Scholars Bulletin ***(Literature)***

An Official Publication of "Scholars Middle East Publishers",
Dubai, United Arab Emirates
Website: http://scholarsbulletin.com/

ISSN 2412-9771 (Print)
ISSN 2412-897X (Online)

Analysis of Network User Behavior on Campus Network

Liu Xiang Wei
PLA University of Foreign Languages, Luoyang Henan, China- 471003

***Corresponding author**
Liu Xiang Wei

Article History
Received: 11.10.2017
Accepted: 26.11.2017
Published: 30.12.2017

DOI:
10.21276/sb.2017.3.12.15

Abstract: This thesis analyzes the behavioral characteristics of network users on the campus network. After capturing and processing communication data and recovering sessions, we can know the group behaviors of the campus users. What's more, the use of the diskless computer among the cadets can also be seen from this analysis. On the one hand, such analysis can offer the theoretical basis for campus network. On the other hand, this work can help the administrative know more about the cadets' study situation on the campus network and make better management. To some degree, the analysis reflects the importance of network users' behavior analysis network management.

Keywords: Network User Behavior, Analysis Behavioral, Characteristics

INTRODUCTION

Data time people life cannot leave the network applications, a comprehensive understanding of the broad masses of users in the network behavior characteristics and analysis, so as to adjust to develop network management planning, the problems existing in the system should be made clear, master network operation environment and processing mechanism, the related network integrated management technology as data the growth of the age growing, the number of Internet use and frequent and network security problem becomes more and more important, also more and more with the corresponding challenging [1].

Network user behavior analysis

Web users

Web users to use the network information to the user. Specifically refers to the various kinds of practice activities such as research, teaching, need to use the network to communicate information and information group and individual [2].

Internet users can be categorized according to the individual and group, other factors can also according to the classification, classification of different views mainly depends on their classification purposes. In this paper, we study the network users mainly depending on the nature of the industry, from the perspective of students, analysis of user behavior characteristics of this group.

The United States researchers according to the user and the network contact time and use of the network users can be divided into four kinds of frequency. This kind of classification method mainly reflects the influence of network development for human life. The first (Netizens) for the netizen, situation of the contact network more at ordinary times life work; The second is practical (Utilitarians), the class user is mainly the network as an everyday vehicle for assistance; And the third for the user or the experimenter (Experimenters), they are for the use of the network is mainly used for network access to relevant information; The last category for novice (Newcomers), also known as a rookie, experience in using of the network is not enough. The network and computer development in the United States earlier than our time, a lot of experience and methods we can still be used to reference. Now in our country the classification of network users can still according to the classification method for reference, working life is dependent on the network, there are also about Internet use few amateurs.

Network user behavior

Network user behavior, from the academic level, refers to according to the definition of active or passive network measure in advance, to summarize the changing rule of the corresponding measurements. Involved in safety management, measurement regulation of network behavior, and the hardware equipment and so on various aspects [3]. Specific include: the definition of present in the use of network resources has certain rules of behavior, can use statistical correlation with feature or characteristic of quantitative or qualitative. In addition, the data stack the layers in the network user behavior also have different and embody characteristics.

Network user behavior analysis

Network user behavior analysis refers to the network user behavior as the basis of comprehensive analysis, to get more effective, more meaningful data values and the corresponding conclusions, thus for further network planning and the next step of work lay the good foundation.

Usually found in anomaly detection, and inhibit the related illegal activities using the most common network user behavior analysis [4]. Before analysis, first of all, to record the user's normal behavior, and on this basis to build a model library, and then carries on the analysis, record the normal behavior of users, and the data into the new database will be collected to form the new database to match the pattern library, if there is abnormal situation in matching, records generated and stored, such records will serve as a warning system to continue the behavior matching.

Can classification perspective, from the level, technology as well as the data source to study.

From the Angle of protocol, the network layer and application layer can begin. According to the classification of different angles and analysis technology is mainly analysis the purpose and demand to make the best choice. When in actual use in some technology alone or separate from a certain Angle analysis of the situation is rare.

Analysis techniques such as principal passive analysis and data mining are common analysis methods, and in practical application, for the sake of a more comprehensive analysis using the collected data, will consider a variety of methods to obtain more specific analysis report.

According to the different analysis of the data source is also commonly used classification methods, data sources such as protocol control information, network traffic, web use record, system and the audit log.

Data acquisition and processing

This thesis first to collect data we need winpcap environment, build environment to build after the success of the preparation of the program, the needed data are caught and further processing and reduction, according to the agreement of the head format were analyzed, and relevant source address, destination address and time, etc., the last to restore the part points out the data processing, in order to get the data reduction [5].

Data capture

Winpcap related content after the completion of setup, data capture code.

- The first thing you need to get the network card interface information. The pcap_findalldevsfunctioncan achieve this function.
- Call pcap_open_live function is gained by the open interface.
- After the inspection on network situation and set filter, to create a new thread to work in the background of data capture and thread invokes the pcap_next_ex function to intercept data, and then to intercept the data for storage, collection part to this end, the data capture the core code is as follows.

```
While(res=pcap_next_ex(pthis->adhandle,&header,&pkt_data))>=0
{
   If (res==0)
       Continue;
Structdatapkt * data=(structdatapkt *)malloc(sizeof(structdatapkt));
Memset(data,0,sizeof(structdatapkt));
If(NULL==data)
   {
MessageBox(NULL,_T(“space is full!”),_T(“Error”),MB_OK);
     Return -1;
    }
}
```

Data reduction

To preprocess the data collected data reduction, classification, sorting and statistics. Computer communication between the layers of data processing program from figure 1, you can see the specific process of protocol stack. We want to deal with reduction of actual communication data, the first thing you need to first layer analysis for data of each layer. When packet parsing, need to understand what each layer protocol of the head, agreement is different, the corresponding first format also differ.

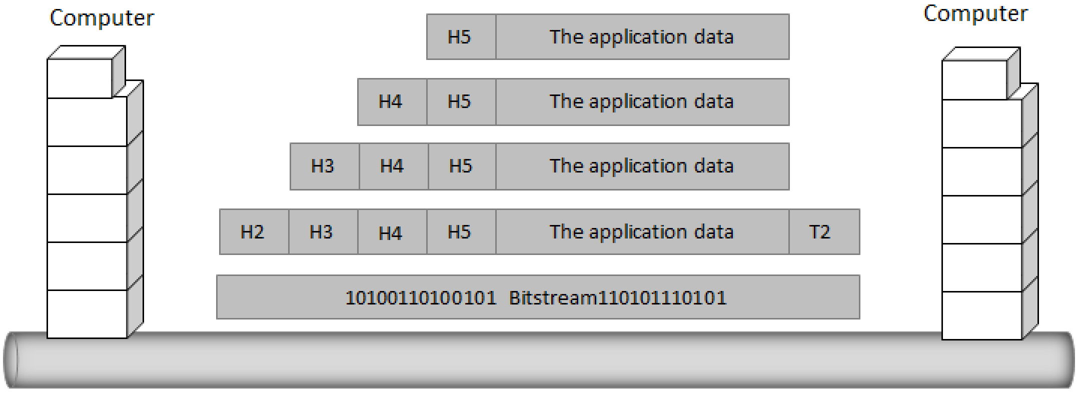


Fig-1: the layers of data transfer process

Each layer protocol has the difference, the category is different also, when parsing packets need according to the characters of different protocols to determine. The flow chart below 2 is represented in the actual program parsing different analytic function invoked.

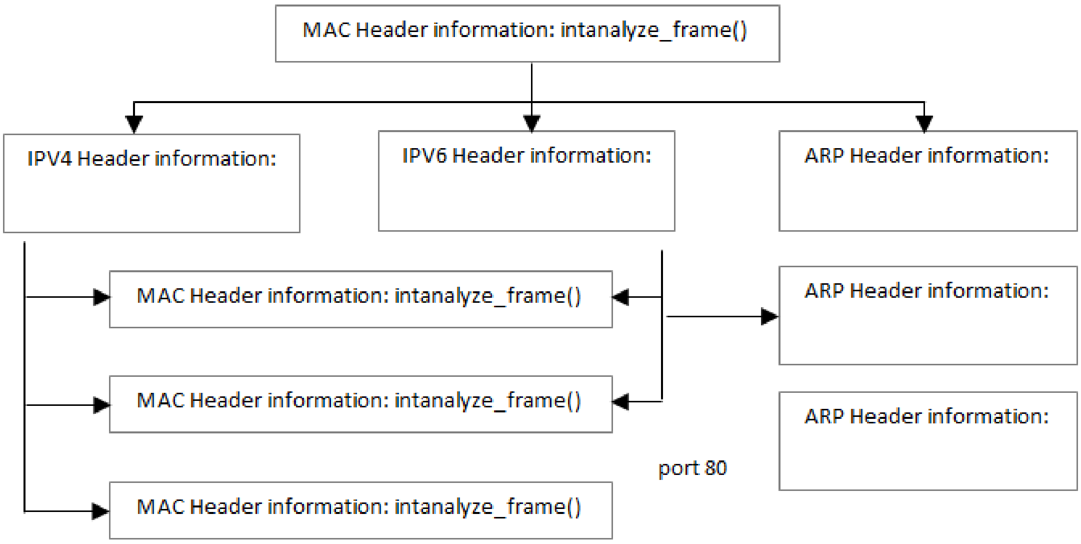

MAC Header information: intanalyze_frame()
IPV4 Header information:
IPV6 Header information:
ARP Header information:
MAC Header information: intanalyze_frame()
ARP Header information:
MAC Header information: intanalyze_frame()
ARP Header information:
port 80
MAC Header information: intanalyze_frame()

Fig-2: analytic function call process

Later in the analysis of the main HTTP packets for reduction and analysis, a preliminary selection reduction data, set the port to port 80. HTTP USES the TCP as the transport layer protocols, in the TCP header analysis as an example, as shown in figure 3 for TCP header structure.

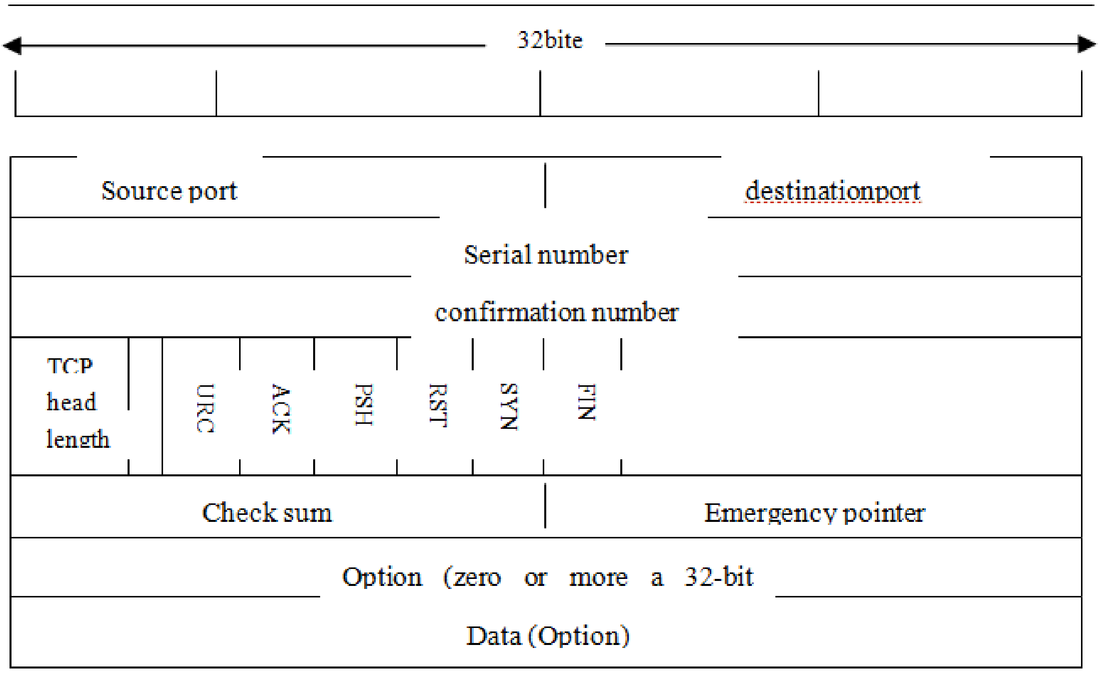
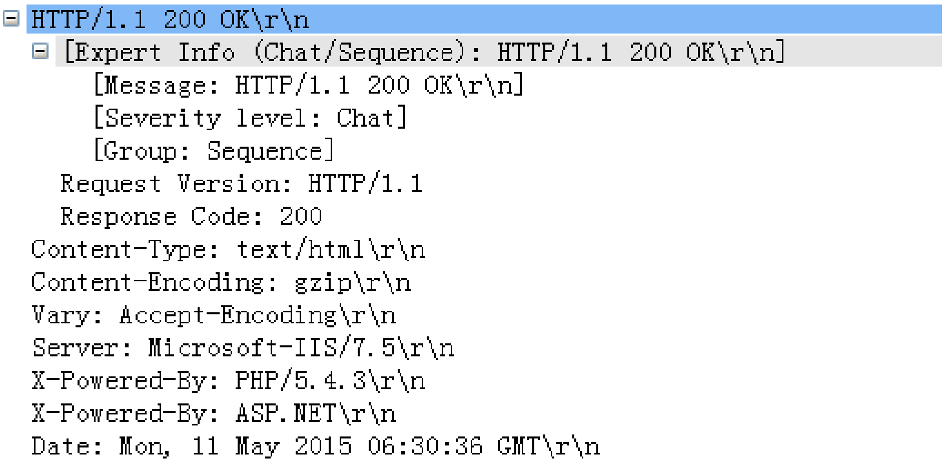


Fig-3: TCP header structure

After the analysis of the head and handle, we need to specifically for session restore data part of the HTTP message. HTTP Request contains (Request) and Response (Response) two kind of message, the difference between sent to the server for the client, start action Request line; The latter for the server to send back to the customer and start behavior status line. Among them, request the content of the message the first line for method (that is, to the operation of the requested object), the requested resource URL, as well as the HTTP version used. Methods usually have the OPTION, the GET, HEAD, POST, CONNECT and so on, different methods of content and the object is different.

Request packet usually followed by a response message, after the response message is worth our concern is a status code, a status code is three digits, can be divided into five categories, including 1 the beginning of the said notice information, such as said it was processed, and 2 indicates success beginning to accept, starting with 3 to determine the position of the requested resource or direction, such as a 304 error result indicates that the requested content is permanently removed, starting with four represent the client side appeared a mistake, such as our common of 401, 401, 5 indicates the server end the mistakes in the beginning, such as shown in figure 4 success back to the customer for a server response message the requested documents.



```
⊟ HTTP/1.1 200 OK\r\n
  ⊟ [Expert Info (Chat/Sequence): HTTP/1.1 200 OK\r\n]
      [Message: HTTP/1.1 200 OK\r\n]
      [Severity level: Chat]
      [Group: Sequence]
    Request Version: HTTP/1.1
    Response Code: 200
  Content-Type: text/html\r\n
  Content-Encoding: gzip\r\n
  Vary: Accept-Encoding\r\n
  Server: Microsoft-IIS/7.5\r\n
  X-Powered-By: PHP/5.4.3\r\n
  X-Powered-By: ASP.NET\r\n
  Date: Mon, 11 May 2015 06:30:36 GMT\r\n
```

Fig-4: response packet data capture

This article requires users to browse the web content mainly from the class message, need to restore the HTML response content, the following for preliminary analysis has been reducing the document content of English characters:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">\r\n
<html xmlns="http://www.w3.org/1999/xhtml">\r\n
<head>\r\n
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />\r\n
[truncated]
<title>\344\270\216\346\242\246\346\203\263\350\200\205\345\220\214\350\241\214\342\200\224\342\200\224\346\23
4\254\347\247\221\347\224\237\345\255\246\345\221\230\351\230\237\346\227\266\351\232\224\344\270\203\345\271
\264\351
<meta name="keywords" content="\346\242\246\346\203\263,\345\267\245\347\250\213\344\270\200," />\r\n
[truncated] <meta name="description" content=" \346\240\241\345\233\255\350\256\272\345\235\233
\345\245\275\350\247\206\351\242\221\345\272\224\350\257\245\345\215\225\347\213\254\345\217\221\344\270\252\
350\264\264\347\232\204\346\217\22
<meta name="generator" content="Discuz! 7.2" />\r\n
<meta name="author" content="Discuz! Team and Comsenz UI Team" />\r\n
<meta name="copyright" content="2001-2009 Comsenz Inc." />\r\n
<meta name="MSSmartTagsPreventParsing" content="True" />\r\n
<meta http-equiv="MSThemeCompatible" content="Yes" />\r\n
<meta http-equiv="x-ua-compatible" content="ie=7" />\r\n
<link rel="archives" title="\346\240\241\345\233\255\350\256\272\345\235\233"
href="http://bbs.plaufl.mtn/bbs/archiver/" />\r\n
<link rel="stylesheet" type="text/css" href="forumdata/cache/style_1_common.css?y7Z" /><link rel="stylesheet"
type="text/css" href="forumdata/cache/scriptstyle_1_viewthread.css?y7Z" />\n
[truncated] <script type="text/javascript">var STYLEID = '1', IMGDIR = 'images/default', VERHASH = 'y7Z', charset =
'utf-8', discuz_uid = 0, cookiedomain = '', cookiepath = '/', attackevasive = '0', disallowfloat = '', creditnotice = '1|\3
<script src="forumdata/cache/common.js?y7Z" type="text/javascript"></script>\r\n
</head>\r\n
```

Clearly, this is the campus BBS of the website, the content of the captured is an HTML document, part of the intercept for web page code file head, which can determine the preliminary web page source encoding to utf-8, if you want to know the specific content of the web browsing, need to < title > after a string of Numbers for reduction. After reduction is as follows:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title> With the dreamer——Undergraduate students regain games champion after seven years
- ‖Institute of dialogue‖ -  The campus BBS  - Powered by Discuz!</title>
<meta name="keywords" content="dream,engineering," />
<meta name="description" content=" The campus BBSGood video should separately send a stick the extraction of the
code ： god - Discuz! Board" />
<meta name="generator" content="Discuz! 7.2" />
<meta name="author" content="Discuz! Team and Comsenz UI Team" />
<meta name="copyright" content="2001-2009 Comsenz Inc." />
<meta name="MSSmartTagsPreventParsing" content="True" />
<meta http-equiv="MSThemeCompatible" content="Yes" />
<meta http-equiv="x-ua-compatible" content="ie=7" />
<link rel="archives" title="The campus BBS" href="http://bbs.plaufl.mtn/bbs/archiver/" />
<link rel="stylesheet" type="text/css" href="forumdata/cache/style_1_common.css?y7Z" /><link rel="stylesheet"
type="text/css" href="forumdata/cache/scriptstyle_1_viewthread.css?y7Z" />
<script type="text/javascript">var STYLEID = '1', IMGDIR = 'images/default', VERHASH = 'y7Z', charset = 'utf-8',
discuz_uid = 0, cookiedomain = '', cookiepath = '/', attackevasive = '0', disallowfloat = '', creditnotice = '1| prestige
```

```
|,2|gold|,3|manual|,4|passion|,5 Special contribution |', gid = parseInt('5'), fid = parseInt('97'), tid =
parseInt('89830')</script>
<script src="forumdata/cache/common.js?y7Z" type="text/javascript"></script>
</head>
```

To this, the whole part of data reduction has been completed.

The data processing

After preliminary collection of data is too large and heavy and complicated, and has certain disorder, although in capturing packets when it set the part stresses the packet filtering rules, there are still many redundant data attributes to work late [6]. If applied to the server side to collect all the campus network user's behavior, the number will be more big, it will have to consider in advance to the analysis of the operation time of work and operation problem of space.

Therefore need to collected data in accordance with the requirements set by the target and further filtering, picking, classification to be suitable for the analysis of the data set. In order to get to the analysis of the effect, according to different analysis purposes, in accordance with the relevant properties are classified storage of data, to remove redundant information, the other to the analysis of follow-up work as the basis, and the data extraction after deposit in the database.

CONCLUSION

This thesis use of BBS based on campus network users as the main example, main campus network data flow, by capturing, aggregation, get the student to the computer using the relevant data, further to restructure the session layer to restore sessions, repass classify data classification and processing of crowd behavior identity, provide the basis for the back of the data analysis. So, convenient manager were analyzed and relevant usage, more intuitive and effective analysis results.

REFERENCES

1. Brachman, R. J., & Anand, T. (1996). The process of knowledge discovery in databases. Advances in knowledge discovery and data mining.
2. Dunham, M. H. (2006). *Data mining: Introductory and advanced topics*. Pearson Education India.
3. Howard, P. E., Rainie, L., & Jones, S. (2001). Days and nights on the Internet: The impact of a diffusing technology. *American Behavioral Scientist*, *45*(3), 383-404.
4. Guanqiang, L., Chen, Y., & Qiang, L. (2004). Analysis of Internet users in China, 5, 43-46.
5. Xiren, X. (2012). Computer network (fifth edition), electronic industry press, 5, 23-33
6. Shuangxi, C., & Deng, X. (2006). Network users information behavior research ShuLve. *Journal of intelligence*, 2, 79-81.