Detecting Technology of Network Storage Covert Channel Based on OPTICS Algorithm

Linkai Huang[†], Linna Zhou[*],and Yunbiao Guo[ξ]
[†]University of International Relations, Beijing, China
E-mail: huanglk@uir.edu.cn Tel: +86-18810168127
[*] University of International Relations, Beijing, China
E-mail: zhoulinna@tsinghua.edu.cn Tel: +86-10-62861171
[ξ]Beijing Institue of Electronics Technology and Application, Beijing, China
E-mail:gybgnx@sina.com Tel: +86-13910016672

Abstract— **Nowadays, with our enhanced technology, network security has gradually been valued by more and more people. Network covert channel, as an important area of network security, was put forward these years. On the one hand, covert channels provide new and safe communication environment for network communication. However, on the other hand, some illegal persons are exploited to spread viruses, trojans, and so on. Therefore, the research on covert channels is particularly important. According to the resource attributes, network covert channels can be divided into two parts, storage-based channels and timestamps-based channels. The paper will base on the storage-based covert channels, and a new detection method based on clustering algorithm will be proposed. According to the clustering analysis of the values of various parts of the data flow packet, the clustering results are graphically displayed. Determine whether there is a covert channel in the packet. The experimental results show that the detection technology has the advantages of high accuracy, simple algorithm, and intuitive result images, and achieves the desired ideal results.**

I. INTRODUCTION

With the continuous development of science and technology, the Internet has entered millions of households. The Internet has become an important tool for communication between people. While facilitating the daily life of people, the Internet has also become the target of frequent attacks by hacking organizations. Therefore, cyber security has gradually been taken seriously by people.

The covert channel hides secret information over the Internet to a redundant field of the network space protocol. It uses data packets as carriers of secret data to carry out covert transmission in a network space channel. The study of covert channels belongs to the category of information hiding in network information security. Its existence has two aspects: on the one hand, it provides new and more secure communication channels for network communications; on the other hand, it is also used by some lawbreakers, who spread viruses, trojans, and other activities that endanger safety. Therefore, in recent years, the study of network covert channels has been highly valued by the military and security departments of the country and has attracted the attention of many researchers. The paper will base on the storage-based covert channels, and a new detection method based on clustering algorithm will be proposed.

II. NETWORK COVERT CHANNEL

A. Definition of Network Covert Channels

In cyberspace, channels are divided into public channels and covert channels. The public channel is a legitimate stream of information that people use in their daily lives. The covert channel uses information hiding algorithm to embed some secret information in the redundancy field of the protocol of the data packet. It passes data packets in a legitimate stream of information to allow covert transmission of secret information under the cover of the open channel.

In terms of computer security, the covert channel is a kind of computer security attack. It creates a capability that enables the communication between two parties to transmit information without allowing of computers' security policy. In 1973, Lampson defined a covert channel as "neither used for information transmission nor used for communication" to distinguish it from the public channel accessed by COMPUSEC [1].

The covert channel can be hidden in the access control mechanism of the high security operating system, so it is called "covert" channel. The reason why it can be hidden is that it does not require the use of legal instructions of the computer system, such as read and write transmission mechanisms, so the hardware security mechanism based on high security operating system cannot detect it. However, in real systems, covert channels are extremely difficult to install and are often detected when monitoring system performs. In addition, there are two other characteristics of covert channel transmission: on the one hand, it has a low signal-to-noise ratio and a low data rate (a few bits per second); on the other hand, it can be manually removed from a high-security system through the great covert channel analysis strategy [2].

Covert channels and hiding information through public channels are completely different methods, and sometimes we also confuse the two. The latter is to use uncomplicated algorithms such as steganography, and to conceal the hidden information into legitimate carriers and to perform legal

transmissions over open channels. Obviously, the information hiding of the public channel by using a method such as steganography does not correspond with the definition of the operation principle of the covert channel [3].

B. Implementation Technology of Covert Channels

There are many ways to classify covert channels. According to shared resource attributes, the most widely used classification method is to classify covert channels into storage-based covert channels and time-stamp-based covert channels. The main research in this paper is the storage-based covert channel.

The storage-based covert channel mainly uses the redundancy field of the protocol control part of the data packet to load the hidden information, to form a new data packet, and to transmit in the network space channel, instead of directly using the conventional fields in the protocol for concealment.

The fields that are often used to store covert channels include:

①Unused IP header fields, such as TOS (Type of Service), DF (Don't Fragment) field in the flag, as shown in Figure 1;

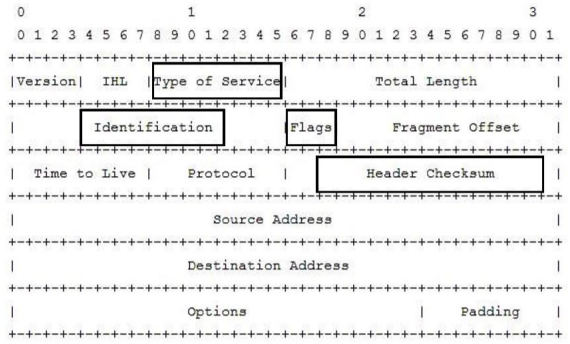


Fig.1 IP header format[1]

②IP header extension and padding fields, such as ID (Identification), and Header Checksum, as shown in Figure 1;

③Flag bit fields of the TCP header, such as Sequence Number, Checksum, and Timestamp in TCP Options, as shown in Figure 2.

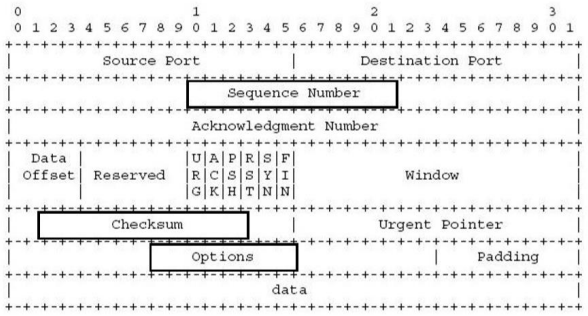


Fig.2 TCP header format[2]

[1] Jon PoStel. RFC791-Internet Protocol Specification. 1981.

[2] Jon PoStel. RFC793-Transmission Control Protocol Specification. 1981.

III. OPTICS CLUSTERING ALGORITHM

Cluster analysis, also known as group analysis, classifies certain samples or a specific indicator according to the principle of "reunion of objects". It divides the sample data into several clusters and makes each cluster have its own specific multivariate statistical analysis method [4].

The OPTICS clustering algorithm is a density-based clustering algorithm. Its full name is Ordering Points To Identify the Clustering Structure [5]. The purpose of this algorithm is to cluster the data in the sample according to the density distribution, and then obtain clusters of different densities. In other words, after analysis and processing of this algorithm, clustering with any density distribution can theoretically be achieved [6].

Before introducing the process of the OPTICS clustering algorithm, we first defined two parameters, namely the radius ε, and the minimum number of points $MinPts_{th}$, and defined the following three concepts:

Core point: The number of points included in the radius of a point is not less than the minimum number of points, then this point is the core point;

Core distance: For the core point P, the distance from its nearest point to the $MinPts_{th}$;

Reachable distance: For the core point P, the reachable distance from any point O to P is defined as the distance from O to P or the core distance of P [5].

The OPTICS algorithm flow chart shown in Figure 3.

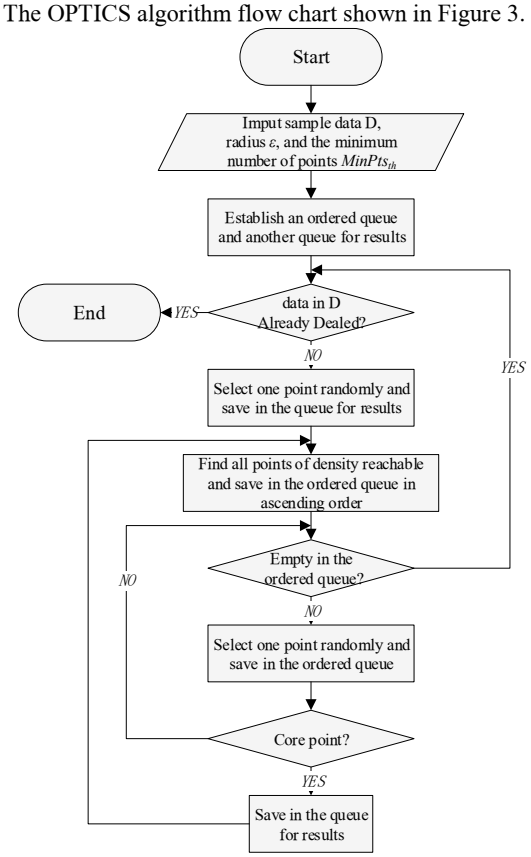


Fig. 3. Flow of the OPTICS algorithm

IV. NETWORK STORAGE COVERT CHANNEL DETECTION BASED ON CLUSTERING ALGORITHM

C. Brief Introduction

The essence of the storage-based covert channel is to hide the secret information into existing, redundant field values in the network protocol and transmit them, such as the TOS field, the DF field, the ID field, Checksum field in the IP header field and serial number field, checksum field, etc. in the TCP header mentioned above. Because each field has a different meaning, after the "tampered" packet will be abnormal. For example, the length in the original data packet, the time interval between the data packets, and the value of the serial number in the TCP packet are all random. However, in the "tampered" data packet, the data packet length, the time interval between the data packets, and the value of the serial number in the TCP packet will show certain rules [7].

According to the idea of clustering algorithm, if some data in a large amount of data presents a certain rule, after applying a clustering algorithm, data with the same rules will be clustered into a cluster. The three parts of the packet length, time interval, and sequence number in the TCP packet in the normal data flow packet are discrete random and irregular. Therefore, when we cluster the three parts of the normal flow, we will obtain separate scatter plots. However, for the flow packets embedded in the hidden information, the three parts of the packet length, the time interval, and the sequence number in the TCP packet show certain rules. Therefore, in the cluster analysis of these three parts of the flow packet embedded with hidden information, several points appear to be clustered into one cluster. That is, the data packets in this cluster contain the hidden information with high probability. Among other outliers, there is a large probability that data comes from normal packets [8].

D. Algorithm flow

The flow chart of detection of network storage covert channels based on clustering algorithm is shown in Figure 4.

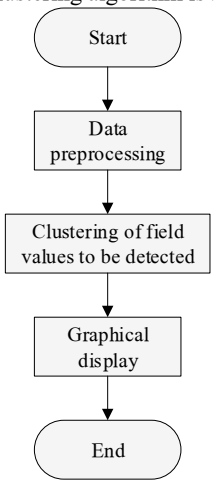


Fig. 4. Flowchart for network storage hidden channel detection based on clustering algorithm.

Our use of OPTICS clustering algorithm to detect the network space storage hidden channel needs to be divided into three steps: First, the data is pre-processed, then the values of the detected fields are clustered, and finally displayed graphically.

a. Data Preprocessing

Data preprocessing is to initially filter the captured data packets and store the numerical information of the fields to be detected into the database. This will facilitate the use of OPTICS for clustering later. For packets captured at high flow levels, Gb levels may be reached in a few seconds, so filtering for data is essential. Since broadcast packets generally do not contain hidden information, broadcast packets are filtered out during initial screening. For the reserved data packets, the data of packet length, time interval, and sequence number information of the TCP packet are extracted and stored in the corresponding column of the database. This facilitates later clustering of data.

b. Clustering of field values to be detected

For the field to be detected, there are three parts: packet length, time interval, and serial number of TCP packet. In the following, the packet length will be taken as an example to introduce the clustering process. The remaining two parts of the clustering process are the same as the packet length [9].

First, we enter the packet length, which is all the sample data. Then we specify the Euclidean distance radius ε and the minimum number of points *MinPts*. Secondly, according to the OPTICS clustering algorithm, we calculate the reachable distance for all data in the packet length and divide the cluster by the reachable distance. That is, each core point is clustered with its distance from the point of ε. We then store the reach of the core point in each cluster into a new database. In accordance with the output order calculated by the OPTICS algorithm, we renumbered the data and entered the data packet's original number and the corresponding packet's packet length data into the database at the same time. The database was used as the data source for the later graphical display.

c. Graphical display

We select the new database saved after the OPTICS algorithm analysis, and use the sequence number output by the OPTICS algorithm as the X axis, and graphically display the packet length, time interval, and sequence number data in the TCP packet as the Y axis.

For the packet length data, if the result of the graphic display is that the packet lengths are all aggregated at a certain value, there is a covert channel with a large probability in the flow. For the time interval, after it is clustered, if the result of the graphical display is that the time interval, the packet length after clustering, and the number of data packets are basically the same, this indicates that there is an implicit channel with a high probability in the flow. For the serial number in the TCP packet, if the result of the graphic display is that the sequence number values are all gathered in 0-128, and there are certain rules at the same time, it means that there is an implicit channel in the flow with a large probability.

V. EXPERIMENTS AND DATA ANALYSIS

To test the effectiveness of this detection technique, this experiment will construct two batches of data flow. The two batched of data flow are detected in our school during the Internet peak period. During the period, we use specified software we developed (referred as SENDER) to build covert channel and transmit secret information. SENDER can hide information in Identification IP packets and Sequence Number in TCP packets when packets flow in the covert channel.

During the experiment, we use SharpPcap which is a network packet capture framework for .NET environments based on the well-known pcap/WinPcap library and provides capture, injection, analysis, and build capabilities to monitor and catch the flow in the environment of the current internet [7].

The number I (Figure 5) is the normal flow of the public channel, and the number II (Figure 6) is the common flow of the covert channel and the public channel. The experiment will analyze and graphically display the OPTICS algorithm for the packet lengths, time intervals, and sequence numbers in the TCP packets for these two batches of flow and compare them.

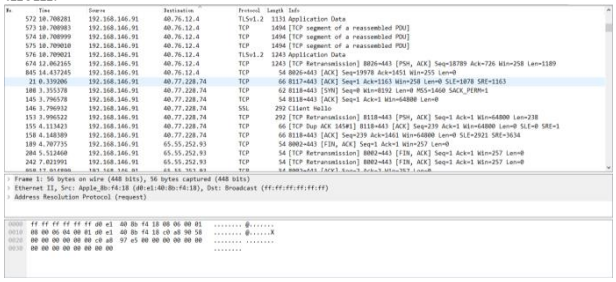


Fig. 5. Flow I

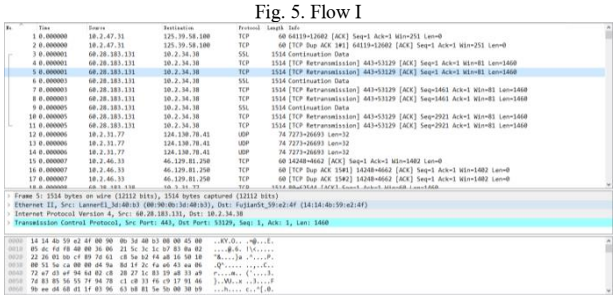


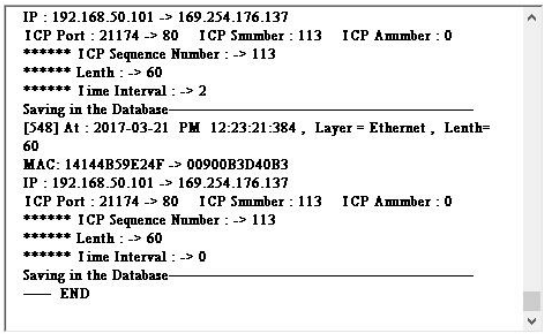
Fig. 6. Flow II

A. Experiment content

a. Data Preprocessing

Since the flow of No. 1 is normal flow, once the data is pre-processed, the amount of data may be too small to complete the follow-up experiment. Therefore, the same number of data packets with flow No. II after data preprocessing are randomly selected for flow No. I.

Data preprocessing for flow II is shown in Figure 7.



```
IP : 192.168.50.101 -> 169.254.176.137
ICP Port : 21174 -> 80   ICP Snumber : 113   ICP Anumber : 0
****** ICP Sequence Number : -> 113
****** Lenth : -> 60
****** Time Interval : -> 2
Saving in the Database----------
[548] At : 2017-03-21 PM 12:23:21:384 , Layer = Ethernet , Lenth=
60
MAC: 14144B59E24F -> 00900B3D40B3
IP : 192.168.50.101 -> 169.254.176.137
ICP Port : 21174 -> 80   ICP Snumber : 113   ICP Anumber : 0
****** ICP Sequence Number : -> 113
****** Lenth : -> 60
****** Time Interval : -> 0
Saving in the Database----------
---- END
```

Fig. 7. Data preprocessing

We open the database for viewing. The flow data feature of I is not obvious, as shown in Figure 8, and the flow data feature of II is obvious, as shown in Figure 9.

ID	len	Snumber	time
112	60		5
113	56		0
114	60		0
115	60		1
116	60		0
117	60		2
118	56		95
119	56		0
120	56		0
121	288	2473133836	12
122	60	2452546715	39
123	144	2569558359	3
124	1494	2473134070	3
125	1494	2473135510	0
126	1494	2473136950	0
127	1494	2473138390	0
128	688	2473139830	0
129	56		49
130	56		0
131	56		0
132	56		0
133	56		0
134	56		0

Fig. 8. Flow I

ID	len	Snumber	time
1	60	113	0
2	60	113	0
3	60	113	0
4	60	113	0
5	60	113	0
6	60	113	0
7	60	119	2
8	60	119	0
9	60	119	0
10	60	119	2
11	60	119	0
12	60	119	0
13	60	119	0
14	60	119	0
15	60	119	0
16	60	119	3
17	60	119	0
18	60	119	0
19	60	119	2
20	60	119	0
21	60	119	0
22	60	119	2
23	60	119	0
24	60	119	0

Fig. 9. The result of the pretreatment of flow data II

b. Clustering of field values to be detected

We analyze the packet lengths of Flow I and Flow II, as shown in Figure 10 and Figure 11.

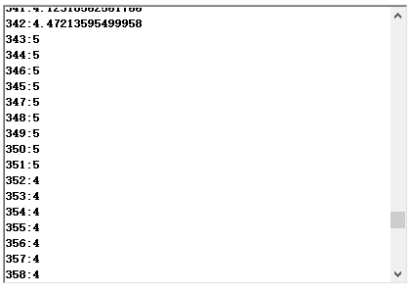


Fig. 10. Cluster analysis of packet length for flow I

Fig. 11. Cluster analysis of packet length for flow II

We perform cluster analysis on the time interval of flow I and II and the sequence number in TCP packets. The principle is the same as the packet length and is not shown here.

c. Graphical display

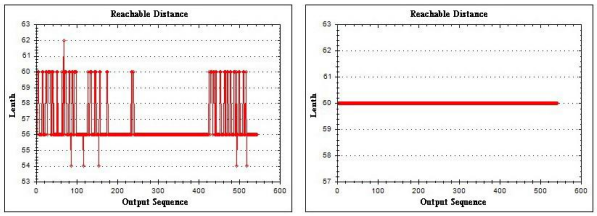


Fig. 12. Graphical comparison of packet length after clustering of flow I (left) and flow II (right)

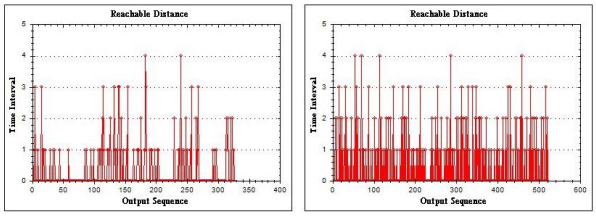


Fig. 13. Graphical comparison display of the clustering of time intervals between flow I (left) and flow II (right)

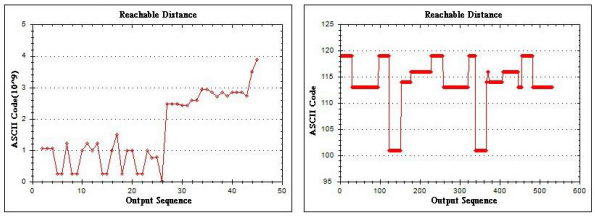


Fig. 14. Graphical comparison of clustering of sequence numbers in TCP packets for flow I (left) and flow II (right)

B. Analysis of experimental results

According to the graphical display, from the packet length and serial number of the TCP packet, we clearly found that flow number II is suspicious flow. From the right figure of Figure 12, the packet lengths of Flow II are all clustered around 60, and according to the right figure of Figure 14, the sequence numbers of the TCP packets of Flow II show a very regular pattern, and are all clustered between 0-128. However, the packet length of Flow I in the left figure of Figure 12 is not completely clustered around a certain value, but rather the packet length of 55 and 60 packets. For the serial number of the TCP packet for flow I in Figure 14, it shows that each point is isolated, and the sequence number has a large magnitude, reaching 109. Therefore, according to the cluster length analysis of the packet length and the TCP packet sequence number, it can be detected to a great extent that there is a storage covert channel in the flow number II.

For the time interval [10], as shown in Figure 13, the time intervals are all clustered between 0-2 and are irregular. It is difficult to judge whether there is a covert channel on the surface. However, we carefully observe the number of packets on the horizontal axis in the left figure of Figure 13 and compare it with the number of packets in the left figure of Figure 12. We found that in the flow I, the number of packets participating in the time interval clustering is obviously less than the number of packets participating in the packet length clustering, which is about 200 packets less. This shows that the time interval in the 200 packets is not in this cluster in the figure, that is, the time intervals in the 200 packets are excluded as outliers by the clustering algorithm. We compare this with Figure 13 on the right and Figure 12 on the right. In flow II, the packets participating in the time interval clustering are basically equal to the number of packets participating in the packet length clustering. Therefore, based on the comparison of the time interval and packet length packet, we can also detect the covert channel stored here.

C. Experiment Summary

According to the above three detection indicators, the packet length and the serial number of the TCP packet are the most easily detected indicators. The time interval needs to be compared with the result of packet length clustering to detect whether there is a covert channel. The process is cumbersome and not as intuitive as the result of the packet length and TCP packet sequence number.

The experiment put forward a way to detect covert channel in the big environment of internet based on OPTICS clustering algorithm. We also analyze the covert channel in three dimensions, lengths, time intervals, and TCP Sequence numbers. According to the results of experiment, we can find that it is obvious to find a covert channel by the graphs above.

In the meantime, we performed multiple experiments to test the accuracy of the model, the reliability of the method and the performance of the algorithm.

In one experiment, the secret information sent by SENDER contained 310 characters, and we detected 309 characters. The accuracy of the model is up to 99%.

As for the reliability of the method, it depends more on purity of the detected internet environment. If we detect in

pure environment which contains covert information only, the detection rate is 100%. However, in high traffic environment, the detection rate is 94%, lower than that in pure environment, but it is also predominant.

As for the performance, it depends more on the magnitude of the flow. The size of the flow we detected in the experiment above is about 40GB. We used many hours to preprocess the data in the first time. Afterwards, we used multithreading to optimize performance, and the time of the whole process was decreased to less than one hour.

VI. CONCLUSIONS

This paper focuses on the research and implementation of storage-based covert channel detection technology, and proposes a storage-based covert channel detection method based on OPTICS clustering algorithm. The detection method first preprocesses the data, and then performs clustering calculations on the three indicators (package length, time interval, and sequence number of the TCP packet), and graphically displays the data obtained by the clustering calculation. We can basically detect the storage covert channel by some features of the image. The experimental results show that this method is suitable for the detection of storage-based network covert channels and has an ideal effect. This method has the advantages of high accuracy, simple algorithm, and intuitive result images. At the same time, the detection algorithm of network storage covert channel based on OPTICS algorithm proposed in this paper has important theoretical and practical significance for the research of network security and has a certain application prospect.

ACKNOWLEDGMENT

This work was finally supported by National Natural Science Foundation of China (Grant No. U1536207 & No. U1636212).

We would like to express our gratitude to all those who have helped us during the writing of this thesis. Linkai Huang gratefully acknowledge the help of my tutor Professor Linna Zhou and the project manager Professor Yueqin Liu and Professor Yansen Zhou. We do appreciate their patience, encouragement, and professional instructions during our thesis writing. Also, we would like to thank Xiangchen Liu, Yuedong Fu and Yifeng Fang, who kindly gave us a hand during the whole project. We are also deeply indebted to Lin Hou in translation studies for her direct and indirect help to us. Special thanks should go to our friends who have put considerable time and effort into their comments on the draft. Last but not the least, our gratitude also extends to our families who have been assisting, supporting and caring for us all of our lives.

REFERENCES

[1] B.W.Lampson, "A Note on the Confinement Problem," *Communications of the ACM*, vol. 16(10), pp. 613-615, October 1973.

[2] Y.J.Wang; J.Z.Wu; H.T.Zeng; L.P.Ding; X.F.Liao, "Covert channel research," *Journal of Software*, vol. 21(09), pp. 2262-2288, September 2010.

[3] Z.Wang, "Covert Channel Technology In Network Data Communication," *Information & Communications*, vol. 08, pp. 228-229, August 2016.

[4] J.Yuan; T.Wang, "Detection Algorithm of Network Storage Covert Channel Based on Cluster Analysis," *Computer Engineering*, vol. 41(9), pp. 168-173, September 2015.

[5] L.K.Huang, "A Traffic Classification Algorithm Based On OPTICS Clustering," *China New Telecommunications*, vol. 08, pp. 38-39, April 2017.

[6] G.X.Fu, "Covert Timing Channel Detection Method Based on Random Forest Algorithm," *IEEE Beijing Section Sichuan Institute of Electronics*. vol. 07, pp. 206-212, October 2017 [Proceedings of 2017 17th IEEE International Conference on Communication Technology (ICCT 2017)].

[7] Y.D.Fu, "A Combined Transform Hidden Channel Based On Network Protocol," *Telecom World* vol. 06, pp. 128, March 2017.

[8] X.J.Wu, *Research on Network Covert Channel Detection Technology.* Jiangsu Province, Nanjing University of Science and Technology, 2012.

[9] Y.Shen, *Research on Network Protocol Hidden Channel Detection and New Construction Scheme.* Hefei Province, University of Science and Technology of China, 2017.

[10] G.Z.Ji; Q.F.Tan, "Research on Implementation and Detection Method of Covert Channel Based on Data Packet Time Interval," *Communications Technology*, vol. 51(01), pp. 189-194, January 2018.