3D Information Visualization Interactive Technology and Its Application in Network Security

Dehui Ma$^{1,*}$ and Qiang Zhang$^1$

$^1$Shandong Yellow River Information Center, Jinan, Shandong, China, 250013

*Corresponding author e-mail: madehui@yrcc.gov.cn

Abstract. With the increase of network complexity, network security incidents are showing a trend of rapid growth. Aiming at the low efficiency of large-scale network security data visual analysis, this paper proposes a heterogeneous tree network security data organization method, which improves the real-time performance of data analysis, and designs a three-dimensional multi-layer spherical space for large-scale networks Visualization model (TMSM). The experimental results show that this method has strong analytical capabilities for large-scale, high-dispersion network security data, and can effectively identify different types of intrusions and facilitate network forensics.

Keywords: 3D Information Visualization Interactive Technology, Anomaly Detection, Large-Scale Network, Network Security, Visual Analysis

1. Introduction
With the rapid growth of network security incidents, many companies and organizations will use host monitoring systems to detect anomalies and attacks to meet the supply and demand requirements of protecting network security. In the operation of network security, a lot of network user information data and network operation information will be generated. We use these data to achieve application user behavior patterns$^{[1-2]}$. As the complexity of the network increases, the longer it takes us to analyze data. Researchers try to use a kind of abstract data for visual representation to speed up the time of information data analysis. This visualization system can handle large-scale network security data$^{[3-4]}$. It also shows the network behavior more effectively. Currently, many visualization systems are used for two-dimensional visualization analysis of data sets. However, as the amount of data increases, two-dimensional visualization can no longer fully detect and analyze data sets, and it is easy to load too much useless information$^{[5-6]}$. When two-dimensional visualization cannot meet the demand, three-dimensional visualization system appears, and at the same time, it can provide users with big data and achieve data clarity and no confusion. Three-dimensional visualization can enhance the sense of space vision. In terms of network security, three-dimensional visualization can enhance special networks and improve network emergency response.

Based on network security, this paper proposes an easy-to-build tree model that can quickly retrieve information and improve interaction efficiency. At the same time, it can better display the characteristics of network activities and propose a new multi-layer spatial three-dimensional
visualization model. Through this model, users can effectively identify different types of intrusion problems and conduct network forensics.

2. Multi-layer spherical space visualization model
In order to solve the problem of mutual occlusion or too dense graphics in the 2D visualization method, which leads to the neglect of important information, a multi-layer spherical spatial visualization model (TMSM) based on 3D technology is proposed to visually analyze large-scale network security data. The model is composed of multiple spheres, each sphere corresponds to a network segment; the sphere embedded in each sphere represents the corresponding network node, and the color of the sphere maps the role of the sphere in the network, such as Web server and mail server. The coordinates of the sphere on the sphere are calculated using the IP address mapping algorithm.

TMSM analysis design includes three analysis modes: global analysis, relationship analysis and detailed analysis. The global analysis is mainly for the analysis of the popularity of the host; the relationship analysis is mainly for the analysis of the relationship between the hosts; the detailed analysis is mainly for the detailed analysis of the data at a certain moment.

(1) Global analysis
Define the heat function of node k at time \( t_i \) as:

\[
f_k(t_i) = \text{Num}(\text{Receive}(\text{IP}_k)_{t_i}) + \text{Send}(\text{IP}_k)_{t_i}
\]

In the formula, \( t_i, t_{i+1}, \cdots, t_{i+1}, t_{i+1} > t_i \) is discrete time points at equal intervals. The heat function describes the amount of data received or sent by the host node.

Assuming that the start time of the time zone is \( t_s \) and the end time is \( t_e \), the radius of the drawn sphere for the k-th node is defined as:

\[
R_k = \lambda_{\text{scale}} \sum_{t \in T(t_s, t_e)} f_k(t)
\]

In the formula, \( T(t_s, t_e) \) is the time period that the user is interested in; \( \lambda_{\text{scale}} \) is the scale adjustment parameter of the global visualization space. In the global analysis mode, the radius drawn by the sphere reflects the popularity of the network nodes.

(2) Relationship analysis
The connection between the spheres is used to indicate the connection between nodes, and different colors are used to indicate whether the current host is biased to initiate a connection or receive data. The relationship between the nodes is mapped to the width of the node line, which is defined as:

\[
W_{i,j} = \lambda_{\text{scale}} N_{i,j}(T(t_s, t_e))
\]

In the formula, \( N_{i,j}(T(t_s, t_e)) \) represents the number of logs of network node i and network node j in time period \( T(t_s, t_e) \).

The three-dimensional volume space is formed by the intersection of six planes. There are two planes perpendicular to the x-axis, y-axis and z-axis. If we want to see the data in a dense space, we need to stretch the six opposite planes. To determine such a space, the positions of the planes perpendicular to the x, y, and z axes are controlled separately, as shown in the two-dimensional cross-sectional view in the figure below. If we only move one surface, all the positions in the quadrant of the same surface will change. But if certain planes are specified to determine such a stretched area, the position change made will be local.

Use a number between 0 and 1 to indicate the position of the plane in a quadrant (0 and 1 define the size of the displayed scene). In general, if a face is set to a new position, the data in this quadrant will be updated by \( O(n) \). The algorithm in this chapter supports local-based updates and keeps the overall relative position unchanged. If we only focus on the data in the P size area, we only need to
update a local space like P, so that the amount of calculation P<<O(n). The data storage method is: use a binary tree to represent such a relative position, call the node of the tree the split surface, and use the split ratio to indicate its relative position. The value of the split ratio is the left side under the jurisdiction of the split surface from the parent node where the surface is located. The position of the node is higher than the value of the position between the left and right nodes.

When the size of the two left and right child areas managed by a parent node are equal, the division ratio is 0.5. The absolute position of the split surface can be obtained by traversing the entire tree from top to bottom according to the value of its split ratio. As shown in Figure Niu 1, $x_5=0.6$, and its absolute position is 0.78. If you want to calculate the absolute position of $x_4$, you can calculate it by the division ratio and absolute position of its parent node:

$$x_4^A = (x_5^A - x_4^A) \times x_4^R + x_4^A = (0.78 - 0.45) \times 0.4 + 0.45 = 0.582$$

The superscript A represents the actual position of the dividing surface, and the superscript R represents the value of its division ratio. This calculation is top-down, starting from the root node of the entire binary tree until the parent node of the split surface is calculated. Therefore, if such a binary tree has n nodes, the time required to calculate the actual position of the split surface is complicated. The degree is $O(\log n)$.

Each area is bounded by two split planes. All child nodes of a parent node exist in the space under the jurisdiction of the parent node. All nodes on the left are bounded by the leftmost node of its parent node and the split plane where the parent node is located. Bound; all nodes on the right are bounded by the parent node and the dividing plane where the rightmost node under the jurisdiction of the parent node is located.

3. Experimental results and analysis
Using C++ development language and OpenGL graphics API, this paper designs and develops a visual analysis system based on the multi-layer spherical space structure. The system interface can be divided into three parts.

3.1. Experimental results
In order to verify the effectiveness of the visualization method in this paper, the prototype system was tested on the experimental platform. The test data adopts the network security data set.

First, use the global analysis method to search all Netflow log data from April 7th to April 15th using the timeline, and visually analyze the characteristics of the data. Through the observation of the timeline, it was found that the Netflow log data had two extreme peaks at 12 o'clock on April 11th and 15:00 on April 15th. Especially at around 12 o'clock on April 11th, the data peak was the most obvious. Then move the timeline, choose the time between 12 o'clock and 13 o'clock on April 11th, observe the TMSM model, and find that there are 3 nodes with much larger radius than other nodes, which means that the amount of data received by this node is much higher than that of other nodes. After carefully observing the attributes of these nodes, it is found that these nodes are all Web servers, and they are all on the same network segment. In order to observe the data sent by the node, through detailed observation of the TMSM model, it was found that about 20 external network hosts had extremely large data transmission volume, so they were added to the watch list to focus their attention. Then using the relationship analysis method, all three web servers interacted with the 20 extranet hosts that were previously on the watch list. Finally, the detailed analysis method is used to search all FireWall log data from April 7th to April 15th using the timeline. Select the FireWall alarm data from 12:00 on April 11th to 13:00 on April 11th on the panel for dynamic visual display. By observing the TMSM model for a period of time, it can be found that there are long-term and huge amount of alarm messages between 3 intranet servers and more than 20 extranet servers, showing very obvious Ddos attack characteristics.

3.2. User analysis
In order to get users' feedback and evaluation on the TMSM system, 21 volunteers selected different performance methods according to their personal preferences. Three representative 2D and 3D visualization systems were selected respectively, namely Color Maps, Parallel Coordinate and Scatter Plots, and 6 common evaluation indicators of visualization systems were selected for evaluation. The choice of method is indefinite. For different 2D or 3D visualization systems, they are unified into 2D or 3D visualization methods in the table. The statistical results are shown in Table 1. It can be found from Table 1 that the 3D visualization method proposed in this paper is more interesting to users than other visualization methods.

**Table 1. User feedback list of prototype system.**

| Test items                          | User acceptance/% | 2D visualization method | 3D visualization method | Visualization method in this article |
|------------------------------------|-------------------|-------------------------|-------------------------|-------------------------------------|
| Performance network activity       | 17                | 33                      | 50                      |                                     |
| Reflect the network topology       | 11                | 27                      | 62                      |                                     |
| Convenient and practical interactive functions | 37                | 20                      | 43                      |                                     |
| Obtain abnormal behavior information | 24                | 38                      | 48                      |                                     |
| Reasonable layout and beautiful design | 42                | 14                      | 44                      |                                     |
| User interested                    | 34                | 31                      | 35                      |                                     |

4. Conclusion

On the basis of analyzing the characteristics of network security data, this paper proposes the use of heterogeneous tree structure for data organization, and proposes a multi-layer spherical space visualization model. Experimental results show that compared with traditional visualization methods, this 3D visualization method has great advantages in network topology performance and abnormal behavior recognition. Future work will further deepen the research topic of this article, mainly researching the visual analysis method of network security data based on sketches.

References

[1] Li Z, An Y, Hu Z A, et al. Preparation of two-dimensional flexible mno2/graphene thin film and its application in supercapacitor [J]. Journal of Materials Chemistry A, 2016, 4(27):1-10.

[2] Zheng K, Zhang X, Zheng Q, et al. Quality-of-experience assessment and its application to video services in lte networks [J]. IEEE Wireless Communications, 2015, 22(1): 70-78.

[3] Bohlin E. Foreign direct investment, information technology and economic growth dynamics in Sub-Saharan Africa [J]. Telecommunications Policy, 2020, 44(1): 101-113.

[4] Kuperman G J, Blair J S, Franck R A, et al. Application of information technology: Developing data content specifications for the Nationwide Health Information Network Trial Implementations.[J]. Wireless Networks, 2015, 21(3): 859-869.

[5] Ide, N, Okada, I, Kojima, K. Nutrition education and interactive communication technology - brief article [J]. International Journal of Neuropsychopharmacology, 2015, 18(2): 5489-5497.

[6] Thompson C C, Amaral G R, Campeão M, et al. Microbial taxonomy in the post-genomic era: Rebuilding from scratch? [J]. Archives of Microbiology, 2015, 197(3): 359-370.

[7] Song C, Xu Z, Zhang Y, et al. Dynamic hesitant fuzzy Bayesian network and its application in the optimal investment port decision making problem of "twenty-first century maritime silk road" [J]. Applied Intelligence, 2020, 3(2): 1-13.

[8] Yang B, Huang T, Li X. A time-series approach to measuring node similarity in networks and its application to community detection [J]. Physics Letters A, 2019, 383(30): 125-138.