Study on Traffic Classification Taxonomy for Multilateral and Hierarchical Traffic Classification

Ji-hye Kim, Sung-Ho Yoon, and Myung-Sup Kim
Dept. of Computer and Information Science
Korea University
Korea
{jihye_kim, sungho_yoon, tmskim}@korea.ac.kr

Abstract— **Internet traffic has rapidly increased due to the increasing use of wireless devices and the appearance of various applications and services. With the rapid increase of** Internet **traffic, the need for Internet traffic classification becomes important for the effective use of network resources. However, the traffic classification taxonomy has received little attention compared to the study of classification methods. In this paper, we propose novel traffic classification taxonomy for multilateral and hierarchical traffic identification. The proposed taxonomy can support multilateral identification based on the proposed four classification criteria: service, application, protocol, and function. In addition, the proposed taxonomy can support hierarchical structure supporting roll-up and drill-down operation to the classification result. We proved the applicability and advantages of the proposed taxonomy by applying it to real campus network traffic**

Keywords-* *Traffic classification; Traffic identification; Traffic Taxonomy;

I. INTRODUCTION

Internet traffic has rapidly increased due to the appearance of various applications and services on Internet, as well as the increasing use of wireless smart devices. As the Internet traffic increases, the need for traffic classification is getting more important in order to prevent a traffic crisis and manage limited network resources. Furthermore, traffic classification can be used for a variety of purposes such as trend analysis of Internet services and user preference in various Internet services.

The establishment of precise traffic classification taxonomy can not only solve the problems listed above but also has many additional benefits. First, it is possible to develop more elaborate traffic analysis systems, because complex and varied analysis methods can be classified within the precise classification taxonomy. Second, it is possible to analyze traffic in the perspective of multiple dimensions. That is, the same traffic can be analyzed in an application dimension or traffic occurrence dimension. Multi-dimensional analysis is useful for understanding and utilizing traffic. Third, it can provide an accurate understanding of traffic analysis results to both end users as well as network managers.

In this paper we propose a classification taxonomy structure, and propose classification criteria and attributes for mutiliteral and hierarchical classification of Internet traffic. We defined classification taxonomy with a tree structure. Also we proposed four classification criteria (Service, Application, Protocol, and Function) for multidimensional classification. Each criterion has hierarchical attributes for utilizing the hierarchical results. For understanding of the proposed classification taxonomy, we explain classification results with an example. Also, we prove applicability of the proposed classification taxonomy by experiment with real campus network traffic.

The remainder of the paper is organized as follows. Section 2 describes related work on the analysis of classification taxonomy for commercial network equipment. Section 3 describes the proposed classification taxonomy in terms of structure, criteria, and attributes. Section 4 describes the experiment for proving the feasibility of the proposed classification taxonomy. Section 5 summarizes this paper and outlines possible future work.

II. RELATED WORK

In this section we explain about the existing classification taxonomy used in commercial traffic analysis system, and the classification methods that we use in this paper for our experiment.

A. Existing classification taxonomy

In our previous study, we analyzed the classification taxonomy used in some commercial network traffic analysis systems and the requirements of classification taxonomies. We defined a dimension as a classification criteria and attributes as a layer in each criterion.

CheckPoint[3] and Fortinet[4] use the application as a single dimension criterion for traffic classification. They use hierarchical attributes made from groups of each application for traffic classification. By using a tag function, they describe additional information for each application. Thus, this method can give a lot of information about each application to users. However, it is hard to utilize and apply this information to various areas of traffic analysis flexibly and extensively.

Paloalto[5] uses the application and its functions as a mixed single dimension for traffic classification. This method has an advantage that users can obtain classification results systematically and use these results easily. However, this classification taxonomy has a disadvantage that it is hard to define the relationship between application and function using mixed dimensions.

[1] This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (20100020728)

III. Proposed Classification Taxonomy

In this section, we propose the basic structure of multilateral traffic classification taxonomy and classification criteria with each attributes which can analyze internet traffic effectively. Also we explain advantages and feasibility of proposed classification taxonomy.

A. Classification taxonomy structure

In this section, we propose a new traffic classification structure which satisfies a variety of classification purposes and results in multilateral and hierarchical classification.

Figure 1 shows the overall structure of the classification taxonomy proposed in this paper. Depending on defined classification criteria, the number of children nodes from the root can be created in the second level, which represents the dimensions of traffic classification. Each classification dimension has hierarchical classification attributes as a tree structure. When we define the level of the classification attributes, we count the number of levels in the classification tree except the root node and its first children nodes. So, the lower level of classification criteria (grand-children node of the root) is the first level in the classification attributes. By using hierarchical classification attributes, the proposed classification taxonomy provides the drill-down or roll-up operation to the classification result, so it can provide multilateral and hierarchical results that users are interested in.

Each classification attribute has variety of elements values but the unit of value should be equal in one root. For example, one classification taxonomy can't support both a byte unit of value and packet unit of value in the same classification taxonomy. Also, conflicts and duplicates of classification results are avoided by applying the following rules to attributes in each classification dimension.

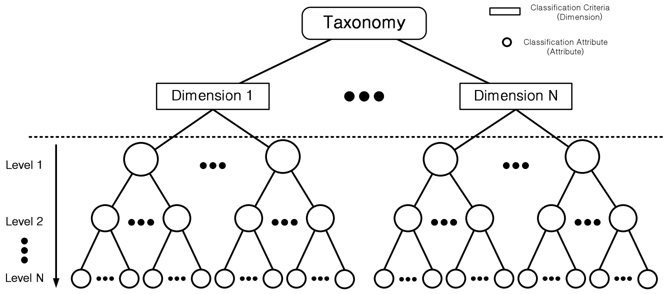
Taxonomy
Classification Criteria (Dimension)
Classification Attribute (Attribute)
Dimension 1
Dimension N
Level 1
Level 2
Level N

Figure 1. Overall Structure of Classification Taxonomy

- The sum of the measured values of classification attribute nodes in level N has to be the same as the sum of measured values of classification attribute nodes in level N-1.
- Sum of the measured values of classification attribute nodes in level 1 has to be the same as the original measured value of the input data from the classification target.

B. Proposed classification criteria and their feasibility

In this section, we describe the four classification criteria that can be used for various purposes and produce multilateral classification results effectively. And we prove the feasibility of each criterion.

Table 1. definition of four classification criteria

Criteria	Definition
Service	All type of IT service accessed by user for specific purpose.
Application	Used application in end host
Protocol	Used protocol when traffic is transferred.
Function	Purpose of traffic occurrence based on user's behavior.

The proposed four kinds of classification criteria (service, application, protocol, and function) have high utilization and give accurate information about traffic due to provision of multilateral classification results for various purposes of traffic classification. By using four kinds of classification criteria, we can get some of the following effects

First, by classifying based on service criterion, we can prevent unclear classification of services which are provided by a number of different end-user applications. For example NateOn messenger service provides the same service in the form of NateOn standalone application as well as Internet browsers. In this case if we use the application criterion for classification, we have to classify the same service traffic as two different applications, for example Internet Explorer and NateOn. Therefore, service criterion for classification has the advantage that the service criterion performs more detailed analysis than the application and protocol criteria.

Second, by using the application criterion we can get more specific information about traffic and human behavior pattern. By using application criterion we can classify same service more detail in the viewpoints of application and protocol.

Third, by using the protocol classification criterion we can reduce the needs for additional analysis of new applications or services that use the same specific protocol. For example, the eDonkey protocol is used by a number of P2P service applications such as Donkey, Pruna, Emule, and so on. These applications are often developed because of the nature of P2P applications. In this case, if we do not use protocol as the classification criterion, additional work to analyze an application when a new application is developed or updated must be performed. So, if we classify based on protocol classification criterion it has the advantage in that we can reduce the needs of additional application analysis.

Fourth, if we classify based on function classification criterion, the traffic classification results can be enhanced by the utilization of a special control field by classifying specific functions in the same service or application. For example, if we classify traffic based only on service or application criteria, we classify traffic that is generated by NateOn file transfers as the NateOn service or application, respectively. However these results cannot distinguish between the file transfer function and other functions in NateOn. So when traffic classification results are used for control purposes, it includes all traffic from the NateOn service or application. However if we classify based on function criteria, we can classify specific function in the same service or application.

The proposed classification taxonomy performs multilateral classification using four criteria so it can provide accurate and fine-grained information about traffic, as showed in Table 2.

Table 2. example of four demension of classification criteria

Criteria	Classification result
Service	Internet Explorer
Application	Google
Protocol	HTTP
Function	Web posting

Each criterion can be optional in a signature. The service criterion has to have more than 2 levels in the attribute hierarchy of Fig.1. And the application and protocol criteria have to have all level it can classify. Also the function criterion has to have more than 1 level.

C. Proposed attributes of classification criteria

In this section we propose a hierarchical structure of attributes for the proposed classification criteria. As mentioned earlier, hierarchical attributes provide details of results utilization. Also we can manage classification results easily by using hierarchical attributes.

In this paper, we propose the rules of hierarchical attributes for each criterion in the following table.

Table 3. definition of attributes for each criteria

Criteria	Level 1	Level 2	Level 3
Service	Service Type	Service Name	Provided Service
	Ex. Multimedia	Ex. YouTube	Ex. Video
Application	Application Purpose	Application Type	Application Name
	Ex. Web	Ex. Browser	Ex. I.E
Protocol	Protocol Type	Protocol Name	
	Ex. Web	Ex. HTTP	
Function	User Purpose	User Action	
	Ex. Multimedia	Ex. Video Strmeaing	

As shown in Table 3, service and application criteria are composed of the three layers, protocol and function criteria are composed of two layers. If we do not use the proposed attributes of each criterion, the management and utilization of classification results becomes difficult. Using 2 levels in service criterion can result in incorrect results. For example, the mail and search services in Google can be defined as Portal→Google→Mail or Portal→Google→Search by using 3 levels of attributes. However, using only 2 levels of attributes can only define Portal→Google, Mail→Google or Search→Google. As such, these results cannot support flexible utilization and detail understanding of classification results. Also, if we do not use hierarchical attributes according to the user's purpose and action in the function criterion, we cannot perform accurate control in the control field. For example, management policy may allow file downloads in NateOn but not uploads. In this case as we cannot distinguish the user's purpose, we cannot control upload and download separately as these 2 activities are simply classified as file transfer only.

Table 4 shows examples of proposed classification criteria and attributes for Yahoo traffic. The example below assumes Yahoo is used through the Internet Explorer browser. In Table 4, traffic type means when traffic occurs. So, on second record in this table, traffic occurs when the user do search on Yahoo. In this case traffic should be classified as Portal→Yahoo→Search in service criterion, Web→Browser→I.E in application criterion, Web→HTTP in protocol criterion, and Web→Search in function criterion. In service criterion, Portal is a group name at level 1 and Yahoo is service name at level 2. Also, Search is the name of the provided service by Yahoo, and it is an e provided service by Yahoo, and it is an element of the third level. Table 4 shows 2 different services in the sixth row and ninth row. They have Market→Basket and ad→adthis. The sixth row traffic occurred when a user use shopping category in Yahoo by named basket shopping mall. In the ninth row, traffic is generated by an advertisement service on Yahoo, this is classified in the ad group in service criterion. Even though it occurred when a user accesses the Yahoo site, this traffic is generated from other advertisement site. Also we leave a blank in the last level for each criterion if we can't classify it accurately.

Table 4. Exampe of suggetsted classification taxonomy

Traffic type	*Service*	*Application*	*Protocol*	*Function*
Yahoo main page access	Portal → Yahoo	Web → Bowser → I.E	Web → HTTP	Web →
Search	Portal → Yahoo → search	Web → bowser → I.E	Web → HTTP	Web → search
Blog	Portal → Yahoo → blog	Web → bowser → I.E	Web → HTTP	Web → comment
News	Portal → Yahoo → news	Web → bowser → I.E	Web → HTTP	Web → news
dictionary	Portal → Yahoo → dictionary	Web → bowser → I.E	Web → HTTP	Education → search
Shopping	Market → basket	Web → bowser → I.E	Web → HTTP	Shopping →
Login	Portal → Yahoo → login	Web → bowser → I.E	TCP → TLS	Managemen → Login
mail	Portal → Yahoo → mail	Web → bowser → I.E	Web → HTTP	Mail →
ad	Ad → adthis	Web → bowser → I.E	Web → HTTP	Web → ad

IV. Application of Classification Taxonomy

In this section we explain about classification results of real signatures for the proposed classification taxonomy criteria and attributes

A. *Classification results*

In this section we prove it is possible to use the proposed classification taxonomy for real enterprise network traffic. Table 7 shows the details of the traffic trace for the experiment.

Table 5. Traffic trace

Flow	Packet	Byte
66,620K (num)	3,417,367K (num)	3,004,436MB

Fig. 3 shows the classification results of Yahoo traffic multilaterally. From Fig. 3, we can get information about which Yahoo service traffic is provided by which application or which protocol and traffic purpose.

B. *Classification results*

In this section we prove it is possible to use the proposed classification taxonomy for real enterprise network traffic. Table 7 shows the details of the traffic trace for the experiment.

Table 6. Traffic classification results

Flow	Packet	Byte
66,620K (num)	3,417,367K (num)	3,004,436MB

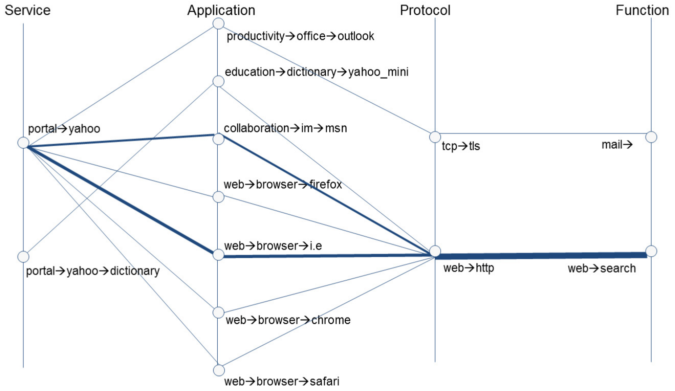

Figure 2. Classification result of Yahoo traffic

Fig. 3 shows the classification results of Yahoo traffic multilaterally. From Fig. 3, we can get information about which Yahoo service traffic is provided by which application or which protocol and traffic purpose. According to Fig. 3, most traffic of the Yahoo service is provided by Internet Explorer or MSN. All traffic except for that from outlook uses the HTTP protocol. We can guess the Yahoo service from outlook to be a mail service on level 3, but the signature we have could only classify as far as Yahoo for level 2.

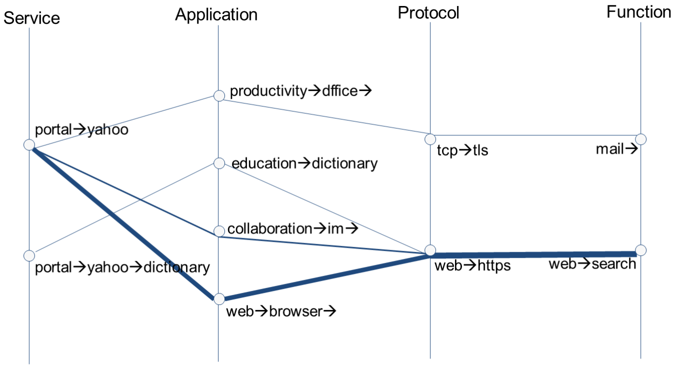

Figure 4. Classification of Yahoo traffic
(using 2 level in application)

Fig. 4 shows the multilateral results when the application criterion uses 2 levels. According to Fig. 4 we can see most traffic from Yahoo comes from the browser group. As a result, we can say the proposed classification taxonomy can be used for real traffic classification. Also we have proved it is possible to classify based on the proposed four criteria and to use the classification results effectively by roll-up and drill-down of hierarchical attributes. By applying four criteria to the traffic, we proved that the proposed classification taxonomy can provide multilateral results and specific information to heighten understanding of user activity.

V. Conclusion

In this paper, we proposed new classification taxonomy for multilateral and hierarchical Internet traffic classification in the form of structure, criteria and attributes. Also we applied our method to a real enterprise network traffic trace to prove the usefulness of our classification taxonomy's criteria and the possible utilization of classification results. In the experiment we showed that we classified 99% of enterprise Internet traffic while in the function criteria we classified only 20% due to the absence of fine-grained signatures for function domain classification. In future, we will study the definition and extraction of fine and accurate signatures for classification taxonomy.

References

[1] Cisco, "2010~2015 Cisco Visual networking Index", 2010.

[2] Ji-hye Kim, Sung-Ho Yoon and Myung-Sup Kim, "Research on Traffic Taxonomy for Internet Traffic Classification," Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2011, Taipei, Taiwan, Sep. 21-23, 2011.

[3] CheckPoint, Available from:< http://appwiki. check point.com/appwikisdb/public.htm>.

[4] Fortigaurd, available form: <http://www.forti guard.com/applicationcontrol/ListOfA plicati ons.html>.

[5] Paloalto networks, Available from:< http:// www.paloaltonetworks.com/applipedia/>.