Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

APPLICATION OF WILDPACKETS ON PROTOCOL CHARACTERISTICS ANALYSIS OF P2P

[1]MIN LI, [2]XIANHUA JIN, [3] PENG ZHANG

[1] Media Research Institute, Anyang Normal University, Anyang, 455000, Henan, China

[2] Media Research Institute, Anyang Normal University, Anyang, 455000, Henan, China

[3] China Academy of Telecommunication of MIIT, Beijing, 100000, Beijing,China

ABSTRACT

WildPackets is the leading provider of network, application performance, and protocol analysis, VoIP monitoring, and troubleshooting solutions. The key of P2P traffic management is constantly updated P2P protocol characteristics library of network traffic management device. In this paper, we analyze the determinants of P2P protocol characteristics, the characteristics change of P2P software and coping strategies, then we focus on discussing the specific ideas and methods of P2P protocols characteristics by WildPackets, and illustrate the reasons for lower efficiency of P2P traffic identification.

Keywords: *P2P; WildPackets, Payload, Protocol Characteristics*

1. INTRODUCTION

With the rapid growth of P2P applications, the illegal P2P services account for a large amount of network bandwidth, it has seriously affected the normal operation of network services. The network traffic management technology which supports seven numbered layers can achieve to limit the illegal P2P business, and protect the service quality of the normal application [1, 2]. The network traffic management device on the market which is for P2P traffic identification is mainly based on protocol characteristics and parameters of the payload header [3]. How to quickly build protocol characteristics library of popular P2P software is the key to monitor P2P applications. Compared to other network protocol analysis tools, WildPackets can quickly help us analyze protocol characteristics of P2P applications in-depth, in order to monitor the illegal P2P applications effectively.

2. P2P TRAFFIC CHARACTERISTICS

2.1 Protocol Characteristics Change of P2P Software and Coping Strategies

Protocol characteristics of P2P software are determined by the characteristics of P2P technology [4]. The behavior characteristics of centralized P2P technology software is to use a fixed TCP port, using the TCP connection is known to a central server, access to shared files on the other end of the directory information, and then connection thus identification of a centralized P2P traffic is mainly port, the central server domain name (or IP address) of the identification. Distributed P2P technology is no longer dependent on a central server, on the other side by all co-responsible for the management of distributed file directory, all the nodes in the network to become truly peers [5]. In order to successfully through network monitoring equipment, such P2P applications commonly used dynamic random port, camouflage port (such as the use of HTTP port 80), or take HTTP as a basis for communication protocols. However, weaknesses in the distributed P2P technology is obvious, the dynamic changes frequently of network nodes will lead to file-sharing requires a lot of query flooding overhead. The characteristics of the distributed P2P technology features determine their agreement must be stable; the establishment of signature recognition agreements cans long-term use. Hybrid P2P technology integrated the advantages of centralized and distributed P2P technology, compared to centralized, weakening the role of the central server; compared to a distributed, self-healing capacity of the topology of the node to be strengthened. However, hybrid P2P network survivability advantages of its own making P2P software can quickly update their agreement features of the field and protocol parameters, to avoid the blockade of network traffic management device. To identify the hybrid P2P application protocol, we must update the protocol characteristics library constantly.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.2 Characteristics Changes of Popular P2P Software

Taking the domestic popular P2P software for an example, Early emule using protocol features and protocol parameters of edonkey, and made some simple changes. Emule protocol followed the introduction of new features and protocol parameter field, retaining some of the characteristics edonkey fields and protocol parameters, so can continue to connect emule edonkey network. Now the fuzzy connection agreement are increased, that is, the characteristics of the protocol and protocol parameters in the Payload field, the position of the head appear randomly, which adds to its agreement features of the field identification difficult, but increased the appropriate strategy. The protocol characteristics of domestic QQ software update version number is mainly caused by the version number of features of the field changes, and other features of the field has not changed, the new version of the QQ file transfer for HTTP as the protocol basis. But QQ constantly adding new services, the identification of QQ must be embodied in a comprehensive classification and recognition, such as text chat, voice chat, video chat, file transfer, live, games, download cyclone classification, etc., that improve the accuracy of the identification. Domestic popular Thunder download using updated protocol characteristics constantly to avoid blocking the network traffic management devices.

3. THE APPLICATION OF WILDPACKETS ON PROTOCOL CHARACTERISTICS ANALYSIS OF P2P

3.1 Network Environment of P2P Communication

Based P2P communication features, and the following network environment: P2P overlay network and physical network mapping; the actual physical network connection (internal network and external network connections; within the network and the network connection; outside the network and external network connection); P2P congestion control algorithm; P2P network client restrictions (firewalls, and network speed limits and other restrictions); P2P encryption applications; P2P software version update changes; P2P software and other open source P2P protocols HTTP, FTP, ICMP and other basic network protocols.

These features are the basis for developing P2P software own protocols, such as QQ, video transport protocol congestion control algorithm which using three different fields and the corresponding protocol parameters based on network transmission protocol features. Emule data transfer protocol using the appropriate length of the data report based on speed of the network environment. Thunder not only uses its own P2P download protocol, supporting HTTP, FTP and other download protocols. Now, the latest version of Flashget is compatible with BT, HTTP, FTP, emule download.

3.2 Analysis Principles

Analysis of P2P protocol characteristics should be based on network environment; it is an integrated, multi-angle analysis. It is the data analysis process which for payload of application layer, data analysis process must be followed by the principle of coarse to fine, from the general to the specific, from the particular to the general promotion, and so forth.

Payload data analysis based on the hierarchical structure of the network, first, the classification of the data transport layer, and in-depth primary end to end connection, analysis characteristics in the Payload, the final sum for all end to end connection agreement in the characteristics of Payload. In the analysis of the end to end link, the first analysis is global characteristics as the entry point for further analysis; and the detailed characteristics of partial protocol characteristics, such as shaking hands, query, response and other interactive information and data transmission protocol information, including its own P2P protocol parameters and the Payload length tail parameters; Finally, the local features into global agreement considered, to get the correct protocol characteristics, and the trends.

3.3 Specific Ideas and Methods

It is divided into three parts: the establishment of experimental environment, the analysis of protocol characteristics, and the protocol authentication feature. The first is to establish a specific test environment, including restrictions on traffic between nodes, the node's communication parameters and functions of the software settings. And then filtering P2P traffic, according to the port, IP address, transport bearer protocol (OSI 3 layer to layer 7 application protocol, TCP / IP network transport layer to application layer protocol), payload characteristics of the head and tail. Finally, the agreement is written in the regular expression feature added to the appropriate backbone network traffic identification device to verify.

3.4 WildPackets Example

The following is the partial results of P2P software flow characteristics which using

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

WildPackets, expressed with regular expressions, see Table 1. Regular expression is exinda flow measuring devices require representation; recognition rate is in a separate flow of P2P software obtained under the experimental conditions. Generally believed that more than 85% recognition rate can be used as commercial applications, the following are some analysis of the experimental results.

Table 1

Type	Regular expression filtering	recognition rate
QQ	^.?\x02.+\x03$	89%
emule	^[\xc5\xd4\xe3-\xe5].?.?.?.?([\x01\x02\x05\x14\x15\x16\x18\x19\x1a\x1b\x1c\x1d\x1e\x1f\x21\x23\x32\x33\x34\x35\x36\x38\x39\x3a\x3b\x3c\x3d\x40\x41\x42\x43\x44\x46\x47\x48\x49\x4a\x4b\x4c\x4d\x4e\x4f\x50\x51\x52\x54\x55\x56\x57\x58\x59\x5b\x5c\x5d\x5e\x5f\x60\x61\x81\x82\x85\x86\x87\x90\x91\x92\x93\x94\x95\x96\x97\x98\x99\x9a\x9b\x9c\x9d\x9e\xa0\xa1\xa2\xa3\xa4\xfe]).*$	85%

The most obvious protocol characteristics of the QQ are the use of "\ x02"at the beginning and "\ x03" ends. Protocol characteristics of Emule are divided into "[\ xc5 \ xd4 \ xe3 \ xe4 \ xe5]" five types, of which "[\ xe3 \ xe4 \ xe5]" is used edonkey protocol eigenvalues, "[\ xc5 \ xd4] " is the expansion of emule eigenvalues; ".?.?.?.?" is length sizes of own protocol payload; handshake followed by the node, the file information, file transfer feature and other control information values, the values of these are in the early version of the software emule, mainly for domestic and international traffic blocked, because emule earlier version of the software almost copied edonkey

QQ not been blocked by network operators because the QQ software is a monopoly, thus QQ focus on increasing more new software features, protocols features almost no change, and therefore protocol characteristics relatively fixed, but the flow of QQ to maintain connection information and advertising information in a large number of nodes, text chat messages almost submerged flow, and thus the recognition rate is 89%, did not reach higher. However, the fate of emule is not the same, the network operators more preferred to block the software download, emule early upset with a simple encryption technology, that protocol obfuscation , and therefore can not filter fuzzy agreement, but now due to emule not limited by the block, protocol obfuscation occupy a secondary position, the recognition rate can reach 85%.

WildPackets as a troubleshooting tool, no special software protocol decoding for the popular P2P, or for a variety of encryption algorithms and decryption functions, and thus WildPackets becomes powerless in the analysis of P2P applications, that must be increased analysis capabilities.

4. CONCLUSION

In short, feature search is a complex task, according to network traffic changes in the environment, P2P software version updates, the rules out of the market, new emerging P2P software protocol features, large-scale application of the coming of encryption, P2P protocol feature recognition will increase the difficulty of the work. It is necessary to develop software for P2P protocol characteristics analysis. However, to relate to such a complex network of ideas and the actual environment and market environment changes, even if the use of data mining, artificial intelligence is just drop in the bucket. Thus the key is artificial recognition-based, locally written software to implement intelligent search and application for simple encryption application crack.

REFERENCES:

[1] KOIZUMI T, YOSHIDAY M, OHZAHATAZ S, et al. "An analysis of the P2P traffic characteristics on file trans-fers between prefectures and between Autonomous Systems in the Winny network", The16th Asia-Pacific Conference Communications (APCC 2010) , *IEEE computer Society*, 2010,pp.147 - 152.

[2] THANH V V, FUKUDA K, CHAN H N, et al. "A study on P2P traffic characteristic evaluation in the WIDE backbone", The 6th International Conference on Information Technology and Applications (ICITA 2009) , The IEEE NSW Section,2009,pp.286 - 291.

[3] OHZAHATA S, KAWASHIMA K, "A study on traffic characteristics evaluation for a pure P2P application"16th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, *IEEE Computer Society*, 2008,pp. 483 - 490.

[4] HAN Y T, PARK H S. "Distinctive traffic characteristics of pure and game P2P

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

applications", the 10th international conference on advanced communication technology, *Institute of electrical and electronics engineering*, 2008, pp. 405 - 408.

[5] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband Internet traffic", *Proc. of the 9th ACM SIGCOMM Conference on Internet Measurement*, 2009, pp. 90-102.