Research on Network Simulation Abstract Technology Based on Simplicity Theory

Jia Wang，Xiangzhan Yu
School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China 150001
e-mail: wangjia.hrb@gmail.com

Jun Yan
School of Computer Science and Technology
Harbin Engineering University
Harbin, China 150001
e-mail: dabiaoyanjun@gmail.com

Abstract—**Given a large scale internet, how can we derive a representative simplifying topology model by network simulation? There are many known algorithms to compute interesting measures (scale down the number of clients, deletion methods, contraction methods, et al.), but most of them only reduce the scale of the network topology, the facticity of network simulation results can't be guaranteed. So when simplifying network topology and reducing the simulation complexity, we should not only consider the reduction of network scale but also consider the actual simulation application and the facticity of simulation result.**

Given that, this paper proposes a network topology simplicity method based on focus-zone with Weight Forest Fire(WFF) , which can greatly reduce the scale of network topology and guarantee the facticity of simulation results. Then through sampling network topology experiment and UDP worm simulation experiment, it can be known that this method can not only maintain the topological properties of original topology but also improve the efficiency of network simulation by 11%. And the maximum error of the packet drop ratio is 7.8%, which occurred at the sampling rate of 0.5.

Keywords-network simulation; network topology simplicity; graph sampling algorithm; worm simulation

I. INTRODUCTION

It is difficult to predict when a large scale network security event will take place and how badly it will cause. Men have to know more clearly how a network breakage happened. But now it is difficult to capture a happening event's features which include its characters, machines it involved, styles and so on. At the same time, it is less precision to perform the event in lab than in reality because of the number of network equipments and the scale of the network. Network simulator will meet the problem. Simulating a network is one of the most important methods to study the network security events as it costs less and more efficient. As the environment of the network getting more and more complexity, we have to simplify the model so that the most important problems will be simulated more quickly and more clearly[1].

A typical simulation model includes three kinds: the topology model, the traffic model and the protocol model. The simplification of network simulation model is to increase the abstract degree of these models. While currently the work focuses on simplifying the network model by deleting some unimportant nodes and collapsing the network. But from the experiment results before, it shows the simulator could not simulate the real network exactly. So it is supposed to combine the real network application with deleting some unimportant nodes and collapsing the network in order to get a better simulation results.

Given that, based on the traditional sampling algorithms, we propose a method called the network topology simplicity method based on focus-zone and a network event contracted model based on topology simplicity. The model uses degressive compute method revolving around the focus-zones, which treats the zones that security event cares as predict-focus, exactly simulates the status of the whole network and then spreads from the focus-zones. We will abstract the simulation granularity to reduce compute complexity according to the distance from focus-zones. To a large extent, this method reduces the network scale, improves the efficiency and guarantees the facticity of simulation results.

The rest of the paper is organized as follows: Section II introduces the network topology model and traditional sampling algorithms from large graph. Section III presents the network topology simplicity method based on focus-zone which reduces compute complexity and guarantees the facticity. The performance of this method is studied using some experiments in Section IV, including sampling network topology experiment and UDP worm simulation experiment. We conclude the paper with a summary of the findings in Sec.V.

II. RELATED WORK

A. The reduction of network topology model

Considering the specific network simulation, the reduction of topology model is a simplified model of a variety of nodes, links, and topology for prototype components to streamline the combination to enhance the performance of network simulation. The current methods of network topology model simplification are mainly: reductive methods (reducing the number of clients), deletion methods, contraction methods, exploration methods, SHRiNK methods and traffic sampling methods. While these simplified methods are mainly used to reduce the size of simulation in order to improve simulation

978-0-7695-3901-0/09 $26.00 © 2009 IEEE
DOI 10.1109/WNIS.2009.10

speed and reduce the cost of simulation storage and computing, they don't perform well on the facticity of simulation results.

In [2], we reduce the complexity of the simulations by using a smaller number of clients as compared to reality while increasing the activity of each client and thus keeping the traffic load approximately constant. The experiment results show that this method could not only gain a significant reduction of the required memory, a factor 4-8, but also a side-effect in simulation speed of up to 30%. And that accuracy of Hurst parameter, end-to-end delay and the traffic load for each link can be guaranteed. Nevertheless, accuracy of the loss rate should be considered in the near future.

In [3,4], we propose three methods: (a) deletion methods, (b) contraction methods, and (c) exploration methods, to "sample" a small realistic graph from a large real network and speed up the network simulation. The experiment results show that it can effectively reduce the nodes of a graph by as much as 70% while maintaining its important properties, such as the average degree of topology graph, power laws, clustering coefficient and so on. This paper doesn't do any correlative network simulation experiment to verify the simulation validity. Obviously, the facticity of simulation results can not be guaranteed only by the topology parameters.

In [5], we take a reduction of the network topology parameters and a sample of the input traffic and feed it into a suitably scaled version of the system, then extrapolate from the performance of the scaled system to that of the original. Through the reduction we can slow down the simulation speed and minish the data packages transmitted. Therefore, the computing cost will be little. Although this method can, accurate network behavior and some details of end-to-end transition can not be guaranteed.

B. Traditional Sampling Algorithms from Large Graph

Conceptually we can split the traditional sampling algorithms into three groups[6]: methods based on randomly selecting nodes, randomly selecting edges and the exploration techniques that simulate random walks or virus propagation to find a representative sample of the nodes. In [6], it shows that the best performing algorithm is "Forest Fire", sampling by which we can nicely maintain the topological properties of original, such as the average degree of topology graph, power laws, clustering coefficient and so on. In Forest Fire (FF) we randomly pick a seed node and begin "burning" outgoing links and the corresponding nodes. If a link gets burned, the node at the other endpoint gets a chance to burn its own links, and so on recursively.

III. PROPOSED METHOD

From the computational complexity aspect, it is a NP problem that simulates the evolution of the state of large scale Internet security. Our task is to set up a novel simulation model, in order to reduce the simulative complexity and guarantee the facticity of network simulation with some abstract method.

A. The network event contracted model based on topology simplicity

The model uses degressive compute method revolving around the focus-zones, which treats the zones that security event cares as predict-focus, exactly simulates the status of the whole network and then spreads from the focus-zones. We will abstract the simulation granularity to reduce compute complexity according to the distance from focus-zones.

B. The network topology simplicity method based on focus-zone

The model based on topology simplicity reduces the complexity and cost of network simulation. But in order to improve the efficiency of network simulation and guarantee the facticity of network simulation, we take consideration of the traditional graph sampling algorithm and the characteristic of real network topology, propose a network topology simplicity method based on focus-zone.

Particularly, we firstly standardize and mark focus-zone for the input original network topology. Then compute the association degree and abstract degree of other subnet and focus-zone through topology association degree model and topology abstract degree model. Finally sample the original network topology through Weighted Forest Fire (WFF) and output the sampled network topology. Through WFF the subnets near the focus-zone will have low abstract-degree and other subnets far will have high abstract-degree. In this way we can simulate accurately the focus-zones and hierarchically simulate other subnets according to their distance, and also reduce the compute complexity of simulation.

1) The model of network topology association degree

In the abstract network topology, we need to ascertain abstract degree using the association degree between two networks. If the association degree between one network and focus-zone is big, we should reduce the abstract degree and simulate it intensively, otherwise, we should increase the abstract degree and simulate it sketchily. We define the association degree as A_{ij} between subnet i and subnet j, A_{ij} is computed as follows:

$$A_{ij} = \alpha W_i/(W_j * \ln(Hop_{ij})) \qquad (1)$$

W_i is the gateway router weight of subnet i, W_j is the gateway router weight of subnet j, and Hop_{ij} is the gateway router hops between subnet i and j, α is the normalized coefficient, $A_{ij} \in (0,1]$.

2) The network topology simplicity model based on focus-zone

In 1999, Faloutsos et al. analyzed triplet data from 1997 to 1998 of NLANR (National Lab for Applied Network Research) and one trace route measure data in 1995, and then found that there exists four power laws on Internet topology[7]. It is inspired from the power laws that in Internet an average property value is not an accurate characterization of this property, we should try to use an index to characterize. Given that, we can hypothesize: the relationship of the abstract degree and association degree can be characterized by some index, as a result:

Power Law 4 (association index A) is: the abstract degree D is proportional to e ^A.

Concretely, if the number of focus-zones of network topology is M, subnet is $\mu_j\,(1 \leq j \leq M)$, the network topology simplicity degree based on focus-zone is D_i, thus D_i is computed as (2):

$$D_i = min\left\{\frac{N_i}{\lambda e^{A_{i\mu_j}}}\right\}(1 \leq j \leq M) \qquad (2)$$

In which, λ is an adjustable parameter, N_i is the number of nodes of subnet i, $A_{i\mu_j}$ is the topology association degree between subnet i and focus-zone μ_j. The greater the abstract degree D_i is, the greater the subnet i abstracted.

3) Forest Fire and its improvement

Traditional Forest Fire sampling method is a kind of "conformation method", that is to say, the generation of topology is a process of "small to large". However in this paper the goal is to sample from topology graph, therefore we improve the Forest Fire into a kind of "sampling method", namely the process of topology is "large to small". Meanwhile, the traditional Forest Fire sampling method doesn't consider the vertex and edge weights, and the importance of nodes and edges of real network topology are different. It is more reasonable to consider the network topology as a weighted graph, as a result, we name a new method as WFF (Weighted Forest Fire) with considering the weight of nodes and edges.

In WFF, we randomly pick a node and begin "burning" outgoing links and the corresponding nodes with a certain probability. The computation of this probability takes the weight of nodes and the distance between nodes into account comprehensively. If a link gets burned, the node at the other endpoint gets a chance to burn its own links, and so on recursively until to the sampling scale.

The algorithm of WFF is:

INPUT: original graph-G, sampling size-size, focus-zone-f

OUTPUT: sampling graph-S

Algorithm:

Step 1: picking a node u in the original graph G randomly

Step 2: picking a neighbor node v which is not burned of u

Step 3: computing the association degree between node u and v as Auv

Step 4: computing the abstract degree of node v as Dv

Step 5: "burning" outgoing link linkuv and the node v with the probability Dv

Step 6: ending the burned nodes u, v and the edge linkuv into the sampling graph S

Step 7: letting u=v, go to Step 2 until to size.

It keeps running recursively until to the sampling scale which equals to size-the number of nodes in sampling graph S.

IV. EXPERIMENTAL EVALUATION

In the following section we present the results of experiments to verify the network topology simplicity method based on focus-zone. First of all, we make a comparison on the topological properties of the sample before and after, and observe that how many topological properties of the original graph can be retained with this method, and then we execute an Internet security event to verify the facticity of simulation results and efficiency. In this paper, the network topology generator selected is BRITE[8], the Internet security event is UDP worm simulation and the simulation tool is GTNetS.

A. Topology model for experiment

We use hierarchical topology model in the experiments. The entire network topology is divided into three levels: the Autonomous System (AS) level, router level and host level, as the following figure shows:

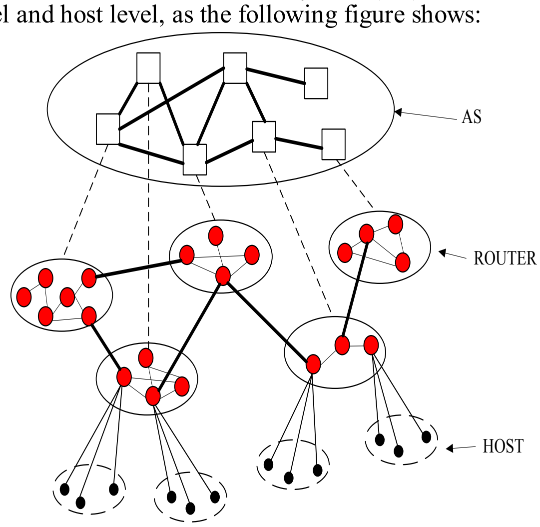

Figure 1. Hierarchical topology model

In Fig.1 the top-level shows the connection between ASes which is represented by a square node. The two adjacent domains are connected through the border gateway router. They are the core nodes of the network. The level below AS is router node, which is represented by the red circle node. Each AS has a certain amount of routers, which are in charge of the packages transition of this region. To reflect the characteristics of scale-free network, the routers are connected with the BA model. And the bottom level is host node which is connected to the gateway router. It is shown with the smallest black circle node.

B. Experiment 1: topology sampling

Based on the comparison between the topological properties of the sample before and after, we measure the following five criterias: average degree of the node, degree distribution, average clustering coefficient, clustering coefficient distribution and bandwidth distribution. And we use hierarchical topology model in the experiment naming the experimental topology as AS-ROUTER, in which AS represents the number of AS and ROUTER represents the number of routers of each AS.

1) Average degree

On average degree, we analyze the topology model 20-100, 20-200 and 20-400, and the Table I shows the data.

TABLE I. Comparison on Average Degree of the Sampling before and after

Topology	40%	50%	60%	70%	80%	90%	100%
20-100	3.53000	3.15200	2.91583	2.77000	2.64250	2.54778	9.76
20-200	3.29500	3.00000	2.80917	2.68393	2.60062	2.52778	9.88
20-400	3.58062	3.40750	3.30146	3.22500	3.17375	3.12111	9.94

From Table I we can see that: the average degree after sampling is obviously smaller than before, which coincides with the sampling algorithm. It is because sampling algorithm uses breadth-first approach to sample and "burning" node by node, when the link of two nodes gets burned, the other links of this two nodes will be ignored. In order to retain the connectivity after sampling, we should reserve the necessary connections. And there is a small percent of nodes with a degree less than 1, as a result, the average degree before sampling is small.

2) Degree distribution

On degree distribution, we analyze the topology model 20-100.In the process of generating 20-100 topology graph, we set the parameter m=5 (Number of neighboring node each new node connects to), theoretically, there is a large percent of nodes with a degree of 5- a "heavy-tailed" distribution. As Fig.2 shows, the degree distribution of model 20-100 before and after sampling is:

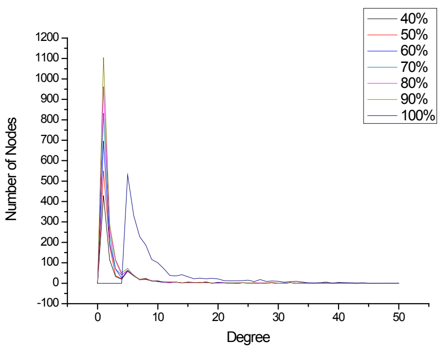

Figure 2. Degree distribution before and after sampling

From Fig.2, we can see that: there is a prominence in the distribution before sampling, that is to say, there is a large percent of nodes with a degree between 5 and 10, which takes on a "heavy-tailed" phenomena. And there are two "prominences" in the distribution after sampling, the small one coincides with the degree distribution before sampling, which shows the degree distribution after sampling retain the properties of original. But there exists a large "prominence" which means there is a large percent of nodes with degree 1. It also coincides with the sampling algorithm.

3) Average clustering coefficient

On average clustering coefficient, we analyze the topology model 20-100, 20-200 and 20-400, and the Table II shows the data.

TABLE II. Comparison on average clustering coefficient of the sampling before and after

Topology	40%	50%	60%	70%	80%	90%	100%
20-100	0.68728	0.66749	0.68068	0.68536	0.68170	0.68665	0.40630
20-200	0.67328	0.67619	0.67134	0.68585	0.68648	0.69304	0.26404
20-400	0.76031	0.76729	0.76742	0.77399	0.78354	0.78615	0.15616

From Table II, we can see that: the average clustering coefficient before sampling is obviously smaller than after, which coincides with the sampling algorithm. In the sampling process, we deal with the original topology according to the extent of concern of the nodes and links, and after sampling we only retain the nodes and links concerned. It is equivalent to a "pruning" to the original topology (wiping off the nodes and links not concerned).

4) Clustering coefficient distribution

On clustering coefficient distribution, we analyze the topology model 20-100 as the Fig.3 shows:

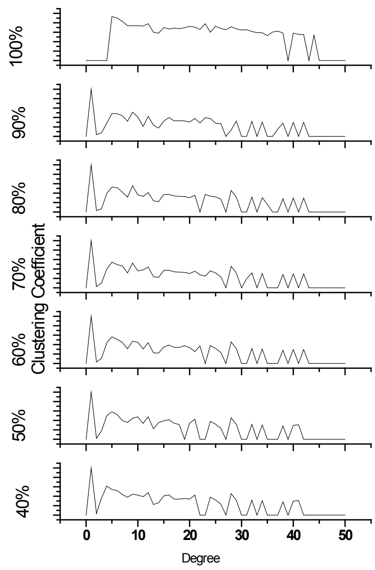

Figure 3. Clustering coefficient distribution before and after sampling

From Fig.3, we can see that: the clustering coefficient before sampling takes on the character-“flat in the middle and volatile at both ends”, and after sampling it almost remains in addition to the flat region which is a little lower than original. It shows that “small group” looks more obviously after sampling as a result of “group burning” (always burning from one node outward).

5) Bandwidth distribution

On bandwidth distribution, we analyze the topology model 20-100 as the Fig.4 shows:

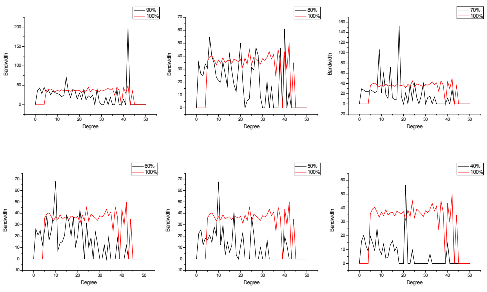

Figure 4. Bandwidth distribution before and after sampling

From Fig.4, we can see that: the difference between sampling before and after is obvious. The higher the sampling degree is, the bigger the difference is. It is because in the original topology bandwidth distribution takes on “heavy-tailed” distribution, while in the process of sampling the routing is random, besides that, bandwidth distribution of the sample is more sensitive. Therefore the difference is obvious.

C. Experiment 2: UDP worm simulation

In UDP worm simulation experiment, the topology model we used is 20-100, namely 20 ASes and 100 routers of each AS. We set each router connected to 64 hosts, thus the original topology has 128000 nodes. In this network we randomly set two focus zones, that is to say, initially a total of 128 host nodes are infected with the worm. The scan rate of UDP worm is 100 times/sec, the payload of each packet sent is 1000 bytes and the simulation time set is 6.00 seconds.

In order to verify the facticity of simulation results, we measure the following five criterias: the infected machine number, the infected machine distribution, the link state, the simulator parameters and the memory usage.

1) The number of infected machine

From Fig.5, we can see that: before and after sampling, the number of host nodes infected has the same general trend as simulation time, that is to say, in the first 2 seconds the number increases with the time and after 2 seconds the number is almost no further increasing. The final number of machines infected reduced with the sampling rate decreases. Meanwhile, the host infection rate after sampling is faster than before, which coincides with the sampling goal-accelerate the simulation speed and reduce the time required for simulation.

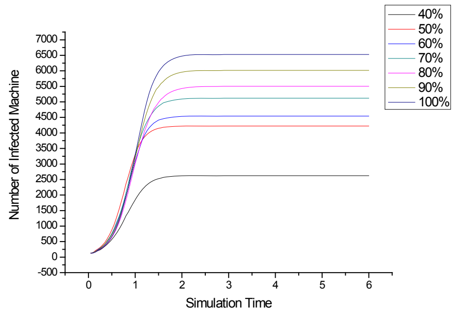

Figure 5. Relation of infected machine number and simulation time

We define infected machine number distortion as $Distortion$, which is computed as follows:

$$Distortion = (Infect_a / Infect_b)/Sample \qquad (3)$$

In which, $Infect_a$ is the infected machine number after sampling, $Infect_b$ is the infected machine number before sampling and $Sample$ is the sampling rate. Table III shows that the distortion is minimum when sampling rate is 40% and maximum when 50%.

TABLE III. Infected machine number before and after sampling

Topology	40%	50%	60%	70%	80%	90%	100%
20-100	2624	4224	4544	5120	5504	6016	6528
distortion	1.00490	1.29412	1.16013	1.12045	1.05392	1.02397	1.00

2) Infected machine distribution

Infected machine distribution describes the distribution of hosts infected with worm in the network topology. It uses focus zone as benchmark and the distance between host node and focus-zone-hop as main measure parameter. As the Fig.6 shows, the distribution curve is generally consistent before and after sampling. But there exists a large “abnormal” when the sampling rate is 40%. Therefore the large rate would have a great impact on distribution.

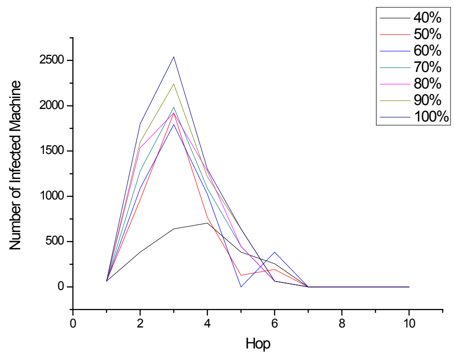

Figure 6. Distribution of infected machine

It is a macro-description in Fig.6, at the micro, the distribution of infected machine in six focus-zone concerned is shown in Fig.7.

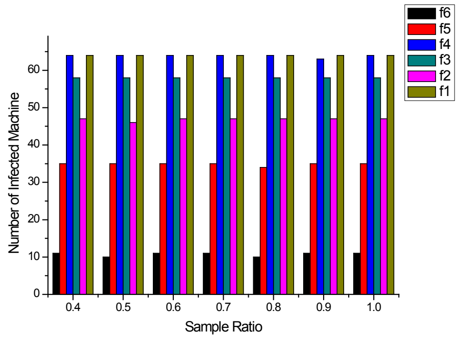

Figure 7. Distribution of infected machine in focus-zone

From Fig.7, we can see that: the number of infected machine before sampling is the same as after sampling. It indicates the sample almost has no impact on focus-ones.

3) Simulator parameters

On simulator parameters, we select the simulation run time, the number of processed events and scheduled events for comparison as the indicators.

From Fig.8, we can see that: with the sample ratio decreasing, the simulation run time almost takes on a linear decline. And it shows the increase on efficiency of network simulation by 11%.

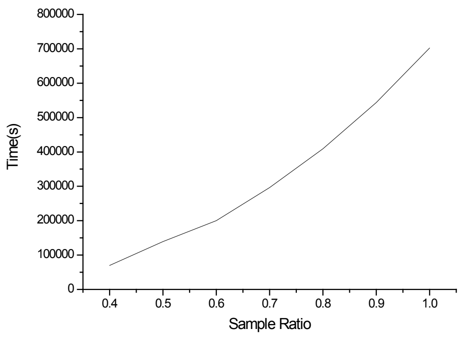

Figure 8. Relation between simulation run time and sample ratio

4) Memory usage

From Fig.9, we can see that: the relation between memory usage and simulation time takes on a "parallel"trend. The memory usage increases with the simulation run time increasing, which shows the number of host nodes infected also increases and a corresponding increase in the packet, so the demand for memory will be more and more. To some extent, this explains the facticity of the sample results.

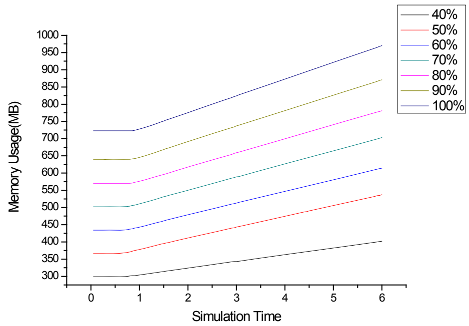

Figure 9. Relation between memory usage and simulation time

V. CONCLUSION

Over the past decade, the Internet developed with an alarming speed and the network size grown exponentially. With the outbreak of network security events become more frequent, the researchers could not wait to predict the state of the network through network simulation. While the computing cost for large scale network simulation is too huge to accept. Thus our work addresses some subtle questions: How to reduce the computing cost for large scale network simulation? How to simulate factually? What do we mean by "realistic"? How do we measure the deviation from realism?

Our contributions can be summarized in the following points:

● We propose an idea: simplifying the network simulation model, using the focus-network security events to sample and increase sample degree, in order to reduce the simulation time and cost.

● Based on the traditional sampling algorithms, only from this point of view: remain the relevant properties of original topology, we propose a method called the network topology simplicity method based on focus-zone. When simplifying network topology and reducing the simulation complexity, we not only consider the reduction of network scale but also consider the actual simulation application, thus improve the facticity of simulation results.

● We propose the network topology association degree model and simplicity degree model, and based on which, we improve the FF algorithm and name a new method as WFF with considering the weight of nodes and edges. Through WFF, we can sample hierarchically, that is to say, the simplicity degree of region near the focus zone is smaller and away from it is greater.

● Inspired from the topology sampling and worm simulation experiment, we give a list of "realism criteria". From the experiment we show that after sampling it can remain topology properties with WFF method. UDP worm simulation verified the network topology simplicity method based on focus-zone can not only guarantee the facticity of simulation results but also increase the efficiency of network simulation by 11%. And the maximum error of the packet drop ratio is 7.8%, which is occurred in the sampling rate of 0.5.

In practical application, we only get the study phase of research results. In order to further improve the performance of network simulation, future work can focus on the following points:

● The research on the lowest percentage of sample, which can guarantee the authenticity of network simulation and simulation efficiency.

● The abstract topology on distributed network simulation. Increasing the process ability of hardware is one of the most intuitive way to improve the performance of network simulation. As a result of strong computing power of distributed system, we can do some research on it. Firstly, according to association degree, we can divide a large scale network topology into some parts. Secondly, in each part, the simulation nodes get sampling and finally form the simulation results.

ACKNOWLEDGMENT

This paper is supported by the National 863 plans projects, China (Project No. 2007AA01Z442, 2007AA01Z406)

REFERENCES

[1] D. M. Nicol, M. Liljenstam and J. Liu. Advanced Concepts in Large-scale Network Simulation. In Proceedings of the 2005 Winter Simulation Conference, Orlando, Florida, 2005:153~166

[2] K. Below, U. Killat. Reducing the Complexity of Realistic Large Scale Internet Simulations. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), San Francisco, USA, 2003:3818~3823

[3] V. Krishnamurthy, M. Faloutsos, M. Chrobak, L. Lao, J. H. Cui and A. G. Percus. Reducing Large Internet Topologies for Faster Simulations. In Proceedings of IFIP Networking 2005, Waterloo, Ontario, Canada, 2005:328~341

[4] V. Krishnamurthy, J. Sun, M. Faloutsos and S. Tauro. Sampling Internet Topologies: How Small Can We Go? In Proceedings of the International Conference on Internet Computing, Las Vegas, Nevada, USA, 2003:577~580

[5] R. Pan, B. Prabhakar, K. Psounis and D. Wischik. SHRiNK: A Method for Enabling Scaleable Performance Prediction and Efficient Network Simulation. IEEE/ACM Transactions on Networking. 2005,13(5): 975~988

[6] Jure Leskovec, Christos Faloutsos. Sampling from large Graphs. Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Philadelphia, PA, USA

[7] Faloutsos M, Faloutsos P, Faioutsos C. On power-law relationships of the Internet topology. ACM SIGCOMM Computer Communication Review, 1999, 29(4):251~262

[8] Medina A, Lakhina A, Matta I, Byers J. BRITE: An approach to universal topology generation. In: Proc. of the MASCOTS 2001. Washington: IEEE Computer Society, 2001. 346~353.

[9] P. Huang. Enabling Large-scale Network Simulations: A Selective Abstraction Approach. Ph.D thesis, University of Southern California. 1999:51~54

[10] Z. Hao, X. Yun and H. Zhang. An Efficient Routing Mechanism in Network Simulation. In Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation, Singapore, 2006:150~157