Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 827376, 7 pages
http://dx.doi.org/10.1155/2013/827376

Research Article

A Topology Visualization Early Warning Distribution Algorithm for Large-Scale Network Security Incidents

Hui He, Guotao Fan, Jianwei Ye, and Weizhe Zhang

Department of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

Correspondence should be addressed to Weizhe Zhang; wzzhang@hit.edu.cn

Received 30 July 2013; Accepted 20 August 2013

Academic Editors: J. Shu and F. Yu

It is of great significance to research the early warning system for large-scale network security incidents. It can improve the network system's emergency response capabilities, alleviate the cyber attacks' damage, and strengthen the system's counterattack ability. A comprehensive early warning system is presented in this paper, which combines active measurement and anomaly detection. The key visualization algorithm and technology of the system are mainly discussed. The large-scale network system's plane visualization is realized based on the divide and conquer thought. First, the topology of the large-scale network is divided into some small-scale networks by the MLkP/CR algorithm. Second, the sub graph plane visualization algorithm is applied to each small-scale network. Finally, the small-scale networks' topologies are combined into a topology based on the automatic distribution algorithm of force analysis. As the algorithm transforms the large-scale network topology plane visualization problem into a series of small-scale network topology plane visualization and distribution problems, it has higher parallelism and is able to handle the display of ultra-large-scale network topology.

1. Introduction

With the network's application in various areas of human life, network security draws more and more attention all over the world. Network security problems such as computer virus and hackers' illegal intrusion lead to important information leaks and may even cause the network paralysis. The accidents have caused huge economic losses to various countries and many companies and even endanger the security of the countries and regions. Just in the first half of 2004, nearly 2 million hosts were attacked by major worms like Mydoom, RPC loopholes, and LSASS loopholes in China [1, 2].

The study of early warning system and intrusion detection technology has already carried out in many countries. These systems monitor the illegal intrusion in some important economic, political, and military networks. They play important roles in the protection of network security, the early detection of intrusion, and the control of virus' spread. There is no suitable intrusion detection and early warning system for large-scale network in China at present. In order to support the information system and adapt to the requirement of information warfare, it is necessary to develop the large-scale network intrusion detection and early warning system. It has very important significance to improve the network system's emergency response capabilities, alleviate the cyber attacks' damage, and strengthen the system's counterattack ability.

The large-scale network system's plane visualization is realized based on the divide and conquer thought. First, the topology of the large-scale network is divided into some small-scale networks by the MLkP/CR algorithm [3]. Second, the subgraph plane visualization algorithm is applied to each small-scale network. Finally, the small-scale networks' topologies are combined into a topology based on the automatic distribution algorithm of force analysis. As the algorithm transforms the large-scale network topology plane visualization problem into a series of small-scale network topology plane visualization and distribution problems, it has higher parallelism and is able to handle the display of ultralarge-scale network topology.

2. Related Works

Some international research institutions have been engaged in the study of this aspect. In 1999, the Information Assurance

```
Begin
G ← CutDown(G_0); //cut G_0 and get the graph backbone G
k_partition(G);        //divide G into k sub graphs
for i ← 1 to k
   Plane_Visualize(G_i);    //plan visual each sub graph
   G_i = AttachLeaf(G_i);        //attach the leaf nodes to the backbone
end for
layout(G_1, G_2, G_3, ..., G_k);    //distribution each sub graph
G_0 ← connect(G_1, G_2, G_3, ..., G_k);    //connect each sub graph and get the final plane visual graph G_0
end
```

Algorithm 1: Plane visual algorithm.

Advisory Council (IAAC) conducted a project called Threat Assessment and Early Warning Methodologies for Information Assurance [4, 5]. It mainly develops and evaluates the analysis methods for threat assessment and early warning. The research goal is to prove that quantifiable threat assessment and early warning are feasible, which lays foundation for further application research. The achievements are as follows: (1) to prove the feasibility of threat outline's generation and describe the threat outline from the attacker's motives, intentions, capabilities, and behavioral patterns and (2) to argue the feasibility of indicating and alarming computer attacks from substate stage actor behavior.

The study has several limitations. It focused on network's external threats. The considered attacker type is limited to sub state stage attacker; the state or national agent stage attackers are not considered. Many theories and technologies involved in the project are not mature and still need further research and development. Another related project is Information Warfare Attack Assessment System [4], developed by the International Centre for Security Analysis (ICSA) of British Kingps College London between 1997 and 2000. This project presented the concept of information warfare attack's threat assessment, indication, and warning, as well as the conceptual framework of an open information source decision support system. Its goals are as follows: (1) to evaluate the information warfare threat caused by different attackers, (2) to provide information warfare attack's indication and warning, and (3) to predict the enemy's behavioral path. The above two projects are of great relevance.

In China, researches have been done on attack detection technology and feature information extraction methods of network security strategic early warning system [6, 7]. Overall, only a small number of domestic agencies are working on network security early warning system, and there is no much open technical literature available. And existing intrusion detection systems just simply submit alarm information to administrator by a format of record. Administrator can hardly get the distribution state of current network's abnormalities from the boring records, and it is also not conducive to deal with abnormalities in time. Under this background, this paper proposes a large-scale network security incident early warning system which combines active measurement and anomaly detection, aiming at macro-early-warning for the outbreak of wide range network events based on network topology. And we will focus on the visualization of the early warning system. Through this system the administrator can get the security events' distribution state intuitively in graphical form.

Therefore, this paper comes up with a plane visualization problem and probes into how to locate large-scale network topology on plane and get good visual effect at the same time.

3. Topology Distribution Visualization Technology for Large-Scale Security Incidents

3.1. Plane Visualization Algorithm Framework. According to the balanced subproblem thought of the divide and conquer algorithm, the original topology should be decomposed with the following requirements: (1) to decompose the original topology into k subtopology with the same scale less than N; each subtopology is described as $G_i(N_i, V_i)$, so $N_i < N$; (2) to make the edges in each subtopology as few as possible and keep each subtopology independent as far as possible, so that each subproblem will be independent and at the same time the subtopologies will have good locality; thus the administer can observe a subtopology's information relatively independently; (3) and to ensure that each subtopology is a connected graph, according to the divide and conquer algorithm, as the network topology is an undirected connected graph, so each subgraph is a logical network topology.

As a result, the plane visualization algorithm framework for large-scale network is shown in Algorithm 1.

The key technologies of the algorithm include core router screening technology, undirected graph segmentation algorithm MLkP/CR, the subgraph internal vertex distribution algorithm, quasi-planarity technology, and the subgraph automatic distribution algorithm based on force analysis. This paper focuses on explaining the core router screening technology and the subgraph automatic distribution algorithm based on force analysis.

The scale of each subgraph is defined as N, so the algorithm realizes the plane visualization of undirected graph with $k * N$ scale. The algorithm is able to solve the plane visualization problem of undirected graph with any size, but the value of k is different, $k = |G_0|/N$. In particular, the problem becomes a small undirected graph plane visualization

problem when $k = 1$. The algorithm has high parallelism, and the subgraph plane visualization can be realized in parallel.

3.2. Core Router Screening Technology. As the connection relations of core routers constitute the backbone of the network topology, the plane visualization of the topology backbone is the key to the plane visualization of the network graph. Before dividing the network topology, we can find its backbone and reduce its scale through cutting. The topology is defined as $T = (V, E)$, where $|V| = n$ and $|E| = m$. As T is connected, so the minimum degree of the vertexes in T is 1. Cut down the vertexes with degree 1 in T; a new topology $T' = (V', E')$ is got. The vertexes in T' are the vertexes in T with the degree more than 2. If the number of the vertexes in T with degree 1 is t, then $|V| = n - t$ and $|E| = m - t$. So if t changes sharply, the number of the drawn vertexes will be greatly reduced when this method is used iteratively. If T is equal to T_0 at the beginning; after several iterations we get $T_1, T_2, \ldots, T_n$ and $D_1, D_2, \ldots, D_n$, where $D_k = T_{k-1} - T_k$; namely, D_k is constructed by the vertexes with degree 1 and the edges connecting these vertexes in T_{k-1}. If we can draw T_n at this time, then we can draw T_0 reversely. The method is as follows: according to the definition of D_n, a vertex v in D_n is mapped to a unique vertex s in T_n; select a position around s and draw v. After that, connect s and v. Deal with the vertexes in D_n circularly and get a new topology including all the vertexes and edges in D_n and T_n. As $D_n + T_n = T_{n-1}$, so we can get T_0, namely, T, by calling this method iteratively.

3.3. Subgraph Macro–Automatic-Distribution Algorithm Based on Force Analysis. After the plane visualization of each subgraph is realized, put the subgraphs together and connect them with the edges between them. Then a whole topology, namely, the large-scale undirected graph before divided, is got. Therefore, the plane visualization of large-scale undirected graph is realized. For any two subgraphs G_a and G_b, their associated value is defined as $K(a, b) = \sum e(u, v)$, $u \in G_a \wedge v \in G_b$. If $K(a, b) = 0$; then the associated value is the minimum and there are no edges between the two subgraphs. If $K(a, b) > K(c, b)$, then the associated value of G_a and G_b is bigger than that of G_c and G_b. As K is different between different subgraphs, the subgraphs cannot be put together randomly. Their associated values K are related to their mutual positions, and the cross of the edges can be reduced by putting the subgraphs with higher K together. As shown in Figure 1, there are 5 subgraphs; on putting the subgraphs with higher K together, the cross can be reduced and the connect relation between subgraphs can be shown more clearly.

The goal of the distribution algorithm is to make the distance between the subgraphs as even as possible and make the cross edges as few as possible, namely, trying to put the subgraphs with higher K together.

For a given undirected connected graph $G(V, E)$, it is made up with m subgraphs $\{G_m, G_{m-1} \cdots G_1\}$ and their edges. The distribution space of G is defined as L_{m*m}, a matrix with $m * m$ scale. For $i, j \in \{0 \cdots m - 1\}$, if graph G_k captures the space, then make $L[I, j]$ equal to k $(k > 1)$. If $L[I, j]$ is 0, then the space is not captured by any subgraphs. $X[i]$ presents the start abscissa of the space with row i and $Y[j]$ presents the start ordinate of the space with column j. The abscissa range of $L[I, j]$ is $(X[i] \sim X[i] + G_i \cdot \text{length})$ and the ordinate range is $(Y[i] \sim Y[i] + G_i \cdot \text{width})$.

Now we import the force analysis method to our algorithm. With the above method, the subgraph distribution is transformed into the matrix distribution, and the number of each subgraph indicates the position where it is in the whole graph. Take the distribution matrix as a box, each subgraph waiting for distribution as a quality pellet, and the edges as rubber band. If the value of K is different, then the elastic coefficient of the rubber band is different. The rubber band has a free length. If it is pulled, then tension is generated. There is repulsion between any two pellets. Through this physical system, the subgraphs' placing process in the matrix is transformed into the process that the pellets move in the box according to mechanics laws and ultimately achieve balance. The pellets' positions in the box when balance is achieved are the subgraphs' right positions in the matrix.

The physical formula is defined as follows:

(1) tension formula:

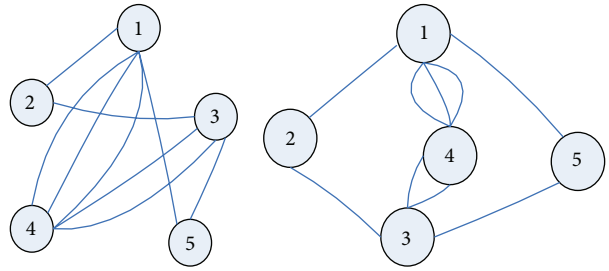
$$\left|\text{tension}\left(v_i, v_j, e_k\right)\right| = \begin{cases} 0, \left(\text{length}\left(e_k\right) \geq \text{distance}\left(v_i, v_j\right)\right), \\ k * \left(\text{distance}\left(v_i, v_j\right) - \text{length}\left(e_k\right)\right), \end{cases} \tag{1}$$

(2) repulsion formula:

$$\left|\text{repulsion}\left(v_i, v_j\right)\right| = \begin{cases} f, \left(\text{distance}\left(v_i, v_j\right) = 0\right), \\ \dfrac{g * \text{mass}\left(v_i\right) * \text{mass}\left(v\left(\text{distance}\left(v_i, v_j\right) \neq 0\right)_j\right)}{\text{distance}\left(v_i, v_j\right)^2}. \end{cases} \tag{2}$$

The pellet's quality is proportional to the subgraph's degree. Compared to the star structure, the greater the center vertex's degree, the heavier its quality and the greater the repulsion, and then the vertexes around it will have more space to distribute. As distribution in the matrix can't be as accurate as that in the physical world, we use a kind of greedy algorithm to distribute the subgraphs. It can avoid the accurate calculation in real physical world and achieve good effect at the same time. Distribute the subgraphs according to the pellets' qualities. As $d \propto |F|/m$, the greater the quality is, the less the displacement is [8]. And at the same time, the greater the degree is, the bigger the influence on the graph's distribution is. Therefore, the subgraph with higher quality is distributed preferentially.

Once the distribution of all the subgraphs is decided, their positions are fixed and will not be changed. As a result, when the next subgraph waiting for distribution is i, the positions of its previous $i - 1$ subgraphs are already certain and we only need to consider the influence of the previous $i - 1$ subgraphs.



(a) Place randomly (b) Place according to K

Figure 1: The influence of the subgraphs' positions on cross-edges.

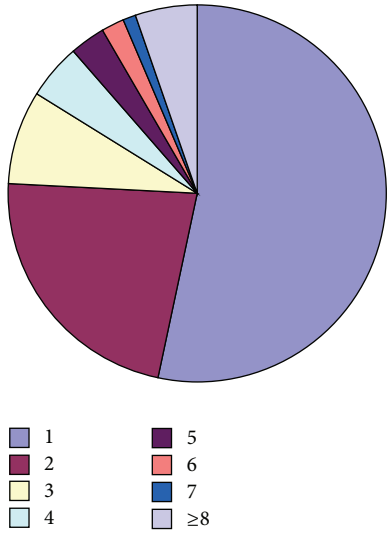


Figure 2: The degree distribution map of all the routers in China.

Table 1: Results of random distribution.

Cluster no.	Coordinate
1	(12, 10)
2	(8, 1)
3	(0, 6)
4	(1, 12)
5	(13, 1)
6	(7, 1)
7	(11, 8)
8	(12, 9)
9	(2, 5)
10	(9, 1)
11	(8, 10)
12	(2, 13)
13	(2, 7)
14	(5, 4)
15	(7, 10)

Here we use the traversal algorithm. Traverse every position of the matrix when it is put in pellet i and find the position where the force is the minimum; it is also the balance and final position of pellet i.

Distribution algorithm based on force analysis is shown in Algorithm 2.

4. The Experimental Results and Analysis

4.1. The Experiment Results of Core Router Screening. The degree distribution of all the routers in China is shown in Figure 2. There are 19847 routers in the network of China. Among these routers, 53.25% are routers with 1 degree. The network scale can be reduced by half through cutting off these vertexes, because they are the leaves attached to the topology backbone, and it is easy to add them to the backbone when the backbone is drawn. Finally, the plane visualization is realized.

4.2. The Experiment Results of the Distribution Based on Force Analysis. The effect of the distribution based on force analysis is related to the effect of the random distribution; the more the two kinds of distributions are a like, the better the algorithm's randomness is, the greater the dispersion degree of the distribution is, and the more balancedly the vertexes distribute. The experiment results of the two kinds of distributions are shown in Figures 1 and 2. The coordinate is the subgraph's row and column position in the distribution matrix.

Next, we compare the effect of the two kinds of distributions by calculating the distance of the data in Tables 1 and 2 and the distribution algorithms' results.

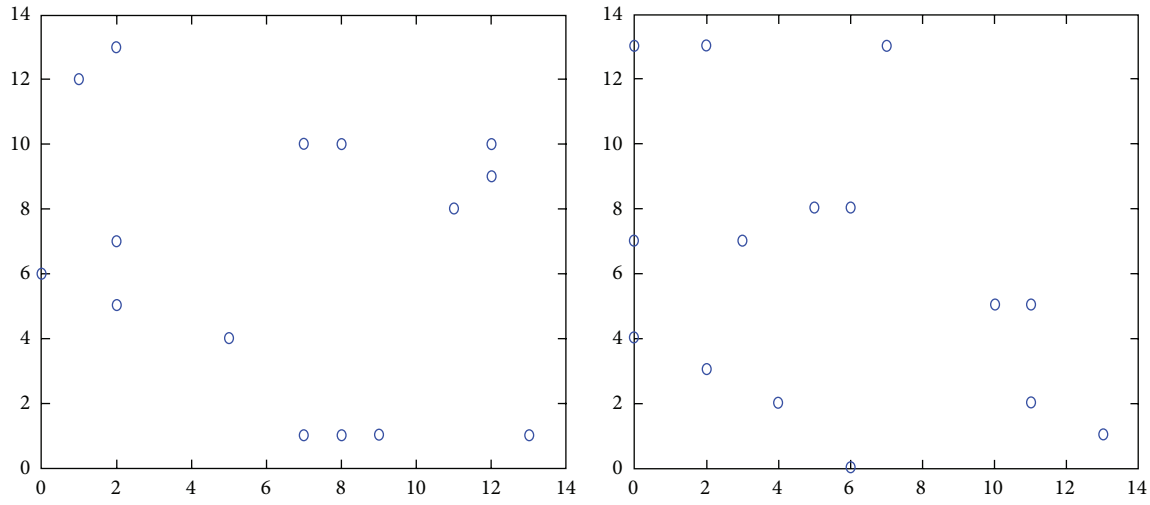
The distribution distance between two subgraphs is defined as

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (3)$$

The ratio of all the distances' mean absolute deviation and average is defined as α and it is calculated by (4). α indicates the well-proportioned degree of the subgraphs. The smaller α is, the better proportionally the subgraphs are distributed.

$$\alpha = \frac{(1/m)\sum \left|d_i - \overline{d}\right|}{\overline{d}}. \quad (4)$$

Table 3 is calculated from Tables 1 and 2. We can see that there is little difference between the two α, which means that the effect of the two kinds of distributions is almost the same and the force makes effective use of the distribution space, as shown in Figure 3.



(a) Random distribution (b) Distribution based on force analysis

Figure 3: Distribution diagram of the subgraphs.

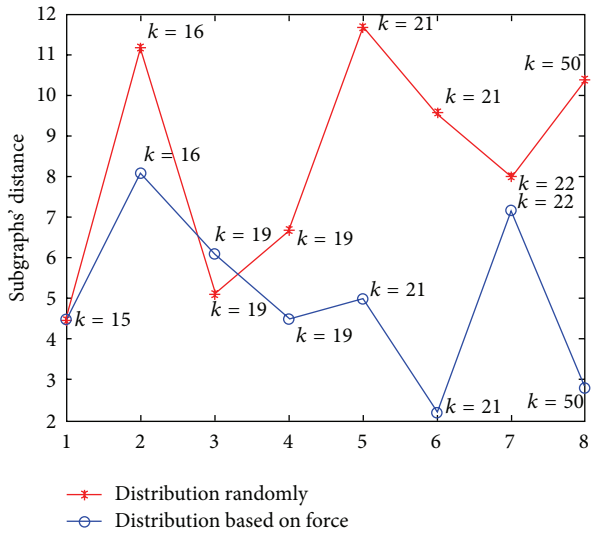


Figure 4: Comparison of the subgraphs' distance when $k > 15$.

Table 2: Results of distribution based on force analysis.

Cluster no.	Coordinate
1	(6, 0)
2	(3, 7)
3	(6, 8)
4	(5, 8)
5	(13, 1)
6	(11, 2)
7	(2, 3)
8	(0, 13)
9	(10, 5)
10	(11, 5)
11	(0, 4)
12	(2, 13)
13	(4, 2)
14	(7, 13)
15	(0, 7)

Table 3: Comparison of the two α.

	Random	Based on force analysis
α	0.0021	0.0028

The distribution based on force analysis can reflect the relationship and related degree of the subgraphs. Calculate parameter K which measures the distance between two subgraphs. From Figure 4 we can see that for the same value of K, the subgraphs' distance of the distribution based on force analysis is smaller than that of the random distribution, which says that the distribution based on force analysis reflects the relationship between the subgraphs. As the value of K becomes greater, the subgraphs' distance of the distribution based on force analysis presents degressive tendency. It means that our algorithm reflects the related degree of the subgraphs; the bigger the value of K, the greater the related degree and the smaller the distance. But the random distribution can't express these characters.

The α of the distribution algorithm based on force analysis is almost the same as that of the random distribution algorithm, meaning that the average degrees of the two kinds of distributions are similar. While the subgraphs' distribution is related to the value of K in the distribution algorithm based on force analysis, the subgraphs with greater K are closer. Therefore, the distribution result of the distribution algorithm based on force analysis is better than that of the random distribution algorithm.

4.3. The Experiment Results of the Network Logical Topology in China. The scale of the network in China is 10^4 stage. It has 19847 points and 24864 edges. Obviously, it is a large-scale

```
Input: sub graphs and their adjacent matrix;
Output: the position coordinate of each sub graph in the topology;
Begin
{ for each sub graph v_i in V
            Distribute coordinate (x_i, y_i) for v_i in the canvas randomly;
            Force(v_i) = 0;
      end for
    while there is sub graph to distribute
      begin
      Select a sub graph v_i with the highest quality from the sub graph set waiting for distribution;
      for each edge (v_i, v_j) connected with v_i
        if v_j is distributed
          Force(v_i) +=  Tension(v_i, v_j);   //calculate the tension
        end if
      end for
      for each vertex v_j in V
        if v_j is distributed
          Force(v_i) += Repulsion(v_i, v_j);   //calculate the repulsion
        end if
      end for
      ForceMin = Force(v_i);
   //reverse to find the balanced position
      for each position (x_j, y_j) in the layout
        if the position is used
            Calculate the resultant force Force(v_i) of this position;
            if Force(v_i) < ForceMin;
              ForceMin = Force(v_i);
              x_i = x_j;          //record the balanced position
              y_i = y_j;
            end if
      end if
    end for
  end
}
```

Algorithm 2

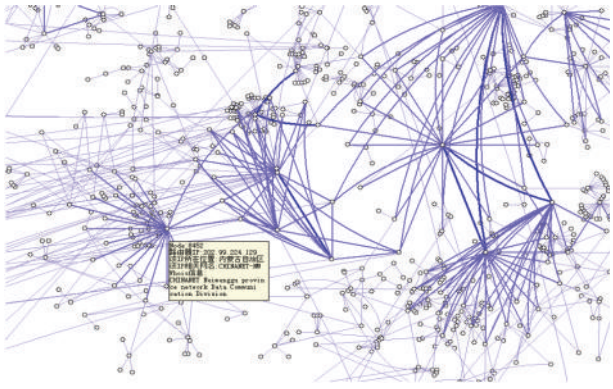


Figure 5: The subgraph display of the network logical topology in China.

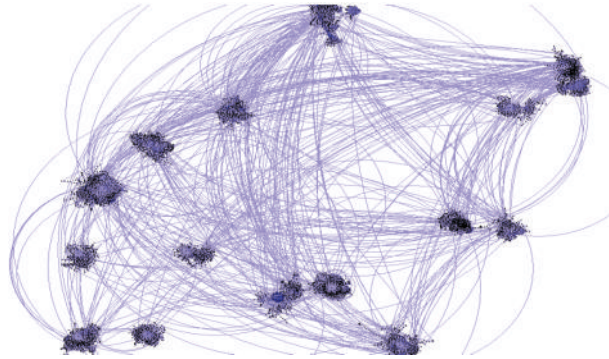


Figure 6: The whole display of the network logical topology in China.

network. The experiment results are shown in Figures 5 and 6.

5. Conclusion

This paper proposes a whole framework of the plane visualization algorithm based on the divide and conquer thought, aiming at solving large-scale network plane visualization problems, and it probes into two of the key algorithms and technologies applied in this algorithm. (1) Core router screening technology: we use this technology to cut down the leaf nodes and get the main stem of the graph. It can reduce the scale of the graph and improve the graph's plane visualization efficiency. Experiment has shown that after

processed by this technology, the scale of the graph can be halved. (2) Subgraph automatic distribution algorithm based on force analysis: with physicalification ideology, we transform the subgraph distribution problem into physical problem and get a reasonable distribution in the distribution space. Compared with the random distribution, experiments show that our algorithm has its superiority.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was partially supported by the National Basic Research Program of China (973 Program) under Grant no. 2011CB302605, the National High Technology Research and Development Program of China (863 Program) under Grants no. 2011AA010705 and no. 2012AA012506, and the National Science Foundation of China (NSF) under Grants no. 61173145 and no. 61100188.

References

[1] G. Karypis and V. Kumar, "Multilevel k-way partitioning scheme for irregular graphs," *Journal of Parallel and Distributed Computing*, vol. 48, no. 1, pp. 96–129, 1998.

[2] Y. Jiang, M. Hu, B. Fang, and H. Zhang, "An Internet router level topology automatically discovering system," *Journal of China Institute of Communications*, vol. 23, no. 12, pp. 54–62, 2002.

[3] K. Taşdemir and E. Merényi, "Exploiting data topology in visualization and clustering of self-organizing maps," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 549–562, 2009.

[4] M. Ankerst, M. M. Breunig, H. Kriegel, and J. Sander, "OPTICS: ordering points to identify the clustering structure," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 49–60, 1999.

[5] B. Krishnamurthy and J. Wang, "Topology modeling via cluster graphs," in *Proceedings of the 1st ACM SIGCOMM Internet Measurement Workshop (IMW '01)*, pp. 19–23, San Francisco, Calif, USA, November 2001.

[6] B. Krishnamurthy and J. Wang, "On network-aware clustering of web clients," in *Proceedings of the Applications, Technologies, Architectures, and Protocols for Computer Communication Conference (SIGCOMM '00)*, Stockholm, Sweden, August 2000.

[7] Z. Peng, E. Grundy, R. S. Laramee, G. Chen, and N. Croft, "Mesh-driven vector field clustering and visualization: an image-based approach," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 2, pp. 283–298, 2012.

[8] F. Bruckmann, F. Gruber, N. Cundy, A. Schäfer, and T. Lippert, "Topology of dynamical lattice configurations including results from dynamical overlap fermions," *Physics Letters B*, vol. 707, no. 2, pp. 278–285, 2012.