ISSN:0975-9646

Internet Traffic Classification: An Enhancement in Performance using Classifiers Combination

[1]Indra Bhan Arya
[1]*Maulana Azad National Institute of Technology*
Bhopal

[2]Rachna Mishra
[2]*Truba Institute of Engineering & IT*
Bhopal

Abstract-Internet traffic classification basically used in many areas such as network management and operation, network design, Quality of Services, traffic control and network security by which network administrator can efficiently handle the network. Traditional Internet traffic classification such as, port number, payload and heuristic, fails to identify the new version of P2P applications. Early version of P2P systems usually use TCP with some fixed ports whereas new version of P2P applications can both use TCP and UDP connections with arbitrary ports. Researchers have applied another technique which is based on statistical features. Machine Learning classification algorithms which are based on statistical features fall into two categories (i) Supervised, (ii) Unsupervised. This work evaluates J48, Random Forest, Naïve Bayes classifiers and classifier combinations like Bagging, Boosting over benchmark datasets. This benchmark dataset are freely available to us. This work presented also evaluates feature selection algorithms to reduce noise and time required for model generation without affecting the performance. This work proposed multilevel classifiers based on the performance of multiple classifiers for specific classes.

Keywords: ML Algorithms, Internet Traffic, P2P applications, Feature Selection, Multilevel Classifiers

I. INTRODUCTION

Internet traffic classification is a requirement for identification of internet traffic applications. It is basically used to network management, network security, and Quality of Services. Accurate network management is possible through accurate traffic classification.

New emerging applications are increasingly evolved in the internet. Traditional methods are not play an important role for identification of internet traffic applications. Many new P2P applications are such as e-DONKEY, KaZaA etc are using dynamic port number, masquerading techniques, and encryption [15].

Researchers are trying other approaches like Machine Learning, which is based on statistical features. In traffic classification, features are computed over first multiple packets coming. Features are alternatively known as attributes of the data or discriminators. An instance is combined form of features in which a special feature is taken as class label feature, which enables to find the accuracy of the classification algorithm. Here accuracy

Means, how classifier correctly classified the test dataset with full features sets or subset of features. Feature selection algorithms play an important role for designing classifier in terms of classification accuracy and computational performance. There are two types of feature selection algorithms (i) Filters, (ii) Wrappers. Machine learning classifiers categories as follow: (i). Supervised, (ii).Unsupervised.

The paper is organized as follows. Section 2 describes datasets. Section 3 discusses feature selection. Section 4 describes methods and section 5 evaluates the performance of machine learning algorithms. Section 6 discusses related work. Section 7 shows proposed work and section 8 setup & results and section 9 we conclude and future work.

II.DATA SETS

Data sets for internet traffic classification may be SNMP based, flow based, byte based, and packet based. Data sets either may be self collected by various links namely campus link, ADSL link, and Backbone link or available to us at various locations such as NLAR and Waikato trace, Auckland IV, and CAIDA [17]. All above locations have benchmark datasets. Benchmark data sets are publicly available to us or on private request to access.

We illustrate our method with pre-classified data described originally in [20]. Training data & testing data are trace1 and trace2 respectively. The test data is available to us after 12 month later.

III.FEATURE SELECTION

Traffic classification use features, set of attributes of each instance, to evaluate the outcome of class. A class is a special attribute of each instance, which shows result of instance. Feature selection algorithms play an important role for ML algorithms. It is not only reduces the features sets but also improve computational performance and classification accuracy. Full feature sets is not good for machine learning classification algorithms because some features is irrelevant and redundant for data set , it also consume lot of time for generating the classifier model and reduce classification speed. Feature selection algorithms consume much time for finding out the relevant features and inefficient for real-time applications. Few real time features need to calculate over packets coming in, which is much faster. The full sets of real time features are listed in [15] and a full description of these features is available in [16].Filters basically use the characteristics of each feature of data set to evaluate the importance of feature with other features. The reduced sets of features are extracted by filters, which can be suitable for any learning algorithms. In contrast, Wrapper algorithm evaluates performance of different subset of features using ML algorithms by which we can find out optimal set of features that are suited for particular ML algorithms.

IV. METHODS

Methods basically used to classify the Internet traffic applications can be categories as follows: (i) Port Number, (ii) Payload Based, (iii) Heuristic methods, and (iv) Machine Learning.

A. *Port Number Analysis*

The port number in TCP and UDP header is match with default port number if match is occurred then packet is a P2P application but recent P2P system may use random port numbers. For example P2P applications such as KaZaA, Lime Wire, and port number may not be registered with IANA. Hence, port based analysis is not suitable for Internet traffic classification.

B. *Payload-based Analysis*

This technique only identifies P2P traffic for which application unique string (Signature) is found in the payload of data packets. This signature-detection approach is process intensive, which performs deep-packet inspection (which is not possible when privacy is required).

Second, Some P2P applications have encrypted data that makes it impossible to detect unique string in payload [9].

Payload analysis requires much computational power because it analyses full payload of packet [28, 29].Third, these technique often requires increased processing and storage capacity [18].

C. *Heuristic-based Methods*

First heuristic based, P2P protocols use concurrent UDP and TCP transport layer but however some other web applications like NETBIOS, DNS also use UDP and TCP transport layer protocols. In second heuristic approach, web servers use multiple parallel connections to other peer machine in order to transfer web pages text and images .In contrast, P2P has one or more consecutive connections, i.e. only single connection can be active at a time. However some popular streaming applications (Window media server, helix server, and Quick Time) not necessarily have parallel connections to the web server. By mistake these data flows would be identified as P2P flows [10]. In third heuristic, P2P traffic uses default ports of P2P applications. However port can be dynamically chosen or when firewall or port-blocking is observed.

D. *Machine Learning Approaches*

Witten and Frank [7] define four basic types of learning (a) Classification (or Supervised Learning) (b) Clustering (or Unsupervised Learning) (c) Association and (d) Numeric prediction. Internet Traffic classification often contains numerical attributes calculated over multiple packets, such as mean packet lengths, total flow length, inter-packet arrival time and so on [27]. ML algorithms which have been used for traffic classification fall into the two categories of Supervised or Unsupervised. ML algorithms are based on the statistical features rather than port number or payload. The multiple packets coming from the same flow is considered as an instance. Each flow is described by the Source IP Address, Source Port number, Des IP Address, Des Port Number and Protocol Type number. A flow also referred to as a connection. A connection has bidirectional exchange of packets between two nodes. The statistical features may be packet-inter-arrival time, packet duration, packet length etc. The set of statistical features is denoted by each flow. Classifier can uniquely identify the traffic classes depending on its statistical feature values. In Supervised learning algorithm, a class value is already known through the training datasets and predicts the class value of the test datasets. Naive Bayes, Decision Tree, Neural Network, Bayesian Neural Network are supervised learning algorithm. In Unsupervised Learning, Instances or flows are assigned to different clusters according to similar features values. They need not to be labeling specific class before learning process as in supervised. EM, AutoClass and K-Means are unsupervised learning algorithms [5].

V. MACHINE LEARNING PERFORAMANCE

The efficiency of ML algorithms can be measure by following metrics: precision, recall, and overall accuracy. A sample has four prediction outputs, in which two are correct and rest are incorrect. TP (True Positive) , where instance is actually P2P and it is predicted as P2P and TN(True Negative), where instance is actually non-P2P and it is predicted as non-P2P , and other two FP (False Positive) , where instance is non-P2P and it is predicted as P2P and FN(False Negative) , where instance is P2P and it is predicted as non-P2P.

Precision is defined as true positive to the true positive and false positive.

$$\text{Precision (P)} = TP / (TP+FP)$$

Recall is defined as true positive to the true positive and false positive.

$$\text{Recall (R)} = TP / (TP+FN)$$

Accuracy is defined as the sum of all True positives and True Negative to the total number of test instances. This measures the overall accuracy of the classifier. Precision and recall are per-class measures.

Overall accuracy is related to precision in that it measures the average precision of all classes [2].

$$\text{Accuracy} = (TP+TN) / (TP+ TN+FP+FN)$$

In [7] a multiclass prediction, the test data sets is often denoted by confusion matrix with row and column. Where row and column are actual class and predicted class respectively. Correct instances are measured from diagonal of confusion matrix. The confusion matrix is an useful tool for analyzing how well your classifier can recognize instances of different classes [8].

Table 6.1 Confusion Matrix

		Predicted Classes	
		P2P	Non-P2P
Actual Classes	P2P	**TP**	**FN**
	Non-P2P	**FP**	**TN**

VI. RELATED WORK

7.1 *Comparison of clustering vs. supervised techniques*:

Jeffrey Erlen et al. [2], this paper evaluate the overall accuracy, precision and recall using Naïve Bayes and AutoClass Machine Learning algorithms. The two data sets which are publicly available traces is collected from the university of Auckland (NLANR).The performance of Naïve Bayes and AutoClass algorithms was evaluated on the two 72-hour AuckIV and Auck VI subset of traces. This paper shows nine application classes (HTTP, SMTP, DNS, SOCKS, IRC, FTP Control, POP3, and LIMWIRE). In Naïve Bayes classifier, on an average, the precision and recall for six application classes out of nine classes were above 80%. Where as for AutoClass classifier, the precision and recall values for all classes were above 80% and the average precision values for six classes were above 90% and recall values for seven classes were above 90%. The average overall accuracy of AutoClass was 91.2% whereas in the Naïve Bayes classifier was 82.5%. Thus, we found that AutoClass outperforms the Naïve Bayes classifier by 9%. All the analysis is performed on a Dell GX620 pentium 4 3.4 Ghz processor with 1 GB RAM. The Naïve Bayes classifier takes 0.06 seconds for 8000 objects whereas AutoClass took 2070 second to generate the classifier model. The build time of Naïve Bayes classifier was better than AutoClass. Above results shows that unsupervised learning machine is good approach without requiring the training data to be labeled beforehand.

Zhao et al. [1] in this paper, five supervised and unsupervised algorithms are evaluated using dataset1 & dataset2 as training and testing data respectively. Dataset 1 has 0.4% P2P traffic whereas dataset2 is 45% of P2P traffic. Dataset has ten classes with 249 attributes. Supervised learning algorithms C4.5 and Random forest achieve greater than 96% accuracy and Naïve Bayes shows poor performance. This paper also showed that the build time for real-time features are lowest as compared to another feature sets. Classifier's accuracy has became degrades when proportion of P2P traffic is increased.

7.2 *Comparison of different supervised ML algorithm*

William et al. [4] and Nguyen et al. [5], This paper describe the performance of C4.5, NBK, NBD, NBT, and Baysian Netwok in terms of classification accuracy and computational performance using full features sets and different reduced features sets. Feature selection algorithms, CFS subset, CON subset using the different search methods namely, greedy forward and best first search, and full feature set are used.

Four of them ML algorithms have 95% accuracy using full sets of features, and reduce feature subsets shows little change accuracy. It is difficult to differentiate ML algorithms basis on the accuracy, precision and recall using different subsets of features. Due to similarity of accuracy, the paper shows significant difference for computational performance. C4.5 has highest classification speed as compared to other ML algorithms. This results show that C4.5 is a good candidate for real-time classification tasks. NBK has lowest classification speed followed by NB Tree, Bayes Net, NBD, and C4.5.For above experiment, 3.4 GHz Pentium 4 Workstation running SUSE Linux 9.3 has been used. The highest classification speed measure was 54700 per second for C 4.5 algorithm. And the NBK take lowest classification speed as compare with other algorithms. This paper shows that feature selection reduce the dataset as well as improve the computational performance.

Bijan et al. [9], overall success rate of model using C4.5 algorithm depends on the two factors: attribute sets selection and the number of records. Attribute set1 took as Protocol, Length, Source TP, Destination IP and set2 take Protocol, Length. The overall success rate was 99.98% using set1 over 32000 instances and 88.42% for 2007 instances. Where as success rate were 79.23% using set2 over 32000 instances and 85.05% for 2007. The performance of success rate was degraded without IP addresses. The build time can be improved using IP addresses over 32000 instances. Another success rate depends on the attributes selection and ratio of P2P/non-P2P traffic. The results shows that higher the disproportion P2P/non-P2P traffic with set1 specify better success rate in contrast with set2 degrade the success rate.

The conclusion of this paper show that the classifier can be implemented with in the administrative domain of the individual service provider's network and need to be continuously updating that can handle new P2P traffic.

7.3 *Comparison of different clustering algorithms*

Erman et al. ([18], [5]), In this paper, performance of three unsupervised algorithms: K-means, DBSCAN, AutoClass are evaluated using two benchmark datasets which are publicly available at university of Auckland and self collected through the University of Calgary. The clustering algorithms have ability to produce group of objects into each cluster for specific type of traffic application using unlabelled training data. The application classes for Auckland IV dataset were DNS, FTP-control, FTP-data, HTTP, POP3, NNTP, LIMEWIRE, IRC, and SOCKS. The results as, on average, the overall accuracy were 92.4% and 88.7% for Auckland and Calgary datasets respectively for AutoClass Classifier. On average, AutoClass produce 167 clusters for Auckland IV dataset and 247 clusters for Calgary dataset respectively. For K-means cluster the overall accuracy were 49% and 67% for Auckland IV and Calgary data sets respectively. The overall accuracy improved slightly as the number of clusters increases. The overall accuracy for DBSCAN was 75.6% and 72% for Auckland IV and Calgary dataset respectively.Raimir et. at. [13], proposed E-Ratio and Boxplot diagram methods for feature selection. The feature selection was based on those variables that present large F-Ratio values.

VII. PROPOSED MULTILEVEL CLASSFIERS APPROACH

In multilevel classifier, various classifiers are combined to enhance the performance. The classifiers combined are preferably complimentary. This work evaluates various classifiers over the benchmark data set [16]. In multilevel classifier at level zero we have classifier, which has highest recall & precision values maximum number of classes, in comparison of other classifiers. The classes evaluated at level k are not evaluated again at level k+1 or more higher levels. The remaining classes of test data i.e. the classes which were not evaluated at level zero are evaluated in next or other higher levels.

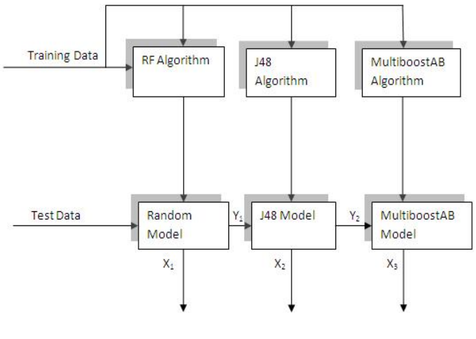

Figure 4.1 Multilevel Classifiers

Where X1= {service}, X2= {FTP-Control, FTP-Data, Interactive}, and X3= {WWW, Mail, FTP-PASV, Database, and P2P} and Y1= {WWW, Mail, P2P, Database, FTP-Control, FTP-Data, FTP-PASV, Interactive}, Y2= {WWW, Mail, P2P, Database, FTP-PASV}.

VIII. EXPERIMENT SETUP AND RESULTS

Throughout this paper, data collected by the high-performance network monitor described in [21]. This site is a research-facility host to about 1000 users connected to the Internet via a full-duplex gigabit Ethernet link. Full-duplex traffic on this connection was monitored for each traffic sets. The campus has Internet via a full-duplex gigabit Ethernet link. The monitor was located on this connection to the Internet. Each traffic set consists of a full 24-h weekday period in both link directions.

.

8.1 Analysis Results

Table 9.1 Recall & Precision values for Multilevel Classifiers

	Random Forest		J 48		Multiboost 48	
Classes	**P**	**R**	**P**	**R**	**P**	**R**
WWW	0.967	.999	1	.999	.999	1
Mail	.898	.898	.857	1	.998	1
FTP-C	.993	.962	1	1	0	0
FTP-P	.999	0.996	.995	.574	1	.98
Attack	0	0	0	0	0	0
P2P	.947	.778	.983	.956	.983	.963
DB	0	0	1	.99	1	.986
FTP-D	.984	.707	1	1	0	0
MM	0	0	0	0	0	0
Ser	1	1	0	0	0	0
Int	0	0	1	1	0	0
Gm	0	0	0	0	0	0

For experiments with trace 2 as test dataset performance of C4.5 algorithm is not good for Mail, FTP-P, P2P, Service classes. The performance of J48 classifier for P2P and non-P2P classes increases by using bagging and boosting which reduces error due to variance. The performance of random tree classifier drops for the classes FTP-C, Mail, FTP-D, Services, Interactive and P2P when trace 2 is used as test data. The performance of P2P class more drops than other classes. The overall accuracy for multilevel classifiers is 99.8471%. The highest accuracy for non-multilevel classifier for MultiboostAB classifier combinations is 99.842%. Multilevel classifiers improve accuracy. It takes more build time as compared to a single classifier.

Table6.12 Build time for Multilevel Classifiers

Classifier	Build Time(sec)
RF	178.99
J48	935.79
MULTIBOOSTAB	11755.49

IX.CONCLUSION & FUTURE WORK

Five classifiers namely J48, Random Tree, Random Forest, Bagging and boosting algorithms are evaluated over single benchmark dataset. Proposed multilevel classifiers give better performance than single classifier. Performance of classifier for P2P class increases by using classifier combinations using Bagging and Multiboosting. Multiboosting outperforms the Bagging approach. In future, other benchmark data sets like Auckland IV can be evaluated for various machine learning algorithms

REFERENCES

[1]. Zhao Jing, Huang X, "Real-time feature selection in traffic Classification", The Journal of China Universities of Posts and Telecommunications 2008:68-72.

[2]. Jeffrey Erman, Anirban Mahanti, and Martin Arlitt, "Internet traffic Identification using Machine Learning techniques", *in proc. of 49th IEEE Global Telecommunication Conference 2006(GLOBE COM 2006),* San Francisco, USA, December2006.

[3]. Min Zhang, Wolfgang John, State of the Art in Traffic classification: A Research Review", PAM 2009 April 1-3, 2009, Soul Korea.

[4]. Nigel Williams, S Zender, and G Armitage, "A preliminary performance comparison of Five Machine learning algorithms for practical IP Flow Classification", *SIGCOMM Computer Communication Review, vol.36, no. 5, pp. 5-16, 2006*

[5]. T.Nguyen and G. Armitage, "A survey of techniques for Internet traffic classification using machine learning", *IEEE Communications Surveys, and Tutorials, 2008.*

[6]. Dash M, Liu H, "Consistency-based search in feature Selection", *Artificial Intelligence, 2003, 151(I-2): 155–176.*

[7]. I. Witten, and E. Frank, *Data mining: Practical machine learning tools and techniques with java Implementations (Second Edition).* Morgan Kaufmann Publishers, 2005.

[8]. Jiawei Han and Micheline Kamber, *Data mining concepts and Techniques*, Second Edition 2009.

[9]. Bijan Raahemi, Ahmed Hayajenh, and Peter Rabinovitch, "Peer -to-peer IP Traffic Classification Using Decision Tree and IP Layer Arrtibutes", *IGI Publication Jan 2007.*

[10]. Marcell, Trang Dinh, Dang,Andras, "Identification and Analysis of Peer-to-Peer Traffic", Journal of communication, 2006.

[11]. Andrew W Moore, and Denis Zuev, "Internet Traffic Classification using Beysian Analysis Techniques", *in ACM International Conference on Measurement and Modeling of Computer Systems(SIGMETRICS) 2005,* Banff, Alberta, Canada June 2005.

[12]. Tom Auld, S.F. Gull, and Andrew W Moore, "Bayesian neural network for Internet traffic Classification", *IEEE Transactions on Neural Networks, Vol. 18, No.1, pp 223-239, Jan2007.*

[13]. Raimir Marcus Jose Gabriel, "An Internet Classification Methodology based on Statistical Discriminators", IEEE Transaction 2008.IEEE.

[14]. Jin Li, Microsoft Research Peer-to-Peer Multimedia Applications, MM'06 October 23-27 ACM 1-59593-447-2/06/0010.

[15]. Abuagla Babikar, Sulaiman Mohd nor, Neer, "Real time online Flow-based Internet traffic Classification using machine learning (C4.5)", *International Journal of Engineering, Vol. 3 issues 4.*

[16]. A. W. Moore and D. Zuev, "Discriminators for use in flow-based classification. Technical Report, Intel Research, Cambridge 2005.

[17]. CAIDA, An overview of traffic classification, 2009, http://www.caida.org/research/*trafficanalysis/classification-overview/*

[18]. J. Erman, A. Mahanti, and M. Arlitt, "Traffic classification using clustering algorithms", *in MineNet*

2006: Proceeding of the 2006 SIGCOMM workshop on mining network data. New York, NY, USA: ACM Press, 2006, pp. 281-286.

[19]. Waikato Environment for Knowledge Analysis (WEKA) 3.6.2, http://www.cs.waikato.ac.nz/ml/weka/.

[20]. A. W. Moore and D. Papagiannaki, “Toward the accurate identification of network applications,” in *Proc. 6th Passive Active Meas. Workshop (PAM)*, Mar. 2005, vol. 3431, pp. 41–54.

[21]. A.Moore, J. Hall, C. Kreibich, E. Harris, and IPratt,“Architecture of a network monitor,” in *Passive Active Meas. Workshop (PAM)*, La Jolla, CA, Apr. 2003, pp. 77–86.

[22]. J. Erman, A. Mahanti, M. Arlitt, I. Cohen, and C. Williamson.Off_ine/Online Tra_c Classi_cation Using Semi-Supervised Learning. Technical report, University of Calgary, 2007.

[23]. O. Chapelle, B. Sch□olkopf, and A. Zien, editors.Semi-Supervised Learning. MIT Press, Cambridge, MA, 2006.

[24]. J. Erman, A. Mahanti, M. Arlitt, I. Cohen, and C. Williamson, “Semisupervised network traffic classification,” ACM International Conferenceon Measurement and Modeling of Computer Systems (SIGMETRICS)Performance Evaluation Review, vol. 35, no. 1, pp. 369–370, 2007.

[25]. http://en.wikipedia.org/wiki/Random_forest#External_links.

[26]. http://en.wikipedia.org/wiki/C4.5_algorithm.

[27]. Nguyen T T T, Armitage G. Training on multiple sub-flows to optimize the use of machine learning classifiers in real-world IP networks. IEEE LCN, 2006:369-376.

[28]. [28]. M.S.Kim, H.J.Kang, J.W. Hang, 2003, Towards Peer- to-Peer traffic analysis using flows, working paper obtained from the Distributed processing and Network management Laboratory. Pohang University of Science& Technology, Republic of Korea.

[29]. Robin Sommer and anja Feldman, Saarlamd University, Germany NetFlow: Information loss or win? ACM measurement workshop, 2002.