Signature Management System to cope with Traffic Changes in Application and Service

Kyu-Seok Shim[1], Sung-Ho Yoon[1], Mi-Jung Choi[2] and Myung-Sup Kim[1]
Dept. of Computer and Information Science, Korea University[1]
Dept. of Computer Science, KangWon National University[2]
Sejong, Korea
{kusuk007, sungho_yoon, tmskim}@korea.ac.kr[1], mjchoi@kangwon.ac.kr[2]

***Abstract*—Today, the number of applications using network service has been increasing. Also, many applications have changed their traffic pattern frequently due to various reasons. Nevertheless, network managers tend to stay with old signatures. But they should update with new signatures to detect the modified application traffic. The extraction of signature is work to demand a lot of time. And it is difficult to continuously and timely extract the new signature for all applications. In this paper, we propose a noble signature management system which automatically extract new signatures detecting the modified traffic and delete old signatures no longer used. The proposed system analyzes traffic with existing signatures and extracts new signature automatically for updated traffic. For automatic generation of new signatures, we uses a sequence pattern algorithm. Also, the proposed system analyze usage of the old signatures to remove them when they are not used any more. We proved the feasibility and applicability of the proposed system by showing that that detection rate of all application was increased.**

Keywords—Network; Signature; Payload; Traffic; Automatic Generate; Update; Management;

I. Introduction

The purpose of network management makes the best use of network resources and provides services smoothly to users [1,2]. The most important thing in network management is network traffic monitoring. Network traffic monitoring is to determine whether any application and service. Essentials for network traffic monitoring is signature of each application. Signature means unique characteristic of traffic from each application. Network manager can monitoring the amount of each application by using signatures.

Today, the number of applications using network service have been increased. Also, many applications have modified their traffic pattern frequently due to various reasons. Therefore, network management is not easy from the perspective of network manager. Nevertheless, network manager have detectable signature, they should extract new signature for detecting the modified to application traffic. In this paper, proposed system extracts new signatures that are possible to detect modified traffic and delete old signature that is not used to analysis.

Automatic signature generation is the field that is being studied in a variety of ways [4-9]. Network manager should detect in a short time for the new application. However, signature is required in order to respond to new application. The original method for signature extraction requires the hand of the network specialist. This original method has various problems.

First, it is spent a lot of time and cost. Signature generation should check the large amount of traffic. Especially, payload signature is difficult to extract by checking the payload in traffic. Second, many applications has been encrypted for security of personal information and bypassing. Searching the common string is a difficult work on encrypted traffic. Third, the extracted signature is different by proficiency of network specialist. A lot of experience and skills are required to extract the correct signature. Therefore, it is a difficult work to extract the objective signature.

The study on the automatic signature generation is going to solve the following problems. The proposed system to input the traffic on the particular application and the currently existing signature add the automatic generated signature and delete the dispose of target signatures. The dispose of target signatures are the signatures that is not used for analysis. This signature is that was used in the past traffic analysis but that was not used in the current traffic due to the deformation of traffic.

The paper is organized as follows. Section 2 describes related work. In Section 3, we propose the signature management system. Section 4 refer to the experimental result of using the system. Section 5 concludes the paper with suggestions for future research.

II. Related Work

The signature for traffic analysis exist in a wide range that based on the unique characteristics of the traffic. Types of signature are based on port number, based on statistic information and based on payload. The signature based on port number analyzes the traffic by using the port number in the header information of the traffic. The signature based on statistic information analyzes the traffic by using the size, the location and the time of the traffic. The signature based on payload analyzes the traffic by using the payload of the traffic

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2015R1D1A3A01018057) and by Brain Korea 21 Plus (BK21+).

The signature based on port number uses the port information specified by the IANA (Internet Assigned Number Authority) [15][16]. Port-based signature is possible that rapidly analyzes the traffic by using the less memory. But many applications set up the random port number in order to pass through the firewall and IPS devices. Also, port-based signature is meaningless because the user establish the port number or the application come settings random port number at each execution.

The signature based on statistic information analyzes traffic by using the size, the location, and the time of the traffic [17]. But Statistics-based signatures are not only difficult to generate signatures but also to gives low accuracy in real-world analysis.

The signature based on payload is method for analyzing the traffic by using payload in the traffic. Payload-based signature is configured with a set of substrings. This method identify the traffic by matching the substring and the payload. Payload-based signatures have been the most widely studied because its high rate and accuracy in traffic analysis.

However, payload-based signature is very difficult to extract. The network specialist extracts the common string in payload in order to extract the payload signature to an existing method. This method is different the extracted signature according to proficiency of network specialist. In order words, the objectiveness of the signature is reduced. Also, this method is a lot of work that needs to invest the time.

Therefore, the study of automatic signature generation is continuous. The method for automatic signature generation extract the common string based on payload of packets. Currently, the method for automatic signature generation is being studied in a variety of ways, such as LCS algorithm and Smith-Waterman algorithm. In this paper, the proposed system uses a sequence pattern algorithm used in data mining [10, 11].

LASER (LCS-based Application Signature ExtRaction) was transformed LCS (Longest Common String) algorithm into purposes application traffic signature extraction [6]. The method for signature extraction using LASER find the common string using backtracking from matrix that compares the two strings. Therefore, this method has a high computational complexity and time complexity.

Smith-Waterman algorithm is also frequently used in the automatic signature extraction [5, 8]. Smith-Waterman algorithm is presented for the purpose of determining the degree of similarity of the DNA. The method for signature extraction using Smith-Waterman algorithm find the optimum common string using backtracking from matrix that compares the two strings. When compared to the LASER, the largest difference is method for backtracking. Smith-Waterman algorithm has a low time complexity and computational complexity than the LASER. But, Smith-Waterman algorithm needs a considerable amount of time and calculations.

Autosig method for automatic signature extraction extract the common string possibly signature [7]. And signature is generated by structuring the extracted string. When a common string is extracted as possible signature, many string are calculated. For example, if the 4-lengh string is extracted from 20-length string, then 16 strings are extracted and are calculated. Therefore, Autosig have a lot of memory usage and processing time because the number of many substring and wide range.

The existing methods have a number of limitations for applying to the real-world traffic, because the two string match. The existing methods need to pre-processing that determined the sequence of traffic and to post-processing that integrated the resulting string as a rule. Also, this methods decrease reliability that this environment may be changed to generated a signature that is generated is different.

III. Signature Management System

In this paper, the proposed system is signature management system to cope with the modified traffic for specific application. This system is composed of three steps. First step is the traffic capture. When collecting the traffic from a particular application, if there is a lot of noise not only the traffic of a particular application, correct signature is difficult to be extracted. Therefore, this system is needed to process for collection the traffic of pure application. The second step is traffic analysis and signature generation. This step analyzes the traffic of pure application using existing signatures. The step outputs the analysis result and some of them are the unresolved traffic. The analysis result are used to remove unused signature. Also, new signature is generated to using unresolved traffic. The third step is signature update. Signature of the dispose target is removed from original signature and new signature is added.

A. Collecting to Traffic and Generating to pure Traffic

In this paper, the first stop of the proposed system is the traffic capture. This step is to leave only the traffic of a particular application from the all traffic collected. Traffic capture is very import in network traffic analysis. If there are many noise in the collected traffic, signature and analysis result are not correct. Figure1 illustrates the processing for traffic capture. First step in traffic capture is to collect packets. The user collects all traffic of the managed application time to generate the signature. But the system collects not only the managed traffic but also another traffic. If so, as well as the signature of target application is not extracted, the signature accuracy is decreased.

The collected traffic type of packet is converted to flow in the basic analysis unit. The flow is a set of packet with the same Source IP, Source Port, L4 Protocol, Destination IP, and Destination Port. Also, the flow is defined a set of packets in the reverse direction (Source IP = Destination IP and Source Port = Destination Port). Therefore, this study analysis result is to measure the completeness in the perspective of flow. The flow-level means that if one of the packets in the flow to have the string to determine the application, to determine all of packet in the flow with the application. Preprocessing of traffic is remove to abnormal flow from the captured traffic flow. In this paper, abnormal flow is defined as follows:

1. The traffic flow that is transmitted local network to local network.

2. The traffic flow that is transmitted remote network to remote network.

3. The traffic flow that is not TCP or UDP protocol.

After the abnormal traffic is removed, all traffic are normal using belonging to TCP or UDP protocol. Finally, we use the TMA to analyze the only particular application traffic [14]. TMA provide socket information that uses process in running from host to target server. TMA provide to Process name, IP address (local, remote), Port number (local, remote), State (start, continue, end, server), Protocol, and Path. This information is collected from End-host. And this information compare to collected normal traffic. Therefore, it is possible to collect the actual process information that application use to IP-Port. After TMA be executed for collect to traffic, the user collects traffic while running the management target application. Comparing the collected result from TMA with the traffic result that the user are collected, it is removed that traffic is not the management target traffic.

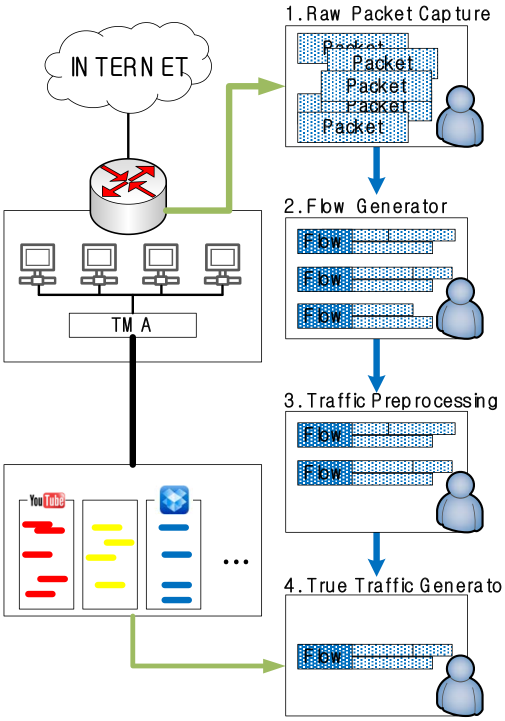

Figure 1. Process of collecting to traffic and generating to pure traffic. (Step 1-4)

B. *Automatic Signature Generation*

In this paper, the second step of the proposed system is the automatic signature generation. In this step, the identifier analyzes pure traffic using to original signature as in Figure 2. Unresolved traffic carried out preprocessing of signature generation. The sequence which are the basis of the signature extraction is generated from traffic data to perform step of pre-processing. The sequence is used in automatic signature generation.

Identifier analyzes pure traffic using original signature in order to determine whether to use the original signature. This step outputs two results. One result is the completeness of each signature. The completeness of each signature is needed to signature management step. Another result is the unresolved traffic. If completeness is not exceed threshold value that defined in the system, signature is extracted using the unresolved traffic. Before extraction the signature, the step of signature extracting should be removed meaningless part from payload. Therefore, this system separates payload by fields. The field is a section that be able to separate payload of each packet. For example, HTTP protocol is able to separate URL, HOST, USER, DATA, and etc. This process removes the part cannot be used for signature, decreases system overhead, and increases extracted signature accuracy.

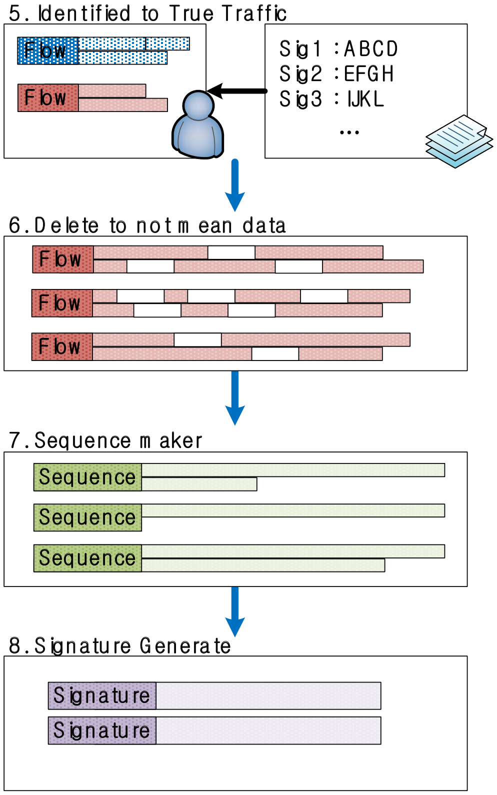

Figure 2. Process of separating to unused traffic and extracting signature. (Step 5-8)

After removing meaningless part through the separation for each fields, the sequence is generated. The two sequence are generated from single bi-direction (forward, backward) flow. The string of sequence is constituted payload for each packet. Also, the sequence is constituted *host_id* or *file_id* because it is calculated support value in the step of signature generation.

Signatures are extracted from generated sequence. The system uses the modified sequence pattern algorithm which is one of the association rule algorithm. The sequence pattern algorithm is a method that draw association between each data based on the occurrence frequency for the data. In this study, this algorithm be used for the search common string based on the association character or string and the frequency.

In this paper, the sequence pattern algorithm is used to signature extraction. Figure 3 represents an example of a signature extraction. The sequences are generated from the flow, which is composed of packet. And signature will be removed if does not satisfy the minimum support value are extracted from length 1. Signature is extracted through a processing, such as.

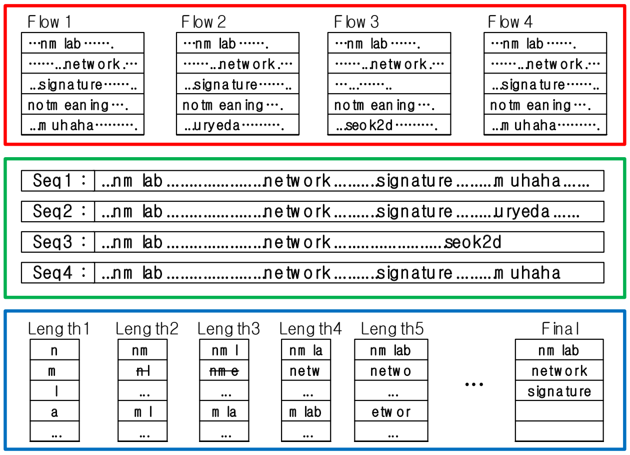

Figure 3. Processing of making sequence from flow data and extracting signature from sequence.

This algorithm is the method for outputting a set of signature that satisfied minimum support value from the set of sequence. If algorithm of signature extraction is performed, signature (length1) that is extracted from set of sequence are saved set of signatureL_1. As the length increase by one the signature of any length are extracted. The extracted signature are saved set of signatureL_k.

However, the signature that is not satisfied minimum support value is deleted from the newly generated set of signature. The signature that is not satisfied is not only failure to satisfy the signature extraction qualification but also failure to satisfy the extended signature support value.

Processing repeats extraction of signature and deletion of signature under of minimum support value repeats while increasing the length by one until no longer new signature extraction. The final step identify the containment of the extracted signature. And the included signature is deleted. Finally, the set of signature is provided the next step.

```
Input : SequenceSet = { $S_1, S_2, S_3, \ldots S_s$ } , minsupp
Output : SignatureSet = { $Sig_1, Sig_2, Sig_3, \ldots, Sig_g$ }
signatureExtractor(SequenceSet, minsupp)
    foreach sequence S in the SequenceSet do
        foreach character a in the sequence S do
            $L_1 = L_1 \cup a$ ;
        End
    End
    k=2;
    while $L_{k-1} = \emptyset$ do
        foreach signature sig in the $L_{k-1}$ do // delete under minsupp
            for i=1 to s do
                if ($S_i$ include sig) then
                    count = count + 1;
                end
            end
            if ( (count/s) < minsupp) then
                $L_{k-1} = L_{k-1} - Sig$;
            end
        end
        $L_k$ = candidate_gen ($L_{k-1}$) // extract length-k content
        k++;
    end
    SignatureSet = $\forall L_k$
    deleteSubset(SignatureSet)
    return SignatureSet
```

Algorithm 1. Signature extracting algorithm.

C. Signature Management

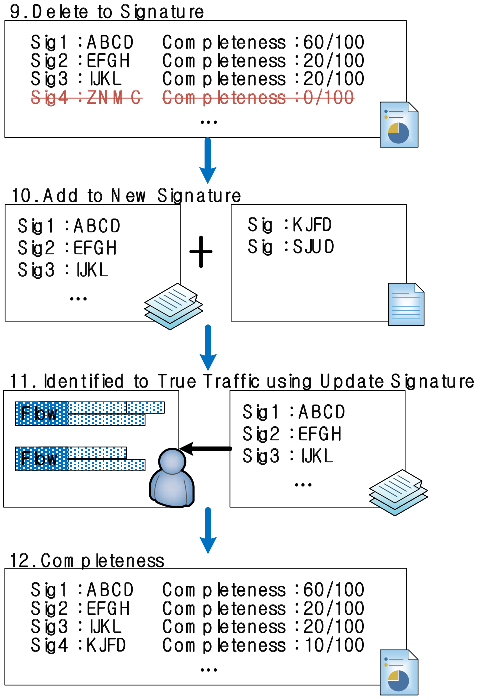
9.Delete to Signature
Sig1 : ABCD Completeness : 60/100
Sig2 : EFGH Completeness : 20/100
Sig3 : IJKL Completeness : 20/100
~~Sig4 : ZNMC Completeness : 0/100~~
…
10.Add to New Signature
Sig1 : ABCD
Sig2 : EFGH
Sig3 : IJKL
…
+
Sig : KJFD
Sig : SJUD
11. Identified to True Traffic using Update Signature
Flow
Flow
Sig1 : ABCD
Sig2 : EFGH
Sig3 : IJKL
…
12.Completeness
Sig1 : ABCD Completeness : 60/100
Sig2 : EFGH Completeness : 20/100
Sig3 : IJKL Completeness : 20/100
Sig4 : KJFD Completeness : 10/100
…

Figure 4. Processing of removing to unused signature and verification to new signature. (Step 9-12)

In this paper, the final step of proposed system is signature management. This step carry out four processing as follows in Figure4. It is removing the signatures that are not used in the analysis, and adding to the signature that is extracted. By again analyzing traffic to the updated signature, this system verify the updated signature.

When the fifth course of the traffic to the existing signature analysis of the present system, we record the analysis rate for each signature. After check the analysis rate for each signature, we remove the signature that is not used for analysis. The next step is to add extracted the new signatures in step 2 of the present system to list of the target application signature. The system analyze the traffic of the target application using to updated signature list.

IV. Experience and Result

In this section, shows the traffic analysis result using the proposed signature management system in chapter 3. First, shows the amount of Flow, Packet, and Byte of experimental data. Also, shows the completeness and the number of signatures of before and after performing signature management system. The table 1 shows the amount of Flow, Packet, and Byte of experimental data. The amount of each application traffic created by the pre-processing and removing the abnormal traffic is the amount of pure application traffic.

Table1. The amount of Flow, Packet, and Byte for application traffic

Application	Flow	Packet (10^3)	Byte (MB)
Nateon	741	20	10.6
Youtube	470	230	208
Facebook	170	10	6.6
Torrent	2,613	269	235.5
Ebay	1,414	108	84.8
Yahoo	936	47	31.2
Skype	769	18	5.6

The amount of traffic of each application and service have difference such as table 1. The torrent of typical application for file sharing is the most application traffic occurred during 7 applications. The Youtube of typical application for streaming service occur a large amount of traffic. Also, the Youtube has heavy flow format because is not P2P services such as torrent. 7 applications and services are representative of each field. Each field is messenger, streaming, file sharing, social network service, shopping, portal and video chatting.

The table 2 shows the number of exist signatures of each applications. Also, after implement signature management system, show the number of signature for each applications and the number of removed signature from existing signatures. And shows the number of new signatures. The number of signature of non-encrypted application from original signature is much more than can be seen that the number of signature of encrypted application. This is reason difficult that the extraction of signature of the encrypted application. Therefore, the application of encrypted low analysis rate.

The number of original signatures and analysis of rate are sufficient because the continuously updated signature as if NateOn and Torrent. But the other applications are hard continuous signature update. Therefore, the other applications are the low not only the number of signature but also the analysis of rate. In particular, the original signatures of Youtube did not analysis even anyone flow because of the recent traffic was encrypted. However, after doing the signature management system, 32 signatures were added, and all the original signatures were deleted.

The number of Facebook signatures were not much because representatives of the encrypted application is based very difficult to extract the signature. However, this system was able to extract the 11 signatures by automatically extracting signatures. Also, the other applications are able to analyze the traffic but, the amount is not much. Therefore, this system has been added the more signature.

Table2. The number of Signatures for each applications

Application	#original signature	#update signature	#delete signature	#new signature
Nateon	45	253	12	220
Youtube	14	9	14	9
Facebook	1	11	1	11
Torrent	31	287	0	256
Ebay	4	961	0	987
Yahoo	2	97	0	95
Skype	1	143	0	142

Table3. The analysis rate of differences existing signature and updating signature.

Application	Completeness (%) (origin signature)			Completeness (%) (update signature)		
	flow	pkt	byte	flow	pkt	byte
Nateon	70.04	87.11	90.49	100	100	100
Youtube	0	0	0	79.36	83.38	82.18
Facebook	0	0	0	62.94	60.54	55.18
Torrent	91.95	66.59	65.98	93.30	93.53	93.56
Ebay	9.34	3.80	2.63	96.61	96.97	97.92
Yahoo	4.57	0.81	0.33	87.61	92.17	91.20
Skype	1.42	0.53	0.21	50.33	58.46	79.85

The table 3 shows the completeness and the number of signatures of before and after performing signature management system. Although the rate of increase in the difference analysis of each application, the analysis rate increased in all applications.

In Table 3, the rate of analysis for all applications has been improved. Also, Youtube and Facebook that could not analyze the original signature were analyzed possible.

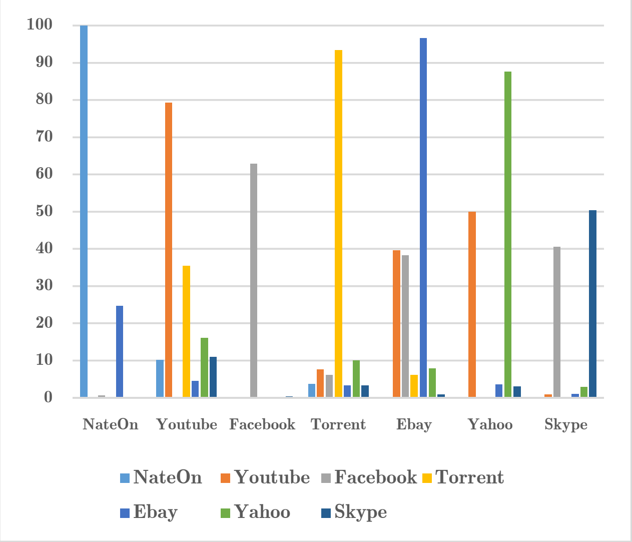

Figure 5. The result of analysis to the other applications using extracted signatures.

Finally, the graph 1 shows the traffic analysis rate of another applications for verification to accuracy of new signature. The signatures used this experiment are new signatures except for the existing signatures.

When analyzing the traffic of each application to another application signatures, the graph shows the result of the analysis. Therefore, analyzing only the signature of application is correct signature. But the traffic of Skype much analyzed by the signatures of Facebook. Because the Skype uses the features of Facebook.

V. Conclusion

This paper proposed the signature management system for solving the problem cannot be detected by traditional signature. This system is a system for detecting application traffic changes over time. The network administrator may be able to eliminate the difficulties of network management by allowing the automatic signature update because the constant signature updates is not possible. The system was demonstrated by comparing the analysis ratio of the existing signature and the updated signature. In addition, the newly generated signature is demonstrated by analyzing the accuracy with other traffic.

Future, we will not enter the specific application traffic, we are expected to develop a system that can be applied in real time in real network. The system also can extract network attack and new application by updating the signature in real-time by applying a real network. Also, the signature should be extracted with high accuracy. And research is required to reduce the number of signatures.

References

[1]Y. Wang, Y. Xiang, W. L. Zhou, and S. Z. Yu, "Generating regular expression signatures for network traffic classification in trusted network management," Journal of Network and Computer Applications, vol. 35, pp. 992-1000, May 2012.

[2]B. Park, Y. Won, M. S. Kim, and J. W. K. Hong, "Fine-grained traffic classification based on functional separation," International Journal of Network Management, vol. 23, pp. 350-381, Sep 2013.

[3]snort. Available: https://www.snort.org/

[4]H.-A. Kim and B. Karp, "Autograph: Toward Automated, Distributed Worm Signature Detection," in USENIX security symposium, 2004.

[5]J. Newsome, B. Karp, and D. Song, "Polygraph: Automatically generating signatures for polymorphic worms," in Security and Privacy, 2005 IEEE Symposium on, 2005, pp. 226-241.

[6]B.-C. Park, Y. J. Won, M.-S. Kim, and J. W. Hong, "Towards automated application signature generation for traffic identification," in Network Operations and Management Symposium, 2008. NOMS 2008. IEEE, 2008, pp. 160-167.

[7]M. Ye, K. Xu, J. Wu, and H. Po, "Autosig-automatically generating signatures for applications," in Computer and Information Technology, 2009. CIT'09. Ninth IEEE International Conference on, 2009, pp. 104-109.

[8]X. Feng, X. Huang, X. Tian, and Y. Ma, "Automatic traffic signature extraction based on Smith-waterman algorithm for traffic classification," in Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on, 2010, pp. 154-158.

[9]C. MU, X.-h. HUANG, X. TIAN, Y. MA, and J.-l. Qi, "Automatic traffic signature extraction based on fixed bit offset algorithm for traffic classification," The Journal of China Universities of Posts and Telecommunications, vol. 18, pp. 79-85, 2011.

[10]R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. 20th int. conf. very large data bases, VLDB, 1994, pp. 487-499.

[11]R. Agrawal and R. Srikant, "Mining sequential patterns," in Data Engineering, 1995. Proceedings of the Eleventh International Conference on, 1995, pp. 3-14.

[12]C.S Park, J.S Park and M.S Kim “Automatic Payload Signature Generation System” The Korean Institute of Communications and Information Sciences, Vol.38B No.08, pp.615-622, Aug, 2013

[13]S.H Yoon, J.S Park, H.M An and M.S Kim “Traffic Behavior Signature Extraction using Sequence Pattern Algorithm” Conference proceeding paper, in Proc. KICS Int. Conf. Commun. 2014 (KICS ICC 2014), pp.996-997, Jeju Island, Korea, June 2014

[14]S.H Yoon, H.G Roh, M.S Kim “Internet Application Traffic Classification using Traffic Measurement Agent” Conference proceeding paper, in Proc. KICS Int. Conf. Commun. 2008 (KICS ICC 2008), pp.618, Jeju Island, Korea, Jul 2008.

[15] IANA, IANA port number list, Retrieved 5, 24, 2013, from http://www.iana.org/assignments/service-names-port-numbers/servicenames-port-numbers.xml.

[16] J. Zhang and A. Moore, “Traffic trace artifacts due to monitoring via port mirroring,” in Proc. End-to-End Monitoring Techniques and Services (E2EMON), pp. 1-8, Munich, Germany, May 2007.

[17]Hyun-Min An, Jae-Hyun Ham, Myung-Sup Kim, "Application Traffic Classification using Statistic Signature," Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2013, Hiroshima, Japan, Sep. 25-27, 2013.