing.

Table 2 presents the comparison between our proposed approach, hybrid naïve Bayes Tree and PORT-SCAN. As compared to these methods in term of false positive, our proposed approach has acceptable FP which is 2.8%. Moreover, our classifier is speed up the process of the classification as compared to the result for NBTree (which is 416s) and port-based (which is 4s) using same dataset. This improvement is a result of reducing the number of features.

TABLE 1 THE EVALUATION RESULTS

Partition	TP	FN
Training	98.00%	2.00%
Testing	97.90%	2.10%

TABLE 2 COMPARISON OF OUR METHOD, HYBRID NBTREE AND PORT-BASED METHOD

Methods	TP	FP	TN	FN	Time
Port based	97.7%	5.2%	94.8%	2.3%	4.09
NBTree	99.5%	0.3%	99.7%	0.5%	416.32
The proposal	97.9%	2.8%	97.2%	2.1%	2.00

V. CONCLUSIONS

Feature selection methods can significantly improve the computational performance of traffic classification. In this paper, we proposed a set of features for internet traffic classification using machine learning by analyzing different types of features to show their strength over the others. The experimental results indicate that construction classifier obtains a higher computing performance and accuracy. The accuracy and testing time for our proposal classifier are 98.4 % and 2.18 second, respectively.

REFERENCES

[1] Jamil, H.A., Bushra M. A , Ahmed Abdalla , Ban M. K , Sulaiman M. Nor , Muhammad N. , Improving P2P Network Traffic Classification with ML multi-classifiers. International Journal of P2P Network Trends and Technology (IJPTT), 2014. Volume - 4(Issue - 2).

[2] AH, H. and H.A. Jamil, Enhance the accuracy of Machine Learning Internet Traffic Classifier by Applying Datasets Validation Issues and Using a Hybrid Classifier. 2013.

[3] Ibrahim, H.A.H., S.M. Nor, and H.A. Jamil. Online hybrid internet traffic classification algorithm based on signature statistical and port methods to identify internet applications. in Control System, Computing and Engineering (ICCSCE), 2013 IEEE International Conference on. 2013. IEEE.

[4] Jamil, H.A., et al., Selection of online Features for Peer-to-Peer Network Traffic Classification, in Recent Advances in Intelligent Informatics. 2014, Springer International Publishing. p. 379-390.

[5] MA, B., et al. Multi-stage Feature Selection for On-Line Flow Peer-to-Peer Traffic Identification. in Asian Simulation Conference. 2017. Springer.

[6] Kira, K. and L.A. Rendell. The feature selection problem: Traditional methods and a new algorithm. in AAAI. 1992.

[7] Narendra, P.M. and K. Fukunaga, A branch and bound algorithm for feature subset selection. Computers, IEEE Transactions on, 1977. 100(9): p. 917-922.

[8] Koller, D. and M. Sahami, Toward optimal feature selection. 1996.

[9] Erman, J., et al. Semi-supervised network traffic classification. in ACM SIGMETRICS Performance Evaluation Review. 2007. ACM.

[10] Dash, M. and H. Liu, Feature selection for classification. Intelligent data analysis, 1997. 1(1-4): p. 131-156.

[11] Molina, L.C., L. Belanche, and À. Nebot. Feature selection algorithms: A survey and experimental evaluation. in Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on. 2002. IEEE.

[12] Bins, J. and B.A. Draper. Feature selection from huge feature sets. in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. 2001. IEEE.

[13] Bermejo, P., et al., Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking. Knowledge-Based Systems, 2012. 25(1): p. 35-44.

[14] Foithong, S., O. Pinngern, and B. Attachoo, Feature subset selection wrapper based on mutual information and rough

sets. Expert Systems with Applications, 2012. 39(1): p. 574-584.
[15] Cadenas, J.M., M.C. Garrido, and R. Martínez, Feature subset selection Filter–Wrapper based on low quality data. Expert Systems with Applications, 2013. 40(16): p. 6241-6252.
[16] Moore, A.W. and D. Zuev. Internet traffic classification using bayesian analysis techniques. 2005. ACM.
[17] Bernaille, L., et al., Traffic classification on the fly. ACM SIGCOMM Computer Communication Review, 2006. 36(2): p. 23-26.
[18] Jun, L., et al. P2P traffic identification technique. in Computational Intelligence and Security, 2007 International Conference on. 2007. IEEE.
[19] Yang, Y., et al. Solving P2P traffic identification problems Via optimized support vector machines. in Computer Systems and Applications, 2007. AICCSA'07. IEEE/ACS International Conference on. 2007. IEEE.
[20] Auld, T., A.W. Moore, and S.F. Gull, Bayesian neural networks for internet traffic classification. Neural Networks, IEEE Transactions on, 2007. 18(1): p. 223-239.
[21] Moore, A.W., D. Zuev, and M. Crogan, Discriminators for use in flow-based classification. 2005, Technical report, Intel Research, Cambridge.