INTELLIGENT AUTOMATION & SOFT COMPUTING, 2017
http://dx.doi.org/10.1080/10798587.2016.1267445

The Research of Address Message of an Unknown Single Protocol Data Frame

Zheng Jie[a,b] and Li Jianping[a]

[a]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China; [b]Chongqing College of Electronic Engineering, Chongqing College Town, Chongqing, China

ABSTRACT
Network protocols are sets of standards for certain network communications. The identification and analysis of network protocol are of significance to network management and security. There are various technologies of protocol identification, but in the process of identification protocols, in order to simplify the identification process and improve the efficiency of protocol identification, unknown mixed multi-protocol needs to be separated into single protocol so as to make further identification. This paper presents an efficient method to determine the single protocol address message based on the previous research of separating unknown mixed data frame into single protocol. By this way, the data frames of single protocol are split into point-to-point data frame according to the address; consequently, the final identification of unknown protocol can be realized. Moreover, the method was evaluated by analysis of the ARP and TCP data; this method is able to find the more than 2/3 of address information.

KEYWORDS
Protocol identification; Separate protocol; Single protocol; Data frame; Address message

1. Introduction

Wireless network has been an essential approach for accessing network. With the development of wireless network and broad application of wireless access technology, mobility and flexibility of wireless network have brought about the great increase of the number of wireless network users (China internet information center, 2012). However, owing to the vulnerability and covert communication of wireless network protocol (WNP) are likely to induce potential risks, in order to facilitate the healthy and rapid development of wireless network, it is urgent to enhance relevant management and constant optimization. Unknown protocol technology is able to find out the underlying unknown protocol in a wireless network environment. It provides effective technology and data support to the management and control of wireless network. Besides, it also can restrain excessive exclusive protocols in present wireless network and optimize wireless network environments.

Port mapping was performed to realize protocol identification in the early period (Zhu, 2008) on the assumption that the programming of most of operating systems and software are conducted by strictly obeying the request for comments (RFC). Default communication port is set prior to the protocol standards are released. In the meantime, all users must obey the standards, for instance, the port 80 used in HTTP and Port 21 in FTP. The earliest research was made on port based protocol identification technology (IANA), which can identify the register port protocol in Internet Assigned Numbers Authority (IANA) (Liu, Huang, & Chen, 2004). However, with the development of network protocol, large numbers of protocols apply on free ports. In such context, some new protocols exhibit new features as follows: a) The communication without employing fixed port; b) private protocol communication by reusing public port and c) using transmission tool of public known protocol. Those novel features have shown apparent limits of the port mapping based protocol identification techniques and lead to the fact such technology fails to adapt to present network environments. Researchers (Chen & Wang, 2013; Kim, Won, & Hong, 2005) illustrate the reasons of causing invalidity of such technology and propose that the accuracy of this technology is found to be lower than 50%.

The leading feature of deep packet inspection (DPI) (Liu, 2010; Sen, Spatscheck & Wang, 2004a) is that it scans and matches the data package of network message more deeply (Schiller, Binkley, & Harley, 2009). The realizing principle of the technology is basically consistent with the common inspection based on tagged words; while it inspects more deeply than the common one and therefore shows a higher inspection rate. Besides, DPI flexibly applies the advantages of other inspection technologies. Generally, simple port identification is performed first in the DPI system and tagged word based identification is conducted for the messages only when port identification fails. In addition, gateway inspection at application-level is integrated in the DPI to solve peer-to-peer (P2P) problems for separating control flow from the data stream.

Sen, Spatscheck & Wang (2004b) determined protocol feature string at application-level by analyzing relevant documents and actual message flow of five P2P protocols. Based on the method, the protocol feature string at the application-level can be accurately identified. However, it is merely applicable for the identification of protocols with public rules. Wang Yipeng, et al. (2012) put forward a protocol identification method based on semantics. By using the method, the original data sets are segmented through 3-gram, and then the words are spilt according to semantics. Afterwards, the units in other formats are screened and the tagged word of protocol is generated by applying LDA algorithm. Finally, the clustering and the extraction of the protocol format are realized. However, the method

CONTACT Zheng Jie hlx2015@yeah.net

shows complex computation and takes a long time. In addition, the extraction of keywords at semantics –level is more feasible for text protocols or binary protocols with known semantics. Wang, Zhang, Wu, et al. (2013) proposed a keyword based extraction approach, which is applicable for both text protocols and binary protocols. According to the method, the input data of binary protocol are segmented and spliced based on the byte to obtain the keywords of specific protocols. Whereas there is no standard principle for segmenting spliced long strings in a practical application.

Theoretically, by applying load based protocol identification, protocol features can be obtained by analyzing the protocol rules and actual interactive messages. The method is the most accurate algorithm in the field so far. The studies (Kang, Kim, & Hong, 2003; Van der Merwe, Caceres, & Chu, 2000 and Wang, Zhang, & Wu, 2013) based on the method presented less than 10% of false positive rate. But among the existing algorithms, this kind of algorithm shows highest spatial and temporal complexity, which increases with the increasing number of protocols to be identified. In the application of load based protocol identification, the development of protocols to be identified has to be monitored all the time, which causes a large workload for updating. Therefore, this kind of algorithm is commonly utilized in protocol identification with little amount of protocols. Even in such context, it still costs a remarkable workload.

In general conditions (except encrypted flow identification), by adopting DPI, the identification rate can be improved to 95%. The major advantage of DPI is that it can accurately identify P2P flow using uncertain flow, (which is the development trend of P2P application). Aims at existing P2P application, therefore DPI is less effective in identifying the updated version or new protocols. To solve the problem, old rules have to be revised and new rules have to be added. As a result, it cannot be extended widely. In addition, when the data at application-level of network are mined using DPI, users' privacy may be violated. Therefore, many users contradict the method. For encrypted data facing network, they are difficult to be identified using the tagged word based identification. As the core technology of DPI is tagged word based inspection technology, encrypted data are difficult to be identified using DPI.

Aho and Corasick proposed a pattern matching algorithm in Bell laboratory based on deterministic finite automata (DFA) in 1975 (Li, Wang, & Wang, 2007). It is a classical algorithm for pattern matching. The method converts character comparison into state transition using deterministic finite automata. Pattern matching is a basic algorithm. Exact pattern matching algorithm includes single pattern matching (He, Wang, & Zha, 2010) and multi-pattern matching algorithms (Zhu, 2012). The former is used to find certain patterns in character strings. When multiple patterns need to be matched, each pattern has to be traversed using the algorithm. The later can match multiple patterns through one traversal and therefore shows high matching efficiency. Besides, the multi-pattern matching algorithm is applicable for the matching of a single pattern as well. As network protocol becomes more complex, a feature is likely to match or partially match multiple rules. In such context, based on single matching, the matching algorithm has to be operated every time in the matching of each rule. Obviously, it is less efficient. Therefore, multi-pattern matching algorithm presents significant advantages in protocol identification (Zhang, Ye, & Chen, 2010).

Based on the research focus of wireless network, a wireless protocol identification approach was put forward by referring to the association feature of bit stream data at data link layer studied in previous work. By utilizing the method, unknown mixed multi-protocol was separated to single protocols for further identification (Wang, Wu, Li, & Zhang, 2015). Then the address message was found. Finally, the data frames of the single protocols were spilt into point-to-point data frames according to the address. The method can provide technical support for guaranteeing the security, management, and quality of service (QoS) of a wireless network.

2. Preparation

2.1. Nomenclatures

Medium/Media Access Control (MAC) address is a series of unique symbols for identifying the network card and is generally known as the physical address of the network card. It is designed to represent the identifier of each website and denoted using a hexdecimal number with six bytes (48 bits).

IP address is known as a 32 bit of address assigned to each host computer linked to the internet. According to TCP/IP protocol, IP address is expressed using a binary number with 32 bits (4 bytes).

Protocol fingerprint is a unique series for identifying a protocol. It represents the message feature of protocol data. The series exists in each message of a protocol and can be used to indicate the protocol in data stream.

Local Area Network (LAN) refers to a computer group interlinked by multiple computers in an area. It has functions including file management, share of application software and printer, schedule in the group, communication services such as email and facsimile communication, etc.

Waikato Environment for Knowledge Analysis (WEKA) is a free and non-commercial (commercial data mining software Clementine produced by SPSS company on the contrary) machine leaning and data mining software based on JAVA environment with open source.

Defense Advanced Research Projects Agency (DARPA) is the department for organizing, coordinating, and managing national key scientific and technological projects and the research of military high-tech of the United States Department of Defense. It mainly takes in charge of the development and application of high-tech.

2.2. Principle for Finding Address Message of Data Frame of a Single Protocol

To conduct further identification, after the data frame of unknown mixed multi-protocol is separated into a single protocol and then the obtained clusters are evaluated as credible data frames using evaluation algorithm, the data frame of a single protocol needs to be split into point-to-point data frames according to address message. So, it is essential to find the location of the address message.

Suppose that the input protocol data frame of the algorithm displays the following characteristics:

(1) There are source address and destination address information (physical address or IP address) in the data frame;
(2) The number of bytes of address message is unknown. Assuming that the address message is not less than

2 bytes in length (it is possible that the address only has 1 byte).

In single protocol context, suppose that there are n (large enough) data frames of the protocol:

(1) The columns in the data frames are found out. In these columns, the species number of character is in the range of 1~K (default value: 256). K is a variable parameter;
(2) Suppose that there are S columns obtained, each of which is processed cyclically so as to select column pairs to compose a set R;

If more than w% (default value: 60%) of characters in S_i column appear in different locations of S_j and vice versa, S_i and S_j are included in the set R;

(3) The columns in set R are the candidate sets for the address columns;
(4) If there are more than 2 columns in the set R, the adjacent columns are spliced;
(5) The corresponding address couples are computed when w is set in the range of 10–90;
(6) The optimal solution is acquired by comparing the obtained address couples.

Parameters K and w can be set accordingly.

3. Computation of Address Message of Protocol Data Frame

(1) Data input is realized as follows: The segmented binary data frames are converted to corresponding hexdecimal formats. Based on this, a two-dimensional matrix of n rows and m columns is established by applying 2 bytes as the processing unit. In the matrix, each element is a hexdecimal character corresponding to two bytes and represented by character string.
(2) The object of minimum processing element is defined as TwoByte, which illustrates the following attributes:

```
Class TwoByte
  { /**the row of the byte*/
  public int row = 0;
  /**the column of the byte*/
  public int line = 0;
  /**the frequency of the byte appearing in the column*/
  public float frequence = 0f;
  /**the content of the byte*/
  public String twoByte = "";
  /**the serial number set of rows containing the byte in the column */
  public HashSet<Integer>alist = new HashSet<Integer>(); }
```

(3) Afterwards, a two-dimensional array of n rows and m columns of TwoByte was constructed. The content of every two bytes of the input data frame is assigned to the twoByte domain of the object of TwoByte and the row and column of the character string are recorded.
(4) The two-dimensional array of TwoByte is traversed cyclically to collect the appearing frequency of each character string in each column and the rows containing the character string. Then the appearing frequency is assigned to the num domain of TwoByte and the rows containing the character string are added in the alist set of TwoByte. In this way, the frequency of each character string in the column and the row numbers of the data frame containing the character string are obtained.
(5) Threshold values min_numOfperLine (default value is 1) and max_numOfperLine (default value is 256) are set to screen the species numbers (in the range of min_numOfperLine to max_numOfperLine) of character string of the column, which is applied as the input in next step.
(6) Suppose that S columns are obtained. Each column is cyclically processed and then the threshold value w% (default value: 60%) and the result set R are set. Such column pairs are selected and included in the set R.

If more than w% of characters in S_i column are found in different locations of S_j column and vice versa, S_i and S_j are included in the set R.

(7) The address couples in the set R are the columns containing the required candidate addresses. If there are more than 2 columns in R, the adjacent columns are spliced.
(8) To find out the location of the address more accurately, w is set in the range of 50~95. Then the optimal solution is obtained by comparing the address couples in the set R.

4. Results and Analysis

4.1. Experimental Conditions

In a simple intercommunication setting, a classification method for single protocol was put forward based on feature string. In the method, the n-gram segmentation and splice of data frame and feature selection algorithm are realized through JAVA code and cluster algorithm is conducted using k-means algorithm of the WEKA based on WIN7 operating system using the computer with Pentium(R) Dual-Core CPU E5800 and 2 G of memory.

4.2. Experimental Data-set

DARPA data-set provided by Lincoln Laboratory was applied as the data-set. The data-set is divided into two parts, namely inside.tcmpdump and outside.tcpdump. The former, which contains 1,130,829 dataframes was adopted as the experimental data-set. To verify the validity of the above algorithm, 2,000 ARP data frames and 10,000 TCP data frames were utilized. The experimental results are described below.

4.3. Experimental Results

Address message determination using 2,000 ARP data frames

Data input was performed as follows: Tthe front 42 bytes (that is, the data frames are least 42 bytes in length) of the 2,000 ARP data frames were adopted with 2 bytes as the minimum processing unit, and total 21 columns were obtained.

Table 1. Address Couples Extracted from ARP Data Frames.

W%	Candidate address columns	Address couples	Spliced address couples
50%	[1, 2, 3, 4, 5, 11, 12, 13, 14, 17, 18, 19]	(1,4)(1,12)(1,17) (2,5)(2,13)(2,18) (3,11) (4,1)(4,12) (4,17) (5,2)(5,13)(5,18) (11,3) (12,1)(12,4)(12,17) (13,2) (13,5)(13,18) (14,19) (17,1)(17,4)(17,12) (19,14) (18,2) (18,5)(18,13)	[1 2,4 5];[1 2,12 13];[1 2,17 18] [4 5,12 13]; [4 5,17 18] [12 13,17 18] [17 18 19,12 13 14]
60%	[1, 2, 4, 5, 12, 13, 14, 17, 18, 19]	(1,4)(1,12)(1,17) (2,5)(2,13)(2,18) (4,1)(4,12)(4,17) (5,2) (5,13)(5,18) (12,1)(12,4)(12,17) (13,2)(13,5)(13,18) (14,19) (17,1)(17,4)(17,12) (19,14) (18,2)(18,5)(18,13)	[1 2,4 5];[1 2,12 13];[1 2,17 18] [4 5,12 13]; [4 5,17 18] [12 13,17 18] [17 18 19,12 13 14]
70%	[1, 2, 4, 5, 12, 13, 14, 17, 18, 19]	(1,4)(1,12)(1,17) (2,5)(2,13)(2,18) (4,1)(4,12)(4,17) (5,2) (5,13)(5,18) (12,1)(12,4)(12,17) (13,2)(13,5)(13,18) (14,19) (17,1)(17,4)(17,12) (19,14) (18,2)(18,5)(18,13)	[1 2,4 5];[1 2,12 13];[1 2,17 18] [4 5,12 13]; [4 5,17 18] [12 13,17 18] [17 18 19,12 13 14]
80%	[1, 2, 4, 5, 12, 13, 14, 17, 18, 19]	(1,4)(1,12)(1,17) (2,5)(2,13)(2,18) (4,1)(4,12)(4,17) (5,2) (5,13)(5,18) (12,1)(12,4)(12,17) (13,2)(13,5)(13,18) (14,19) (17,1)(17,4)(17,12) (19,14) (18,2)(18,5)(18,13)	[1 2,4 5];[1 2,12 13];[1 2,17 18] [4 5,12 13]; [4 5,17 18] [12 13,17 18] [17 18 19,12 13 14]
90%	[1, 2, 4, 5, 12, 13, 14, 19]	(1,4)(1,12) (2,5)(2,13) (4,1)(4,12) (5,2)(5,13) (12,1)(12,4) (13,2)(13,5) (14,19) (19,14)	[1 2,4 5];[1 2,12 13] [4 5,12 13]

Table 2. Comparison of Addresses Determined Based on ARP Data and Actual Data.

	Destination MAC address	Source MAC address	Source MAC address column	Destination MAC address column
Identified data	2 3 4 5	8 9 10 11	24 25 26 27	34 35 36 37
Actual data	0 1 2 3 4 5	6 7 8 9 10 11	22 23 24 25 26 27	32 33 34 35 36 37

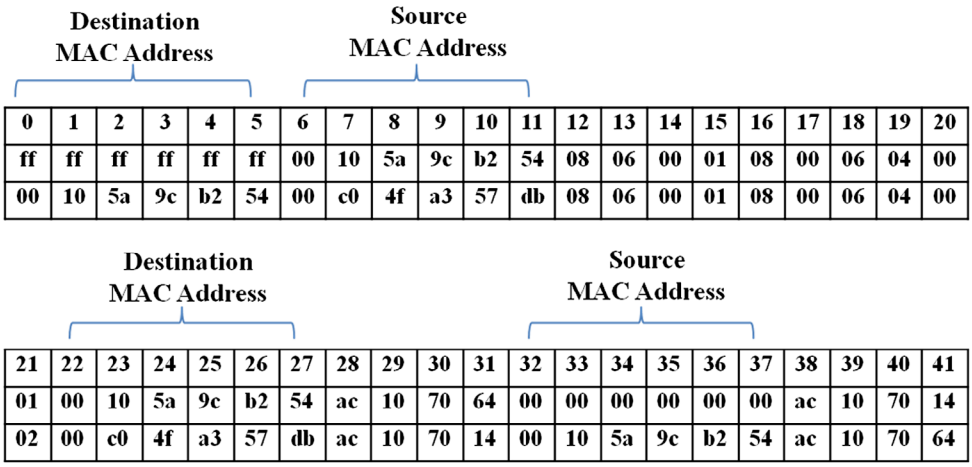

Figure 1. Address Messages of Two ARP Data Frames.

min_numOfperLine=1 and max_numOfperLine=256. The results with w varying from 50 to 90 are illustrated in Table 1 (column number is started from 0).

The spliced address couples in Table 1 indicate that 1 2, 4 5, 12 13, and 17 18 in the program are address columns from column 0. They correspond to columns including 2 3 4 5, 8 9 10 11, 24 25 26 27, 34 35 36 37 in the input data.

4.4. Result Analysis

The results in Table 1 demonstrate that the address columns of ARP data frames determined using the above algorithm are 2 3 4 5, 8 9 10 11, 24 25 26 27, and 34 35 36 37. Figure 1 illustrates two ARP data frames. According to the format of ARP data frame, the figure indicates that the column of 0 1 2 3 4 5 is the destination MAC address, column 6 7 8 9 10 11 is the source MAC address, 22 23 24 25 26 27 is the source MAC address column, 28 29 30 31 is the IP address column of the sender, 32 33 34 35 36 37 is the destination MAC address column, and 38 39 40 41 is the IP address column of the receiver. By analyzing the structure of ARP data frames and comparing the address columns determined and the actual address columns of ARP data frames in Table 2, the correctness of the experimental results was verified.

The comparison of the data in Table 2 indicates that though not all the address columns are found through the algorithm, 2/3 of an address column can be identified for each determined address field. Then by processing the address message, the determined address message is sufficient for separating these data frames to point-to-point data.

4.5. Address Message Determined Using 10,000 TCP Data Frames

Data was input first, the front 60 bytes (that is, the data frames have 60 bytes at least) of the 10,000 TCP data frames were utilized with 2 bytes as the minimum processing unit. A total of 30 columns were obtained.

min_numOfperLine=1 and max_numOfperLine=256. The value of w was in the range of 50~90. The results are shown in Table 3.

As illustrated by the spliced address couples in Table 3, 0 1 2, 3 4 5, 13 14, and 15 16 in the program are address columns, which correspond to columns 0 1 2 3 4 5, 6 7 8 9 10 11, 26 27

Table 3. Address Couples Extracted from TCP Data Frames.

W%	Candidate address column	Address couple	Spliced address couple
50%	[0, 1, 2, 3, 4, 5, 13, 14, 15, 16]	(0,3) (1,4) (2,5) (3,0) (4,1) (5,2) (13,15) (14,16) (15,13) (16,14)	[0 1 2,3 4 5] [13 14,15 16]
60%	[0, 1, 2, 3, 4, 5, 13, 14, 15, 16]	(0,3) (1,4) (2,5) (3,0) (4,1) (5,2) (13,15) (14,16) (15,13) (16,14)	[0 1 2,3 4 5] [13 14,15 16]
70%	[0, 1, 2, 3, 4, 5, 13, 14, 15, 16]	(0,3) (1,4) (2,5) (3,0) (4,1) (5,2) (13,15) (14,16) (15,13) (16,14)	[0 1 2,3 4 5] [13 14,15 16]
80%	[1, 2, 4, 5, 13, 14, 15, 16]	(1,4) (2,5) (4,1) (5,2) (13,15) (14,16) (15,13) (16,14)	[1 2,4 5] [13 14,15 16]
90%	[13, 14, 15, 16]	(13,15) (14,16) (15,13) (16,14)	[13 14, 15 16]

Table 4. Comparison of Addresses Determined Using TCP Data Frames and Actual Data.

Address name	Determined data column	Actual data column
Destination MAC address	0 1 2 3 4 5	0 1 2 3 4 5
Source MAC address	6 7 8 9 10 11	6 7 8 9 10 11
IP address of sender	26 27 28 29	26 27 28 29
IP address of receiver	8 39 40 41	8 39 40 41

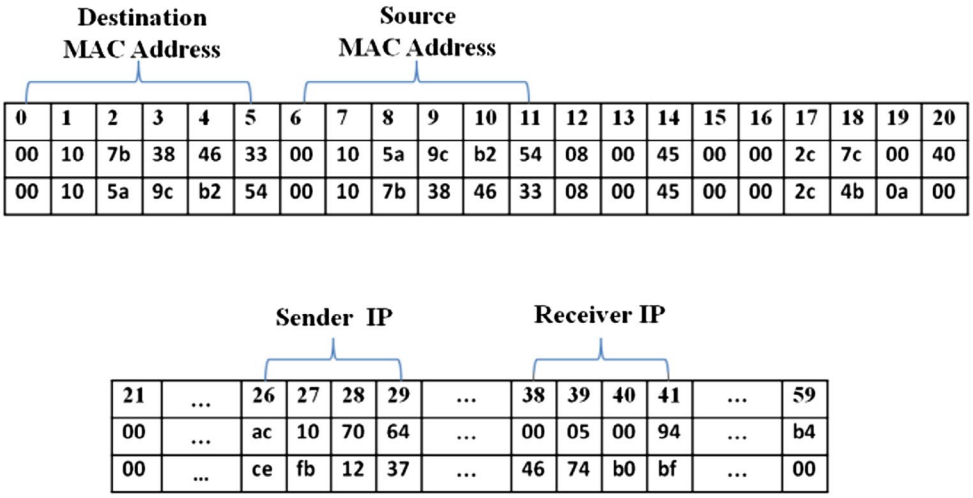

Figure 2. Address messages of two TCP data frames.

28 29, 38 39 40 4128 29, and 38 39 40 41 in the input data. Two TCP data frames are demonstrated in Figure 2. According to the format of TCP data frame, it can be seen that the column 0 1 2 3 4 5 is the destination MAC address, 6 7 8 9 10 11 is the source MAC address, 26 27 28 29 is the IP address column of the sender, and 38 39 40 41 is the IP address column of the receiver.

The address columns of actual TCP data frames and the identified address columns using the method are listed in Table 4. By comparing the data in Table 4, the authors found that all the address columns identified using the algorithm is the address columns of TCP data frames. These determined address columns can be used as the basis for separating the data frame to point-to-point data.

5. Conclusions

On the basis of separating unknown mixed multi-protocol into single protocols, an approach for searching the address message of unknown single protocol was proposed. By using the method, the address message of the unknown single protocol can be found out efficiently, so as to provide basis for splitting single protocol to point-to-point data. By doing so, the unknown mixed multi-protocol can be identified. Analysis on ARP data frames verified that the address columns determined by using the method matched with the more than 2/3 of actual address columns. For TCP data frames, the determined address columns completely matched the actual ones. The obtained address columns can be applied as the basis for splitting the data frame into point-to-point data. In this way, more accurate protocol identification can be realized and redundant features can be reduced further.

Acknowledgement

This work is supported by the Science and Technology Development Fund of China Academy of Engineering Physics (Grant No. 2012A0403021), NSAF Joint Found (Grant No. U1230106) and the Development Program of China National information security (Grant No.2013F098).

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Zheng Jie received Diploma and Ph.D. degrees from University of Electronic Science and Technology respectively. He is currently an associate professor in Chongqing College of Electronic Engineering. His current research interests include software, computational methods and web mining. He has authored or co-authored more than 20 academic papers and books.

Li JianPing, received a B.S. degree in Applied Mathematics from Chongqing University, Chongqing, China, in 1986, an M. Eng. degree in Software Engineering and an M. Sci. degree in Computational Mathematics from the Graduate School of Xi'an Jiaotong University, Xi'an, Shanxi, China, in 1989, and a Ph.D. degree in Computer Science from the Graduate School of Chongqing University, Chongqing, China, in 1998. His current interests include wavelet theory and applications, fractal, image processing, pattern recognition, information security.

References

Chen, C.C., & Wang, S.D. (2013). An efficient multicharacter transition string-matching engine based on the Aho-Corasick algorrithm. *ACM Transactions on Architecture and Code Optimization, 10*, 1–22.

China internet network information center. (2012). Report on the development situation of China Mobile Internet. Author.

He, W., Wang, R.G., & Zha, Q.M. (2010). A Novel Fast Moving Algorithm for Single Pattern Matching. *Journal of Hefei University of Technology, 33*, 665–669.

Kim, M.S., Won, Y.J., & Hong, J.W.K. (2005). Application-level traffic monitoring and an analysis on IP networks. *ETRI Journal, 27*, 22–42.

Kang, H.J., Kim, M.S., & Hong, J.W.K. (2003). *A method on multimedia service traffic monitoring and analysis// self-managing distributed systems*. Heidelberg, Germany: Springer.

Li, H.W., Wang, X.W., & Wang, P.Q. (2007). Ethernet protocol identification algorithm based on pattern matching. *Computer Engineering and Applications, 43*, 143–145.

Liu, R.T., Huang, N.F., & Chen, C.H. Kao C.N. (2004). A fast string-matching algorithm for network processor-based intrusion detection system. *ACM Transactions on Embedded Computing Systems, 3*, 614–633.

Liu, J.X. (2010). *The design for a real-time P2P traffic detection system based on DPI and DFI*. Qinhuangdao: Yanshan University.

Schiller, A.C., Binkley, J., & Harley, D. (2009). Botnets: The killer web app. [S.l.]: Syngress.

Sen, S., Spatscheck, O., & Wang, D.M. (2004a). Accurate, scalable in network identification of P2P traffic using application signatures. Proc of the 13th International World Wide Web Conference, 512–521.

Sen, S., Spatscheck, O., & Wang D. (2004b). *Accurate, scalable in-network identification of p2p traffic using application signatures*. Proceedings of the 13th International Conference on World Wide Web. New York: ACM. 512–521.

Van der Merwe, J., Caceres, R., & Chu, Y. Sreenan Cormac, J. (2000). Mmdump: A tool for monitoring Internet multimedia traffic. *ACM SIGCOMM Computer Communication Review, 30*, 48–59.

Wang, Y., Zhang, N., & Wu, Y. et al. (2013). *Protocol specification inference based on keywords identification//advanced data mining and applications*. Zhejiang, China. Springer Berlin Heidelberg, 443–454.

Wang, Y. Yun X., Shafiq, M. Zubair., Wang, L., Liu, Alex X., Zhang, Z., Yao D., Zhang, Y., Guo, L. (2012). *A semantics aware approach to automated reverse engineering unknown protocols*. ICNP 2012: 20th IEEE International Conference on Network Protocols, Austin, TX, USA: IEEE. 1–10.

Wang, Y., Wu, Y.M., Li, F., & Zhang, N. (2015). Protocol identification association analysis in mobile network environment. *Application Research of Computers, 32*, 243–248.

Zhu, S.Y. (2008). *The study on protocol identification technology*. Changsha: National University of Defense Technology.

Zhu, J.J. (2012). YE M. Multi-pattern Matching and Application of Improved Algorithm to Protocol Identification. Video. *Engineering, 36*, 60–63.

Zhang, Z.Y., Ye, W.C., & Chen, Y.H. (2010). Technology of stateful inspection based on the multi-pattern matching. *Electronic Measurement Technology, 33*, 98–101.