Empirical Analysis of Application-level Traffic Classification using Supervised Machine Learning

Byungchul Park[1], Young J. Won[1], Mi-Jung Choi[1],
Myung-Sup Kim[2], and James W. Hong[1]

[1] Dept. of Computer Science and Engineering, POSTECH, Pohang, Korea
[2] Dept. of Computer and Information Science, Korea University, Jochiwon, Korea
[1] {fates, yjwon, mjchoi, jwkhong}@postech.ac.kr
[2] tmskin@korea.ac.kr

Abstract. Accurate application traffic classification and identification are important for network monitoring and analysis. The accuracy of traditional Internet application traffic classification approaches is rapidly decreasing due to the diversity of today's Internet application traffic, such as ephemeral port allocation, proprietary protocol, and traffic encryption. This paper presents an empirical evaluation of application-level traffic classification using supervised machine learning techniques. Our results indicate that we cannot achieve high accuracy with a simple feature set. Even if a simple feature set shows good performance in application category-level classification, more sophisticated feature selection methods and other techniques are necessary for performance enhancement.

Keywords: Internet application traffic identification, traffic measurement and analysis, machine learning, supervised algorithm

1 Introduction

Accurate traffic classification based on application is an important step toward providing an informative snapshot of networks with respect to various network operation and management aspects such as traffic shaping, monitoring, capacity planning, charging, QoS management, and so on.

Traditional Internet traffic classification approaches rely on well-known port mapping and application signature mapping [1, 2]. However, the newer generation of P2P applications is incorporated with various sneaking-by strategies such as ephemeral port allocation and relay node to avoid detection and filtering. Moreover, emerging network applications such as Skype are eluding detection by packet payload encryption or plain-text ciphers [3]. Privacy legislation related to payload inspection is also a problem in this approach.

To overcome the limitation of port-based classification and signature-based classification, machine learning techniques that use statistical information of the transport layer have been introduced recently [4, 5, 6, 7]. These classification methods

depend on the fact that different applications have different communication patterns or behaviors.

In this paper, we narrow the classification level down to the specific application level, in contrast to previous research, and we explore the possibility of supervised machine learning technique for enabling traffic classification by empirical evidence. In addition, we point out the possible problem of traffic classification using machine learning techniques.

The remainder of this paper is organized as follows. Section 2 presents an overview related work with problem domain. In Section 3, selected supervised learning algorithms, feature sets, and data used in our wok are covered. Finally, we conclude our work with some suggestions of possible future work in Section 4.

2 Problem Domain

Many researchers have attempted to use machine learning algorithms for traffic classification [4-10]. Erman *et al.* [4] used three different unsupervised algorithms (K-Means, DBSCAN, Autoclass) and other researchers used various supervised learning algorithms. Each researcher observed that these algorithms show good performance for traffic classification in terms of classification accuracy (maximum 92~93%).

The classification target of these works is somewhat different from our research goal. Although the common aspect of these studies may be that machine learning techniques were used in traffic classification, the depth of the classification level in them is different. Prior works classified network traffic based on the application type (Bulk, Database, Mail, Web, P2P, etc) or application protocol (DNS, RTP, HTTP, POP3, etc). While our objective is to target more sophisticated application-level classification using supervised machine learning algorithms, previous works relied on known and limited means of layer-3 or above protocol formats.

3 Machine Learning Algorithm and Feature Set

To verify the performance of supervised learning in application traffic classification, we selected four different algorithms: J48, REPTree, Multi-layer Perceptron (MLP), and BayesNet. The first three algorithms showed very high accuracy in previous works. MLP is one of the supervised learning algorithms that showed good performance in other data mining applications.

We selected features that were commonly used in other traffic classification works: 1) Source IP address, 2) Destination IP address, 3) Source port number, 4) Destination port number, 5) Total bytes transferred, 6) Connection duration, 7) Packet size (min/max/mean/standard deviation), 8) Packet inter-arrival time (min/max/mean/standard deviation). The target application traffic was aggregated to a flow that is a collection of packets sharing identical 5-tuple (source IP, destination IP, source port, destination port, and protocol) information. After aggregating traffic data into the flow, the features were extracted from the flow.

3. 2 Target Applications & Data Collection

We selected 14 popular applications (Table 1). As described in Section 2, our traffic classification targeted more sophisticated application-level classification and not the application-category level; therefore, several applications belonging to the same category were selected as target applications.

The ultimate goal of traffic classification research is to classify the backbone network traffic; therefore, we collected testing data form POSTECH's backbone network.

Table 1. Traffic summary of training and testing data.

	Application	Training data		Testing data	
Type	Name	# of flow	Size (MB)	# of flow	Size (MB)
Web	Clubbox	231	165	165	102
Storage	ParanDisk	223	278	148	145
	TotoDisk	526	367	247	192
	EnDisk	246	447	550	384
Web	HTTP	463	27	2045	15
App.	HanGame	998	24	47	7
	SayClub	2870	10	85	3
FTP	AlFTP	518	1,011	231	342
P2P	BitTorrent	11,000	708	4,000	256
	Fileguri	571	506	408	407
	Soribada	8803	339	5960	273
	Gample	10,943	990	547	4,032
Game	Starcraft	15	0	6	0
Etc.	MS Remote Desktop	86	28	34	12
	Total	37,493	4900	14,473	6170

Table 1 shows the traffic trace summary of each application. In the case of Starcraft, very small amount of traffic data was generated over duration of 30 minutes; consequently, the total amount of traffic was less than 1 MB.

4 Conclusion

From the experiment, we made the following observations. (1) Even if we narrow the level of classification to that of a specific application level, the overall accuracy of the classification is still compatible (80~90%) with the accuracy of the application category-level of classification that was conducted in previous research. (2) However, precision and recall of each class fluctuated from 0 to 1 depending on the application.

From the results obtained by measuring precision and recall, we could conclude that the machine learning algorithms and the feature set that we used are not efficient for application-level traffic classification, even if these algorithms and the feature set performed very well in classifying application categories due to the difference in the level of classification. Based on the above discussions, we concluded that a feature selection algorithm and a modified machine learning algorithm are essential to overcome these hurdles. There has been much research on feature selection [11, 12], and finding an effective feature selection for traffic classification will help enhance the accuracy of classifiers. For future work, we plan to find the best possible feature set and modify machine learning algorithms so that they are suitable for application-level traffic classification.

Acknowledgments. This work was partly supported by the IT R&D program of MKE/IITA [2008-F-016-01, CASFI] and the EECE division at POSTECH under the BK21 program of MEST, Korea.

References

1. Sen, S., Spatscheck, O., Wang, D.: .Accurate, Scalable In-Network Identification of P2P Traffic Using Application Signatures. WWW 2004 Conference.
2. Karagiannis, T., Broido, A., Faloutsos, M., Claffy, K.C.: Transport layer identification of p2p traffic. Internet Measurement Conference (IMC), 2004.
3. Park, B.-C., Won, Y.J., Kim, M.-S., Hong, J.W.-K.: Towards Automated Application Signature Generation for Traffic Identification. In: Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS 2008), Salvador, Brazil, April 2008, pp. 160--167.
4. Erman, J., Arlitt, M., Mahanti, A.: Traffic Classification Using Clustering Algorithms. SIGCOMM'06 Workshops, Pisa, Italy, Sep. 2006, pp. 281--286.
5. Zander, S., Nguyen, T., Armitage, G.: Automated Traffic Classification and Application Identification using Machine Learning. In: Proceedings of the IEEE Conference on Local Computer Networks, Sydney, Australia, Nov. 2005, pp. 250--257.
6. Thuy, T., Nguyen, T., Armitage, G.: Training on multiple Sub-Flows to optimize the use of Machine Learning classifiers in real-world IP Networks. IEEE Conference on Local Computer Networks, Tampa, Florida, USA, Nov. 2006, pp. 369--376.
7. Park, J., Tyan, H.-R., Kuo, C.-C.J.: GA-Based Internet Traffic Classification Technique for QoS Provisioning. International Conference on Intelligent Information Hiding and Multimedia, Pasadena, California, USA, Dec. 2006, pp. 251--254.
8. Moore, A.W., Zuev, D.: Internet Traffic Classification Using Bayesian Analysis Techniques. SIGMETRICS'05, Banff, Alberta, Canada, Jun. 2005, pp. 50--60.
9. Erman, J., Mahanti, A., Arlitt, M.: Internet Traffic Identification using Machine Learning. IEEE Global Telecommunications Conference, California, USA. Nov.-Dec. 2006, pp. 1--6.
10. Williams, N., Zander, S., Armitage, G.: A Preliminary Performance Comparison of Five Machine Learning Algorithms for Practical IP Traffic Flow Classification. SIGCOMM Computer Communication Review, Oct. 2006, pp. 7--15.
11. Battiti, R.: Using mutual information for selecting features in supervised neural net learning. IEEE Transactions on Neural Networks. Vol. 5, No.4, July 1994.
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research, 2003.