Research on network topology generation algorithm based on node layout

LI Teng-fei*, LI Qiang, YU Xiang, WU Dai-yue

Electronic Engineering Institute，Hefei AnHui，China

ltfstudy@sina.cn

Keywords: Network Topology, Node Location, Node Contact, Layout Algorithm

Abstract: Network topology layout is one of the key technologies in the research of network topology structure. How to make the node overlap less and distribute evenly in the layout is an important and key problem in the research of network topological structure. Establishment of network topology generation algorithm based on node layout, design the node layout model, through node location layout and node linkage adjustment to achieve uniform layout, construction network topology generation algorithm, to achieves the effect from the physical topological structure to the logical topology beautiful display. The experiment shows that the network topology generation algorithm based on node layout has lower node coincidence ratio, point-edge crossover rate and better evaluation index, and the layout effect is better than other algorithms.

1. Introduction

Network topology map is an important means and method of network configuration, fault diagnosis, security management[1] and so on, which is the connection between the network topology and the network node. To the actual running network system, how to arrange the network nodes rationally according to the result of network topology discovery, and to display the network topology map correctly, clearly and aesthetically with the graphics processing technology, is not only the need of the user to understand and comprehend the network topological structure correctly, but also a hot issue in the network topology research.

Currently, the network topology layout methods that have been acquired include: 1) Simple Network structure layout[2]: Using the simple structure of tree[3] or ring to define the network, the network is distributed according to this kind of structure, and a strong network layout is obtained. The algorithm is simple in layout and easy to implement, but it is only suitable for network with few nodes and simple connection, so it is not easy to be applied in practice. 2) Distributed network structure combination layout[4]: The complex network is divided into several simpler network structure, respectively, using simple network structure layout method divide and conquer, and then according to each simple network structure of the interaction between the overall layout. The algorithm has perfected the adaptability of traditional layout algorithm to complex network, but it is not ideal for the network layout with very complex interactive relationship. 3) Force-oriented layout algorithm[5][6][7]: Give each node physical properties, two nodes with the distance between the repulsion, give each line to physical properties, each of the two nodes have a certain gravity. The algorithm keeps the nodes moving in the force process until the final approximate force balance is reached. The algorithm is suitable for simple network structure, but when the number of nodes is larger or the network structure is complex, many nodes are difficult to balance, some of them have cross problems, and the layout effect has a strong dependence on the initial position of the nodes. Overall, the existing layout algorithm has less nodes and simpler connection, which can realize the topology layout of the network, but the node overlap is basically unavoidable when the nodes are too much, and the force-oriented layout algorithm is more adaptable than other algorithms. So we based on the force-oriented layout algorithm, combined with the network topology related concepts and requirements[8], set up a node layout model, design network topology generation algorithm.

Based on the analysis of Network topology layout algorithm, the network topology layout is used as the goal, and the problem of the distance between the midpoint and the edge crossing is studied.

This paper establishes the node layout model, establishes the criterion of network topological layout, solves the problem of node distribution and node distribution, determines the coordinate between nodes, coordinates between nodes, edges and edges, adjusts the coordinate relation of these three, and quantifies the layout effect[9]. Design network topology generation algorithm, The topology map of the detected network is generated automatically, and the automatic generation of the node layout model to the network topology map is realized.

2. Node Layout model

2.1 Node Layout Model Description

Node layout model is based on node information and node contact information, using node distribution rules to compute node location, and adjusting the layout model through node connection. The node layout model can be described by a ternary group:

$$NDM=(ITD，CNDP，CNLA)$$

Which:

a) ITD(Information To Detect):Get the node information and node contact information in the actual network, including: IP address, open service, vulnerability, node contact, etc.

b) CNDP(Calculating Node Distribution Position):By formula calculation, the location of the node being obtained is laid out.

c) CNLA(Contact Node Layout Adjustment):The distribution of nodes in the form of the edge of the node to increase the contact, if the point is too close or edge intersection, you need to adjust the node position, reduce the problem.
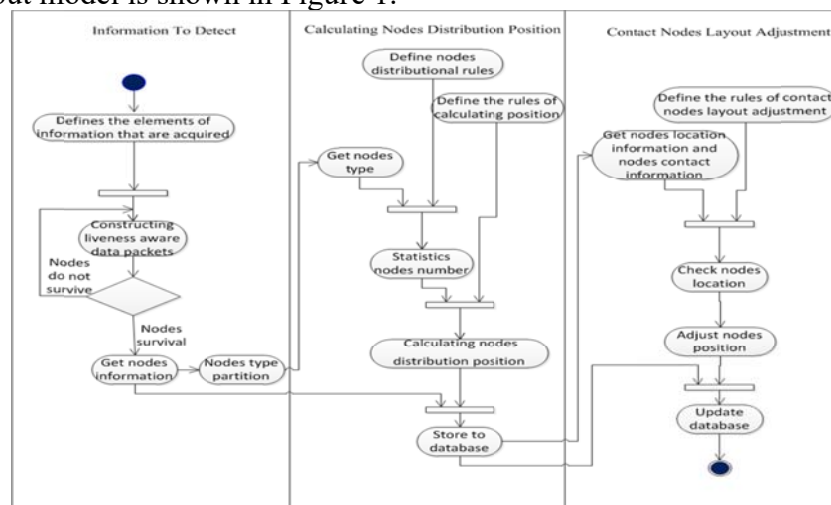
The node layout model is shown in Figure 1.

Figure 1. Node Layout Model

2.2 Information To Detect

Information To Detect is a component of the node layout model, which is used to obtain the information elements in the target network and to realize the function of determining the surviving nodes and obtaining the information elements in the target network. Information detections are obtained by sending packets. The packet consists of an address probe segment, a port scan segment, an operating system scan segment, an open service segment, a return value segment, and so on, which is formatted as shown in Figure 2.

←Address detection section→		←—Port scan section—→			←—OS scan section—→		←Open service section►		◄Return value section→	
Originator IP address	Destination IP address	Number of open ports	Port type	Special port	OS type	OS version	Type of service	Special service	Group object value	Return object value

Figure 2. Packet format

Through port scanning, address scanning[10] to obtain the node's IP address, open port, service and other information and fill in the packet, if the node is alive, the sender will receive the response

sent by the node grouping, and return the object value[11]; otherwise, the sender receives no response grouping for the node. The data packets are parsed, the contents of the packets are decomposed, the node types are divided according to the specific design requirements of the above data packets, and the node type partition table is shown in table 1.

Table 1. Node type partition table

Node type	IP Address	Open Special port	Operating System	Open Special Services	Return object value	Grouped object values
Router	Multiple	No	Yes	No	1	No
Switch	No	No	No	No	2	78
Computer	Single	No	Yes	No	2	76
Server	Single	21	Yes	ftp	2	76
Printer	Single	90	Yes	No	2	76

2.3 Calculating Nodes Distribution Position

Calculating Node Distribution Position is the location distribution of the acquired node in the layout area. According to the connection number of different network ports, the nodes are divided into single and multiple connecting nodes, and then according to the node distribution rules, the position of each node in the network topology map is calculated. Through the node layout position calculation, fixed each node position, will obtain the node abstraction layout to the graphical interface, realizes the logical layout effect.

The node distribution rule refers to the way in which nodes are arranged according to the node type, in which the single connection node, that is, the computer, the server, the printer nodes are distributed on the circumference of R radius, and the multiple connection nodes, that is, the switch and the router are distributed on the circumference of r radius. The above two circumferences are centered on the geometric center of the layout area, and the calculation of R and r is shown in Formula 1 and Formula 4.

A single connection node is a network node in which only one network is connected with other devices, and a multi-node node is a network node connected by multiple network ports to other devices; A layout area is an area that depicts a network topology.

The layout method of a single connection node is centered on the center of the layout area, according to the circle, the radius is proportional to the number of nodes, the maximum cannot exceed the shortest side of the layout area, assuming that R is the radius length of a single connection node's circumference, w is the shortest side length of the canvas, and c is the number of single connection nodes, the calculation formula is:

$$R = \frac{w}{k_1} \times c \qquad (1)$$

Calculating the transverse and ordinate of a single connection node by using the position calculation formula, i indicates the ith single connection node:

$$x = R \times cos(\frac{i \times 2\pi}{c}) \qquad (2)$$

$$y = R \times sin(\frac{i \times 2\pi}{c}) \qquad (3)$$

The layout method for multiple-connection nodes is centered on the center of the canvas of the layout, according to the circle, the radius is proportional to the number of nodes, the maximum can not exceed the radius of the circle of the single connection node, r is the circle radius of the multiple-connected nodes, and R is the radius length of the circumference of the single connecting node, and C is the number of multiple connecting nodes, the calculation formula is:

$$r = \frac{R}{k_2} \times C \qquad (4)$$

By using the coordinate transformation formula to compute the transverse and ordinate of the multiple-join nodes, j indicates the jth multiple-connection node:

$$x = r \times cos(\frac{j \times 2\pi}{C}) \qquad (5)$$

$$y = r \times sin(\frac{j \times 2\pi}{C}) \qquad (6)$$

2.4 Contact Nodes Layout Adjustment

Contact Node Layout Adjustment is to add nodes in the form of the nodes after the Calculating Node Distribution Position is complete. To optimize the layout effect, the node position layout needs to be adjusted. The linkage between nodes is divided into two kinds of situations: edge crossing and point edge distance too close[12]. When there is a connection between nodes, that is, the connections between nodes through the edge. If the node and the edge distance are too close, the coordinates of the node will change, and the degree of change is inversely proportional to the distance. When the distance between the node and the edge is greater than some distance m, the node coordinates change to zero. If the node projection point is on the edge, it is thought that the distance between the point and the edge may be too close, for this reason, design the formula of the degree of node's coordinate change. S is the degree of the node's coordinate, D is the straight distance from the node to the edge, and n is the constant. At the end of the layout, the node and the edge will maintain a certain distance between the node and the edge to avoid the distance too close:

$$|S| = \begin{cases} 0 & D > m \\ n \times |m - D| & D \le m \end{cases} \qquad (7)$$

At layout time, if a side intersection occurs, that is, two nodes on one side are projected on either side of the other, then the two edges may cross. If the intersection occurs, the coordinates of one node are symmetrically transformed along the other side[13]. Assuming that the equation of one of the edges is $y = kx + t$ (The equation is determined by the two nodes of the edge.), the node coordinates to be transformed are (a, b), then the transformed coordinate formula is as follows, and the equations can be solved:

$$\begin{cases} y - b = -\frac{x-a}{k} \\ \frac{(y+b)}{2} = k\frac{(x+a)}{2} + t \end{cases} \qquad (8)$$

3. Network Topology Generation

Network topology generation is the result of node distribution model, according to node location and node connection, design algorithm, generate display network topology map.

3.1 Algorithm idea

The topology diagram is composed of nodes and edges, so the number of nodes in the unit range cannot be too large[14] for the layout to be beautiful. If the number exceeds a range, the node will move outward, the coordinates will change, and conversely, it will move inside. Finally, the balance is achieved, that is, the best layout. Therefore, the network topology generation algorithm is designed based on node layout model and network related knowledge[15].

Calculates the location according to the node layout model and sets the initial node density. When the density of the node is greater than the initial density, that is, the distance between the nodes $Distance(v_i, v_j)$ is less than or equal to the initial density $Length$, the coordinates of the nodes do not change; when the density of the node is less than the initial density, that is, the distance between the nodes $Distance(v_i, v_j)$ is greater than the initial density $Length$, the coordinates The degree of variation is proportional to the difference between the distance between nodes and the density of the initial setting, and the proportional coefficient is the change coefficient k of the edge.

$T(v_i, v_j, e_k)$ for the side e_k to Node v_i, v_j coordinate change degree, and the node set density $Length$ and edge change coefficient k, the formula is designed as follows[6], where the k value and the layout of the size of the area is proportional to the number of nodes (v) inversely proportional to:

$$|T(v_i, \ v_j, \ e_k)| = \begin{cases} 0 & Length \geq Distance(v_i, \ v_j) \\ k * |Distance(v_i, \ v_j) - Length| & Length < Distance(v_i, \ v_j) \end{cases} \quad (9)$$

$R(v_i, \ v_j)$ indicates that the node v_i, v_j distance is too close, resulting in the degree of node coordinate change, and the number of nodes is proportional to the distance between the nodes of the square is inversely proportional to the formula as follows:

$$|R(v_i, \ v_j)| = \begin{cases} f & Distance(v_i, \ v_j) = 0 \\ \frac{g*Degree(v_i)*Degree(v_j)}{Distance(v_i, \ v_j)^2} & Distance(v_i, \ v_j) \neq 0 \end{cases} \quad (10)$$

When the two nodes coincide, that is $Distance(v_i, \ v_j) = 0$, the coordinates of two nodes should be redefined, and the relative positions of two nodes are determined by the defined coordinates, That is, the horizontal and ordinate of a node v_i respectively plus the horizontal and ordinate of the f, v_j nodes, then minus f respectively.

d is the change of the unit coordinate value, that is, when the distance between the nodes of the overall degree of change $F(v_i)$ is not zero, the node's horizontal, ordinate will change:

$$d * \frac{|F(v_i)|}{Degree(v_i)} \quad (11)$$

After the node location is determined, draw the node and connect it. In order to ensure the uniformity of the line, each node to determine the four-directional connection point, respectively, up and down. If all the lines are connected to the same connection point, the layout is dense, affecting the display effect, and for this reason the θ value is computed for the two nodes associated, and θ indicates the point of connection point, and the position of the connection point is determined, and the calculation formula is as follows.

$$\theta = tan^{-1} \frac{|x_1 - x_2|}{|y_1 - y_2|} \quad (12)$$

If $\theta \in [45°, \ 135°)$, connect the upper boundary of the node area, if $\theta \in [-45°, \ 45°)$, connect the right boundary of the node area, if $\theta \in [-135°, \ -45°)$, connect the lower boundary of the node area, if $\theta \in [135°, \ -135°)$, Then the left boundary of the node area is connected.

3.2 Algorithm Description

The network topology generation algorithm is based on the algorithm idea and combines the result of the node layout model to abstract the target network. The emphasis is on the design and display on the basis of the node layout model, and the problems in other algorithms, such as the overlap of nodes, the distance between points and the edges, are improved.

Input: A file strXMLFilePath that stores node information and node contact information after the node layout model is processed

Output: Network topology map

Step 1: Define node linked list m_listnodes and node link list m_listlinks

Step 2: From the file to obtain the Survival node and node contact information, respectively stored in the ListNode linked list and the Listnodelink linked list;

Step 3: Scan the node list obtained in step 2 and assign coordinates randomly to each node scanned, The node's coordinate and node information is stored in the new node chain table M_listnodes.

Step 4: Scan Step 1 to get the node link list, calculate the degree of each node, and store the connection information in the new node contact list m_listlinks;

Step 5: Calculates the overall change degree of each node's coordinates $F(v_i)$, draws the node

Step 6: Compute the connection angle of the node and draw the line.

3.3 Algorithm Analysis

Assuming the total number of nodes is N, including terminal host node, router node, printer

node, switch node, etc., in which terminal host node number i, router node number j, according to the design algorithm, with two-layer cycle, multiplication and division operation as a key operation, its time complexity is $T(N) = O\left(\left(\frac{N-i-j-1}{j}\right)^2\right) < O(((N-j)/j)^2)$.

Comparison of time complexity of tree layout algorithm $T(N) = O(N)$, time complexity of force-oriented layout algorithm[16] $T(N) = O(N^2)$, we know that this algorithm is less significant than the other two algorithms T (N), And with the increase of N、j, the decrease of $T(N)$ increases.

3.4 Example

In order to verify the accuracy of the proposed network topology generation algorithm, the laboratory is selected to build the network environment for experimental verification. Use Microsoft Visual C + + 6.0 software programming to implement this algorithm under Windows platform. Build a 20-node LAN environment in the lab, as shown in Figure 3 (a), which includes five types of nodes for computers, servers, routers, switches, and printers. The layout effect in the actual environment, as shown in Figure 3 (b), is automatically displayed in a layout, and is not adjusted by manual intervention. It can be seen from the graph that the network topology generation algorithm can obtain the network topological information accurately in the real environment, and there are no problems of node coincidence, Edge Crossing and point edge distance in the layout effect.
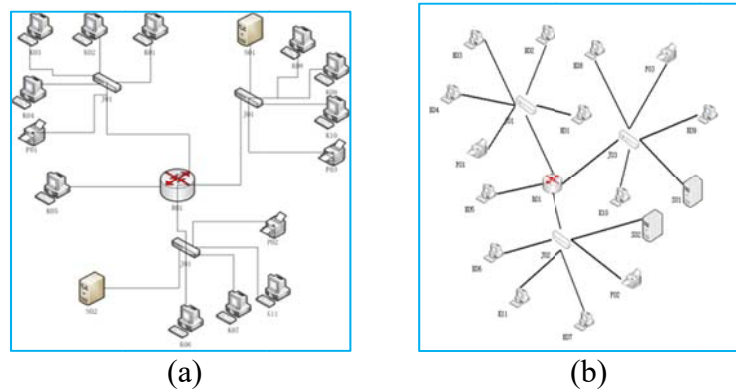

(a)　　　　　　　　　　　　　　　(b)

Figure 3.Actual LAN topology and layout renderings

4. Experimental Verification

Using the above algorithm idea, the network topology generation program based on node distribution is designed, which is tested under the LAN environment. The main goal is to test the network topology layout algorithm to obtain the accuracy of node information and the advantages and disadvantages of layout effect, by defining the node coincidence ratio (PC), the Edge crossover rate (SC) and the evaluation exponent (EI), the generated network topology map is evaluated.

1) Node Coincidence Rate (PC)

The node coincidence rate (PC)refers to the ratio of the number of nodes (r) to the total number(v) of the network topology generated after the layout, and the larger the node coincidence ratio, the more the number of nodes in the generated topological map, the worse the layout effect is.

$$PC = \frac{r}{v} \qquad (13)$$

For this reason, four kinds of LAN with 20, 80, 120 and 200 nodes are built, and the tree layout algorithm[17], force-oriented layout algorithm[5][6][7] and the algorithm (TBS algorithm) designed in this paper were used to test，10 tests were repeated for each environment, check the generated network topology map node coincidence, statistical results to take the average, and the coincidence rate of each algorithm to draw the scatter and fitted curves, further observe the algorithm's node coincidence rate in the future trend of change, as shown in table 2, Figure 4.

Table 2. Comparison of three algorithm layout experiment results(PC)

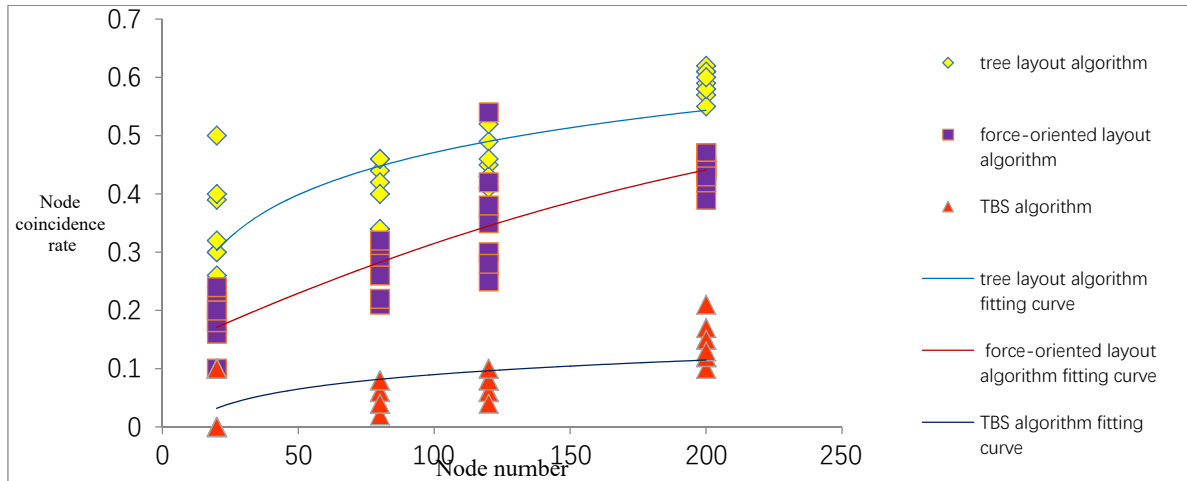| Nodes Algorithm	20	80	120	200
Tree Layout Algorithm	0.39	0.44	0.49	0.52
Force-oriented Layout Algorithm	0.21	0.27	0.35	0.41
TBS Algorithm	0	0.04	0.06	0.08

Figure 4. Matching curve of node coincidence ratio

By analyzing the data of the experimental results, the tree layout algorithm and the force-oriented layout algorithm are coincident in the case of fewer nodes, but there is no coincidence of the algorithm. With the increase of the number of nodes, the node coincidence rate of tree layout algorithm and force-oriented layout algorithm is increasing rapidly, and the algorithm of this paper is only slowly increasing. The fitting curve is the scatter data based on the experiment, and the future trend of the node coincidence rate of each algorithm is described. From the curve, we can see that the layout algorithm designed in this paper is much smaller than the other two algorithms in the node coincidence rate，and as the number of nodes increases, the effect is more obvious. The reason is because the algorithm designed in this paper has done the corresponding calculation and processing to the node coincidence situation.

2) Edge Crossover Rate (SC)

Edge Crossover Rate (SC): The ratio of the number of crossed edges (c)and the total number of edges (B) in the network topology generated after the layout.

$$SC = \frac{c}{B} \qquad (14)$$

When the edge crossover rate is larger, it indicates that the more crossed edges in the generated topology, the worse the effect of the layout is. In order to visually compare the layout effect, the coincidence ratio experiment is used to achieve the same environment repeat 10 test tests for each environment, check the intersection of the generated network topology map, calculate the average value of the statistic results, and draw the scatter plots and fitting curves for the edge crossover rate of each algorithm, and further observe the future trend of the edge crossover rate of the algorithm. , as shown in table 3 and Figure 5.

Table 3. Comparison of three algorithm layout experiment results(SC)

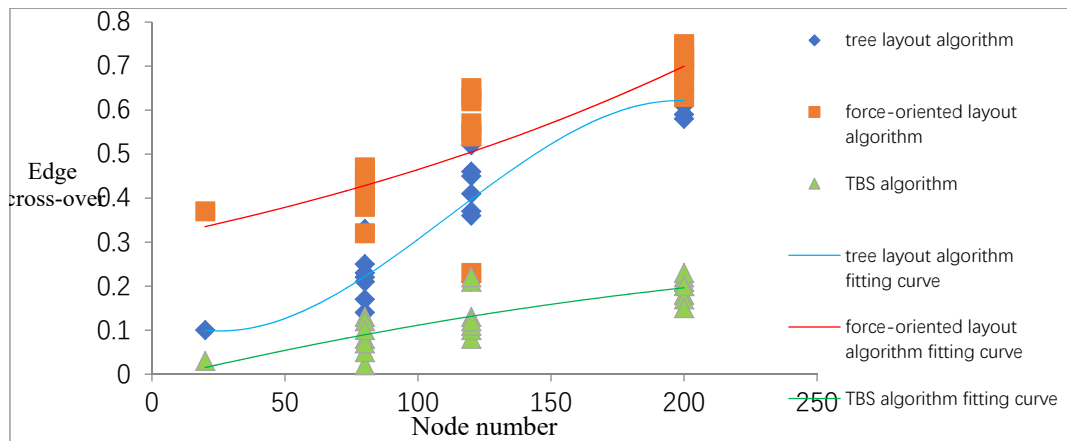| Nodes Algorithm	20	80	120	200
Tree Layout Algorithm	0.11	0.23	0.32	0.59
Force-oriented Layout Algorithm	0.36	0.41	0.44	0.62
TBS Algorithm	0	0.04	0.08	0.11

Figure 5. Comparison of curve-fitting curves with edge-crossover ratio

By analyzing the data of the experimental results, the tree layout algorithm and the force-oriented layout algorithm appear in the case of fewer nodes, while the edge crossover is not present in the algorithm. With the increase of the number of nodes, the edge crossover rate of tree layout algorithm and force-oriented layout algorithm is increasing rapidly, and the algorithm of this paper is only slowly increasing. From the curve, we can see that the layout algorithm designed in this paper is much smaller than the other two algorithms, and the effect is more obvious as the number of nodes increases. The reason is because the algorithm designed in this paper has done the corresponding calculation and treatment to the side-intersection situation.

3) Evaluation Index

Evaluation Index[2] EI is an evaluation of the overall layout effect.

$$EI = (\alpha, \ \beta) \qquad (15)$$

$$\beta = \frac{\frac{1}{m}\Sigma|e_i - \bar{e}|}{\bar{e}} \qquad (16)$$

Among them, α is defined as the number of edge intersections in the network topology map, and when α is smaller, the edges of the network topology map are less and the layout is clearer. β is defined as the ratio of the mean absolute deviation and the mean of the length of all edges in the network topology graph, and the $\bar{e}$ is the arithmetic mean of all edge lengths. β represents the uniformity of the vertex distribution, the smaller the β, the more evenly the vertex distribution in the graph. The comparison of the evaluation Index EI is as follows: First compare α, α size determines the size of EI; if alpha equals, then β, the size of β determines the size of EI. The smaller the EI, the better the topology layout.

In order to visually compare the layout effect, and the node coincidence rate experiment is consistent with the environment, each environment is repeated 10 times test, each test on the overall layout effect of the calculation, the statistical results to take the average, the experimental results as shown in table 4.

Table 4. Comparison of three algorithm layout experiment results (EI)

Nodes / Algorithm	20	80	120	200
Tree Layout Algorithm	(0，0.13)	(1，0.23)	(15，0.38)	(48，0.47)
Force-oriented Layout Algorithm	(1，0.21)	(4，0.19)	(29，0.23)	(75，0.42)
TBS Algorithm	(0，0.07)	(0，0.09)	(2，0.11)	(10，0.17)

The results of experimental data show that the algorithm is more homogeneous than the other two. The reason is that the layout algorithm is designed to consider the overall layout effect, through formula calculation, focus on the problem of node overlap, edge intersection, so that the layout effect is better.

5. Conclusion

In this paper, node layout model is designed to realize node information acquisition, node location distribution calculation and edge layout calculation, and the transformation of actual network to abstract node is completed. Then the force-oriented layout algorithm and the network topology related concepts are combined to design the network topology generation algorithm on the basis of the node layout model, and automatically generate the network topology map. Compared with other algorithms or models, the network topology layout algorithm is designed to focus on the global layout, overcome the problem of edge crossing and edge intersection, which is caused by the local node distribution uniformly, and through the quantization of the parameters, the layout effect is clarified, which makes the layout more scientific and reasonable.

The experiment verifies the layout effect of network topology generation algorithm, and the network topology layout algorithm has lower node coincidence ratio, edge crossover rate and better evaluation index, regardless of the result of the actual display or the evaluation criterion constructed. In the next phase of the study, when the large-scale nodes are automatically laid out, the computer shows that one screen may not show all the layouts, and it is proposed to adopt the method of multi-screen expansion and local amplification to improve the application range of the algorithm.

Reference

[1] Kokar M M.Situation awareness：issues and challenges：The Seventh International Conference on Information Fusion，Sweden，2004[C].

[2] Cheng Yuan, Yan Wei, Li Xiaoming, Topology layout algorithm based on the repulsion-tension model[J]. Computer Engineering, 2004，30(3)：104-106.

[3] M O，S C，M M.Fusion of disparate identity estimates for shared situation awareness in a network centric environment[J].Information Fusion，2006，7(4)：395-417.

[4] Liu lili. Li yeqing, Topology layout algorithm based on network hierarchy[J]. Command Information System and technology, 2015，6(5)：35-39.

[5] Jiang dong. Research on network topology discovery and display technology[D]. Beijing University of Posts and Telecommunications, 2012

[6] Noack A.An energy model for visual graph clustering：Proc of the 11th International Symposium on Graph Drawing，2004[C].

[7] Lv liang, Lu xinze, Li sudan, etc. Topological graph layout algorithm based on extended mechanics model[J]. Computer Application Research, 2010，27(7)：2713-2715.

[8] Yong Z，Xiao-Bin T，Xiao-Lin C，et al.Network security situation awareness approach based on Markov game model[J].Journal of Software，2011，22(3)：495-508.

[9] Zhang chang, Xie jun, Hu guyu, etc. Design and implementation of topology visualization scheme for complex networks [J]. Computer Technology and development. 2014，24(12)：78-82.

[10] Li zhuhong, Zhao canming, Shi gun, etc. Link Layer Topology discovery algorithm in dynamic network environment [J]. Application of computer system, 2015，24(10)：122-128.

[11] He Yingjie, Wang huiqiang, Zhou renjie, Real-time network topology discovery oriented to network situational awareness [J]. Computer Engineering, 2009，35(24)：127-129.

[12] XW L，HQ W，HW L，et al.Fusion-Based cognitive awareness-control model for network security situation[J].Journal of Software，2016，27(8)：2099-2114.

[13] Pan wenxi, Liang wei, Chen mo, etc. Design of real-time network topology discovery system based on OSPF and BGP [J]. Computer Application, 2007，27(12)：2995-2998.

[14] Zhang zhanguo, Liu shufen, Tiebao, etc. A physical network topology discovery algorithm based on STP protocol [J]. Computer Engineering, 2008，34(6).

[15]Abhijit G ， Anish P.Solving Markov Decision Processes with Downside Risk Adjustment[J].International Journal of Automation and Computing，2016(3)：39-50.

[16] Liang sheng, Wan yangsuo. Heuristic network topology layout algorithm based on node attribute [J]. Computer Engineering and application, 2016，52(20)：122-126.

[17]Liu jie. Research on visualization of network topology[D].Xi'an Electronic and Science University,2013.