The design of the recognition and control system of the P to P flow based on Linux

Mingji Yang [a], He Wang [b] and Wanzhu Jiang [c]

School of Harbin University of Science and Technology, Harbin 150080, China

[a]yangmingji@126.com, [b]wh50161@sina.com, [c]jiangwanzhunl@126.com

Keywords: the P to P flow, flow recognition, flow control, DPI.

Abstract. This paper architectures the system that recognizes and controls the P to P flow on the platform that is based on the existing network software platform. It introduces a two-layer P to P flow recognition means, and it also designs and realizes the basic control of the P to P flow on the Net filter structure of Linux. The P to P flow will be recognized when the flow passes through the system, and the system that focuses on the P to P flow will limit the flow and govern the bandwidth according to the strategy.

Introduction

The development of the P to P application opens a new chapter for the download mode of the Internet, it makes the traditional interaction mode between client and single server transmit to the mode that any client can become the server. This mode fastens the download speed of the users and releases the visiting pressure of the server. However, it may bring the problem of consuming the bandwidth and blocking up the network flow. How to recognize and control the P to P flow has already been a new topic of the research. This paper takes advantage of the existing network software platform, and designs a recognition and control system of the P to P flow. By analyzing the mainstream recognition technology of the P to P flow, it reduces the resource waste of the system and especially designs the elementary flow recognition mode and the deep flow inspection mode. This system is based on the Netfilter firewall of Linux, and uses the data packet of the netfilter to obtain and transmit mechanics to realize the purpose of controlling the flow.

The common P to P flow recognition technology

At present, the common and mature flow recognition technology concludes the port recognition technology, the deep packet inspection technology and the deep flow inspection technology. We find that the P to P data flow and the common data flow have different forms by doing a large number of experiments and detecting the P to P flow. And the result concludes the information as follow. Firstly, the uplink bandwidth and downlink bandwidth of the P to P data flow are equal. Secondly, the source IP address is equal to the destination IP address.

The port recognition means. The flow recognition means based on the port is the common means, which only takes advantage of the information of the head of the packet and uses different ports to recognize based on different protocols. Most of the currently existing network flow capturing and protocol analyzing software, such as Ethereal and Wireshark, are all based on the default ports to recognize. For example, the number 80 and the number 443 ports that the TCP flow uses should be signed as the Web flow according to the government rules of the port of IANA. The table below lists the common P to P protocol and corresponding ports.

The DPI (Deep Packet Inspection) technology. DPI is the deep packet detection technology. It adds the analysis of the payload of the application layer except the analysis of IP in the information of the packet head of the IP data, the port and the protocol type. By analyzing the load content of the application layer, we extract the feature information that is related to the protocol. It can recognize which application it is even the behavior of the user by using these information [1]. This paper uses the recognition technology which is based on the tagged word among the DPI recognition

technologies. This technology is mainly to analyze the unique tagged word information that different protocols have. These features conclude the key words, the packet length, IP, the port, the position of the key words and the unique information of the protocol in the other interaction data flow. We can make sure what the protocol is by using these tagged words.

Table 1 the common P to P protocol and corresponding ports

Protocol	Default ports	Transmission Protocol
LimeWire	6346 / 6347	TCP/UDP
Bearshare	6346	TCP/UDP
eDonkey	4662	TCP
EMule	4662	TCP
	4672	UDP
Bittorrent	6881-6889	TCP/UDP
WinMx	6699	TCP
	6257	UDP
Kazza	1214	TCP/UDP
DirectConnect	1412	TCP

The DFI (Deep Flow Inspection) technology. The DFI is the deep flow inspection technology, and this technology can be seen as a part of the DPI technology that is the behavior mode recognition technology. The DFI technology does statistics on the flow behavior of the communication based on the time axis, which mainly concludes the statistical module as follow. They are the packet length sequence, the assemblage of the packet length, the range of the packet length, the average of the packet length, the sum of the simple packet length, the sum of the round packet length, the repetition of the packet length, the statistics of number of the packet and the receiving and transmitting ratio of the message.

Based on the DPI technology and the DFI technology, the table number two makes the comparison between these two technologies.

Table 2 the comparison between the DPI technology and the DFI technology

	DPI	DFI
Inspection content	The tagged word, the port and the packet length of the application layer	The packet length, the direction of the packet, the ratio of the packet and the number of the packet
Speed	Slow	Fast
Accuracy	High	Low
Real-time	Strong	Weak
Encryption	Cannot be recognized	Can be recognized

The correctness of the flow recognition means based on the port is decreasing. The reasons of this situation are as follow. Firstly, many P to P applications do not use static port number, but use the dynamic port number. Secondly, these applications use common service ports to do the camouflage of the protocol. Some P to P applications use number 80 port to disguise the P to P flow as the HTTP flow to avoid the limit of the firewall and obtain the visit to the hitemct. Thirdly, the data packet on the network layer will be encrypted under some conditions. The protocol TCP will be blurred by the head of the DPI, and we cannot know the true port. If we want to recognize the P to P flow precisely in this system, we need to do the elementary recognition of the port and the deep recognition that combine the DPI technology and the DFI technology.

The design and realization of the flow recognition system

According to the analysis of the mainstream flow recognition technology, this paper uses the two-layer data flow recognition mode and the elementary flow recognition module. It mainly uses the port recognition technology and the P to P flow character, and separates the P to P flow and the non P to P flow. Then the system goes into the next recognition module that is the deep data packet

recognition module. This module mainly uses the DPI technology and the DFI technology to deeply inspect the data packet.

The picture below is the block diagram of the flow recognition module,

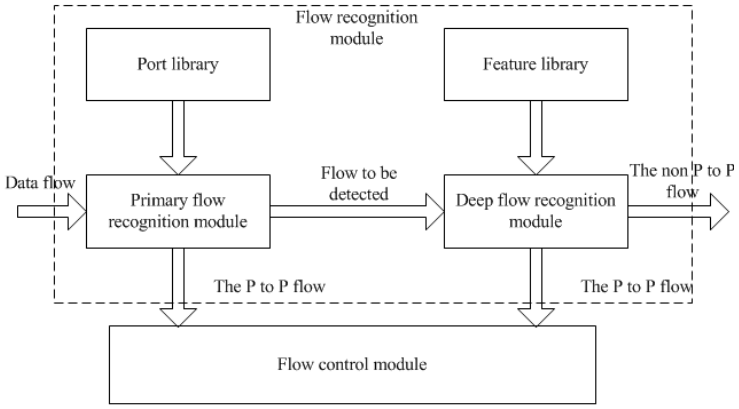

Fig. 1 the flow recognition module

The process of the recognition of flow,

Firstly, the data flow goes into the elementary recognition module and inspects the port of the elementary recognition module that recognizes data flow. Then we match the port with the port library.

Secondly, the flow is seemed as the P to P flow and is sent to the flow control module if the matching is successful.

Thirdly, the data flow will be sent to the deep recognition module if the match is not successful.

Then, the data flow will separate the P to P flow and the non P to P flow in the deep recognition module.

Finally, the P to P flow recognized by the deep recognition module will enter the control module.

The elementary recognition module. The elementary recognition module takes the port recognition means by using the common and known port of the P to P flow in the Internet, and constructs the port library. When the flow enters the elementary recognition module, it will firstly do the matching among the ports. If it is the common P to P flow, it will directly enter the control module and it will save the checking time of the known P to P flow.

The deep recognition module. The deep recognition module is mainly based on the combination of the DPI and DFI means, the establishment of the feature library in this module is the key to the implementation of this module.

The construction of the feature library. The feature library is the module matching base of the flow, the analyst of the protocol uses the DPI technology and the DFI technology to extract the feature in the protocols. They write the grammar into the rule files based on the feature library. Then they compile the rule files a structural file that is easily searched, which is a feature library. The feature library is a combination of the protocol rules that are concluded by the sample flow feature of the application, which is the data base when the deep flow recognition does the module matching. The feature library concludes four feature rule module, which are AC rule, Regex rule, FB rule and Compare rule.

The design and realization of the flow control system

The system is constructed on the netfilter firewall of the Linux, which is an operating system. The system uses the obtaining and transmitting mechanics of the data packet of the netfilter to realize the purpose of controlling the flow. The netfilter is the assemblage of a serial of hook functions in the kernel, it provides the interfaces. Users or developers make the function as the callback functions in the interface parameters, and attach the designed module to the hook function to make the function run on the Linux as the kernel function. By using the Hook mechanics of the netfilter, we can obtain and trace the data packet. Also, we can realize the control of the data packet. This paper also uses the

netfilter mechanics to realize the flow control module [2]. The places where the distribution situation and control module of the Hook point of the netfilter are placed are showed below.

The Netfilter sets five Hook inspection points, and every inspection function is below.

Table 3 Five Hook inspection points and corresponding function

Name of inspection point	Description of the function Scheme 1
NF_IP_PRE_ROUTING	After the data packet entering the network layer, the abnormality inspection of the head of the data packet will be done at this point. The transformation of the source address will happen at this point.
NF_IP_LOCAL_IN	The data packet sent to the host will be routed to this point. The data packets will be filtered at this point and be used to keep the security of the system. The firewall of the system is usually constructed at this point.
NF_IP_FORWARD	Those data packets needed to be transmitted will be routed to this point. The data packets filtered and transmitted here can be used to construct the firewall of the network. The FORWARD chain in the Iptables firewall governs the packets that pass through the router of the system, and all the behaviors happen at this point.
NF_IP_LOCAL_OUT	The machine programs are sent to other host packet and are tested here.
NF_IP_POST_ROUTING	The packets needed to go through the system or to be sent by the system itself are all pass through this point, and this point has the transformation function of the destination address.

Because this system needs to control the flow, this system mainly deals with the flow transmitted by the server.

When the flow enters the server through the point NF_IP_PRE_ROUTING, the system needs to do the data transmission. If it finds that the destination address of the message is the same as the address of the host, the message will be sent to next hook function, the point NF_IP_FORWARD [3]. After the point NF_IP_FORWARD finishing dealing with the message, the message will be sent to the NF_IP_POST_ROUTING hook. The flow recognition module and control module are both constructed on the chain of the netfilter, this system uses the function library of the libnetfilter queue. The libnetfilter queue is a user state library, which provides the data packet of kernel state that is put into the queue of the API [4]. By calling back the function, it can make the program recognize and deal with the packet catch by the kernel. Then it will deal with the return value of the function which can decide where the data packet will go in next step. This step is based on the recognition result and the strategy of the governor of the network.

Summary

This paper demonstrates the function of the system and the design of the frame. It also tests the function of the realization module. This paper basically meets the standard of the system. We will do and complete the distal control module based on this paper to realize the purpose of remote controlling of this system, and show the results of the system on the anterior.

References

[1] J.F. Peng: *The research on the key technology of the P to P flow recognition* (Ph.D. Beijing University of Posts and Telecommunications, China 2011).p.5.

[2] Y. Liu: The research and design of the deep packet inspection (MS. Guizhou University, China 2008).p.14.

[3] Michael Rash: The firewall of Linux (MS. Beijing University of Posts and Telecommunications Publishing Company, China 2009).p.2.

[4] http://www.netfilter.org/projects/libnetfilter_queue/index.