Research Journal of Applied Sciences, Engineering and Technology 5(13): 3560-3566, 2013
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: July 27, 2012 Accepted: September 08, 2012 Published: April 15, 2013

Research on the Technologies of Instant Messaging Software Detection

[1]Hao Zhang, [1]Guangli Xu, [1]Jiongzhao Yang and [2]Jianmin Li
[1]Qinggong College, Hebei United University, Tangshan 063000, China
[2]Institute of Electronics, Chinese Academy of Sciences, Beijing 100000, China

Abstract: In order to solve the security problems of instant messaging, a new Instant Messaging Software (IMS) intrusion detection method has been proposed in this paper. Intrusion Detection is an important component of the security of the Network, through the critical information of the network and host system, it can determine the user's invasion of illegal and legal acts of the user misuse of resources and make an adequate response. Instant messaging software provides a platform for the communication of information. It's convenient for people to communicate, at the same time enterprises and institutions, companies, families; local area network has security implications. Such as access to user's personal information and the company's trade secrets; dissemination of pornography, reactionary remarks; provide attack corridors for Trojans and worms. Therefore, study the detection and blocking techniques of instant messaging software to protect the security of local area network has great application value. According to the detection method of IMS, detection and blocking architecture of IMS is presented and developed at the Windows platform. The rule base store in the configuration file, users can add their own IMS, detection and blocking the rules, users can also add other applications, features, so that the system has better scalability and practicality.

Keywords: instant messaging, intrusion detection software, network security, Win cap

INTRODUCTION

With the rapid development of Internet, instant messaging software such as Ten cent QQ in China, MSN, China Mobile's Fetion and others have been used by more and more net users, it is not only as a chat tool, but also to be a device to provide voice, video and data transmission services (Joe, 2005). However, the harm caused by instant messaging software (Yu *et al*., 2008) cannot be ignored: firstly, it can cross through the firewall, which will cause the firewall lost its protective effect; secondly, LAN users use data of this kind of software, video chat and other services, which take up a lot of bandwidth and affect the normal speed of the network; last, but not least, it provided a platform for obscene, reactionary remarks. It is very important to research on detection of instant messaging software according to network infrastructure and communication features of instant messaging software, which will be a meaningful thing for local area network security and information security.

Zhang (2007) proposed secure analysis and design of instant messenger software, Zhang (2007) proposed research of secure instant messaging add-in based on Microsoft MSN and this article is mainly research detection of Ten cent QQ.

With the numbers of the users of instant messaging software increasingly more and more and the problems in terms of security has become prominent increasingly. It includes the respects are as follows:

- Instant messaging software allows the user to select or to automatically attempt to the communication port and even use some mechanism to bypass the firewall. In this case, the firewall in loses its protective effect and that these communications through a firewall is likely to be exploited by attackers on the network and do damage to computers.
- Due to the safety defects of the instant messaging software itself, it may give hackers convenience, causing data theft, distributed denial os service attacks, virus and worm attacks further.
- Due to each country's different political and cultural background and legal constraints, instant messaging software for people to bring these conveniences, but also open the door to pornography, reactionary remarks and cult heresy, on the security and stability of the state and society constitute certain hazards.
- Using instant messaging software to transfer files, video chat and other services, consume a large amount of network bandwidth, affecting the normal speed of the network.
- Instant messaging software provider add the control program in the software, as long as the user's local

Corresponding Author: Hao Zhang, Qinggong College, Hebei United University, Tangshan 063000, China

machine, add a few keywords to search and filtering, can be critical information back to the server, resulting in a company or user sensitive information leakage. Many vital government departments, military, commercial companies have realized that the security of instant messaging software and prohibited the use of instant messaging software. At present, detection and blocking of instant messaging software technology is not mature and still be able to pass through the firewall.

In order to protect the security of the LAN and prevent harm from the instant messaging software, we need to use the blockade of technology to block it. Different instant messaging software uses a different and is not open agreement and their means of communication are not same, such as the using of P2P technology and HTTP tunneling technology. Blocked premise is to detect the presence of instant messaging software and then take the appropriate blockade technology software according to the different communication technologies. The current detection technology is to detect server IP address and port, or monitor the running processes. Blockade commonly used firewall and use a firewall for the purpose of detection will affect firewall performance and increase network latency, sometimes up to less than blockade. Therefore, the introduction of intrusion detection technology to detect the LAN instant messaging software and then using different communication technologies according to the characteristics of instant messaging software, take different blockade policies.

NETWORK ARCHITECTURE AND NETWORK PROTOCOL ANALYSIS

Instant messaging software in the user level (Pathcha and Park, 2007) the client and server data is encapsulated in the transport layer TCP and UDP protocols to communicate. Means of communication and communication protocols for instant messaging software built on the basis of analyzing data on the instant messaging software package, to be resolved, the packet is necessary for the Ethernet physical and data link layer, Internet layer and transmission layer protocol. The packet capture technology is the premise of the resolution packets. Therefore, here we will mainly discuss and analysis network architecture, network protocols and network packet capture.

TCP/IP architecture: TCP/IP (Niu, 2007) is standard protocol of the Internet. With the rapid development and popularization of Internet, TCP/IP has become the most widely used protocol all over the world in the Internet. TCP/IP protocol is a heterogeneous networked communication protocol; it also applies to a local area network to achieve different kinds of Internet communication between computers. Architecture of TCP/IP is shown as Fig. 1. The layers 1 to 4 are application layer, transport layer, network layer and physical and data link layer.

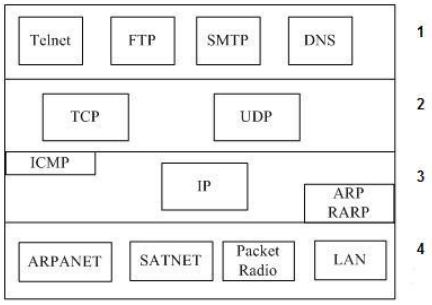

Fig. 1: TCP/IP architecture

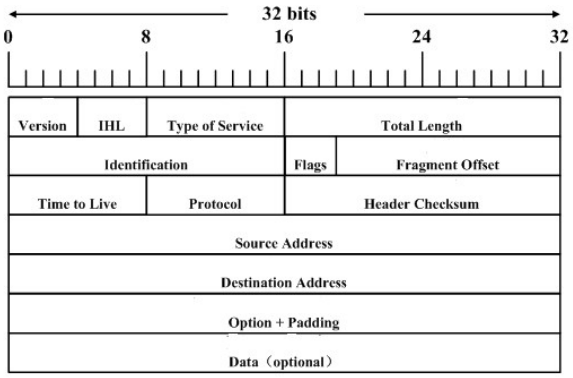

Fig. 2: The IP format

Analysis of network protocols: Internet Protocol IP. The IP protocol is the Internet layer of the most important protocol, which is responsible for the communication subnet range between hosts across the Internet to communicate with each other. IP datagram of the IP header and data, is shown in Fig. 2, the first 20 bytes of the IP head is fixed behind the option field is variable length. The length of the data portion is variable, all UDP, TCP, ICMP and IGMP data transmission in the IP datagram format. The IP header protocol field stored in the data format of the data fields. TCP protocol, such as 6, 17, represent the UDP protocol.

IP protocol format is defined as follows:

```
struct ip__header
{
#if defined (WORDS-BIGENDIAN)
UCHAR ip__version: 4,
      /* version*/
ip__header__length: 4;
      /* header lenth*/
```

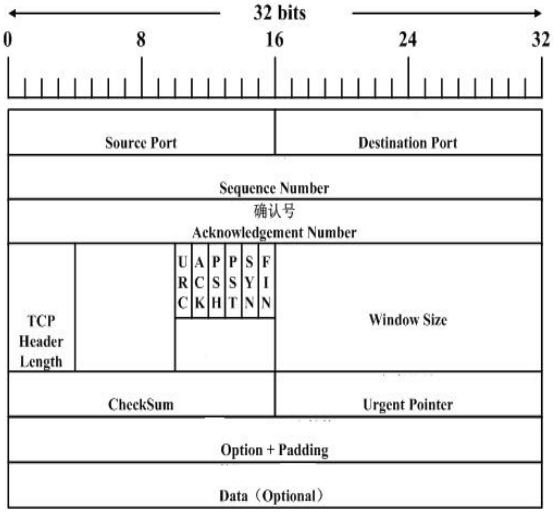

Fig. 3: The TCP format

```
#else
 UCHAR ip__header__length: 4, ip__version: 4;
#endif
UCHAR ip__tos;
      /* Qos*/
USHORT ip__length;
      /* Lenth*/
USHORT ip__id;
      /* ID*/
USHORT ip__off;
      /* offset*/
UCHAR ip__ttl;
      /* Survival time *
UCHAR ip__protocol;
      /* type*/
USHORT ip__checksum;
      /* check sum*/
struct in__addr ip__souce__address;
      /* source ip address*/
struct in__addr ip__destination__address;
      /* Destination IP address*/
}
```

Transmission Control Protocol TCP. TCP unit of data transmitted between two devices, called packets, TCP packets from the head and a data portion. Shown in Fig. 3, 20 bytes before the head is a fixed part, behind the "Options" field is variable length. The data portion of the length is variable, the data portion, including instant messaging software, data or data of the DNS domain name system.

TCP protocol format is defined as follows:

```
struct tcp__header
{
   USHORT tcp__source__port;
                              /* source port*/
```

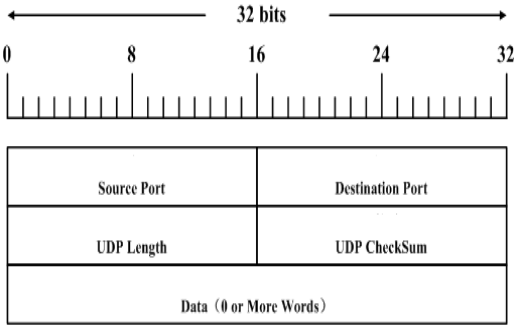

Fig. 4: The user datagram format

```
 USHORT tcp__destination__port;
          /* destination port*/
 ULONG tcp__sequence__lliijjiiaannmmiinn;
    /* Serial number */
 ULONG tcp__acknowledgement;
          /* acknowledgement*/
 #ifdef WORDS__BIGENDIAN
 UCHAR tcp__offset: 4,
                    /* offset*/
   tcp__reserved: 4;
                              /* not use*/
 #else
   UCHAR tcp__reserved: 4,
          /* not ues*/
   tcp__offset: 4;
                              /* offset*/
 #endif
 UCHAR tcp__flags;
                              /* flag*/
 USHORT tcp__windows;
                              /*windows size*/
 USHORT tcp__checksum;
                    /* check sum*/
 USHORT tcp__urgent__pointer;
                    /* urgent pointer*/
}
```

User Datagram Protocol UDP. UDP is a simple connectionless protocol, applied to a simple request - answering queries and fast communications and other occasions. Uses UDP to send packets, the interaction between the sender and took over the side is much less than that of TCP. UDP protocol for multimedia and multicast. Figure 4 the UDP protocol format is divided into two parts of the head and data. Head accounted for 8 bytes; there are no variable-length options. The data portion of the variable length data section includes the data of the real-time communication software, data or DNS domain name system.

UDP protocol format is defined as follows:

```
struct udp__header
{
```

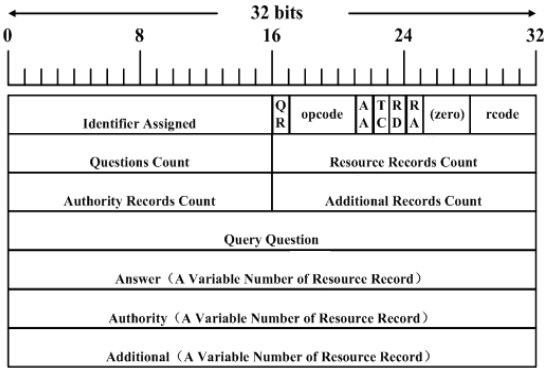

Fig. 5: The DNS data format

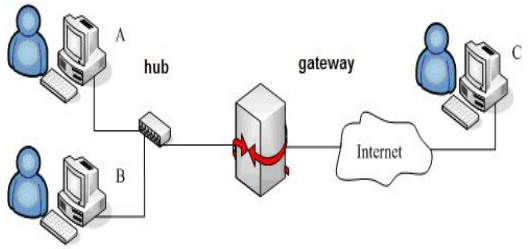

Fig. 6: The protocol analysis hardware environment

```
    USHORT udp__source__port;
                        /* source port*/
    USHORT udp__destination__port;
                /* destination port*/
    USHORT udp__length;
                                /* length*/
    USHORT udp__checksum;
                        /* checksum*/
    UCHAR qq__udp__protocol;
                                /* type*/
}
```

Domain Name System DNS. DNS that is responsible for the Internet domain names into appropriate IP address, the Internet service. Computer called a DNS server with DNS functionality. Because the IP address in digital form, it is difficult to remember, so people only memory address symbolic domain name, DNS server, the domain name into an IP address. Queries fields and answer field contains the domain of instant messaging software and the corresponding server IP information. DNS server listens on port 53 in UDP or TCP protocol, generally encapsulated in UDP packet transmission, when too much data to be transferred (more than 512 bytes) to establish a TCP connection, data is encapsulated in TCP packets transmission. DNS data format is shown in Fig. 5.

This paper's Hardware environment is shown as Fig. 6: 3 computers, an Ethernet hub station, a gateway. The LAN consists of PC A and B and a hub, computer C is the external network PC.

Software environment, including instant messaging software QQ2008, MSN, Kaspersky Internet Security 7.0 and WinPcap 4.0.1 and network protocol analysis tool Ethereal (Ethereal 0.99.0). All clients are using Windows XP SP3 operating system. The client A uses Ethereal, the adapter of capture packets in Ethereal is set to promiscuous mode and three clients, closeing other applications in the network. A can capture and record all transfers within the LAN packets. To rule out the other unrelated packets, only concerning to capture and record instant messaging software packets by setting Ethereal filters the rules. QQ, MSN and Kaspersky Internet Security 7.0 beta are installed to the client B and C.

QQ protocol is an application layer protocol; it is divided into two formats, depending on the transport layer protocol. But the meaning of the protocol field of the same name is the same.

The head field is unique QQ protocol in the TCP packet format, accounting for two bytes uncertainty.

The identifier field occupies one byte; fixed content is 0x02 and used to distinguish the other application layer protocol.

The version number field version of the QQ, different clients have different version numbers. QQ2007 official version 0x111D, in QQ2008 Hesuiban, 0x115B QQ2009 official version to 0x1645. When communicating with the server, but also represents the server version number or mark.

The command field is two bytes; different commands represented with a different number, indicating that the client or server-side operating command.

The sequence number field occupies two bytes of data sent to the successful transmission of the response data is received the same sequence number; the sequence number of the sender and recipient of the data received the same serial number. The serial number may be randomly generated.

QQ number field is the sender to send data QQ number, QQ number of the recipient of the data to be sent are stored in the encrypted data. The data packets received by the client does not include the QQ number field, In this case, the sender's QQ number has been server processing and placed in the encrypted data field.

QQ identified as 0x02, according to misuse detection model detection rules library is established. Using analysis technique to unpack captured network protocol by WinPcap, parse out the application layer data and then match the detection rules library, so that can be detected instant messaging software to run and to reduce the false negative rate, to improve the detection performance. According to this principle, write your own detection system and data collection to save.

- Instant messaging software in UDP protocol communications, most of the client data are transmitted via UDP client to send or receive data

to be received or sent a confirmation packet. Client port can dynamically change when a port communication failure, the client automatically select another unused port for communication. QQ majority of the data packet is transmitted in the UDP protocol. QQ protocol in the format of the UDP protocol, the detection system will be able to detect its presence

- Instant messaging software communications in some special occasions with TCP protocol, QQ client cannot use the UDP protocol to communicate; it will automatically select the TCP protocol to transmit data. Communication port with 80, 443 and 8000, QQ members use port 443 communication. QQ has many TCP servers and open these ports to client communication. These servers handle these data, or data to transfer server processing. Client information through the TCP transport, its principle is the need to use data transmission through UDP TCP protocol for transmission, but it is also fixed in the application layer protocol format has its unique identity. The detection system is also able to detect the presence of the client. However, because the use of users of QQ is more, when everyone using TCP transport network burden will be very large, so it is not in a special environment, QQ client will use UDP protocol communications.
- Instant messaging software to play the role of data relay communications through a proxy server, proxy server. HTTP proxy to HTTP proxy client access, proxy browser to access the web; SOCKS5 proxy simply passes the packet and does not care what kind of application protocols HTTP request SOCKS5 proxy server than other types of The proxy server much faster. HTTP proxy and SOCKS5 proxy is somewhat different, but the effect is the same as the general proxy server provides the QQ data of forwarding
- Instant messaging software using a Web log, is to use HTTP tunneling technology web instant messaging software. Login packets of instant messaging software to send the packet to another form embedded in the HTTP protocol to another protocol format. Through the fixed location of the data to determine the HTTP protocol matches the rule base method cannot detect the data packet of the instant messaging software, because the instant messaging software keyword has no fixed position, but it's the keywords still exists in the application layer data and by string matching method Find the position of the keyword in the application layer and then determine whether the packet of instant messaging software. The QQ client since logged begins every 60s to the login server sends a heartbeat to indicate that it is in the active state. Therefore, even if the user of the client does not send and receive any information, the client will report regularly to the server to send packets, the detection system can detect its presence. In summary, the method of instant messaging software package communication protocol in the transport layer protocol for transmission, or a different communication protocol encapsulated in the HTTP tunnel detection, instant messaging software communication protocol instant messaging software can detect exist.

WINDOWS PLATFORM PACKET CAPTURE ANALYSIS

WinPcap system is the better data packet capture and filtering procedures under the Windows platform, WinPcap network system will not affect the normal speed of the Internet and detection function of the system is based on this technology.

WinPcap (windows packet capture) is a free public network access system under Windows platform. It is for the Win32 platform network packet capture and analysis of open source libraries. The package includes a kernel mode filter for network packet capture and filtering functions, also known as NPF (Net group Packet Filter) packet driver; a low-level dynamic link library (packet.dll), which provides developers a low-level programming interface; a high level does not depend on the library (wpcap.dll), which provides developers to develop a higher level interface, the dynamic link library is independent of the system. Figure 7 is WinPcap's structure and location of the operating system.

WinPcap provides the following functions: for capture the raw data packet, whether it sends or receives data packets or change data packets between the host in the net; before the packet is sent to the application, it should be filtered specified packet in accordance with the users' some special rules; the original packet is sent to the network; to collect and process statistical network traffic flow information. WinPcap is only in the addition of a bypass data link layer processing on the sending and receiving data and then filtering, buffering packets and other related processing. It did not affect network performance. WinPcap packet provides provide a platform to capture data packets, the interface of WinPcap should be used, it can capture and filter out wanted packets. Some related functions of WinPcap packet capture are as follows:

Get a list of network adapters connected to the function:

pcap__findalldevs ();

Open the selected network adapter interface function:

pcap__open__live ();

Compile BPF filter rules function:

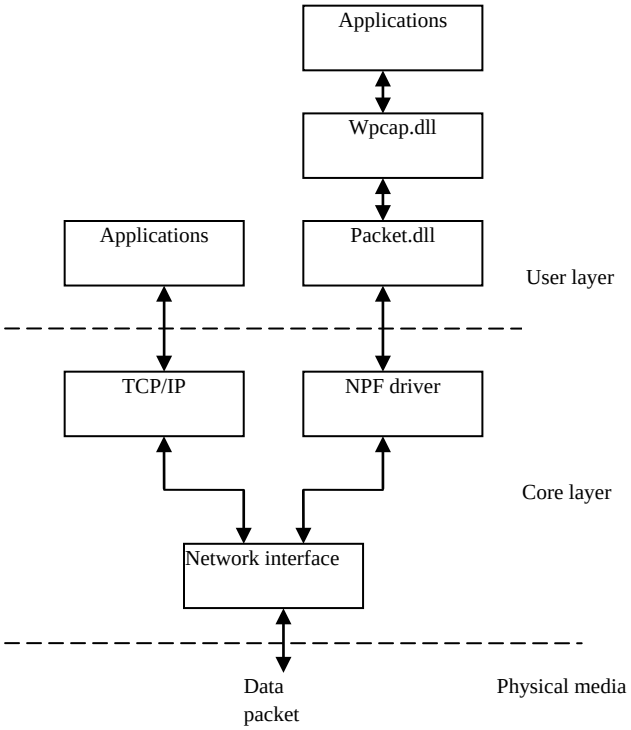

Fig. 7: Architecture of WinPcap

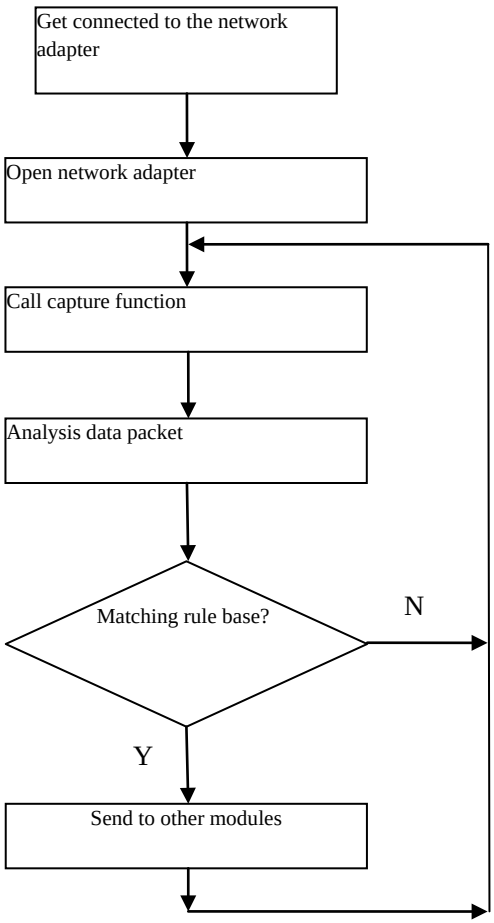

Fig. 8: The flow chart about detection

pcap__compile ();
set filter rules function:
pcap__setfilter ();
LAN type of search functions:
pcap__datalink ();
Network packet callback function:
pcap__loop ().

DETECTION MODULE DESIGN

The main task of detection module is to test instant messaging software, combining method of the intrusion detection technology and testing instant messaging software communication protocol, which can meet the functional requirements of detection module. The test procedures on detection module, IP packet unpack and universal detection technology in the detection module of the application are introduced as follows:

Instant messaging software testing process: A network capture program packet, including the user and core part. Kernel is responsible for the capture data from the network and it can also filter data at the same time; the task of user part is to perform packet formatting, protocol analysis, data conversion and processing and so on. Users' part also can filter some packets. Network packet capture program programming processes is shown as Fig. 8.

i. Select the network adapter. Find all connected network adapter and select the network interface to capture data. IP addresses can be used to represent the equipment, or use string to represent this device.
ii. Capture program initialization. Mainly for setting the selected network interface, such as the length of the packet capture is set to 65536, time out is 1 second and so on. To open the network adapter in promiscuous mode, capture all packets of other hosts. Because of the necessary analysis of all data packets, so you cannot set the kernel filtering rules.
iii. Capture the data. Through callback function or directly through the packet capture function to get the captured packets.
iv. Analysis of the data packet. Using protocol analysis techniques to capture data packets according to the head of the packet layer and to find application layer data, using misuse detection to detect application layer data packet.
v. Test the rule base line data packet and send to other processing modules. The data storage module stored rule base line test data, statistical information module on the packet IP address and port number and other information for statistical analysis, some of the information displayed to the detection program interface.

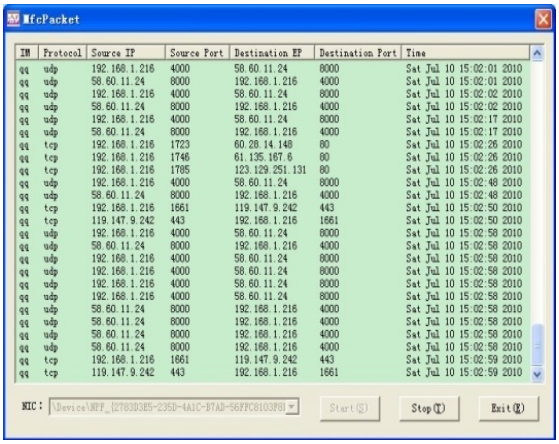

Fig. 9: Running interface of detecting system

where (iv) and (v) can be achieved by the multi-threading, it can increase the efficiency of detection and analysis system and reduce the probability of packet loss at the same time.

In order to detect more kinds of instant messaging software, the system stored the rule base in the configuration file. This instant messaging software testing method that is updating configuration file can increase scalability and availability in the system.

Adding a few simple rules in the configuration file can achieve WebQQ detection. Figure 9 is a detection system on the LAN instant messaging software testing process part of the results.

CONCLUSION

In this study According to the detection method of IMS, detection architecture of IMS is presented and developed at the Windows platform. The rule base store in the configuration file, users can add their own IMS, detection the rules, users can also add other applications, features, so that the system has better scalability and practicality

REFERENCES

Joe, L., 2005. Best practices for instant messaging in business. Network Security.

Niu, X., 2007. Design and implementation of security instant messaging system based on TCP/IP protocol. M.S. Thesis, Information Technology, China.

Pathcha, A. and J.M. Park, 2007. Network anomaly detection with incomplete audit data. Comput. Netw., 51(13): 3935-3955.

Yu, W., S. Chellappan, X. Wang, 2008. Peer-to-Peer system-based active worm attacks: Modeling, analysis and defense. Computer Communications

Zhang, B., 2007. Secure analysis and design of instant messenger software. Comput. Appl., 27: 223-224.