Improved Processing Speed of Traffic Classification based on Payload Signature Hierarchy

Ji-Hyeok Choi, Myung-Sup Kim
Dept. of Computer and Information Science
Korea University
Korea
{jihyeok_choi, tmskim}@korea.ac.kr

***Abstract*— The rapid development of the Internet has necessitated effective network management. Traffic classification is the process or protocol used to manage network traffic. Signature-based classification is the best known traffic analysis technique. This technique, however, has the negative aspect that processing speeds slow as the number of signatures increases. In this paper, we propose a novel method that elevates the processing speed by using a signature hierarchy. In addition, we propose a representative signature concept that reorganizes the conventional one-dimensional signature methodology. Signatures can be structured through representative signatures in a hierarchical structure. The traffic classification system can then classify traffic based on structured signatures. Using this new matching algorithm, traffic can be classified faster than with conventional methods. The feasibility of our proposed system has been demonstrated with experiments on campus traffic data.

Keywords- Traffic Classification, Representative Signature, Structured Signature.

I. INTRODUCTION

The importance of effective network management has become critical with the rise in the popularity of the Internet and its related services. Capturing the traffic related to an application is very important. Signature-based analysis is the most common method of determining actual traffic. This method can distinguish the unique characteristic of each application [2,3]. The problem with this procedure, however, is that the speed of the analyzing process slows as the number of signatures increases [1]. This paper proposes a method that can manage signatures effectively using the concept of a representative signature. In addition, a speed boosting technique has been suggested for traffic classification applications that use the hierarchical signature structure. Conventional traffic classification systems consider signatures as one layer, and match them one-dimensionally [3]. The speed of one-dimensional matching slows as the number of signatures increases [5]. To overcome this problem, we layer the flat signatures, and match the traffic from the top layer.

This paper is written in the following order. In Section II, the matching structure of the traffic classification system is described. In Section III, the concept of the representative signature and its hierarchy method is explained. This section also describes the algorithm for applying the layered signature into the traffic classification system. Experiments using the proposed system are described in Section IV, and the results and future work are discussed in Section V.

II. RELATED WORK

A conventional traffic classification system is shown in Fig. 1. First, the traffic is collected with a collecting program, and a flow generator transforms the traffic to a flow. The transformed flow begins matching the signatures chronologically. As the number of signatures increases, this structure inherently implies that the time required for matching will increase [4].

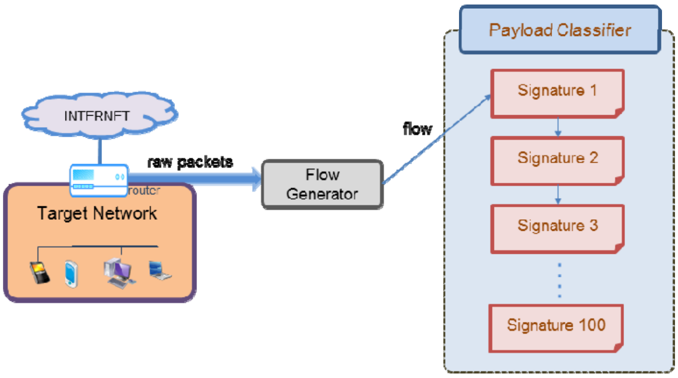

Figure 1. Traffic Classification System Matching Structure

To solve this problem, we propose the hierarchical signature structure shown in Fig. 2. This hierarchical signature structure has two advantages. First, effective management is possible. With a flat structure, the signatures are placed randomly in the memory. It is difficult, therefore, to manage each application. With the hierarchical structure, however, the signatures can be managed independently by the application, because the nodes are separated into children and parents through their representative signature. Second, the analysis speed is significantly faster than with the conventional method. A detailed explanation of the speed improvement is given in the next section.

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (2010-0020728) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A2007483).

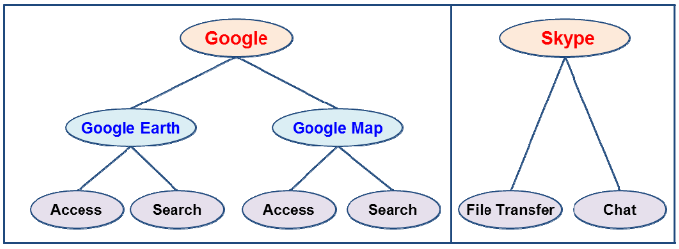

Figure 2. Google and Skype hierarchy structure

III. Signature Hierarchy and Representative Signature

Each application has its own signature hierarchy with up to three levels. For example, services from Google, such as Google Earth and Google Maps, have a broad range of applications. Each service, then, may be considered as an application, and a layer is created. As Google Earth and Google Maps are already categorized as Google, there are a total of three levels. On the other hand, unlike Google, Skype has no additional service applications, and so it has only two levels. Figure 2 shows the hierarchical structure for Google and Skype.

Google Earth, Google Maps, and Skype each have two signatures. There are a total of six signatures, therefore, in the first level. First-level signatures are those in use when an application is being accessed. For instance, when you use Google Earth, there may be one signature when the user accesses Google Earth, and another when a specific area is searched. A signature is defined as a first-level signature when a function is used in a single application.

Second-level signatures are the representative signatures of first-level signatures. The representative signatures in the higher levels are created by determining a common factor among the occurrences in the lower level—for example, when the user accesses Google Earth and searches a specific area. By finding a common string in the various signatures when two different functions are used, a common factor may be found. If a common string exists, a representative signature can be created for the corresponding application.

A third-level signature is also called a representative signature. The method of finding such signatures is the same as for second-level representative signatures. The third-level signature can be created by finding a common factor between the strings of the second-level representative signatures, for example, from Google Earth and Google Maps.

In other words, a signature is called a representative signature when, as a parent signature, it includes more than two child nodes. Using this representative signature methodology, a single application can be transformed into two or three levels of signatures.

The layered signatures have a different matching structure to the conventional flat structure when applied in the traffic classification system. Figure 3 shows the matching algorithm with the proposed hierarchical signature structure.

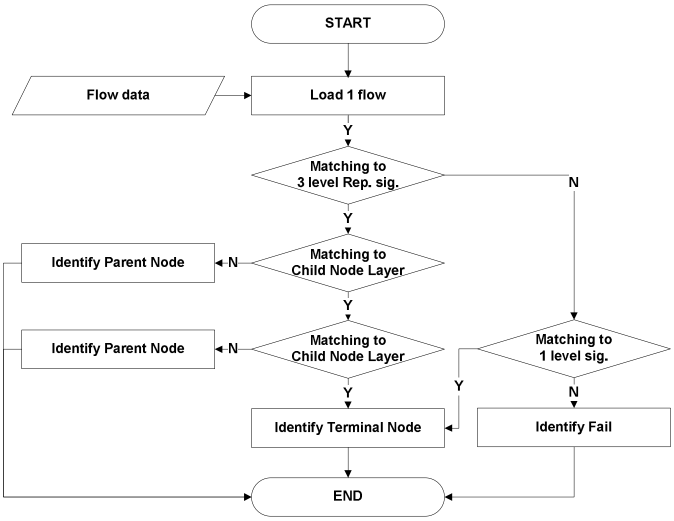

Figure 3. Matching Algorithm

First, when a flow is entered as an input, the corresponding flow begins the matching procedure with the third-level representative signatures. If the flow is matched, the matched third-level signature begins the second matching procedure with the child nodes at the lower level. If the second matching is successful, the flow initiates the third matching with the child nodes at the lowest level. When the third matching has successfully completed, the last child node, which is the first-level signature, becomes the flow signature that matches the flow input. If the matching procedure is successful at the parent mode, but fails at the child node, the last signature existing at the parent node becomes the flow signature from the input. Assume that the user calls the search function of Google Maps, as shown in Fig. 3. The final result would be a flow that corresponds to the search function of Google Maps. This flow would be identified because its signature exists in the present traffic classification system. The third-level representative signature, called Google, initiates the first matching in the identification process. As the third-level representative signature, in this case, includes first- and second-level signatures, the first matching with the third-level representative signature will be successful. Next, Google Earth and Google Maps begin the second matching. The second matching will be successful, because Google Maps has two child nodes that represent access and search. Finally, the child node of Google Maps will initiate the third matching. The corresponding flow would be identified as a signature that indicates a three-level search, because the child node of Google Maps has a search function. If the user used the log-in function for Google Maps, however, the flow would be identified as the representative signature located in the second level, as it does not have a signature at the first level.

If the user accesses the Google Calendar application, the flow cannot be identified by either the first or the second signature. However, it will be identified by the third-level representative signature, as it contains the Google string.

The analysis speed using the proposed algorithm in the traffic classification system would be faster than with the conventional flat system.

IV. EXPERIMENTAL RESULTS

In this section, we demonstrate that the proposed hierarchical signature structure is significantly faster than the conventional method.

Table 1. Applications of Experiment

Application name	Signature hierarchy		
	First layer	*Second layer*	*Third layer*
Naver	1	13	42
Google	1	7	32
Afreeca	-	1	10
Gomtv	-	1	9
Melon	-	1	7
Total	2	23	100

Table 1 shows the list of applications used in the experiment. Naver and Google applications consist of a three-level structure. Afreeca, Gomtv, and Melon are two-level structures. The experiments were carried out with 100 signatures and five applications. The traffic trace was formed by collecting a total of 5,000 ground-truth flows for the five applications.

Table 2. Matching Speed

Matching Result	Flat structure	Hierarchical structure
Number	293,729	69,534
Time (s)	1.79	0.26

Table 2 shows the matching numbers and times for the flat and hierarchical structures. The 5,000 ground-truth flows were input to the traffic classification system. Two experiments were conducted. The first measured the number and time for matching using the flat structure. The second experiment measured the number and time for matching using the hierarchical structure. The flat structure gives approximately 4.2 times as many matchings as the hierarchical structure. In terms of matching time, the hierarchical structure was 6.8 times faster at processing and analyzing than the flat structure.

Table 3. Accuracy and Completeness

Analysis Result	Flat structure	Hierarchical structure
Accuracy (%)	100	100
Completeness (%)	53.79	62.23

Table 3 shows the accuracy and completeness of the flat and hierarchical structures. Because the same ground-truth flow was used, both the flat and hierarchical structures show 100% precision. The hierarchical structure produces an 8.5% higher recall value, however. The hierarchical structure has additional signatures along with its base signatures. The analyzed flow with the representative signatures, therefore, resulted in the increased recall value.

V. CONCLUSION AND FUTURE WORK

The number of signatures for traffic classification has increased as the number of applications has grown. However, the current traffic classification system slows considerably with an increasing number of signatures.

This paper proposed an efficient method to manage the signatures and improve the traffic classification speed. Using the proposed method, traffic classification is faster than with the conventional traffic classification system. Additional traffic classification volume processing is also possible using the representative signature.

In future work, we will develop a signature hierarchy system based on the method proposed in this paper. Furthermore, we will certify the validity of our method by applying it to various traffic data cases.

REFERENCES

[1] Myung-Sup Kim, Young J. Won, and James Won-Ki Hong, "Application-Level Traffic Monitoring and an Analysis on IP Networks," ETRI Journal, Vol.27, No.1, Feb. 2005, pp.22-42.

[2] Jun-Sang Park, Jin-Wan Park, Sung-Ho Yoon, Young-Seok Oh, Myung-Sup Kim, "Development of Signature Generation System and Verification Network for Application Level Traffic Classification," Conference of Korea Information Communication Society, Apr. 23-24, 2009, pp. 1288-1291.

[3] Liu, Hui Feng, Wenfeng Huang, Yongfeng Li, Xing "Accurate Traffic Classification," Networking, Architecture, and Storage, 2007. International Conference.

[4] Park Jun-Sang, Park Jin-Wan, Yoon Sung-ho, Kim Myung-Sub, "Performance Improvement of Application-level Traffic Classification Algorithm using Payload Signature," The Korean Institute of Communications and Information Science, June. 23-25, 2010, pp.482.

[5] Sungho Yoon, Myung-Sup Kim, "Research for Performance Improvement of Internet Traffic Classification System," The Korean Institute of Communications and Information Science, Feb. 8-10, 2012.