社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

A proposal of Evolvable Network Designing Approach with Topological Diversity

Lu CHEN†, Shin'ichi ARAKAWA†, Hideyuki KOTO††,

Nagao OGINO††, Hidetoshi YOKOTA††, and Masayuki MURATA†

† Graduate School of Information Science and Technology, Osaka University,
1-5 Yamadaoka, Suita, Osaka 565–0871, Japan
†† KDDI R&D Laboratories, Inc.,
2-1-15 Ohara, Fujimino, Saitama, 356-8502, Japan

Abstract As environments surrounding the Internet become more changeable, a design approach is needed that requires less equipment to scale up networks against the traffic growth under various environmental changes. Here, we propose an evolvable network design approach where network equipment is deployed without predetermined purpose rather than for a preplanned purpose. We use mutual information on node degree to measure the topological diversity of networks, and maximize topological diversity in the network design by minimizing the mutual information. Evaluations show that, compared to networks with ad-hoc design method, networks constructed by our design approach can efficiently uses the network equipment among various environments.

Key words power-law network, router-level topology, topological structure, mutual information, network heterogeneity, degree distribution, node failure

1. Introduction

The Internet now plays a critical role as a social infrastructure. As Web services become more popular, the environment surrounding the Internet is more changeable. It is estimated that the traffic growth is 1.4 times per a year in Japan, but it is the total traffic growth, and traffic in some places increases even more. That is, once a new service attracts many users, traffic around the server providing the service increase rapidly. Since these changes are hard to predict, a new network design method should be introduced to tackle long term environmental changes.

Currently, operators of ISP networks usually add link capacity and routers in an ad-hoc way. For example, they add link capacity when link utilization exceeds a certain threshold, or they introduce new routers when already-existing routers become unable to accommodate traffic from those enhanced links. Such an ad-hoc design leads to an increasing amount of equipment. This, in turn, leads to problems arising from technical limitations of routers/links such as processing speed or transmission capacity in the near future. Because environmental changes are hard to predict, trying to solve an optimization problem that includes environmental uncertainty is infeasible. Hence, a design approach that uses less equipment to improve a network in response to various environmental changes is urgently required. In this paper, we discuss whether this could be achieved by constructing a network that can easily adaptable to deal with new environments; we call this property evolvability.

Evolution and evolvability have been studied for a long time in biology [?]. The heart of evolution in living species is the presence of genetic diversity at the DNA-level and the adaptability of genetic diversity through natural selection in particular environments. Species that better adapt to their environment survive and pass on their genetic characteristics to the next generation during the evolution. Various species exists today as a result of evolution over billions of years, under many kinds of environments.

Information-theoretic interpretations of an evolutionary process can be used to understand adaptation and evolution in complex systems as described in Prokopenko et al [?]. In general, mutual information is defined as the differences between the heterogeneity and correlation of some variable. The mutual information of a system can be used to characterize the degree of evolution. The mutual information of system components increases as evolution progresses, since the correlation, which represents constraints between components from the system perspective, becomes stronger as the system is specialized to the environment. Then, the unspecialized system, which has low mutual information, has a potential to evolve in various ways, while a specialized system, which has high mutual information, is more constrained and less able to evolve. For example, Solé [?] used mutual information to analyze topological characteristics of complex networks. The mutual information used in [?] is the difference between the heterogeneity in degree distribution and the degree-degree correlation, which is also known as assortativeness [?] appeared in the network structure. They showed that a software network with high mutual information is the result of an engineering process rather than natural selection. Then we expect from [?] and [?] that by using mutual information, we can construct evolvable networks robust against short-term environmental changes including equipment failures.

In information networks, nodes or links are often added for a particular purpose: for example, aggregating or relaying traffic. However, because it is specialized to that purpose, nodes and links added in such a way can be effective in only the environment to which they were introduced; when the environment changes, that equipment may become underutilized, and a large amount of equipment need to be added to follow the new environment. Following the insights from work in biology and complex systems, an information network topology that has a reduced degree of specialization can be expected to enhance the evolvability of a network; when the environment changes, equipments in an old environment can be more efficiently used for new environments as it is not specialized for an environment. Hence, a design approach that reduces the degree of specialization can be expected to enhance the evolvability of a network. Hereafter, we will describe a network having a topology with low degree of specialization as having "topological diversity". It was shown in [?] that router-level topologies characterized by degree-degree correlation [?] leads to high mutual information. Following [?], we will use the mutual information proposed in [?] to strengthen topological diversity, and show the advantages of our design method in terms of its response to environmental changes, by which we mean unpredictable equipment failures.

The rest of this paper is organized as follows. Section 2. explains our proposed design approach that minimizes the mutual information. The evaluation of evolvability is shown in Sec. 3.. We show accumulated equipment during the network growth in Sec. 3. 1, and show how the designed network can easily adaptable to dealing with new environments in Sec. 3. 2. Finally, we conclude our paper in Sec. 4..

2. Network Design Approach by Minimization of Mutual Information

We describe our proposed design approach, which we call EVN (EVolvable Network) design approach. Fundamentally, the purpose of our EVN design approach is to reduce the mutual information on remaining degree so that the designed network has topological diversity.

Mutual information of remaining degree is studied by Solé et al. in [?]. The measurement indicate correlation of degrees between pairs of linked nodes. Remaining degree is the number of edges leaving a node, other than the one that connects the pair. Using the distribution of remaining degree q, the mutual information on remaining degree, $I(\mathrm{q})$, is defined as,

$$I(\mathrm{q}) = H(\mathrm{q}) - H_c(\mathrm{q}|\mathrm{q}'), \tag{1}$$

where $H(\mathrm{q})$ is the entropy of the remaining degree distribution and $H_c(\mathrm{q}|\mathrm{q}')$ is the conditional entropy of the remaining degree distribution.

Note that EVN is not designed to satisfy particular design constraints, for example, performance constrains or budget constraints. Therefore, networks designed by our EVN design approach may not be comparable in terms of its optimality with highly "engineered" networks which are specialized for the particular design constraints. Instead, as we will see later in this paper, the network with topological diversity designed by our approach is evolvable, that is, it can easily adaptable to deal with new environments without requiring a lot of additional equipment.

When designing a network, we should consider various design constraints such as network performance or budget constraints. In this paper, we do not explicitly consider the validity or effectiveness of a particular design constraint; instead, we consider whether our design approach is evolvable or not. For this reason, the following assumptions are introduced. The initial topology is given and nodes are added incrementally. The number of links m added with a new node is fixed. Note that these assumptions should be relaxed in the real situation of the network maintenance, but we expect that the characteristics obtained by our approach are not different much. Furthermore, for simplicity, we assume in this paper that topology is the only information we use to decide where to attach a new node, and physical distance is not considered here, and these assumptions should be relaxed in the future work.

Set an initial topology be $G_0(V_0, E_0)$, where V_0 and E_0 are initial sets of nodes and links. Then, our design approach adds a node and links to the topology at each step by the following algorithm. At each step, we add a single node and the number of links introduced for each node addition is denoted as m. Also, let $G_k(V_k, E_k)$ be the topology obtained by kth step of the algorithm, then it has k additional nodes and km additional links from the initial topology, i.e., $|V_k| = |V_0| + k$ and $|E_k| = |E_0| + km$.

(1) Calculate the entropy $H_{k-1}(\mathrm{q})$ of $G_{k-1}(V_{k-1}, E_{k-1})$.

(2) Add a node (denoted by w) to $G_{k-1}(V_{k-1}, E_{k-1})$.

  (a) Decide m different nodes for setting m links connected to the new node w.

  - For this purpose, first enumerate all of the topologies for all the possible cases of m links addition, and calculate the entropy $H(\mathrm{q})$ and the mutual information $I(\mathrm{q})$ for each topology. Note that we simply use notation q here, but formally, it should depend on the topology including a new node and links.
  - Choose m nodes that minimize mutual information while making the entropy greater or equal than the entropy $H_0(\mathrm{q})$.

  (b) Connects a node w and m links, and obtain $G_k(V_k, E_k)$.

In each node addition, we add m links such that the entropy $H_k(\mathrm{q})$ of the new topology is greater or equal to the initial $H_0(\mathrm{q})$.

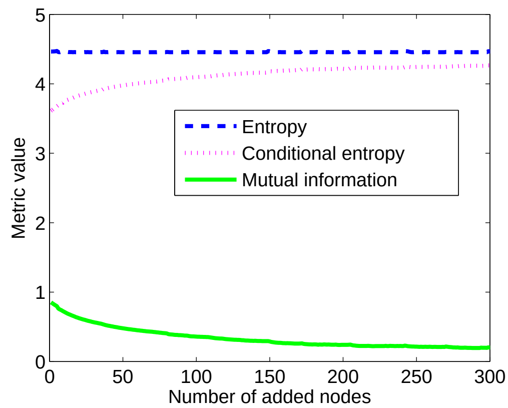


Figure 1: Values of entropy, conditional entropy and mutual information obtained by EVN design approach

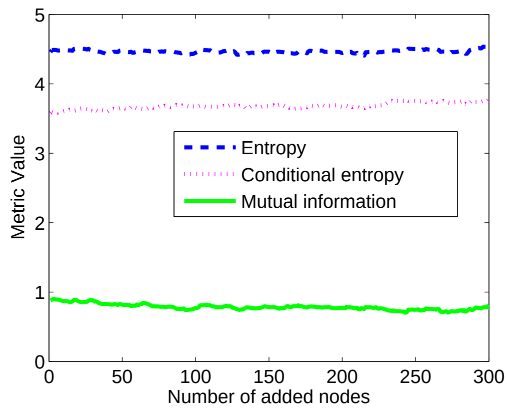


Figure 2: Values of entropy, conditional entropy and mutual information obtained by modified FKP-based network design method

The reason why this entropy–restriction is included is that the reliability of a network is improved by increasing the entropy of degree distribution [**?**], where Wang et al showed that increasing the entropy of the degree distribution of a scale-free network will lead to high reliability against random node failures. Note that, although $H(\mathrm{q})$ measures the heterogeneity of the remaining degree distribution, the distribution is derived from the degree distribution, so the entropy of the remaining degree distribution should not be decreased after the node addition.

Figure 1 shows the values of entropy, conditional entropy and mutual information obtained by EVN design approach. We use the AT&T topology as an initial topology $G_0(V_0, E_0)$. The AT&T topology we used is a measurement result obtained by Rocketfuel tool [**?**]. It has 523 nodes and 1304 links. Then, we apply design approach with the number added nodes n to be 300, that is, we iterate 300 steps of our design approach. Also, we set $m = 2$, i.e., we add two links per each step of node addition. The reason why two links are added per each node addition is not to let the average degree of the designed networks significantly different from the average degree (2.49) of the original AT&T topology. Because it is not possible to know the number of links added per a node addition in reality, and we just assume here that the average degree will not change largely in the near future. We can see from the result that mutual information of the initial topology is around 1.0, and the entropy is around 4.5. As the number of added nodes increases, the mutual information decreases and the entropy of remaining degree distribution is kept high by our algorithm, as expected.

3. Evaluation of Design Approach for Evolvability

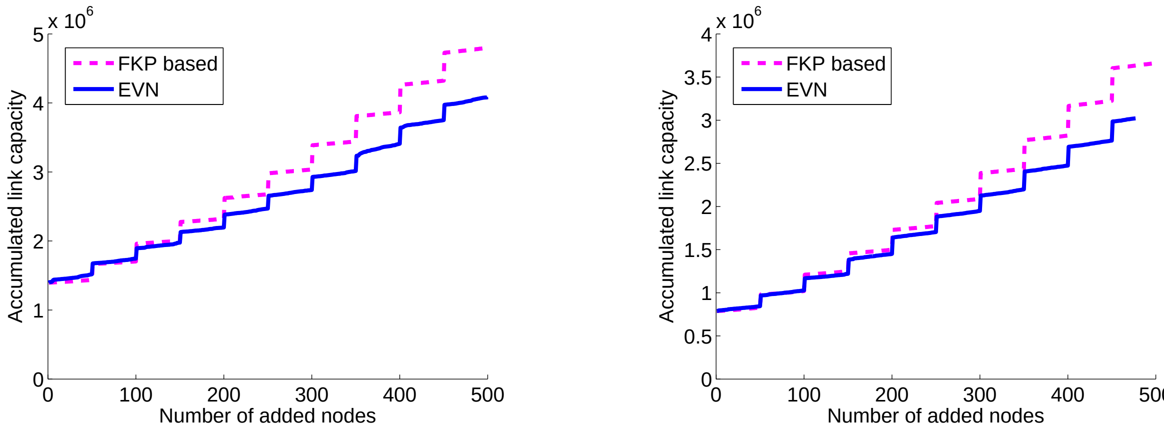
In this section, we show the evolvability of designed networks, that is, how networks with topological diversity can easily be designed and adapted to meet environmental changes. For comparison, we could use a "purely ad hoc method," in which we add nodes and or links at the place where capacity is in short supply. However, instead of using such a method, we consider a more intelligent approach that takes into account some optimization, for a fairer comparison. Though many complicated network design method can be considered, we consider the FKP model [**?**] here, in which nodes and links are incrementally added such that a new link connected to a new node is added to keep minimizing the weighted sum of physical distances and hop distances. The reason why we consider FKP model is that it includes primitive principles for designing information network. Hereafter, we call the topology growth method based on FKP model, FKP-based design method. Please see Appendix 1. for the detail of FKP-based design method.

Figure 2 shows the entropy, conditional entropy and mutual information during the network growth by the modified FKP-based design method. We use the AT&T topology as the initial topology, and set the number of added nodes n = 300 (i.e., the topology obtained after 300 steps) and the number of links for each step m = 2. The locations of nodes at the city-level are obtained from [**?**], and re-scale the latitude and longitude of each city down to $[0, 1]^2$, by letting the southernmost node and the northernmost node to be 0 and 1 for latitude, and the easternmost node and the westernmost node to be 0 and 1 for longitude. We can see from the result that entropy, conditional entropy and mutual information are unchanged during the network growth. This is because a principle of growth of FKP model is to minimize the distance metric (Eq. (A·1)) and is unchanged during the network growth. Mutual information is around 1.0 and is kept high, which means the topological diversity is kept low by the FKP-based network growth model. On the contrarily, that of a network grown by the EVN design approach becomes low, which means topological diversity is kept high.

3.1 Evaluation of Accumulated Capacity

We, first, evaluate equipment accumulated during the network growth without environmental changes. In the designing process, we assume that there is an enhancement of equipment needed to cope with single node failure. The reason for considering the enhancement is to see how designed networks absorb surged traffic arising from node failure. Equipment we consider here is the total capacity of links under the same number of node addition and link addition for the EVN design appraoch and for the FKP-based design method.

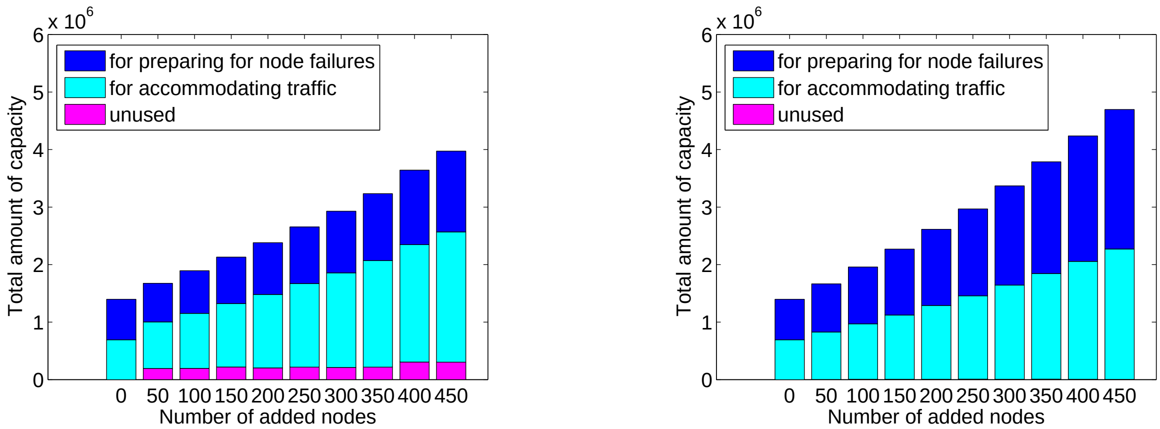
Hereafter, we denote $G_k^{EVN}(V_k, E_k)$ as the topology of the network obtained after k step (with k nodes addition) and $m = 2$ for



(a) AT&T topology

(b) Sprint topology

Figure 3: Accumulated link capacity



(a) EVN design approach

(b) FKP-based design method

Figure 4: capacity for preparing for node failures, capacity for accommodating traffic, unused capacity

the EVN design approach. In what follows, we will simply use G_k^{EVN} instead of $G_k^{EVN}(V_k, E_k)$. Similarly, we will use G_k^{FKP} as the network obtained by the modified FKP-based design method with $m = 2$. We also introduce C_k^{EVN}, which is the total capacity of G_k^{EVN} obtained by

$$C_k^{EVN} = \sum_{e \in E} C_k^{EVN}(e), \qquad (2)$$

where $C_k^{EVN}(e)$ represents the capacity of a link e. In evaluation, the capacity of each link is decided such that the link can accommodate the traffic against every pattern of single node failure in the topology G_k^{EVN}. Shortest path with equal hop path splitting [?] is applied for calculating the capacity. The traffic demand is set to one unit between all of node pairs in G_k^{EVN} for simplicity.

The link capacity is re-designed to cope with the increase of traffic in every node addition and to cope with single node failures in every 50-node addition. The link capacity is incremental, i.e., if a link capacity $C_{(k-1)}^{EVN}(e)$ is enough to accommodate the traffic at G_k^{EVN}, we do not reduce the link capacity but set $C_k^{EVN}(e) \leftarrow C_{(k-1)}^{EVN}(e)$. The initial link capacity, $C_k^{EVN}(e)$, is also calculated to cope with every pattern of single node failue. Also, $C_k^{FKP}(e)$, the total capacity of G_k^{FKP}, was obtained in the same way as explained above.

Figure 3 shows the total link capacity of G_k^{EVN} and G_k^{FKP} dependent on the number of added nodes k. The initial topology is set to AT&T topology (523 nodes and 1304 links) for Fig. 3a and is set to Sprint topology (467 nodes and 1280 links) for Fig. 3b. The Sprint topology is also a measurement result obtained by Rocketfuel tool [?]. Both of figures indicate that our EVN design approach requires less amount of link capacity than the FKP-based design method.

Table 1: Average of additional capacity needed to cover a node failure

	EVN	FKP
Additional capacity	6.0535×10^3	5.8868×10^3

To see how the network with topological diversity can scale up with fewer equipment in more detail, we show three kinds of link capacity, capacity for preparing for node failures, capacity for accommodating traffic, capacity unused based on difference of link capacity between before- and after- 50 nodes addition. Figure 4a shows the result of the EVN design approach, and Fig. 4b does the result of FKP-based design method. Comparing Figs. 4a with 4b, we can clearly see that FKP-based design method requires more capacity for preparing for node failures, while capacity for accommodating traffic is almost the same. This is caused by the overlap in equipment placement in each single node failure. Table 1 shows average of additional capacity needed to cover one pattern of single node failure. It is calculated for G_{450}^{EVN} and G_{450}^{FKP}. Here, the additional capacity is the capacity needed to cover one pattern of single node failure other than that needed only for accommodating traffic. We can see from Table 1 that G_{450}^{EVN} needs larger amount of capacity to cover one pattern of single node failure in average compared to G_{450}^{FKP}. However, it needs fewer amount of capacity to cover every pattern of single node failure. This is because topology generated by EVN design approach is unspecialized to a single environment. Therefore, it can efficiently uses the network equipment placed for other single node failures.

3. 2 Reused Facilities for Unexpected Environmental Changes

In the last subsection, we showed that a network with topological diversity requires fewer capacity to dealing with new environments. Thanks to the unspecialized design of topology, the most of link capacity are reused for the new environment. The evaluation of last section, however, only assumed that amount of link capacity is designed against single node failure.

This subsection evaluates evolvability: the ability to reuse capacity in response to unexpected environmental change other than the single node failures. However, since unpredicted environmental change is hard to define, we use a scenario of unpredicted environmental changes following the evaluation presented in [**?**]. We regard a single node failure between nodes as the environment assumed in designing a network. Then, we consider a scenario in which the same kind of environmental change occurs but the scale of environmental change is large. Here, we choose two simultaneous node failures for the evaluation scenario. Note that, the amount of traffic demand we assume is same as that assumed in Sec. 3. 1. Although actual traffic demand is different, our intension here is to show how the designed network reuses existing capacity in response to unexpected environmental change. Thus, we use unit traffic demand for simplicity.

For evaluation, we introduce a *reuse ratio*, r_k, of topology after k node addition defined by

$$r_k = \frac{F_k^{reused}}{F_k^{new}}, \tag{3}$$

where F_k^{reused} represents the amount of capacity that can be reused among the capacity already been deployed, and F_k^{new} represents the amount of capacity that *was* required to deal with unpredicted environmental changes for k-th network, i.e., the network with the number k of nodes added. r_k ranges from 0 to 1.0. As r_k close to 1.0, capacity that are already placed can be reused for unpredicted environmental change. On the contrary, more capacity are required to deal with the unpredicted environmental change as r_k decreases.

We evaluate the reuse ratio under two node failures in both G_k^{EVN} and G_k^{FKP}. The reuse ratio depends on the topology and failed nodes (denoted as n_1 and n_2). Thus, we refine reuse ratio r_k of G_k^{EVN} as $r_k^{EVN}(n_1, n_2)$. $r_k^{FKP}(n_1, n_2)$ is also used for reuse ratio r_k for G_k^{FKP}.

Figure 5a shows the results of $r_k^{EVN}(n_1, n_2)$ for all case of two-node (n_1, n_2) failures and Fig. 5b shows that of $r_k^{FKP}(n_1, n_2)$. Note that we again use the AT&T topology as initial topology. In these figures, the horizontal axis represents the rank of reuse ratio in an ascending order, and we show the results of reuse ratio by changing k. Looking at reuse ratios from rank 1 to 200, ones obtained by the EVN design approach are higher than those of the FKP-based design method, and this tendency becomes clearer as k increases. It is due to a result of the increase of topological diversity. Because alternative paths for a single node failure would be less likely to be biased on some links, capacity used for coping with single node failures is spread around the network. Therefore, even when a severe two node failure occurs, the required alternative paths could be provided mostly by reusing the capacity already in place. On the other hand, when the topology is less diverse, paths would be likely to be biased on some links, so the capacity for coping with single node failures is also biased. Therefore, when a severe two-node failure occurs, alternative paths would use links that have less capacity in place other than the biased links, which leads to a lower values of reuse ratio.

We can also observe an optimality of the EVN design approach from the figure. The number of two-node (n_1, n_2) failure patterns for which $r_{250}^{EVN}(n_1, n_2)$ is less than 1 is 32,291, and the number of that for which $r_{250}^{FKP}(n_1, n_2)$ is less than 1 is 7,557. It means that networks grown by the EVN design approach are less able to accommodate traffic completely. However, in the EVN design approach, because most values of $r_{250}^{EVN}(n_1, n_2)$ are almost 1, it can be covered by slightly more over-provisioning of links.

4. Conclusion and Future Work

In this paper, we have proposed a design approach based on minimizing mutual information to strengthen topological diversity and make the network evolvable. We have shown that a network grown using our design approach can grow with less capacity compared to a network grown using a method based on the FKP model. Furthermore, we have shown that capacity introduced for one environment can be used in another environment, thereby a network grown using our design approach experiences overlap between equipment placement in an old environment and a new one.

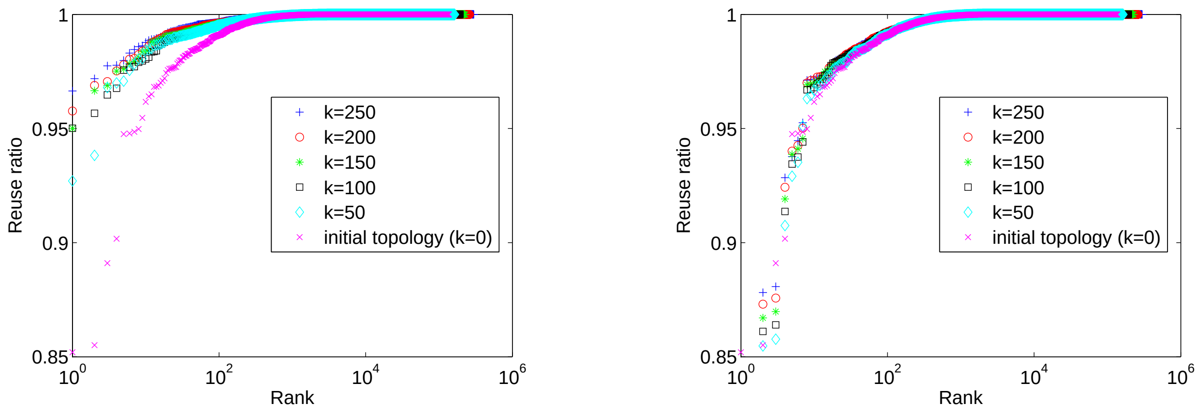
Several problems are left for future research. First, the design approach of this paper considers mutual information only and does not consider the physical lengths of connecting links. We believe that there is a trade-off relationship between mutual information and physical distance when connecting nodes, which is left for future investigation. Second, we have considered topological diversity here, but diversity at a higher-level, such as the diversity of link capacity distribution may help improving evolvability.

References

[1] C. Dovrolis and J. T. Streelman, “Evolvable network architectures: What can we learn from biology?,” *ACM SIGCOMM Computer Communication Review*, vol. 40, pp. 72–77, Apr. 2010.

[2] M. Prokopenko, F. Boschetti, and A. Ryan, “An information-theoretic primer on complexity, self-organization, and emergence,” *Complexity*, vol. 15, pp. 11–28, Sept. 2009.

[3] R. Solé and S. Valverde, “Information theory of complex networks: On evolution and architectural constraints,” *Complex networks*, vol. 650, pp. 189–207, Aug. 2004.



(a) Reuse ratio for G_k^{EVN} (b) Reuse ratio for G_k^{FKP}

Figure 5: Evaluation result of a scenario of failures of two nodes

[4] M. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, p. 208701, Oct. 2002.

[5] L. Chen, S. Arakawa, and M. Murata, "Quantifying network heterogeneity by using mutual information of the remaining degree distribution," *International Journal On Advances in Systems and Measurements*, vol. 6, pp. 214–223, July 2013.

[6] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the Internet's router-level topology," *ACM SIGCOMM Computer Communication Review*, vol. 34, pp. 3–14, Oct. 2004.

[7] B. Wang, H. Tang, C. Guo, and Z. Xiu, "Entropy optimization of scale-free networks robustness to random failures," *Physica A: Statistical Mechanics and its Applications*, vol. 363, pp. 591–596, May 2006.

[8] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with rocketfuel," *IEEE/ACM Transactions on Networking*, vol. 12, pp. 2–16, Feb. 2004.

[9] A. Fabrikant, E. Koutsoupias, and C. Papadimitriou, "Heuristically optimized trade-offs: A new paradigm for power laws in the Internet," *Automata, Languages and Programming*, vol. 2380, pp. 110–122, July 2002.

[10] A. Nucci, S. Bhattacharyya, N. Taft, and C. Diot, "IGP link weight assignment for operational Tier-1 backbones," *IEEE/ACM Transactions on Networking*, vol. 15, pp. 789–802, Aug. 2007.

[11] J. Whitacre, P. Rohlfshagen, A. Bender, and X. Yao, "Evolutionary mechanics: New engineering principles for the emergence of flexibility in a dynamic and uncertain world," *Natural computing*, vol. 11, pp. 431–448, Sept. 2012.

Appendix

1. Network Design Method based on FKP Model

The FKP model proposed by Fabrikant et al. [?] incrementally adds nodes and connects existing nodes at which physical distance and hop distance metrics are minimized.

In the original FKP model, the first node n_0 is set to the root of the topology. Then, a new node incrementally arrives at a random point in the Euclidean space $[0,1]^2$. After a new node n_i arrives, the FKP model calculates the following quantity for each node n_i already existing in the network:

$$\alpha \cdot d(n_{new}, n_i) + h(n_i, n_0), \qquad \text{(A·1)}$$

where $d(n_{new}, n_i)$ denotes the physical distance in the Euclidean space $[0,1]^2$ between n_{new} and n_i, and $h(n_i, n_0)$ denotes the hop distance between n_i and a root node n_0. The root node is prespecified for calculating the hop distance. In this paper, we set the maximum degree node in $G(V,E)$ as n_0. The parameter α determines the importance of physical distance. If α takes a low value, each node tries to connect to higher degree nodes; $\alpha = 0$ is an extreme scenario that creates a star-topology. If α takes a high value, each node tries to connect to their nearest nodes. A topology with high α is shown to behave like a random topology. A power-law degree distribution appears at moderate values of α.

For comparing with our method, we modified the FKP model as follows. Given a topology $G_0(V_0, E_0)$ and physical location of nodes, our modified version of the FKP model adds a node and m links for each node addition in k-th step by the following algorithm in order to obtain $G_k(V_k, E_k)$;

（1） Map the physical location of nodes V to the Euclidean space $[0,1]^2$

（2） Divide $[0,1]^2$ into 20×20 areas, and calculate a node existing ratio in each area. The node existing ratio of a area is defined as the number of nodes exist in the area over the total number of nodes.

（3） When a new node n_{new} arrives, determine the area of the node with probability proportional to the node existing ratio.

（4） Calculate a distance metric defined by Eq. (A·1) for each existing node n_i.

（5） Select m nodes in an ascending order by the value of distance metric. Then, add node n_{new} and links between n_{new} and the selected nodes to the topology.

The modifications to the original model we made in the above are as follows. First, the physical location of the added nod is determined with a probability proportional to the node existing ratio (Step (ii) in the above). The reason is that, because routers are often added to areas where traffic demand is large, an area attracts more traffic as the routers exist more in the area. Second, we add multiple links per a node addition so that the average degree of the designed networks can be controllable (Step (v)).

In evaluation at Subsection 3. 1 and Subsection 3. 2, the parameter α was set to 10.0, where the average hop distance is lowest under the condition that the entropy $H(\mathrm{q})$ is moderate so as not to obtain a star-like topology.