A Methodology for P2P Traffic Measurement Using Application Signature
Work-in-Progress

Liu Bin
Network and Computing Center
Huazhong University of Science and Technology, Wuhan, China
086-027-87541441,430074
bliu@mail.hust.edu.cn

Li ZhiTang
Network and Computing Center
Huazhong University of Science and Technology, Wuhan, China
086-027-87541441,430074
leeying@mail.hust.edu.cn

Tu Hao
Network and Computing Center
Huazhong University of Science and Technology, Wuhan, China
086-027-87541441,430074
tuhao@mail.hust.edu.cn

ABSTRACT
P2P application has become tremendously popular over the last few years, now accounting for a significant share of the total network traffic. P2P protocols used arbitrary ports to camouflage its traffic to avoid detection and recognition with standard measurement tools. This paper presents a measurement system which reliably detects and measures P2P traffic in real-time. The methodology is based on using application level signatures implemented by Netfilter extension as well as packet classifier. The results from test in a large university network indicate that this approach not only can significantly improve the P2P traffic volume estimates but also provided the performance required for practical applications.

Categories and Subject Descriptors
C.2.3[**Computer-Communication Network**]:Network operations - *Network monitoring.*

General Terms
Measurement, Algorithms .

Keywords
P2P, traffic measurement, Netfilter

1. INTRODUCTION
Peer-to-Peer (P2P) system has become extremely popular over the last years. The ability to accurately measure the network traffic associated with P2P applications is crucial for a broad range of network management tasks including traffic engineering, capacity planning, provisioning, service differentiation and cost reduction. The earlier P2P systems mostly used default network ports for communication and traditional techniques like port-based classification of applications can be used for P2P traffic measurement. However, a significant share of P2P traffic nowadays no longer uses the default TCP and UDP ports (port hopping) or is often camouflaged as http traffic. The basic concept of these techniques is to change the transport layer behavior of the P2P applications, such that a purely port based packet filter is not capable of detecting and restricting the related traffic. Some new approaches have been proposed applying signature detection at the application level based on IDS. However, the protocol decoding phase of IDS is I/O intensive and requires large amounts of CPU power and RAM memory. Therefore, these approaches are usually used to analyze and characterize the P2P traffic, and would not be suitable if it were to be used as part of real-time measurement. To overcome this drawback, a measurement system for P2P traffic has been designed and implemented in this paper. The system is comprised of two major components, the packet classier and application level signature matching module. The packet classier classifies TCP stream into P2P stream and NON-P2P Stream. Signature matching module based on Netfilter extensions implemented the accurate identification of P2P application in the presence of packet payload information. This approach has the advantage that it can measure P2P traffic accurately in real-time with standard hardware and can easily be extended to new signatures and other protocols.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference name: Infoscale 2007, June 6-8, 2007, Suzhou, China
Copyright number (LaTeX \crdata{}): 978-1-59593-757-5

2. METHODOLOGY
Since P2P used dynamic port to camouflage its traffic port, it is obvious that standard measurements based mainly on layer 3 and 4 information are not sufficient to measure P2P traffic accurately. This led to the idea of using application layer signature. The measurement infrastructure is comprised of two major components. The first component is a packet classifier which is used to classify TCP stream into P2P stream and NON-P2P stream. The second component is an application level matching module based on Netfilter extensions. The structure of the measurement is shown in Figure 1. The measurement system acts as a traffic sink for the traffic mirrored on the monitor port of a switch. For all incoming packets, the relevant layer3 and layer4 information is hashed and compared to a hash table containing information on already identified P2P connections. If an unknown hash value is found, the packet is fed into the pattern matching process . In case a P2P signature is found, a new mark is created and the packet is marked with it. All marked packets are counted and the data relevant for statistical analysis is collected.

Packet classifier is done with a hash function. The hash value for incoming packet is computed using source and destination IP addresses, the source and destination port. This value is then used

as an index into a hash table, where containing information on already identified P2P connections. While inserting an entry, it is searched in the hash table. If an unknown hash value is found, a new entry is created and added to the hash table, and packet is fed into the application layer signature matching process. If the new entry would overflow the hash table limit, the oldest entry within the hash table is dropped. The hash function has to be cheap to compute and distribute flows well over the hash table. We propose the following function to compute a 16 bit hash:

Let $S_1 . S_2$ $S_3 . S_4$ denote source IP address , and $D_1 . D_2 . D_3 . D_4$, destination IP address and $P_1 . P_2$, source port, and $K_1 . K_2$ be the destination port. S_i , D_i , P_j , K_j (i=1, 2, 3, 4; j=1,2;) is treat as 4 bitfield. Hash value represent as H is computed as follows.

S= $S_1 \oplus S_2 \oplus S_3 \oplus S_4$;

D= $D_1 \oplus D_2 \oplus D_3 \oplus D_4$;

P= $P_1 \oplus P_2$;

K= $K_1 \oplus K_2$;

H=(S&0x0f)<<12+(D&0x0f)<<8+(P&0xf0)+(K&0xf0)>>4 ;

IPP2P , a Netfilter extension, which compares the data payload of each packet with key values provided by a signature database is used to identify all kinds of P2P traffics by its signatures. IPP2P used not only string matches but also bit sequences and layer 4 connection information. The following is an example to detect eDonkey traffic.

iptables -t mangle -A PREROUTING -m ipp2p --edk -j MARK --set-mark 1

iptables -t mangle -A POSTROUTING -m mark --mark 1 -j ACCEPT

Packets of eDonkey get marks (01), then we add rules to the POSTROUTING chain and use the iptables byte and packet counters to sum up eDonkey usage.

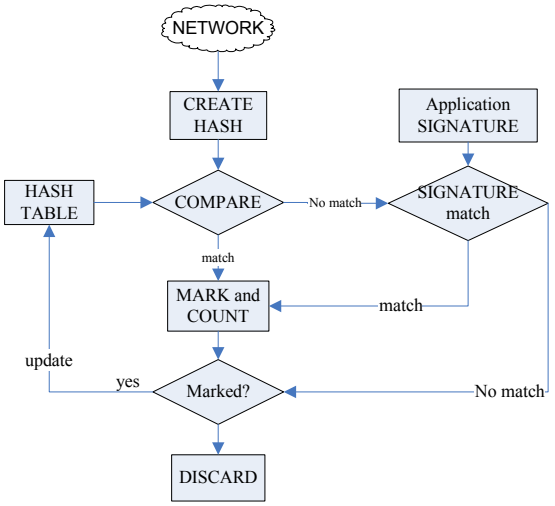

Figure 1. Structure of the measurement system

3. EXPERIMENTS

To validate and evaluate the feasibility of our P2P measurement system, we performed a test in a real life environment. Port-based and Signature-based methods have been compared with the same period of network traffic. Note that the choice of what goes in the list of default ports would impact the port-based method's performance. For this experiment, we selected the list of ports which is used by popular eDonkey peer software, including 4661-4667 used by emule. eDonkey protocols signatures used in our study is that the first byte after the IP+TCP header is the eDonkey marker(0xe3). Figure 2 showed eDonkey traffic every 5 minutes using port-based method and signature-based method on 11/21/06. The figure 2 showed that there was a big difference in results between two methods. Obviously, the port-based approach underestimated P2P traffic which indicated eDonkey traffic nowadays is transmitted over a large number of non-standard ports, making default port-based measurement less accurate.

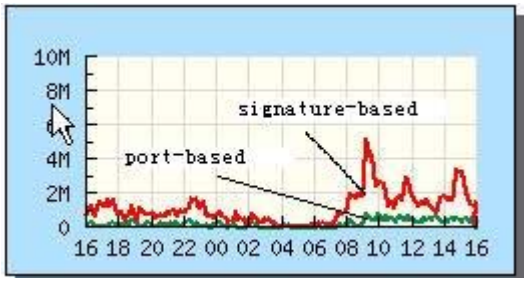

Figure 2. Measured eDonkey traffic

4. CONCLUSION

Traditional traffic to higher-level application mapping techniques such as default server TCP or UDP network-port based disambiguation is highly inaccurate for P2P applications. Some new identification methods using application layer signature based on IDS are proposed. However, the major drawback with these approaches is that it cannot yet do so in real-time. In this paper we have presented a flexible and efficient P2P measurement system using application signature analysis which has been implemented with standard hardware and netfilter extension. We demonstrated the feasibility of this approach in a real network environment and showed that the performance is sufficient to accurately measure high volume traffic on high speed links in real-time.

5. REFERENCES

[1] Sen. S, Spatscheck.O, Wang, D.M, Accurate, scalable in-network identification of P2P traffic using application signatures. Thirteenth International World Wide Web Conference Proceedings, Association for Computing Machinery, New York, May 17-22, 2004 p 512-521.

[2] Bleul.H, Rathgeb. E. P, A simple, efficient and flexible approach to measure multi-protocol peer-to-peer traffic.4th International Conference on Networking, Proceedings, Springer Verlag, Reunion Island, 2005, April p 606-616.