Sky-Scope : Skype Application Traffic Identification System

Sung-Ho Lee, Young-Hoon Goo, Jee-Tae Park, Se-Hyun Ji and Myung-Sup Kim
Dept. of Computer and Information Science
Korea University
Korea
{gaek5, gyh0808, pjj5846, sxzer, tmskim}@korea.ac.kr

***Abstract*— Today, as the network environment increases, various types of traffic patterns generated for each application and service are generated, and traffic analysis methods that can classify traffic applications and services are being studied. In particular, Skype is a VoIP service that is serviced by Microsoft and is currently the most widely used internationally. For this reason, the importance of Skype traffic detection is growing in terms of network management. In order to overcome the limitations of signature and machine learning based detection methods and to more accurately analyze and detect the current Skype traffic pattern, this paper presents a comprehensive Skype traffic detection system that combines pattern, list and signature based application detection methods. The proposed system is applied to various Skype traffic collected through campus network to verify accuracy and detection rate.**

Keyword*— *Skype, P2P, Traffic Classification, Network Management, Analysis

I. INTRODUCTION

Recently, the increase of Internet users and the spread of high-speed network have caused a surge in network traffic. This is largely due to not only traditional Internet services such as WWW, FTP, SMTP and DNS but also the increase in multimedia services. As the traffic increases rapidly, traffic monitoring and analysis becomes more important for effective network management.

The purpose of this paper is to propose a system that can accurately detect the application traffic by analyzing traffic characteristics of Skype application without relying on existing methodology such as payload signature based analysis and machine learning based analysis. Skype traffic is basically encrypted and dynamic port numbers are assigned when installing the application, and it is not possible with a general analysis methodology to accurately detect Skype traffic without using a common protocol. However, if we can identify the dynamic client ports for each Skype application host, we can easily classify most traffic and use it in various ways such as measuring and controlling application usage of each host.

In this paper, We developed a detection system called *Sky-Scope* that detects Skype traffic based on list, pattern, and signature application detection methods, we were able to accurately detect Skype application traffic.

II. RELATED WORK

Although many existing studies have proposed DPI-based, machine learning-based and packet Header-based schemes for application traffic classification, classification accuracy is not high enough for Skype applications or left as future research.[2,3,4,5]

[2] And [3] proposed a payload-based classification method for traffic analysis, but it is difficult to classify Skype applications because the data is encrypted. Payload-based classification methods are difficult to classify if the data part is encrypted. However, when using the automatic signature generation system, high quality application signatures can be generated even in the encrypted data, and these signatures can become important clues in traffic classification.

[4] And [5] propose a method of classifying P2P application programs using machine learning techniques. However, the proposed methodology requires a separate set of data for learning and shows that the classification accuracy is not high enough for Skype applications. The system proposed in this paper does not require a separate data set for learning and has the advantage of accurately classifying and extracting dynamic port numbers at the same time.

[6] Proposes a classification method using packet size distribution. However, it applies to Skype version 3.0, and it has the disadvantage that it is difficult to extract each port of dynamic Skype host. In this paper, we propose a classification method using some packet sizes, but it is applicable to the latest version of Skype. By extracting each port of Skype host, it is possible to measure various traffic of

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 Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government(MSIT) (No. 2017-0-00513)

Skype application in the enterprise network. In the proposed system, {IP, port} of each Skype user is extracted for the traffic classification, and at the same time classification is performed accurately.

[7] proposes a detection methodology using UDP sequence signatures and IP bands that can be detected in the login session of Skype application. However, the UDP sequence signature is only applicable to the old version of the Skype application, and the detection method based on the IP bandwidth is also low in accuracy. The proposed system also uses the IP band based detection method, but it is applied together with the other three detection methods.

III. Proposed Skype Traffic Detection System

In this paper, we classify Skype traffic into four types according to port and protocol used as shown in Fig. 1 for accurate detection of Skype application traffic.

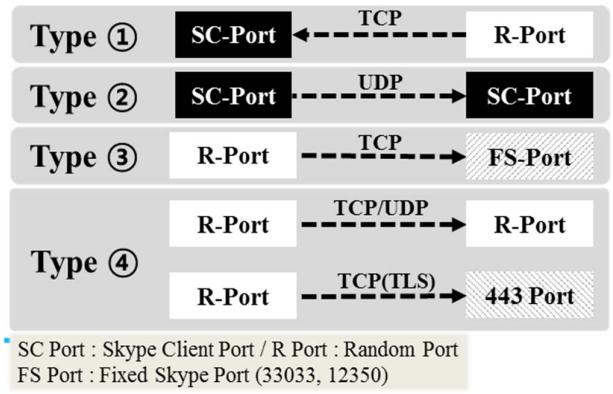

Figure 1. Four types of Skype Traffic

Type 1, 2, and 3 flow is a type of flow that must be detected because it communicates over the Skype fixed port and SC-Port(Skype-Client Port), which is the dynamic port of the Skype application client.

Type 4 flows are flows that occur mostly in the process of using Skype service. In particular, because most of the data is generated through arbitrary port (R-Port / R-Port) UDP connection, Type 4 flow must also be detected for service session detection.

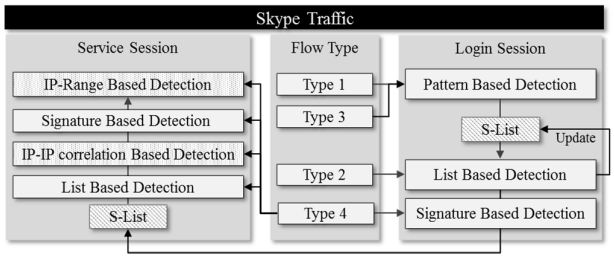

Figure 2. Skype Traffic Detection Method

On the other hand, Type 4 TLS flow is a common flow that occurs in both login and service sessions of Skype application. It is used mostly in other applications as well as Skype application. Therefore, in order to detect accurate Skype application traffic, TLS flow It should be able to detect accurately.

The overall Skype application traffic detection method follows the process of detecting the login session first and then performing the service session detection process as shown in Fig. 2.

III-1. Skype Login Session Detection

As shown in Figure 2, the first step in detecting Skype application traffic is Skype login session detection. The reason why the login session is important is that the login session analysis can find the SC-Port of the Skype-enabled host. It can detect TCP flow (Type 1) and all UDP flow (Type 2) of login traffic occurring in Skype application through SC-Port and build list (S-List) by identifying Skype-enabled host. We apply pattern-based detection method to detect such SC-Port.

In order to define patterns in Skype application login traffic, we collected about 20 sets of login traffic and defined patterns based on the packet size distribution and 5-Tuple information that always occurs in the set.

Figure 3 shows the main connections between the Skype Client and the Super Node during the login process of the Skype application. The SC-Port of the Skype-enabled host can be detected through the connection corresponding to 'SC-Port Detection' in the lower right of Fig. 3.

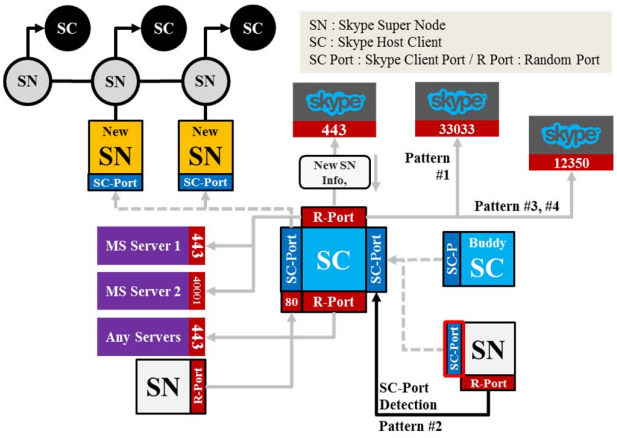

Figure 3. Connections of Skype Login Traffic

The {SC-IP, SC-Port} information of the detected Skype host is registered in the Skype list (S-List). The S-List is a list of IP and port information of Skype-enabled hosts and is used later in the list-based detection method. When the information of the host using Skype is registered in the initial S-List through the pattern-based detection method, the list and the signature-based detection process are sequentially applied.

III-2. Skype Service Session Detection

Performs service session detection after login session detection for Skype application traffic. Type 4 flows occur in the Skype application's service session.

When we analyze the service traffic of the Skype application, it is confirmed that 80% of the actual call data between the receiver and the caller is transmitted through the UDP session. Therefore, in order to detect the service session of the Skype application, it is most important to detect the UDP flow that transmits the actual data.

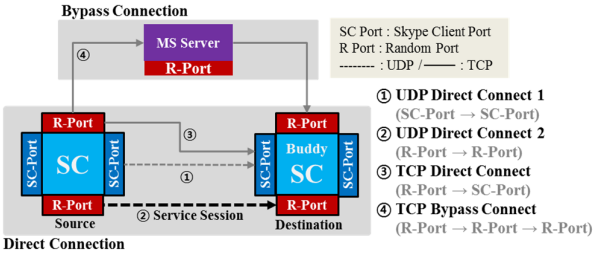

Figure 4. Connection of Skype Service Traffic

Fig. 4 shows the main flow of Skype application service traffic during voice / phone call. The flows that are generated are classified into two types: direct connection type is direct connected flow between the receiver and the sender, and bypass connection is flow bypassing the Skype server or Super Node.

The direct connection flow is divided into three types. There is a UDP connection using the SC-Port (1) and the R-Port (Random Port) (2). The TCP connection (3) between R-Port and SC-Port is irregular and is excluded from detection.

There is a TCP connection using R-Port and R-Port in the bypass connection flow. Among the four types of service flows, the flow that can be judged as the service session is (2) flow. When analyzing the service traffic of Skype application, the flow that transmits most of the UDP flows that transmit actual data is flow (2) in the direct connection flow. Therefore, it is possible to detect the service session of Skype application if UDP flow based on R-Port / R-Port (2) is detected.

However, the problem is that the UDP flow performs R-Port based communication. If SC-Port is used, it can be detected by list-based detection method, but because it uses arbitrary port, list-based detection method cannot detect service session.

In order to detect the service session, the IP-IP correlation method using the correlation between IPs and the signature-based detection method should be applied together.

The IP-IP correlation method is a method of detecting flows based on the 2-Tuple (Source IP, Destination IP) information of previously detected flows.

If the signature-based method is used together, the Skype service session flow can be accurately detected except for the type 2 error flows that may occur when only the IP-IP correlation detection method is applied.

The proposed Skype application detection system applies not only pattern, list, and signature based detection method but also additional correlation and IP band based detection method to improve accuracy. The ultimate reason for applying various detection methods is to detect more precisely and flexibly the various flow patterns that occur in Skype applications.

IV. EVALUATION

The Sky-Scope system proposed in this paper is an offline system for accurately detecting Skype application traffic. The system consists of 4 programs as shown in Fig. 5, and the traffic detection program consists of 3 modules.

The preprocessing process corresponds to programs 1 to 3. Upon completion of the preprocessing process, the Skype traffic detection program performs the detection of the Skype application traffic. The detection program consists of three modules.

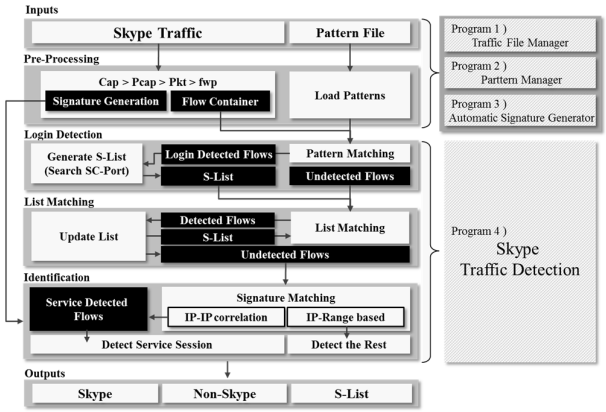

Figure 5. Sky-Scope System Structure

The Login Detection module uses the pattern-based detection method for detecting the host logged into the Skype application. The most important point in the Login Detection process is to detect the SC-Port Detection flow in Figure 2.

The SC-Port can be searched through the pattern matching of No. 2 in Table 1. Therefore, it is possible to find all Skype application login hosts that exist in the collected traffic through the Login Detection process. Based on the IP (SC-IP) and port (SC-Port) information of the detected login host, initial S-List is generated.

The List Matching module detects the Skype application traffic by receiving the initial S-List generated by the Login Detection module. It detects Skype application for all UDP flow (Type 2) communicating with Skype host registered in S-List, detects unregistered SC-IP and SC-Port information, and updates to S-List.

Identification module performs IP correlation and IP band based detection based on signature based detection.

In order to verify the validity of the proposed Skype application traffic detection system, we conducted a verification experiment on the Skype application. The experiment is performed by collecting four types of Skype traffic as shown in Fig. 6.

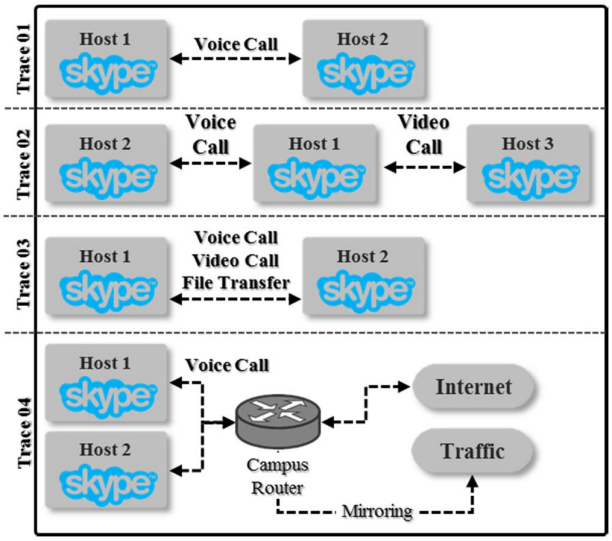

Figure 6. Captured Skype Traffic Types (Second Experiment)

Each traffic collected for experiment is TMA / S(Traffic Measure Agent/Server) validated Skype application ground truth traffic.

When the experiment traffic is applied to the proposed Sky-Scope system, it verifies whether SC-IP and SC-Port of the host are correctly detected and confirms Duration, Flow, Packet, Byte size of the login session and Buddy List information of the host. Also, confirm that the data flow detection and service functions are classified for the service session, and check the SC-IP and SC-Port information of the communication host.

The results of the experiment are shown in Table 1. The detection results of applying the traffic collected through the process of Fig. 6 to the Sky-Scope system showed an analysis rate of over 95% in all traffic traces.

Trace #	Identified Skype Traffic Completeness		
	Flow	Pkt	Byte
Trace 01	100% (243/243)	100% (1,854/1,854)	100% (941K/941K)
Trace 02	100% (390/390)	100% (2,659/2,659)	100% (1,351K/1,351K)
Trace 03	99.5% (670/673)	99.8% (4,798/4,807)	99.9% (2,741K/2,742K)
Trace 04	96% (1,382/1,440)	98.7% (11,728/11,882)	99.8% (8,429K/8,446K)

Table 1. Experiment Result

This undetected flow (FN) can be detected when adding an IP band or updating a signature. Or additional duration-based detection methods are also planned for now, and will be detectable on the Sky-Scope system in the future.

V. Conclusion and Future Work

In this paper, we describe the structure and detection algorithm of Sky-Scope system for Skype application traffic detection and verify the validity of the system through a simple experiment. Through the pattern, list, and signature based detection methods, the proposed system can analyze Skype application traffic more variously and more accurately compared with other application detection systems. In addition, the experimental results using the actual Skype application traffic also show high application detection rate. Unlike previous systems, we designed and verified a system that can detect the host through the login session of Skype application and detect the service session and type.

However, because there are Skype application flows that can not be detected at present, we plan to improve these problems through system improvement and stabilization in the future and apply additional detection methods. Also, since the current system operates based on the traffic collected in the off-line, the load on the system is relatively small. We plan to improve the performance of the system by developing it into a system that operates in the network in real time.

References

[1] S. W. Park, H. S. Lee, M. J. Choi and M. S. Kim,"Real-time Identification of Skype Application Traffic using Behavior Analysis", KICS, vol.36, No.2, Feb. 2011, pp.131-140.

[2] Risso, F., Baldi, M., Morandi, O., Baldini, A., Monclus, P.,"Lightweight, payload-based traffic classification: An experimental evaluation", In Proceedings of IEEE International Conference on Communications ICC, 2008, pp.5869-5875.

[3] N. Cascarano, L. Ciminiera, F. Risso,"Improving Cost and Accuracy of DPI Traffic Classifiers", 25th ACM Symposium On Applied Computing(SAC 2010), March. 2010, pp.643-648.

[4] H.Liu, W.Feng, Y.Huang, X.Li,"A peer-to-peer traffic identification method using machine learning", in International Conference on Networking, Architecture, and Storage, NAS, July. 29-31. 2007, pp.155-160.

[5] Zhu Li, Ruixi Yuan, and Xiaohong Guan,"Accurate Classification of the Internet Traffic Based on the SVM Method" Proc. of the IEEE International Conference on Communications, Jun. 24-28. 2006, pp.1373-1378.

[6] Ying-Dar Lin, Chun-Nan Lu, Yuan-Cheng Lai, Wei-Hao Peng, Po-Ching Lin,"Application classification using packet size distribution and port association", Journal of Network and Computer Applications, Vol.32, Sep. 2009, pp.1023-103.

[7] Z. Yuan, C. Du, X. Chen, D. Wang, Y. Xue,"SkyTracer: Towards fine-grained identification for Skype traffic via sequence signatures", Computing, Networking and Communications (ICNC) 2014 International Conference, Feb. 3-6. 2014.