Online Self-learning Internet Traffic Classification based on Profile and Ontology

[1,*]Chengjie GU, [1]Shunyi ZHANG, [2]Xiaozhen XUE
[1]*Institute of Information Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2]*Department of Computer Science, Texas Tech University, Texas 79415, USA*
jackiee.gu@gmail.com

doi:10.4156/jcit.vol6. issue4.10

Abstract

Internet traffic classification plays important roles in numerous areas such as network management, traffic engineering, QoS provisioning etc. Prior traffic analysis is an essential requirement for existing classification schemes to classify unknown traffic. To overcome the drawback of the previous classification scheme to meet the requirements of the network activities, we propose online self-learning Internet traffic classification based on profile and ontology. We evaluate our proposed method through the experiments on real network traces. Experiment results illustrate this method can reason from existing knowledge on traffic classification for achieving an automatic traffic classification with high accuracy.

Keywords*: Traffic Classification, Self-learning, Profile, Ontology*

1. Introduction

The rapid development of P2P applications has enriched the performance of Internet in recent years. Internet traffic will increase 46% annually from 2007 to 2012 according to measurement study in the literature [1]. The large amount of P2P traffic and the rapid growth on the usage of P2P applications have led to network congestion and traffic hindrance because of the excessive occupation of the network bandwidth [2]. Accurate traffic classification helps to identify the application utilizing network resources, and facilitate the instrumentation of QoS for different applications.

Internet traffic classification could be easily realized by reading port numbers in the early Internet [3]. The new P2P applications often change their behavior using different strategies to camouflage their traffic in order to evade detection. For example, they use dynamic ports in their connections or ports from other well-known applications. Several payload-based analysis techniques have been proposed to inspect the packets payload searching for specific signatures [4]. Although this solution does can achieve high classification accuracy, it can’t work with encrypted traffic or newly P2P applications [5]. At the same time, Internet traffic classification method based on flow statistics shows effective performance in this field. Substantial attention has been invested in data mining techniques and machine learning algorithms using flow features for traffic classification.

While Internet traffic classification methods based on flow statistics offer various degrees of successes, there are several limitations. 1) Prior traffic analysis is an essential requirement for existing classification schemes to classify unknown traffic. 2) Many proposed classifiers can’t solve the online traffic classification problems faultlessly.

To address the above-mentioned problems by exploring an online self-learning traffic classification method, we take two aspects to improve the accuracy and speed of this method for network traffic classification. 1) Firstly, a new classification method is required to be able to manage the knowledge of existing classification schemes, how to use the context and ontology means to manage traffic classification knowledge should be solved. At the same time, it is necessary to investigate how to reason from existing knowledge on traffic classification for achieving an automatic identification capability. 2) In order to achieve early classification, we demand the classifier to classify traffic flows early in the connection using the first p packets of flow.

The remainder of this paper is structured as follows. Related work is represented in Section 2. Section 3 discusses the ontological frame of traffic classification in detail. Section 4 illustrates how to construct online self-learning traffic classification based on profile and ontology. Section 5 presents the experimental results and analysis. The conclusion and potential future work are listed in Section 6.

2. Related Work

The port-based traffic classification relies on well-known port number to classify different Internet applications according to the ports registered in the IANA [6]. But this approach is often inaccurate, due to the dramatic increase in network applications using random ports or in tunneling through HTTP. In order to deal with the disadvantages of the above method, payload-based classification method is proposed to inspect the packet payload. Payload-based classifiers rely on the application specific signatures in the payloads, but they need highly computational costs [7].

The host-behavior-based approach is developed to capture social interaction observable even with encrypted payload [8]. However, this method such as BLINC can't classify exactly the applications which are theoretically from different groups but with similar behavior [9-10].

Machine learning technique which is a powerful tool in data separation in many disciplines aims to classify data based on either a priori knowledge or statistical information extracted from raw dataset. The branch that appears to solve the limitations of the network traffic classification methods is flow statistics analysis based on machine learning (ML) [11-18]. Nguyen et al. [11] provided context and motivation for the application of ML techniques to IP traffic classification, and reviewed some significant works. Machine learning algorithms are generally divided into supervised learning and unsupervised learning. Unsupervised learning essentially clusters flows with similar characteristics together [12-14]. The advantage is that it does not require training, and new applications can be classified by examining known applications in the same cluster. Erman et al. [12] compared the performance of unsupervised machine learning algorithms in traffic classification. Since our main focus is on evaluating the predictive power of a built/trained traffic classifier rather than on detecting new applications or flow clustering. Also, Erman et al. [13] evaluated the performance of two clustering algorithms, namely K-Means and DBSCAN, in Internet traffic classification. The result indicated that K-Means was one of the quickest and simplest algorithms for clustering of Internet flows. Bernaille et al. [14] used a simple K-Means clustering algorithm to perform classification by using only the first five packets of the flow, aiming at applying on the real-time classification. Supervised learning requires training data to be labeled in advance and produces a model that fits the training data [15-18]. Moore et al. [15] used a Naive Bayes classifier which was a supervised machine learning approach to classifying internet traffic. But only 65% accuracy rate, which was not good enough to classify Internet traffic. Williams et al. [16] conducted a comparison of five machine learning algorithms which were widely used to classify empirical study of Internet traffic. Among these algorithms, C4.5 achieved the highest accuracy in their results. Auld [17] proposed supervised machine learning based on a Bayesian neural network to classify the traffic with higher accuracy and better stability, but it was not capable for real-time applications. Ma et al. [18] used C4.5 decision tree to classify Internet traffic. This method could identify traffic of different types of applications with high accuracy, by collecting some features at the start of the flow. Table 1 shows a summary of a variety of traffic classification method.

Table 1. Summary of a variety of traffic classification method

	Port-based	**Payload-based**	**Transport layer behavior**	**Flow statistics**
Key features	Ports and protocols	Payload	Communication behavior	Flow information
Advantages	Fast, easy-to-use	Accurate	Identifying P2P flows.	Highly accurate.
Limitations	Port-masquerading	Complication	Tuning overheads.	High computational overheads,
Accuracy	Low	High	Average	Average
Overheads	Low	High	Low	Depends on ML algorithm
granularity	Fine-grain	Fine-grain	Coarse-grain	Coarse-grain

3. Ontological Frame of Traffic Classification

While previous traffic classification methods offer various degrees of successes, there are several limitations. 1) The existing classifiers cannot classify traffic self-learning, because they cannot always support the automatic classification functionality. 2) The previous method cannot utilize the existing major classification techniques at the same time. 3) Many proposed classifiers can't solve the online traffic classification problems rapidly. To deal with these problems, we follow the idea on ontology presented in [19] to define an ontology that consists of six elements: {F, FA, R, AR, H, X}, where F refers to a set of features; FA represents attributes for each feature; R represents a set of relationships; AR represents the attributes for a set of relationships; H stands for a feature hierarchy and X represents a set of axioms. Traffic classification ontology mainly focuses on features F and subclass of relationships R.

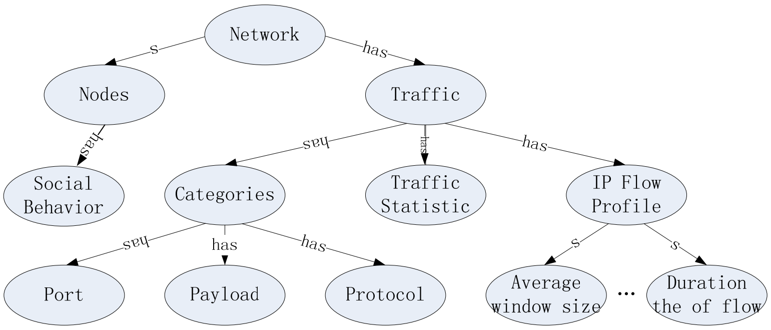

Figure 1. Internet network traffic classification ontology

As shown in Figure 1, network in this ontology is root and includes two concepts: nodes and traffic. Traffic has three concepts which are traffic category, IP flow profile and traffic statistic. The application such as "WWW" has various relationships with some concepts: social behavior, category, traffic statistic and IP flow profile. A flow profile can generally be described by a set of flow statistical features. A network traffic category is closely related to it is IP flow profile. And other classification concepts such as traffic statistic can also be extended to other sub-domain ontology.

3.1. Traffic classification description using context and ontology

Internet traffic classification is the process of classifying an individual Internet application or a group of applications of interest. Internet traffic classification methods implement the computation process by using prior knowledge and the classification algorithm. The classification algorithm can be fully expressed by using the ontology means. For example, traffic classification based on payload typically needs signature information which includes some bits information, the information is the signature value for a particular application. Then, traffic classification algorithm is realized by comparing input data with the prior knowledge. Therefore, we need some mechanisms to represent the traffic knowledge. One is used to represent prior knowledge and input data and the other is used to describe algorithms on traffic classification. In this paper, we will mainly discuss the use of context and ontology.

The concept of context has been widely used in many fields [20]. A context is defined as any information or knowledge that can be used to customize the situation of an entity [21]. The implementation of contexts is realized by network context sensors. For instance, an entity can be one specific IP flow or a network node and its related contexts can be inputs information or prior analysis requirements etc.

The concept of ontology has been extensively used for various knowledge management purposes in different areas such as flow classification and semantic web. Ontology is defined as the explicit specification of concept used to help programs and humans share knowledge [22]. In general ontology includes structured knowledge statements that describe the concepts of a domain and their relationship

[23]. As shown in Figure 2, the connection between context and ontology is linked. Another important term related to ontology is the knowledge base, which is formed by concept instances. According to the knowledge statement from ontology, we can further infer low-level contexts to yield the high-level data for the traffic classification [24]. Through the use of context and ontology, the rising quantities of traffic classification knowledge can be effectively managed in a shared and reusable manner.

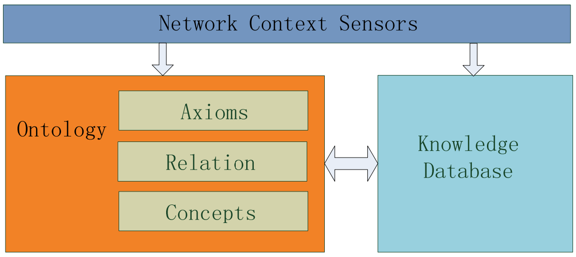

Figure 2. Relationship between context, ontology and knowledge database

3.2. IP flow profile

A flow profile can generally be described by a set of flow statistical features. The IP flow profile of each application is unique, which is affected by various flow statistical factors. In this paper, the accuracy of IP flow profile model is improved by using some features, all of which are closed related to traffic categories.

To build IP flow profile, the first issue is to determine what features should be chosen to construct profile and how to collect these features. As we have mentioned, application profiles are a set of flow statistical features and each flow statistical feature summarizes the common properties shared by a number of flows in the application, which include transport-layer properties and statistical properties. We focus on properties of bidirectional flows, also known as connections. The reason for this is that we need to differentiate the statistical observations on the request and the response direction, which are different in many applications. We use the Sequential Forward Selection (SFS) method to find the optimal feature set {number of bytes in backward direction, number of packets in backward direction, average window size, number of bytes in forward direction, number of packets in forward direction, mean forward packet length, mean backward packet length, duration of the flow}. Then, we can choose this optimal feature subset to model the IP flow profile.

3.3. Relationship between IP flow profile and traffic category

Given IP flow profile *O*, $X = \{x_1, x_2, \cdots, x_n\}$ is defined as a set of flows. A flow instance x_i is characterized by a vector of attribute values, $x_i = \{x_{ij} \,|\, 1 \le j \le m\}$, where *m* is the number of attributes, and x_{ij} is the value of the j^{th} attribute of the i^{th} flow. In some cases, an application can be classified by analyzing the optimal feature set. In the traffic classification context, examples of attributes include flow statistics such as duration and total number of packets. Also, let $Y = \{y_1, y_2, \cdots, y_q\}$ be the set of traffic classes, where *q* is the number of classes of interest. Classifiers use a training dataset that consists of *N* tuples (x_i, y_i) and learn a mapping $f(x) \to y$. The y_i can represent “P2P”, “Game”, and “FTP”.

4. Online Self-learning Traffic Classification Architecture based on Profile and Ontology

In this section, we propose online self-learning traffic classification architecture based on profile and ontology. As illustrated in Figure 3, network context sensors are responsible for collection of all the needed information as stated in the ontology. Ontology manager and knowledge database are designed to support the seamless knowledge management for traffic classification. IP flow profile engine and self-learning traffic classifier together are to implement automatic traffic identification functionalities. Eventually, the classification output would be applied to network activities such as network surveillance, QoS.

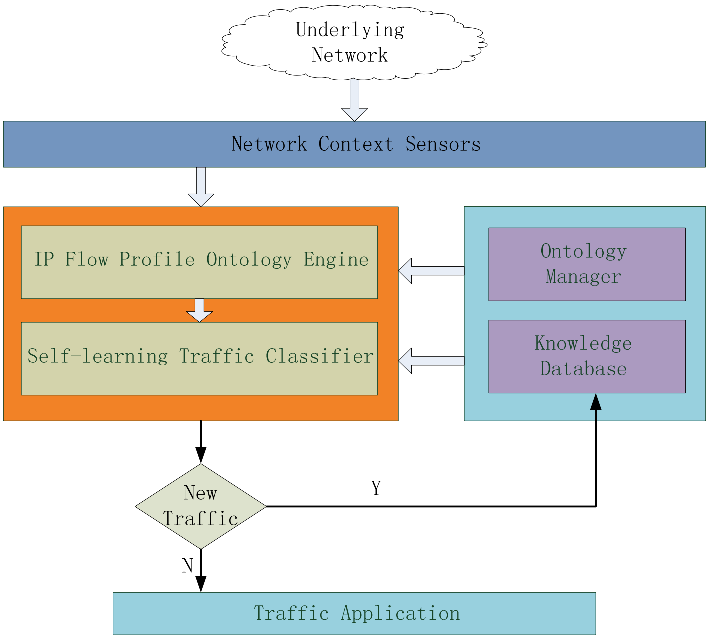

Figure 3. Online self-learning traffic classification architecture based on profile and ontology

Network context sensors collect of all the needed information as stated in the ontology. The flow context sensor collects raw traffic, using the packet capture library, and aggregates the data into flows. Then, the raw data will be transformed into contexts, which will be sent to the ontology clustering engine for the next step in data processing. Generally, it monitors not only the tangible objects, like network nodes, but also the intangible entities. Therefore, sensors are categorized into two types: the network node sensor and flow context sensor. A node sensor senses all the network activities happened on that node.

Ontology manager is the core part of the traffic classification knowledge management. It performs the ontology creation, insertion and maintenance tasks. To successfully achieve these tasks, the key is to have a clear picture about the relationships between different classification techniques. It can manage rising quantities of traffic classification knowledge.

IP flow profile Engine aims to learn the traffic characteristics using the data from network context sensors. The learning algorithm is realized with Sequential Forward Selection (SFS) method, a role of this module is to group contexts into different areas together with the feature statement in the ontology.

Self-learning traffic classifier is designed to perform the automatic classification functionality. It uses the data from IP flow profile engine and knowledge database to implement a profile and category mapping. This module should yield high-level knowledge that can be used to automatically classify traffic, and in particular to reduce the need for manual involvement.

5. Experimental Results and Analysis

5.1. Empirical traces

This subsection describes the empirical traces in our work. The overall network traffic trace consists of three sets: Moore_Set was collected from the experiment of Pro. Moore from Cambridge University;

Handmade_Set was simply labeled by manual classification in our laboratory using payload-based technique or port-based method; Univetsity_Set was collected from Nanjing University of Posts and Telecommunications. To simplify the presentation, we group the applications by category. For example, the P2P category includes all identified P2P traffic from protocols including PPlive, PPstream, BitTorrent, Gnutella, Xunlei and KaZaA.

Moore_Set trace consists of bidirectional network traffic of some biological research institute during 0 to 24 o'clock on Aug 20th, 2003. We extract 10 subsets with an average sampling time of 1680s to form our dataset, which contains 377526 samples of network flow. These samples are divided into 10 types. The application names of each type and the quality as well as the respective proportion of each network flow are shown in Table 2.

Table 2. Statistics of Moore_Set

Type of flow	Application names	Num of flow	Percent(%)
WWW	http, https	328091	86.91
MAIL	Imap, pop3, smtp	28567	7.567
BULK	ftp	11539	3.056
DATABASE	oracle, mysql	2648	0.701
SERVER	ident,ntp,x11,dns	2099	0.556
P2P	kazaa,bittorrent	2094	0.555
ATTACK	worm, virus	1793	0.475
MEDIA	real, media player	1152	0.305
INT	telnet,ssh,rlogin	110	0.029
GAME	half-life	8	0.002
Total	26 applications	377526	100

Each network flow sample of Moore Set is derived from a complete bidirectional TCP flow and contains 248 attributes, among which the first and second attributes are port numbers of source and destination respectively.

Unlike the usual way to obtain traces, we set a local experimental network with around 100 hosts to generate traffic manually to get Handmade_Set. Let each host run the specific application (HTTP, MAIL, FTP, DATABASE, P2P, GAME, etc.) at the same time. Since the applications run in the host is predetermined, it is easy to classify and categorize the traffic flow by the IP address. Table 3 summarizes the applications in our experiments. This set can be used as base truth to evaluate the accuracy of the classifier.

Table 3. Statistics of Handmade_Set

Type of flow	Num of flow	Percent(%)
WWW	1000	12.5
MAIL	1000	12.5
BULK	1000	12.5
DATABASE	1000	12.5
SERVER	1000	12.5
P2P	1000	12.5
MEDIA	1000	12.5
GAME	1000	12.5
Total	8000	100

To facilitate our work, we collect traces in all academic units and laboratories on the campus from the Internet gateway of Nanjing University of Posts and Telecommunications. Univetsity_Set was collected over a span of six months from April 10, 2009 to October 10, 2009. Table 4 summarizes the applications found in the 20 1-hour Campus traces. On the campus network, HTTP, DATABASE, and EMAIL traffic contribute a significant portion of the total flows. On this network, P2P contributes only 0.43% of the flows.

Table 4. Statistics of University _Set

Type of flow	Num of flow	Percent(%)
WWW	4606712	68.18
MAIL	561994	8.32
BULK	11786	0.17
DATABASE	1528681	22.62
SERVER	2876	0.04
P2P	29596	0.43
MEDIA	1698	0.03
GAME	13453	0.21
Total	6756796	100

5.2. Evaluation metrics

To measure the performance of our proposed method, we use three metrics: *accuracy, precision and recall.* In this paper, TP, FP, and FN are the numbers of true positives, false positives, and false negatives, respectively. True Positives is the number of correctly classified flows, False Positives is the number of flows falsely ascribed to a given application, and False Negatives is the number of flows from a given application that are falsely labeled as another application.

Accuracy is the ratio of the sum of all True Positives to the sum of all the True Positives and False Positives for all classes. We apply this metric to measure the accuracy of a classifier on the whole trace set. The latter two metrics are to evaluate the quality of classification results for each application class.

$$accuracy = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FP_i} \times 100\% \tag{1}$$

Precision of an algorithm is the ratio of True Positives over the sum of True Positives and False Positives or the percentage of flows that are properly attributed to a given application by this algorithm.

$$precision = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

Recall is the ratio of True Positives over the sum of True Positives and False Negatives or the percentage of flows in an application class that are correctly identified.

$$recall = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

5.3. Comparing performance among different technology

We compare classification accuracy of online self-learning Internet traffic classification based on profile and ontology with the other classification approaches solely based on port-based approach and payload-based in Table 5. Compared with our proposed method, port-based approach separates network traffic by exploiting the port information, the classification precision is 85.87%, 68.13% for WWW and GAME application respectively. Payload-based approach automatically extracts payload signatures to identify specific flows, and the classification precision is 82.35%, 67.23% for WWW and GAME application respectively. We construct our online self-learning traffic classification based on profile and ontology empirical classifier system to classify Internet traffic using Moore_Set, we get 91.47% and 76.75% for WWW and GAME application respectively. Compared with the other classification approaches, our proposed method based on profile and ontology in this paper can achieve higher recall and precision.

Table 5. Classification accuracy among different traffic classification method

Type of flow	Port-based		Payload-based		Proposed method	
	Recall	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*
WWW	82.51	85.87	82.71	82.35	90.16	91.47
MAIL	81.62	82.54	81.86	78.23	89.31	87.75
BULK	74.16	75.05	74.03	71.01	81.45	80.53
DATABASE	79.65	79.47	76.11	74.53	83.56	84.05
SERVER	79.78	78.51	78.77	75.48	86.22	85.73
P2P	72.21	71.17	74.88	69.88	82.33	79.47
ATTACK	77.67	71.63	70.86	67.19	78.31	76.71
MEDIA	82.82	75.95	79.66	71.62	87.11	81.14
INT	74.29	70.81	72.22	68.08	79.67	77.62
GAME	70.96	68.13	73.26	67.23	80.71	76.75

We calculate the relative computational time, demanded memory and accuracy through the experiments among different approaches. The performance for different approaches is demonstrated in Table 6. We can find that the online self-learning traffic classification based on profile and ontology is able to greatly improve the accuracy, while only minimally impacting computational time and demanded memory. There appears to be a very good trade-off between computational resource and loss of accuracy.

Table 6. Performance among different traffic classification method

Performance	Port-based	Payload-based	Proposed in this paper
Time(s)	10.930	400.356	114.583
Memory(M)	7.362	34.858	15.619
Accuracy(%)	71.87	73.35	91.76

5.4. Impact of number of packets for statistics on classification accuracy

In order to classify the network applications associated with a flow as early as possible, our proposed online self-learning traffic classification based on profile and ontology should classify a flow as soon as possible. A sub-flow is p consecutive packets taken from a full-flow. The sub-flow features have ability to distinguish each application, because the sub-flow contains a sequence exchanging control packets that are pre-defined messages in each application. The statistics information of several packets in sub-flow could distinguish network traffic from Internet traffic accurately with the least p. We conduct experiments to determine the appropriate packet number p. For our experiments, we classify network traffic using flows from University _Set and Handmade_Set respectively.

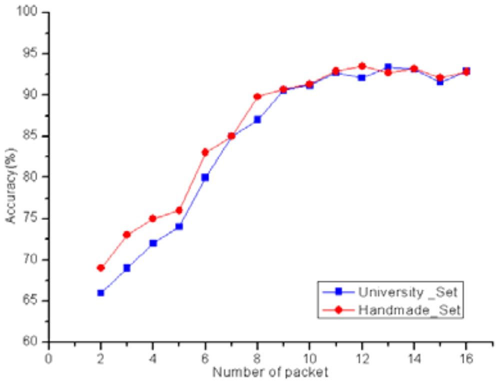

Figure 4. Impact of number of packets for statistics on classification accuracy

The experimental result indicates that the statistics of the first 11 packets could classify traffic with high accuracy in both traces in Figure 4. At the same time, the classification accuracy improves marginally using more than 11 packets. Considering our goal to detect network traffic rapidly with high accuracy, we choose 11 packets for online network traffic classification.

5.5. Classifying unknown P2P traffic

The purpose of this section is to verify whether the proposed method is robust or flexible enough to detect unknown P2P traffic. We then test our classification method using P2P flow from Handmade_Set. The main applications in our experiments include Xunlei, PPlive, PPStream, BitTorrent, Kougou, eDonkey and unkonwn P2P. Table 7 shows statistics of P2P flow from Handmade_Set.

Table 7. Statistics of P2P flow from Handmade_Set

Application	Num of flow	Percent(%)
Xunlei	100	10
PPlive	200	20
PPStream	100	10
BitTorrent	200	20
Kougou	100	10
eDonkey	200	20
Unkonwn	100	10
Total	1000	100

With the method proposed in this paper, the results in Figure 5 show that average Precision and Recall in Handmade_Set are 85.61% and 81.89% respectively, which indicates that P2P traffic can be effectively identified. Specifically, the experimental results show that Precision and Recall is 84.86% and 77.63% for unknown P2P application respectively, which indicates that the methods can classify unknown P2P traffic with considerable accuracy. This method can manage existing traffic classification knowledge and reason unknown traffic application for achieving a self-learning traffic classification with high accuracy. Therefore, the methods are robust enough to classify unknown P2P traffic based on profile and ontology as well as known P2P traffic successfully.

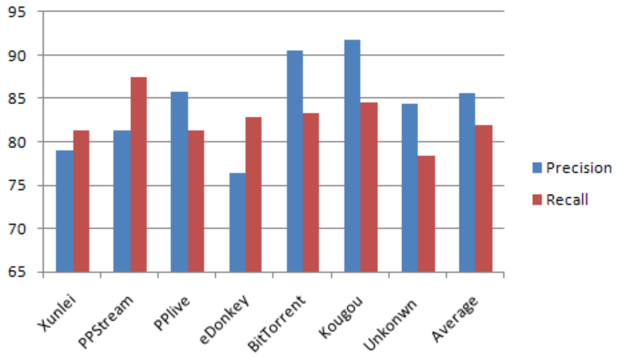

Figure 5. Classification accuracy of unknown P2P traffic

6. Conclusions

As playing important roles in many areas such as traffic engineering, service class mapping, network management etc, network traffic classification recently has attracted a great deal of interest. One of the challenging issues for existing classification methods is that they need prior manual traffic analysis to classify unknown traffic, which is infeasible to cope with the fast growing number of new applications. In this paper, we propose an online self-learning traffic classification based on profile and ontology, which is realized by managing traffic classification knowledge with the use of ontology, while developing the self-learning model on traffic classification according to ontology. Experiment results demonstrate this classification technology can online classify the network traffic effectively. Moreover, we also need more experiments to find out how to deploy this traffic classification method into a large network to improve network management and QoS service.

7. Acknowledgement

This work is supported by National High-Tech Research and Development Plan (863) of China (No.2006AA01Z232, No.2009AA01Z212, No.2009AA01Z202), Natural Science Foundation of China (No.61003237), Natural Science Foundation of Jiangsu Province (No.BK2007603), High-Tech Research Plan of Jiangsu Province (No.BG2007045).

8. References

[1] Antonio Martin, Carlos Leon, Felix Biscarri, "Intelligent Integrated Management for Telecommunication Networks ", International Journal of Advancements in Computing Technology, vol. 2, no. 2, pp. 158-171, 2010.

[2] Qindong Sun, Qian Wang, Jie Ren , "Modeling and Analysis of the Proactive Worm in Unstructured Peer-to-Peer Network ", Journal of Convergence Information Technology, vol. 5, no. 5, pp. 111-117, 2010.

[3] Fazal Wahab Karam, Terje Jensen, "A Survey on QoS in Next Generation Networks", Advances in Information Sciences and Service Sciences, vol. 2, no. 4, pp. 91-102, 2010.

[4] Haffner P, Sen S, Spatscheck O, Wang D. "ACAS: Automated Construction of Application Signatures". In Proceedings of SIGCOMM'05 Workshops, pp. 197-202, 2005.

[5] Moore A W, Papagiannaki K. "Toward the accurate identification of network applications". In Proceedings of Passive and Active Measurement Workshop, pp. 41-54, 2005.

[6] Constantinou F, Mavrommantis P. "Identifying known and unknown peer-to-peer traffic". In Proceedings of IEE NCA'06 Conference, pp. 93-102, 2006.

[7] T. Karagiannis, A. Broido, N. Brownlee, K. Claffy, M. Faloutsos. "Is P2P dying or just hiding". In Proceedings of IEEE Globecom, pp. 1532-1538, 2004.

[8] T. Karagiannis, K. Papagiannaki, M. Faloutsos. "BLINC: multilevel traffic classification in the dark". In Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications, pp. 229-240, 2005.

[9] S. Sen, O. Spatscheck, D. Wang. "Accurate, scalable in-network identification of p2p traffic using application signatures". In Proceedings of the 13th international conference on World Wide Web, pp. 512-521, 2004.

[10] B.Marco, Mellia, Antonio Pescape, LucaSalgarelli. "Traffic classification and its applications to modern networks". Computer Networks, vol. 53, no. 6, pp. 759-76, 2009.

[11] T. Nguyen, G. Armitage. "A Survey of Techniques for Internet Traffic Classification using Machine Learning". IEEE Communications Surveys and Tutorials, vol. 11, no.3, pp. 37-52, 2008.

[12] J.Erman, A. Mahanti, M. Arlitt, I. Cohen, C. Williamson. "Offline/realtime traffic classificationusing semi-supervised learning". Performance Evaluation, vol. 64, no. 9, pp. 1194-1213, 2007.

[13] J. Erman, M. Arlitt, A. Mahanti. "Traffic classification using clustering algorithms". In Proceedings of SIGCOMM workshop on Mining network data, pp. 281-286, 2006.

[14] Bernaille L, Teuxeira R, Akodkenous I, Soule A, Slamatian K. "Traffic classification on the fly". In Proceedings of ACM SIGCOMM, pp. 23-26, 2006.

[15] A. W. Moore, D. Zuev. "Internet traffic classification using bayesian analysis techniques". In Proceedings of international conference on measurement and modeling of computer systems, pp. 50-60, 2005.

[16] N. Williams, S. Zander, G. Armitage. "A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification". ACM SIGCOMM Computer Communication Review, vol.30, no. 5, pp. 5-16, 2006.

[17] T. Auld, A. W. Moore, S. F. Gull. "Bayesian neural networks for internet traffic classification". IEEE Transaction on Neural Network,vol.18, no. 1, pp. 223-239, 2007.

[18] Yongli Ma, Zongjue Qian, Guochu Shou, Yihong, Hu. "Study of information network traffic identification based on C4.5 algorithm". In Proceedings of 2008 International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1-5, 2008.

[19] Yager Ronald R, Petry Frederick E. "A multicriteria approach to data summarization using concept ontologies". IEEE Transactions on Fuzzy Systems, vol. 14, no. 6, pp. 767-779, 2006.

[20] Brezillon Patrick, Cavalcanti Marcos. "Modeling and using context". Knowledge Engineering Review, vol. 13, no.2, pp. 185-194, 1998.

[21] Bettini Claudio, Brdiczka Oliver, Henricksen Karen. "A survey of context modelling and reasoning techniques". Pervasive and Mobile Computing, vol. 6, no. 2, pp. 161-180, 2010.

[22] Jia Chen, Yue Wu. "Rules-based object-relational databases ontology construction". Journal of Systems Engineering and Electronics, vol. 20, no. 1, pp. 211-215, 2009.

[23] Wimalasuriya Daya C, Dou Dejing. "Ontology-based information extraction: An introduction and a survey of current approaches". Journal of Information Science, vol. 36, no. 3, pp. 306-323, 2010.

[24] Goodwin J Caleb, Russomanno David J. "Ontology integration within a service-oriented architecture for expert system applications using sensor networks". Expert Systems, vol. 26, no. 5, pp. 409-432, 2009.