

Parallel Evolutionary P2P Networking Technique Evaluating Network Topologies Sequentially

Kei Ohnishi* and Mario Köppen*

*Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology
680-4 Kawazu, Iizuka-shi, Fukuoka 820-8502, Japan
Email: {ohnishi@cse, mario@ndrc}.kyutech.ac.jp

Abstract—We previously proposed the parallel evolutionary P2P networking technique to realize an adaptive large-scaled P2P network. This technique divides all nodes composing a large-scale P2P network into multiple node groups, and then reconstructs the network topologies by an evolutionary algorithm in each node group. In this technique, a timing at which a set of network topologies (a population of the evolutionary algorithm) is evaluated and reconstructed by evolutionary operators is the same for all the node groups. However, this simultaneous evaluation and topology reconstruction has been shown to be not effective in terms of adaptation of the network topologies to users' (nodes') demands in our previous study. Therefore, in the present paper we propose a method that evaluates and reconstructs a set of network topologies in a one-by-one manner. Simulation results show that the proposed method improves the performance of the parallel evolutionary P2P networking technique conducting simultaneous evaluation and topology reconstruction.

I. INTRODUCTION

The recent growth of information and communication technologies, for example, progress of communication speed, storage capacity, and processing speed, is remarkable. Under this quantitative technological progress, degree of dependence of humans on networks becomes much higher. For instance, the number of users of the Internet in the world reaches around nine hundred and sixty five million in 2005, while the number in 1991 is just around four million [1]. While the quantitative technological progress is being made continuously, the quantitative characteristics of humans as a leading part of the network society, such as calculation speed, has not changed so far and will not change in future basically. Thus, the difference in quantitative characteristics between humans and networks is becoming larger, so that it would be needed for sustainable growth of humans as well as networks to consider a relationship between humans and networks deeply.

One technique that considers a relationship between humans and computers, between which quantitative difference are also becoming larger, is interactive evolutionary computation [2][3]. The technique implements an evolutionary algorithm as an optimization method in a computer that can quickly conduct fixed procedures by including humans as an evaluation system of optimization problem. This technique enables us to optimize parameters of system whose outputs are evaluated only by humans by means of cooperation between humans and computers. Meanwhile, there is few technique considering a relationship between humans and networks at this moment.

In our previous study, we thought that for sustainable growth of the network society, a network adapting to demands

of humans who are a leading part of the network society is needed. Then, we proposed an evolutionary peer-to-peer (P2P) networking technique that reconstructs network topologies based on fitness values given by nodes (users), which is called *EP2P* hereinafter [4]. However, the basic evaluation of EP2P in [4] did not consider a large size of networks as seen in the real P2P networks. For example, LimeWire [7], which is one of the real P2P file sharing networks and is prohibited to distribute the LimeWire software since October 26, 2010, was reported to hold several millions of users simultaneously. EP2P needs a special node called a super node that plays a role of collecting fitnesses from nodes and executing EA to adaptively change the P2P network topologies, but it can happen that the super node is overloaded as the number of nodes increases and then the P2P network stops working due to that. That is the problem of EP2P.

So, as a solution of the problem of EP2P, we then proposed a parallel evolutionary P2P networking technique, which is called *P-EP2P* hereinafter, that first divides an entire P2P network into several smaller networks to avoid the overload of a super node and then applies EP2P to each of the small networks to make the entire network adaptive [5]. However, even P-EP2P has a problem such that search failure rate becomes higher as the number of divided node groups increases. More precisely, in P-EP2P, timing of evaluating and reconstructing network topologies is the same for all node groups, but the simultaneous evaluation and topology reconstruction causes inefficient adaptation of network topologies to users' demands.

In this paper we propose a new method for evaluating and reconstructing network topologies in P-EP2P to simultaneously achieve load balancing of super nodes and low search failure rate. The proposed method evaluates network topologies in a one-by-one manner with some time interval between two sequential evaluations. In addition, we examine if the proposed method improves conventional P-EP2P in terms of search failure rate when the number of node groups increases.

EP2P and P-EP2P are meant to dynamically and evolutionarily optimize several coexisting topologies of a running P2P network using evaluation values obtained from the actual nodes (users). No such networking technique has yet been reported. However, a number of methods for local topology reconstruction in a sole P2P network topology based on the observation of local network states have been proposed [10][11][12]. EA has been used to optimize the parameter values of a P2P network using fitnesses obtained from a simulation model of the P2P network [8][9]. Unlike EP2P and P-EP2P, this is not an online approach to optimizing the

parameters of P2P networks. In addition to P2P networks, EA has been applied to on-line optimization of communication networks, such as on-line optimization of routing tables of routers in the Internet [13] and that of protocol stacks [14].

The remainder of the present paper is organized as follows. In Section II, we describe P-EP2P that is a target to be improved. Section III proposes a method for evaluating and reconstructing network topologies sequentially. Section IV shows the simulation results of evaluation of the proposed method. Finally, Section V presents conclusions and describes areas for future research.

II. PARALLEL EVOLUTIONARY P2P NETWORKING TECHNIQUE

P-EP2P divides all of nodes into multiple node groups and then applies EP2P to each group. The overview of P-EP2P is shown in Figure 1. The method for evaluating and reconstructing network topologies proposed in the next section is related to a timing at which network topologies are evaluated and reconstructed when applying EP2P to each node group. Thus, understanding P-EP2P is needed to understand the method proposed in this paper. Therefore, we explain P-EP2P in this section.

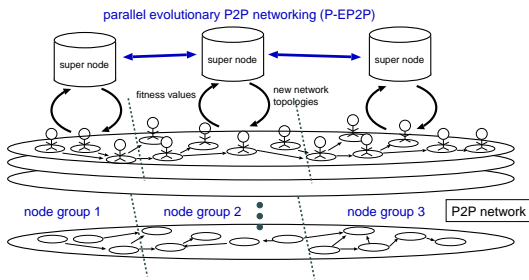


Fig. 1. Overview of the parallel evolutionary P2P networking technique.

A. Network Composition

As shown in Figure 1, a network using P-EP2P is composed of a P2P network that includes several network topologies, in which all of the nodes are included at the same time, and multiple super nodes, in which EA is used to optimize the topologies. P-EP2P first divides an entire network into several smaller networks. Let N_G be the number of node groups, which are obtained by dividing the entire network, and G_k be the number of nodes in the k -th node group ($k = 1, 2, \dots, N_G$). Then, one super node is assigned to each node group.

The actual role of the super node is (1) to determine links for a node in its node group that joins the network for the first time, (2) to reconstruct the network topologies of its node group by executing the EA, and (3) to manage which nodes in its node group join the network at each moment. The super node does not manage which services each node in its node group can provide to other nodes. For example, in a P2P file-sharing network, the super node does not manage which files each node holds. However, the P2P nodes communicate their joining and leaving the network to the super node. Thus, the super node can determine which P2P nodes in its node group

have joined the network and whether these nodes are currently in the network.

B. Fitnesses Assigned by Nodes

A P2P node in each node group uses all network topologies to which it is included for time period T and then assigns a fitness to each of the topologies. The fitness of each network topology is set to zero initially and at every time interval T . Otherwise, each network topology basically increases the fitness by being used by the nodes.

When a P2P node searches the P2P network for P2P nodes that can provide the desired service, this P2P node uses all of the P2P network topologies in which it is included for the search. Therefore, it is possible that within a given allowed number of hops, H_{max} , the P2P node can find the desired service in some topologies while not being able to find the service in other topologies. If the desired service is not found in a certain topology, the fitness of the topology is increased by one. Otherwise, the fitness does not change.

If the above-mentioned search and assignment of fitnesses are conducted for a period of time T , each topology will be assigned a certain fitness in each node group. Then, the topologies with smaller fitnesses are regarded as better in the EA used herein.

In P-EP2P, one super node is assigned to each of N_G node groups and gathers fitnesses only from G_k ($k = 1, 2, \dots, N_G$) nodes that belong to its node group. All L nodes simultaneously belong to N network topologies and are divided into N_G node groups. Therefore, it can happen that the node groups assign different fitnesses to the identical network topology.

C. Representations of Network Topologies

In EA, a solution candidate for an optimization problem is represented in an alternative form. This alternative form is designed by a person who is attempting to solve the problem using the EA and is referred to as a genotype or an individual. Meanwhile, a solution candidate itself is referred to as a phenotype in the EA. In P-EP2P, the P2P network topology is an object of optimization and an individual is an alternative form of a P2P network topology.

Suppose that a P2P network consists of L nodes. The P2P network topology assumed herein is generated by having each of the L nodes make N_C directed links to other nodes. Therefore, an individual is an internal representation of this network topology in the EA. The EA individual used here is shown in Figure 2. As shown in Figures 2, in which N_C is 2, the individual is a one-dimensional vector with $L \times N_C$ elements. An element in an individual is generally referred to as a gene in the EA.

Each node is assigned a serial number as its identifier, and when $N_C = 1$, the identifier corresponds to the index of the vector representing the individual. When $N_C \geq 2$, the identifier corresponds to the index representing each chunk of N_C elements. An element value of the individual represents an identifier of the node to which a focus node makes a directed link. A direction represented by a directed link indicates that a search query can be forwarded only in that direction. Thus, if flooding is used as a query forwarding method, a search

query generated at some node is forwarded node by node in the direction represented by the directed links, and the paths for forwarding the query (flooding tree) are then determined accordingly. However, when data, such as a file, is found during this search, the node having the object transmits the data to the node making the query by means of a direct communication.

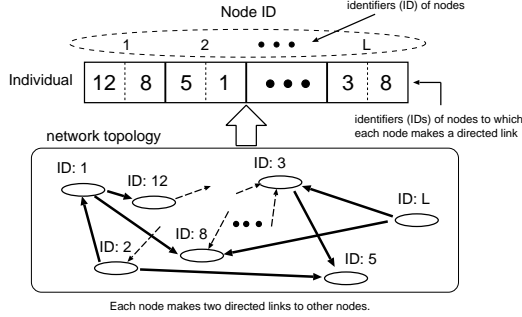


Fig. 2. Representation of a P2P network topology in the EA (EA individual) when $N_C = 2$.

D. Evolutionary Operators

Evolutionary operators are applied to the set of individuals mentioned above, which is referred to as a population, in order to generate a new set of individuals, which is referred to as the new population. The number of individuals held in the EA, i.e., the population size, is N . Evolutionary operators generally include a selection operator, which is inspired by natural selection in Darwinism, a recombination or crossover operator, which models genetic recombination, and a mutation operator, which models gene mutation. The evolutionary operators used in P-EP2P are explained below.

1) *Selection*: The selection operator used herein is a tournament selection with a tournament size of K . The tournament selection randomly selects K individuals from the EA population and selects the individual with the best fitness among the K individuals. This selection procedure is repeated until N individuals have been selected. N is a population size and an even number. In each node group, the tournament selection is conducted using fitnesses assigned to individuals encoding network topologies. Since it can happen that the node groups provide different fitnesses for identical individuals, selected individuals can be different in the node groups.

2) *Crossover*: The crossover operator used in the present study is hereinafter referred to as *node linkage crossover* (NLX). NLX is applied to the selected individuals by the tournament selection in each node group as follows. A range within which NLX is applied in each node group are the vector elements (the loci) that are correspondent to the nodes of that group.

- 1) N individuals selected by the selection operator are divided into $N/2$ pairs of individuals. The selected individuals become parent individuals in this generation.
- 2) The crossover operator is applied to each pair of parent individuals with probability p_c , which is referred to as the crossover rate. Child individuals generated

from each pair of parent individuals are identical to the parent individuals before the crossover operator is applied. Each parent individual has a corresponding child individual.

- 3) For each pair of parent individuals to which the crossover operator is applied, one element is randomly selected from among the L elements of the individual. Recombination is conducted for the selected element with probability p_e .
- 4) For the element to which the recombination is to be applied, which child individual corresponding to one parent individual receives the element values of the other parent individual to be copied on itself is decided randomly.
- 5) After deciding which parent individual provides the element values for recombination, the node (element value) linkage generated by directed links between nodes is copied to the target child individual.

Figure 3 shows an example of NLX with $N_C = 2$. In Figure 3, the third node has been selected as the initial node of the linkage. However, since each node makes two directed links, the third node has two elements that can be referred to by NLX, which, in this example, are 10 and 1. Then, NLX randomly chooses one of the two possible elements and refers to the value of the selected element, which is 10. Next, since the second node of the linkage, which is the tenth node, also has two elements, NLX randomly chooses one of the elements and refers to the value of the selected element, which is 2. In this way, the node linkage is formed. Generally, when $N_C \geq 2$, NLX is performed in the same manner.

However, when there are multiple node groups ($N_G \geq 2$), the way becomes complicated. For example, suppose that a node a_1 in a node group 1 of focus makes a directed link to a node b_2 in another node group 2 and the node b_2 makes a directed link to a node c_3 in another node group 3. Then, we will consider a copy of the linkage among these nodes, $a_1 \rightarrow b_2 \rightarrow c_3$. In case that those nodes belong to the same node group, the linkage of $a_1 \rightarrow b_2 \rightarrow c_3$ can be copied. However, in case that those nodes belong to different node groups, the linkage from the node a_1 , which is in the node group 1 of focus, to the node b_2 , which is outside the node group 1, is copied, but the linkage from the node b_2 to the node c_3 , which is also outside the node group, cannot be copied.

In one attempt of NLX, when the number of times of recombination has not reached N_L yet and a linkage from a node in a node group of focus to a node in another node group appears, a linkage from the node in that other group to some node can not be copied, as mentioned above. In this case, one new vector element is selected from all of the vector elements in the node group of focus to be copied. The node linkage again starts from that selected element (node). One attempt of NLX is finished when the total number of the vector elements copied becomes N_L .

- 6) Repeat Steps (3) through (5) $N_C \times (L/N_G)$ times.

3) *Mutation*: The mutation operator used herein is such that the value at each element (the gene) on the N individuals

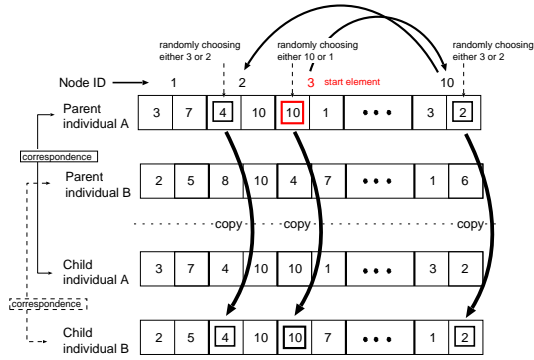


Fig. 3. Example of node linkage crossover (NLX) when $N_C = 2$.

obtained after the node linkage crossover (NLX) is randomly changed to some other possible value with probability p_m , which is referred to as the mutation rate. The element value of the individual represents an identifier of a node to which the node corresponding to the element position is linked, so that the mutation operator changes a node to which the node of focus is linked. Since all the super nodes exchange information on which nodes are present in their networks among them, every element value of the individual can become one of the identifiers for all of the nodes in the entire network by the mutation operator. This mutation operator is introduced mainly for bringing novel genes that did not appear in the initial population. In addition, if we set the mutation rate to be higher, the P-EP2P approaches to a random method.

E. Timing for Topology Generation

The EA population obtained after applying the evolutionary operators is transformed into a new set of P2P network topologies on the super node mentioned above, and the nodes to which each node must make directed links are then communicated to each node in the network. The nodes then make directed links to other nodes according to this information. Nodes that are not present in the network at this moment obtain information on nodes to which they must link upon joining the network.

III. SEQUENTIAL EVALUATION METHOD FOR NETWORK TOPOLOGIES

P-EP2P proposed in our previous study [5] conducts collection of fitness values to network topologies from nodes and evolutionary reconstruction of the network topologies based on the fitness values at the same time for all node groups. In [5], it was suggested from simulation results that change in network topologies of some node group can cause bad effect on fitness values of other node groups, that is, evolution of network topologies that each node group manages does not occur harmoniously. In the conventional P-EP2P, all node groups can give a different fitness value to one network topology, and then each node group can select and modify pieces of different network topologies for the next generation. Then, the entire network topologies are formed by randomly combining such pieces of different network topologies that all node groups selected. There is no basis that this random combination of pieces of different network topologies yields better network

topologies. The problem of P-EP2P mentioned above is also shown in Figure 4.

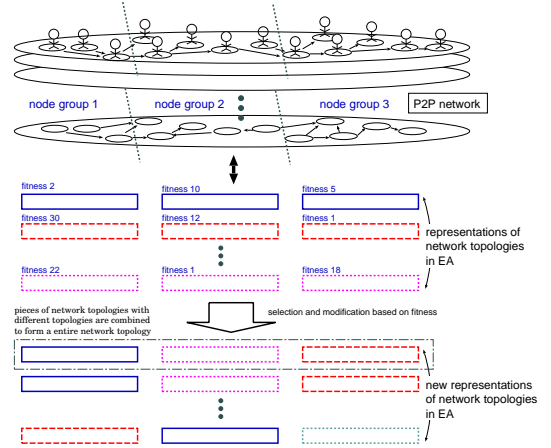


Fig. 4. Problem of the parallel evolutionary P2P networking technique.

Meanwhile, the method proposed in this section evaluates and reconstructs network topologies one by one in a fixed order. This method allows a node group to reconstruct its own network topologies under the condition that network topologies of other node groups are fixed.

Procedures of P-EP2P including the proposed method for evaluating and reconstructing network topologies are as follows: (1) all nodes are divided into N_G node groups, (2) nodes in each node group uses the network topologies for a fixed time period of T , (3) a super node in each node group gathers fitness values from its nodes, (4) only a node group that is now at its turn evolutionarily reconstructs its network topologies, where order of the topology reconstruction for all node groups is determined in advance, (5) returns to (2). For example, suppose that two node groups, A and B, exist. Then, only the node group A first evolutionarily reconstructs its network topologies at time T using fitness values that were obtained from the real use of the network topologies by its nodes from time 0 to T . Next, only the node group B reconstructs its network topologies at time $2T$ using fitness values gathered from time T to $2T$. Then, the turn comes to the node group A again. This procedure is repeated.

IV. SIMULATIONS

In this section, we compare P-EP2P using the sequential evaluation method for network topologies proposed in Section III and the conventional P-EP2P through simulations.

A. Simulation Model and Configurations

To focus only on the examination of the relationship between the number of node groups and the adaptability of network topologies as much as possible, we simplify other things. We consider that all of the nodes are always present in the network without leaving during the simulation period. In addition, it is assumed that all of the nodes excluding a node as a search object search the entire network for only one node to receive some service over the simulation period of time. The node as a search object does not conduct search. The above-mentioned assumptions are not practical, but the network in the

simulation model is static and therefore changes in the network topologies affect the adaptability of the network topologies. A time unit is regarded as the period of time required for all of the nodes to complete one search in turn.

The parameter values of the P-EP2P used in the simulations are listed in Table I. The parameters whose values are changed in the simulations are just the number of nodes L , which takes 10^3 , 10^4 and the number of node groups N_G , which takes 1, 2, 5, and 10. Also, the number of nodes as a search object is just one and all of the nodes excluding the search object conduct only one search in a time unit, so that the search result for every node is always the same until the network topologies are changed by applying the evolutionary operators.

TABLE I. PARAMETERS VALUES OF P-EP2P USED IN THE SIMULATIONS.

Parameter	Description	Value
L	number of nodes (genes)	$10^3, 10^4$
N	number of P2P network topologies (individuals)	30
T	time period for which generated topologies are used	20
H_{max}	allowed number of hops for one search	5
D	number of directed links generated by a node in one topology	1
K	tournament size for the tournament selection	2
p_c	crossover rate	100%
p_e	probability with which recombination is conducted for a selected element in NLX	10%
N_L	length of node linkage in NLX	5
p_m	mutation rate	0.5%
N_G	number of node groups	1, 2, 5, 10
G_k	number of nodes in the k -th node group	L/N_G

B. Results

We observed the search failure rate of the present network topologies during a period of use, which is $T = 20$. Figures 5 and 6 show the time-varying search failure rates for a variety of the number of nodes and the number of node groups, for the case of using the proposed sequential evaluation method and for the case of not using it, respectively. The results shown in those figures are averages over 10 independent simulation runs. The results for $N_G = 1$ in the figures are the same because there is only one node group. In addition, Figure 7 shows the time-varying fitness values that were given to a particular network topology by each node group when $L = 10^3$ and $N_G = 5$.

Comparing Figures 5 and 6, we can observe that P-EP2P using the proposed evaluation method improved fitness values better than P-EP2P not using it for all tested pairs of the number of nodes and the number of node groups. P-EP2P not using the proposed method hardly improved fitness values after approximately time of 1,000. Meanwhile, P-EP2P using the proposed method gradually improved fitness values over all the simulation times. Furthermore, we can also observe from Figure 7 that only in P-EP2P using the proposed method, all the node groups harmoniously improved fitness values for the particular network topology over all the simulation times. Thus, we can say that the proposed method evaluating network topologies sequentially adapts network topologies to users' demands better than the method evaluating them simultaneously.

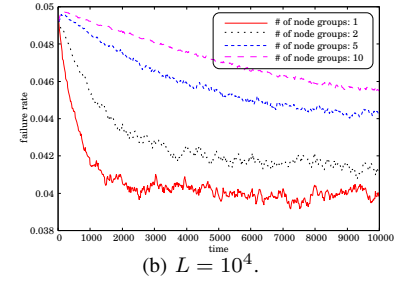
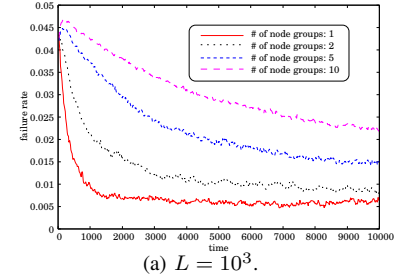


Fig. 5. The time-varying search failure rates for the case of using the proposed sequential evaluation method for network topologies.

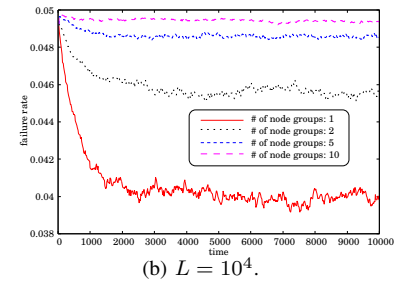
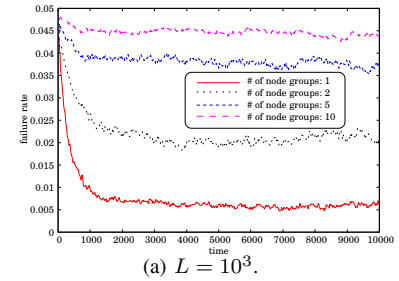


Fig. 6. The time-varying search failure rates for the case of not using the proposed sequential evaluation method for network topologies.

Finally, we compare the proposed evaluation method with its variants in terms of search failure rate. The proposed sequential evaluation method evaluates network topologies one by one. Here, we think that evaluation target is not limited to one network topology but E ($2 \leq E < N_G$) network topologies. Moreover, as order of evaluating the E network topologies, we consider a way to evaluate E network topologies in a fixed order and a way to evaluate E network topologies randomly chosen. Suppose that all node groups

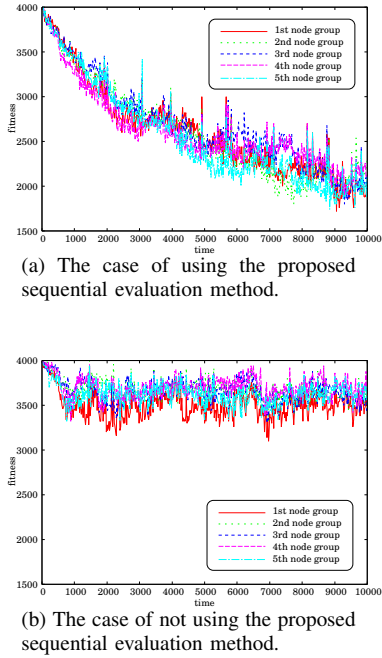


Fig. 7. The time-varying fitness values given to a particular network topology by each node group when $L = 10^3$ and $N_G = 5$.

are assigned serial numbers from 1 to N_G . Then, the way to evaluate E network topologies in a fixed order evaluates node groups with serial numbers $(1 + i) \bmod N_G$, $(2 + i) \bmod N_G$, \dots , and $(E + i) \bmod N_G$ at the i -th evaluation. The way to evaluate E network topologies randomly chosen literally always chooses E node groups randomly from among N_G node groups as evaluation targets. Figure 8 shows results for $E = 2, 3, 4, 5$, $N = 1000$, and $N_G = 10$.

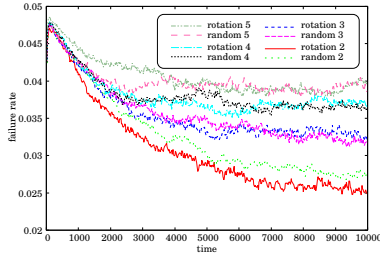


Fig. 8. Results of the way to evaluate E network topologies in a fixed order (label of “rotation E ”) and that to evaluate E randomly chosen network topologies (label of “random E ”).

We can observe from Figure 8 that the smaller the value of E , the better search failure rate. This observation does not contrary to intuition because when $E = N_G$, the evaluation method considered here is the same as the conventional method by which all node groups simultaneously evaluates network topologies. In addition, we can observe that there is almost no difference in search failure rate between the way to evaluate E network topologies in a fixed order and that to evaluate E randomly chosen network topologies. This observation suggests that an essential procedure for lowering search failure rate is to evaluate network topologies of each node group

separately in time.

V. CONCLUDING REMARKS

This paper proposed a new method for evaluating and reconstructing network topologies in the parallel P2P networking technique (P-EP2P) that presented in our previous study, and evaluated the proposed method through simulations. P-EP2P attempts to adapt entire co-existing network topologies to users’ demands by accumulating adaptation of pieces of entire network topologies that several node groups manage. The proposed evaluation method allows the node groups to evaluate and reconstruct their own network topologies sequentially, while in the conventional evaluation method, all node groups do that simultaneously. The simulation results showed that the proposed evaluation method yields better adaptability of network topologies than the conventional one. In future, we will assume more realistic search behaviors of users in simulations and demonstrate the proposed method.

REFERENCES

- [1] ITU Internet Report 2006: digitallife, <http://www.itu.int/osg/spu/publications/digitalife/>
- [2] H. Takagi, “Interactive Evolutionary Computation: Fusion of the Capacities of EC Optimization and Human Evaluation”, Proceedings of the IEEE, 2001.
- [3] M. Sathe, “Interactive Evolutionary Algorithms for Multi-Objective Optimization”, VDM Verlag, 2008.
- [4] K. Ohnishi and Y. Oie, “Evolutionary P2P Networking that Fuses Evolutionary Computation and P2P Networking Together,” IEICE Transactions on Communications, Vol. E93-B, No. 2, pp.317–328, February 2010.
- [5] K. Ohnishi and Y. Oie, “Parallel Evolutionary P2P Networking for Realizing Adaptive Large-scale Networks”, The 2nd Workshop on HEUNET 2011, 6 pages, July 18-21, 2011.
- [6] E. K. Lua, J. Crowcroft, M. Pias, R. Sharma, and S. Lim, “A Survey and Comparison of Peer-to-Peer Overlay Network Schemes,” IEEE Communications Surveys & Tutorials, Vol. 7, No. 2, pp.72–93, Second Quarter 2005.
- [7] “LimeWire,” <http://www.limewire.com/>
- [8] P. Merz and S. Wolf, “Evolutionary Local Search for Designing Peer-to-peer Overlay Topologies based on Minimum Routing Cost Spanning Trees,” In Proceedings of the 9th International Conference on Parallel Problem Solving from Nature (PPSN IX), pages 272–281, 2006.
- [9] K. Walkowiak and M. Przewoniczek, “Modeling and Optimization of Survivable P2P multicasting”, Computer Communications, Vol. 34, No.12, pp.1410–1424, 2011.
- [10] S. G. M. Koo, C. S. G. Lee, and K. Kannan, “A Genetic-algorithm-based Neighbor-selection Strategy for Hybrid Peer-to-peer Networks,” In Proceedings of the International Conference On Computer Communications and Networks (ICCCN 2004), pp. 469–474, 2004.
- [11] M. Srivatsa, B. Gedik, and L. Liu, “Large Scaling Unstructured Peer-to-peer Networks with Heterogeneity-aware Topology and Routing,” IEEE Transactions on Parallel and Distributed Systems, Vol.17, No.11, pp.1277–1293, November 2006.
- [12] E. Pournaras, G. Exarchakos, and N. Antonopoulos, “Load-driven Neighbourhood Reconfiguration of Gnutella Overlay,” Computer Communications, Vol.31, No.13, pp.3030–3039, August 2008.
- [13] M. Munetomo, Y. Takai, and Y. Sato, “An Adaptive Network Routing Algorithm Employing Path Genetic Operators,” Proceedings of the Seventh International Conference on Genetic Algorithms, pp.643–649, 1997.
- [14] P. Imai and C. Tschudin, “Practical Online Network Stack Evolution,” SASO 2010 Workshop on Self-Adaptive Networking, September 2010.