A hierarchical network model for network topology design using genetic algorithm

Chunlin Wang[1], Ning Huang[1,a], Shuo Zhang[2], Yue Zhang[1] and Weiqiang Wu[1]

[1]*School of Reliability and Systems Engineering, Beihang University, Beijing, China*
[2]*China Ship Development and Design Center, Wuhan, China*

Abstract. Network topology design has directly impact on network construction costs and network performance. Majority of current network topology design take the network physical topology parameters into consideration, such as the reliability and cost constraints, ignoring the actual traffic information on the logical network. Moreover, the network traffic exhibits self-similar feature over large time scales, which is complicated and is difficult to predict. In this paper, firstly, a hierarchical network model is proposed that consists of the upper logical topology and the lower physical topology. The logical topology describes the self-similar traffic information on the network and the ON/OFF model is adopted to model the self-similar traffic. The lower physical topology represents the connection relationship between the all kinds of network devices and links. Then, taking advantage of the hierarchical network model, a novel network topology design method based on the genetic algorithm is proposed, which aimed at obtaining the network with minimum delay under certain reliability and cost constraints. Finally, a practical example is presented to verify the effectiveness and the accuracy of our network topology design method. Results show that our method obtains better results than the other methods.

1 Introduction

With the development of network communication technique, the types of network equipment are increasing and the network size is becoming larger. More importantly, the increasing complexity of network traffic appears on the communication network. All these factors have increased the difficulty of network topology design. However, the topology of the communication network is decisive; once improper, the reliability will decrease and the network performance will also decrease [1]. For the communication network, the real-time performance, reliability and construction cost are three decisive indicators which need to be considered in the network design phase [2, 3].

There have been many related researches on the communication network topology design. And these works can be generally divided into the following two categories. The first category focuses on the physical topology design. They often consider the topology design for minimizing the total cost subject to a constraint on the network reliability or for maximizing the network reliability while a constraint on the total cost [4]. For example, Shao [5] proposed a forest research algorithm to optimization of a computer network expansion with a reliability constraint. Altiparmak [6] and Pierre

---

[a]Corresponding author : hn@buaa.edu.cn

[7] introduced a genetic algorithm-based scheme to find an optimal solution when maximized the network reliability subject to certain cost constraint. The above related researches often focus on the physical network, ignoring the network traffic transmission on the network. The second category considers the co-design of both physical topology and logical topology. Such as, Fencl [8] tried to optimize the physical and logical topology simultaneously, and real-time performance and fault-tolerance level are considered in the LAN topology design. Tian [1] took advantage of the bi-level programing theory [9] to optimize the reliability and the real-time performance simultaneously under cost constraint. Usually, such simultaneous design is treated as a multi-objective process [10], which is hard to ensure the ultimate optimal solution.

From the related research, the second category begins to consider the network traffic in the topology design. However, the traffic they described is too simple and the network size is small. In [2, 8, 10] the traffic load as a known constant value in advance and then the traffic matrix is applied to the network. But the traffic on the communication networks is complicated and is difficult to predict [11]. Moreover, the self-similar feature of the traffic has been proved to be the most important feature of network traffic [12, 13], which is ignored in the network topology design.

In this paper, a hierarchical network model is proposed to solve the communication network topology design problem. The lower level is physical topology network. The upper level is logical topology which is the result of self-similar traffic transmission. The self-similar traffic transmission determines the network real-time performance, and the physical topology network is subject to the total cost and reliability constraints. Then, a network topology design method using the genetic algorithm is presented, which aimed at obtaining the network with minimum delay under certain reliability and cost constraints. Finally, an illustrative example is presented to verify the effectiveness of our communication network topology design method.

2 Hierarchical network model

Networks tend to have a two-layer or multi-layer structure in the real-world. For the communication network, the two-layer structure is commonly used in current research [14, 15]. Here, a hierarchical network model is proposed to solve the network topology design problem. The upper layer is the logical topology and the lower layer is the physical topology in our model. The lower physical topology is the actual network, including the physical devices and links. The upper logical topology represents the different traffic flows, which will determine the network performance. The hierarchical network model can be seen in Figure 1.
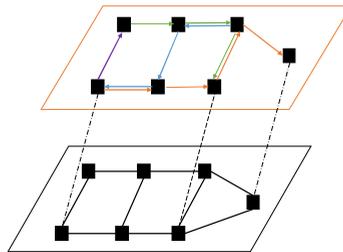

Figure 1. Hierarchical network model.

2.1 Physical Network topology

The lower physical topology describes the connection relationship between all kinds of network devices and links. The communication network can be modeled as an undirected graph G, with N nodes and E edges. Mathematically, it can be depicted by an $N \times N$ adjacency matrix $\{w_{ij}\}_{N \times N}$. If there is an edge between node i and node j, the $w_{ij} = 1$; otherwise $w_{ij} = 0$. For the actual network design, the network resources are often limited due to certain cost constraints. Moreover, the network must meet

some basic reliability requirements, such as several specified nodes can be connected. In this paper, the cost constraint and reliability requirement are both taken into consideration in our network topology design. For the cost constraint, it is reflected in the constraint on the total number of links in the network. That is because the number and the price of the nodes often have been identified before the topology design. Essentially, the network topology design is to determine the connection relationship between each node. Then the total number of links is an important parameter to reflect the cost constraint. For the reliability requirement, all nodes in the network need to be able to communicate with each other, primarily because the connectivity is the basic premise for network operation.

Then, the cost constraint and reliability requirement are two main considerations in our network topology design, which limit the total number of links and require the network connectivity respectively.

2.2 Logical Network topology

Logical network describes the different traffic flows on the physical network and the self-similarity of network traffic has been proved to be the most important feature of network traffic. Different self-similar traffic models have been proposed in related research, such as the heavy-tailed ON/OFF model, wavelets model and α-stable traffic model. Among these models, the heavy-tailed ON/OFF model has been widely used to generate self-similar traffic due to its mathematical simplicity and well explanation of the self-similar feature.

For the ON/OFF model, it assumes that the source alternates between an ON-period and an OFF-period [16]. During the ON-period, packets are generated and sent from the source to the destination at a constant rate *v*, while during the OFF-period, no packets are transmitted. This process can be called as alternating renewal process. Generally, each source is considered as identically independently distributed for the ON-period and OFF-period. Then, it is only need to specify the distribution of durations of the ON-period and OFF-period for each source.

Pareto distribution is a typical heavy-tailed distribution. Let $f(x)$, $F(x)$ denote the probability density function and the cumulative distribution function for the durations of ON-period and OFF-period. The probability density function (*pdf*) of Pareto distribution is:

$$f(x) = \alpha k^{\alpha} x^{-\alpha-1}, 0 < k \le x \tag{1}$$

The cumulative distribution function can be obtained as:

$$F(x) = P[X \le x] = 1 - (k / x)^{\alpha} \tag{2}$$

Where k represents the minimum value of the random variables, α determines the mean and variance of the random variable.

The self-similar traffic transmission determines the network real-time performance. Here, an average packet delay [17] is introduced to evaluate network real-time performance. It can be calculated as follows:

$$T(x) = \frac{1}{\gamma} \sum_{i=1}^{m} \frac{f_i}{C_i - f_i} \tag{3}$$

Where γ is the total arrival rate into the network in bits per second; m is the number of links; f_i and C_i are the assigned traffic load and capacity of link i respectively.

2.3 Problem Formulation

According to our hierarchical network model above, network topology design can be regard as a network topology optimization design problem with certain constraints. The objective function is the

average packet delay which is decided in the logical network. The constraints include the cost constraint and reliability requirement in physical network. That is to say, the network topology should be designed that the objective function is optimal under certain cost and reliability constraints. This problem can be described as:

$$\begin{aligned} Min \quad & T(x) \\ s.t. \, & Cost(G(N,E)) \leq C_{max} \\ & Rel(G(N,E)) \geq R_{min} \end{aligned} \tag{4}$$

Where $Cost(G(N,E))$ is the actual construction cost of links in network, which is reflected by the number of links in later examples; C_{max} is the maximum value of cost; $Rel(G(N,E))$ is all-terminal connective reliability in this paper; R_{min} is the minimum value of network reliability.

This network design problem has been shown to be NP-hard [2]. Then, several heuristic algorithms have been proposed to deal with the specific topology design problems, including Genetic Algorithm (GA), Neural Network, Tabu Search, and Ant Colony Optimization. Especially, genetic algorithm has a great advantage in efficiency and implementation of solving the optimization problems with constraints and multi-objective. In this paper, we employ a genetic algorithm for our network topology design.

3 Network topology design

A genetic algorithm is based on the theory of natural evolution and has the following steps: encoding and decoding, initial population generation, evaluation, selection, crossover, mutation. First of all, a set of the initial individuals is generated, forming the initial population. Then, a series of genetic operators modifies this population repeatedly, including the selection, crossover and mutation. The modified population is the offspring for the current population. At last, find out the optimal solution in the iterative process, thanks to the fitness function.

3.1 Encoding and Decoding

The chromosome that describes network topology is a vector *i* with *l* elements.

$$i = \{i_j : i_j \in \{0,1\}\}, j = 1,...,l$$

Where *N* is the number of nodes in the network. $l = N(N-1)/2$. Vector *i* characterize the upper triangular part of the adjacency matrix *P* because the adjacency matrix is symmetrical. The specific encoding and decoding process can be briefly described from the following example:

The chromosome is the upper triangular part of the adjacency matrix *P*:

Then, the vector $\{1\,1\,0\,0\,0\,1\,1\,1\,1\,0\}$ is the chromosome which can represent the network topology in Figure 2. For the decoding process, the chromosome is converted into a matrix through a reverse process.
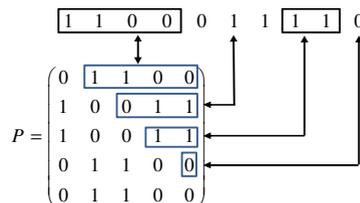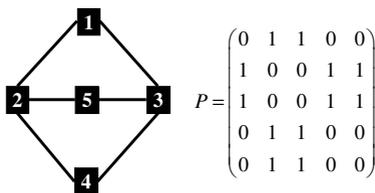

Figure 2. An example topology and adjacency matrix.

Figure 3. The chromosome encoding.

3.2 Initial Population

Generally, the initial population is generated by completely random topologies. However, the random topologies can't guarantee all-terminal are connected for our topology design. Reference [18] proposed a feasible problem solution for the initial population generation. Firstly, design a ring topology connecting all nodes of the network, thus guaranteeing that all initial solutions are feasible. The ring topology is randomly generated for each individual of the initial population. Afterwards, *t* links are added to the ring topology, connecting *t* pairs of randomly selected nodes.

3.3 Fitness Function

We revise the objective function (average packet delay $T(x)$) as a fitness function of the GA, denoted by $Fit(x)$. The fitness function can be expressed as (5) because the lower the objective value, the higher the fitness will be:

$$Fit(x) = \begin{cases} C - T(x), & if \quad T(x) < C \\ 0, & if \quad T(x) \geq C \end{cases} \tag{5}$$

Where C is a relatively large constant value, and the selection of this value can refer to the allowed maximum value of objective function.

3.4 Selection, Crossover and Mutation

In the selection phase, pairs of individuals are chosen for crossover. Usually, individuals are selected based on their fitness value. In this paper, the selection process is based on a roulette-wheel selection scheme, meaning that chromosome with larger fitness value has greater probability of being selected into the mating population.

In the crossover operation, pairs of individuals previously selected are combined, giving rise to another pair of new individuals (offspring). Here, the single point crossover method is selected for the crossover operation. Firstly, a crossover place is randomly selected for the two selected individuals. Then, the two left sides of individuals are copied to offspring 1 and offspring 2, respectively. The right sides of each code shall be exchanged for the offspring1 and offspring 2. An example is displayed in Figure 4.
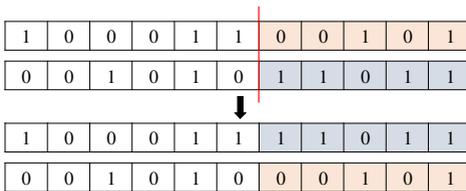

Figure 5. Mutation operation.

Figure 4. Crossover operation.
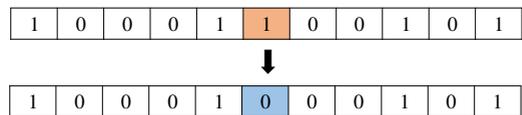
In the mutation operation, a mutated chromosome is chosen at random according to the mutation probability. Then, a simple exchange of 0's to 1's, or vice versa, at random locations of the genetic code. This mutation operation has the goal of increasing the diversity of the population. An example is shown in Figure 5.

After these series of operations (selection, crossover and mutation), the next generation is created. Taking the cost constraint of the physical topology into consideration, some newborn individuals may not satisfy this constraint. So, it is needed to analyse these newborn individuals and discard those that

do not meet the constraints. At last, an optimal network topology can be obtained through certain number of iterations.

4 Numerical results and analysis

4.1 Illustrative Example

The test of our topology design is made for a medium sized network with the number of nodes *N*=16. The related network traffic assumptions as follows: Each node can be regarded as an ON/OFF source which sends certain packets to all other nodes in the network. The packet transmits along the shortest path from the source to destination. And the related parameters for ON/OFF traffic model are: $\alpha = 1.1$, $k = 1$, $v = 1$.

For the test of network topology design, our objective function is the average packet delay $T(x)$ and the physical layer has a number of constraints, including the number of edges in the network $M < 30$, which reflects the network cost constraint, and edge capacity $C_i = 60$.

Lastly, the experiment was done with the following GA parameters: population size $pop\_size = 50$; crossover probability $P_c = 0.4$; mutation probability $P_m = 0.05$; the number of iterations is 200. The constant value $C = 5$ in the fitness function.

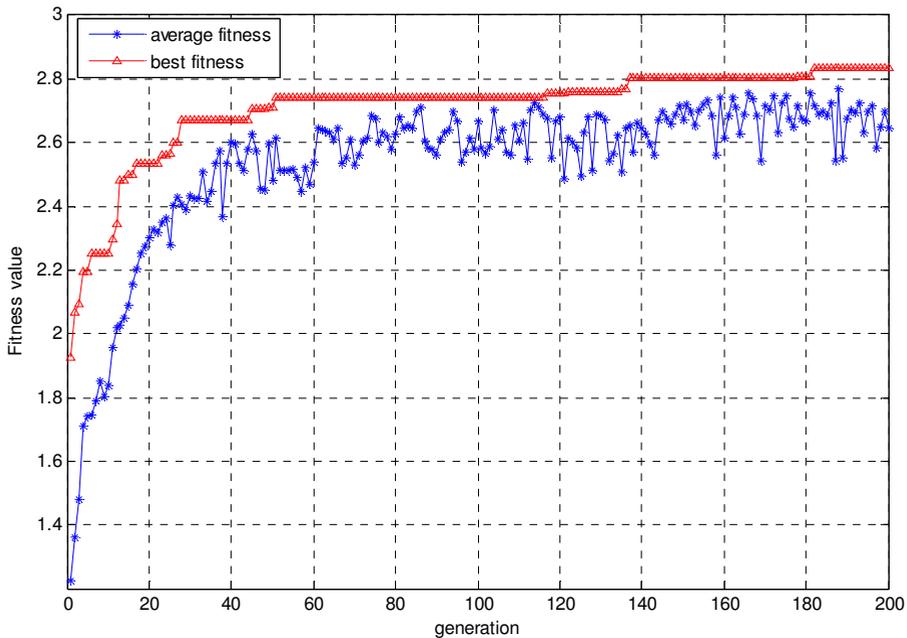The result of genetic algorithm for our topology design is shown in Figure 6.

Figure 6. The results of example.

In Figure 6, the red curve stands for the best fitness value in each generation and the blue curve represents the average fitness value for each generation. It can be seen that the best fitness value reaches the maximum value 2.83 and tends to converge after about 180 generations. Then the optimal chromosome in the iterative process is:

10000001000000110100000010000000001000000010001001000010000
00000100100100111000000010000000110010110000100101000100101

The optimal network topology also can be obtained after decoding the optimal chromosome. The adjacency matrix of the optimal network is as follows:

$$
\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
$$

4.2 Accuracy verification

To illustrate the accuracy of our approach, numerical experiments of the network topology design under different cost constraints are also provided. The number of edges in the network is limited as $M < 30$, $M < 32$, $M < 34$ respectively. And the result is as follows:
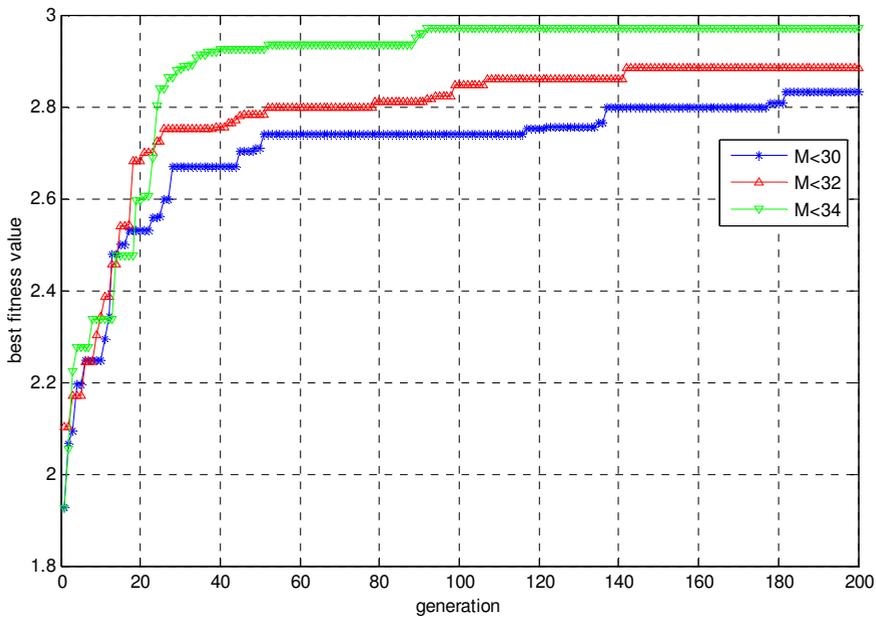

Figure 7. The results for different M.

With the increase of the number of edges *M*, the maximum of the best fitness value is increasing. The optimal solutions under different constraints can be obtained by using our topology design method. And the results of average packet delay *T(x)* for these three optimal topologies are shown in Table 1.

Table 1. The average packet delay

	M<30	*M<32*	*M<34*
average packet delay *T(x)*	*T*=0.0342	*T*=0.0333	*T*=0.0319

As shown in Table 1, with more edges are allowed in the network design (more cost is allowed), the generated network topology has better performance. This conclusion is consistent with the actual situation.

Moreover, several method comparisons also have been provided to demonstrate the accuracy of our topology design approach. Firstly, the classic square lattice network model is introduced to analyse comparatively (seen as Figure 8). Then, the second is the topology generation method proposed in Reference [18], which is adopted for our initial population generation. We guarantee the same network constraints for different approaches and then compare the network performance of different methods. The result is shown in Table 2.
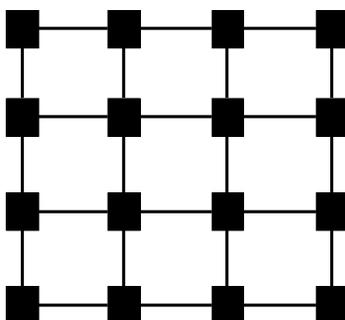

Figure 8. Lattice network model.

Table 2. The comparison results

	Our method	**Lattice network**	**Reference [18]**
average packet delay *T(x)*	*T*=0.0412	*T*=0.0507	*T*=0.04677

It can be seen that the network average packet delay of our design method is the smallest under same design constraints. That is to say, the topology design method proposed in this paper has better effectiveness and practicality compared to the other methods.

5 Conclusion

In this paper, a general hierarchical network model is proposed to describe the problem of communication network topology design. The lower level of the hierarchical model describes the cost and reliability constraints. And the upper level determines the network performance when self-similar traffic is loaded on the network. Based on the genetic algorithm, the topology design solution is presented in the paper, which is aimed at designing the optimal network topology under certain reliability and cost constraints. Finally, a practical example is presented to illustrate our topology design method. What's more, numerical comparative analyses also have been carried out to verify the accuracy of our proposed topology design method. The result justifies that our method has certain guiding significance to communication network topology design. Our future research will focus on some more realistic physical layer constraints, such as considering the distance between nodes and the heterogeneity of nodes.

References

1.   Y. Tian, H. Dong, L. Jia, Y. Qin, and S. Li, Math. Probl. Eng, **2014** (2014)
2.   L. Zhang, M. Lampe, and Z. Wang, Frequenz, **66**,159 (2012)

3. Y. Jiang, R. Li, R. Kang, and N. Huang, International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, 245 (2012)
4. F.M. Shao, X.S. Shen, and P.H. Ho, IEEE Trans. Reliab., **54**, 421 (2005)
5. F.M. Shao and L.C. Zhao, Comput. Math. Appl., **35**, 17(1998)
6. F. Altiparmak, B. Dengiz, and A.E. Smith, IEEE International Conference on Systems, Man, and Cybernetics, **5**, 4676 (1998)
7. S. Pierre and G. Legault, IEEE T. Cybern., **28**, 249 (1998)
8. T. Fencl, P. Burget, and J. Bilek, Control Eng. Practice., **19**, 1287(2011)
9. Z. Gao, J. Wu, and H. Sun,Transp. Res. Pt. B-Methodol., **39**, 479 (2005)
10. C.S. Wang and C.T. Chang, IEEE-ACM Trans. Netw., **16**, 680 (2008)
11. Y. Wang, C. Tian, S. Wang, and W. Liu, Internet Conference of China., **80** (2014)
12. W.E. Lel, M.S. Taqqu, W. Willinger, and D.V. Wilson, IEEE-ACM Trans. Netw., **2**, 1 (1994)
13. V. Paxson and S. Floyd, IEEE-ACM Trans. Netw., **3**, 226 (1995)
14. M. Kurant and P. Thiran, Phys. Rev. Lett., **96**, 138701 (2006)
15. Y. Zhuo, Y. Peng, C. Liu, Y. Liu, and K. Long, Physica A., **390**, 2401 (2011)
16. W. Willinger, M. S. Taqqu, and R. Sherman, IEEE-ACM Trans. Netw., **5**, 71 (1997)
17. E. Szlachcic, IEEE International Conference on Dependability of Computer Systems., 150 (2006)
18. J.F. Labourdette, E. Bouillet, and R. Ramamurthy, IEEE-ACM Trans. Netw., **13**, 906 (2005)