Computer Communications 76 (2016) 101–110

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

An evolvable network design approach with topological diversity

Lu Chen [a,*], Shin'ichi Arakawa [a], Hideyuki Koto [b], Nagao Ogino [b], Hidetoshi Yokota [b], Masayuki Murata [a]

[a] Department of Information Science and Technology Osaka University, 1-5 Yamadaoka, Suita Osaka 565-0871, Japan
[b] KDDI R&D Laboratories, Inc. 2-1-15 Ohara, Fujimino Saitama 356-8502, Japan

ARTICLE INFO

Article history:
Received 20 November 2014
Revised 30 September 2015
Accepted 24 October 2015
Available online 10 November 2015

Keywords:
Router-level topology
Topological structure
Mutual information
Network heterogeneity
Node failure

ABSTRACT

As environments surrounding the Internet become more changeable, a design approach that requires less equipment to scale up networks against the traffic growth arising from various environmental changes is needed. Here, we propose an evolvable network design approach where network equipment is deployed without a predetermined purpose. We enhance topological diversity in the network design by minimizing the mutual information. Evaluations show that, compared to networks built with ad-hoc design method, networks constructed by our design approach can efficiently use network equipment in various environments. Moreover, we show that, even considering the physical lengths of links, the approach of increasing topological diversity can lead to an evolvable network.

1. Introduction

The Internet now plays a critical role as a social infrastructure and, as Web services become more popular, the environment surrounding the Internet becomes more changeable. Actually, it is estimated that traffic grows by a factor of 1.4 per year in Japan. However, this is only the current total traffic growth: traffic in some places increases even more, such as traffic around servers providing a new service which attract many users, and there is no doubt that the environment surrounding the Internet will change even more in the future.

In spite of the upcoming changes, 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, and they introduce new routers when existing routers become unable to accommodate traffic from those enhanced links. However, in a changeable environment, such an ad-hoc design strategy will lead to an increasing amount of equipment. This, in turn, will lead to problems arising from technical limitations of routers or links, such as processing speed or transmission capacity, in the near future. Hence, a design approach that uses less equipment to allow a network to respond to various environmental changes is urgently needed.

In this paper, we discuss whether this could be achieved by constructing a network that can easily adapt to deal with new environments. In information networks, nodes or links are often added for a particular purpose: for example, aggregating or relaying traffic. However, because they are specialized to that purpose, nodes and links added in such a way can be effective only in the environment to which they were introduced; when the environment changes, that equipment may become underutilized, and a large amount of new equipment may be needed to cope with the new environment. Following insights from work in biology and complex systems [1], an information network topology that has a reduced degree of specialization can be expected to enhance the ability to deal with new environments; when the environment changes, existing equipment can be more efficiently used for the new environment as it is not specialized for a particular environment. In this paper, we propose a design approach to reduce the degree of specialization, and show the advantages of our design method in terms of its response to environmental changes, by which we mean unpredictable equipment failures. Hereafter, we will describe a network having a topology with low degree of specialization as having "topological diversity", and the ability to deal with new environments will be referred to as "evolvability".

Some may say that a random network has topological diversity. However, it is not efficient to design an information network as a random network. A well-known disadvantage of a random network is that the average hop distance is larger than that in a scale free network. Because of this, a random network needs a lager capacity to accommodate the same amount of traffic. Therefore, a measure is needed to characterize topological diversity so that one can consider it in conjunction with other factors when designing networks.

* Corresponding author. Tel.: +81 6 6879 4542; fax: +81 6 6879 4544.
E-mail addresses: l-chen@ist.osaka-u.ac.jp (L. Chen), arakawa@ist.osaka-u.ac.jp (S. Arakawa), koto@kddilabs.jp (H. Koto), ogino@kddilabs.jp (N. Ogino), yokota@kddilabs.jp (H. Yokota), murata@ist.osaka-u.ac.jp (M. Murata).

http://dx.doi.org/10.1016/j.comcom.2015.10.009
0140-3664/

The rest of this paper is organized as follows. Section 2 explains our proposed design approach. We explain the measure we use for design in Section 2.1. We then present our approach in Section 2.2 and discuss characteristics of reliability against node failures in Section 2.3. In Section 3, we evaluate accumulated equipment, and evaluate the evolvability by showing how the designed network can easily adapt to new environments. The advantage of our method compared to randomly selected node attachment is explained in Section 4. Section 5 shows that our approach of considering topological diversity is evolvable even if we take account of the physical lengths of links. Finally, we conclude our paper in Section 6.

2. An evolvable network design approach

Evolution and evolvability have been studied for a long time in biology [2]. The core 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: individuals that are better adapted to their environment survive and pass on their genetic characteristics to the next generation. Various species exist today as a result of evolution over billions of years, under many kinds of environment.

Information-theoretic interpretations of an evolutionary process can be used to understand adaptation and evolution in complex systems, as described in Prokopenko et al. [1]. In general, mutual information is defined as the difference between the heterogeneity and correlation of some variables. 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 becomes specialized to the environment. Thus, an unspecialized system, which has low mutual information, has the potential to evolve in various ways, while a specialized system, which has high mutual information, is more constrained and less able to evolve.

Solé [3] used mutual information to analyze topological characteristics of complex networks. The mutual information used in [3] is the difference between the heterogeneity in degree distribution and the degree–degree correlation, which is also known as assortativeness [4], appearing in the network's structure. It was shown in [5] that router-level topologies characterized by degree–degree correlation [6] lead to high mutual information. Following [5], we will minimize the information measure proposed in [3] to strengthen topological diversity. In Section 2.1, we briefly explain the abstract idea of the mutual information measure presented by Solé et al. Our proposed approach using this measure is then explained in Section 2.2.

2.1. Measure used for design

Solé et al. [3] used mutual information on the remaining degree distribution to analyze characteristics of complex networks. Following [3], we briefly explain the definition of mutual information of remaining degree.

Let us consider a network topology with degree distribution P_k, that is, P_k represents the probability that a node has k edges and $\Sigma_k P(k)=1$. Then, the distribution $q(z)$ of the remaining degree z, which is the number of edges leaving the node other than the edge we arrived along, is defined by

$$q(z)=\frac{(z+1)P_{z+1}}{\Sigma_z z P_z}. \tag{1}$$

Using the distribution of remaining degree $\boldsymbol{q}\,(=\{q(z)|1\le z\le N\})$, where N is the maximum remaining degree, the mutual information on remaining degree, $I(\boldsymbol{q})$, is defined as,

$$I(\boldsymbol{q})=H(\boldsymbol{q})-H_c(\boldsymbol{q}|\boldsymbol{q}'), \tag{2}$$

Table 1
Mutual information of example topologies.

Topology	H	H_c	I
Ring topology	0	0	0
Star topology	1	0	1
Abilene-inspired topology	3.27	2.25	1.02
Random topology	3.22	3.15	0.07

where $H(\boldsymbol{q})$ is the entropy of the remaining degree distribution and $H_c(\boldsymbol{q}|\boldsymbol{q}')$ is the conditional entropy of the remaining degree distribution $\boldsymbol{q}$, given the remaining degree distribution $\boldsymbol{q}'\ (=\{q(z')|1\le z'\le N\}$ where z and z' are the remaining degrees of linked nodes). $H(\boldsymbol{q})$ is defined as

$$H(\boldsymbol{q})=-\sum_{z=1}^{N}q(z)\log(q(z)), \tag{3}$$

and $H(\boldsymbol{q})$ always satisfies the inequality $H(\boldsymbol{q})\ge 0$. Within the context of information theory, $H(\boldsymbol{q})$ measures the uncertainty of remaining degree, and it indicates the heterogeneity of remaining degree in the network topology. A network topology with $H(\boldsymbol{q})=0$ is a homogeneous network, and as a network becomes more heterogeneous, the entropy $H(\boldsymbol{q})$ becomes higher. For example, a ring topology is homogeneous whereas the Abilene-inspired topology [6] is heterogeneous in the degree distribution, so it has higher entropy, as shown in Table 1. For reference, we also show $H(\boldsymbol{q})$ for a randomly generated topology. The topology was generated by Random 2 model [7] with 523 nodes and 1304 links, as in the AT&T topology.

The second term $H_c(\boldsymbol{q}|\boldsymbol{q}')$ of Eq. (3) is the conditional entropy of the remaining degree distribution:

$$H_c(\boldsymbol{q}|\boldsymbol{q}')=-\sum_{z=1}^{N}\sum_{z'=1}^{N}q(z')\pi(z|z')\log\pi(z|z'), \tag{4}$$

where $\pi(z|z')$ is the conditional probability

$$\pi(z|z')=\frac{q_c(z,z')}{q(z')}, \tag{5}$$

which gives the probability of observing a vertex with z' edges leaving it, provided that the vertex at the other end of the chosen edge has z leaving edges. Here $q_c(z,z')$ represents the normalized joint probability, that is,

$$\sum_{z=1}^{N}\sum_{z'=1}^{N}q_c(z,z')=1. \tag{6}$$

The conditional entropy, $H_c(\boldsymbol{q}|\boldsymbol{q}')$, always satisfies the inequalities

$0\le H_c(\boldsymbol{q}|\boldsymbol{q}')\le H(\boldsymbol{q})$. $H_c(\boldsymbol{q}|\boldsymbol{q}')$ is 0 for the ring and star topologies for which, if the degree of one side of a link is known, the degree of the node on the other side is always determined. For the Abilene-inspired topology, on the other hand, because of its heterogeneous degree distribution, even if the degree of one side of a link is known, it is hard to determine the degree of the other side of the link. Therefore, $H_c(\boldsymbol{q}|\boldsymbol{q}')$ for the Abilene-inspired topology is higher than that of ring and star topologies. However, $H_c(\boldsymbol{q}|\boldsymbol{q}')$ for the Abilene-inspired topology is lower than that of the random topology although these topologies have almost the same entropy $H(\boldsymbol{q})$. This means that the degree of correlation of two nodes that are connected is more assortative in the Abilene-inspired topology than in the random topology, which agrees with the discussions in [6].

Finally, using the probabilities given above, the mutual information of the remaining degree distribution can be expressed as

$$I(\boldsymbol{q})=-\sum_{z=1}^{N}\sum_{z'=1}^{N}q_c(z,z')\log\frac{q_c(z,z')}{q(z)q(z')}. \tag{7}$$

$I(\boldsymbol{q})$ is high for the star and Abilene-inspired topologies (see the right-most column of Table 1), since information about the degree of a node can be obtained by observing a node connected to it. In contrast, in the random topology, $I(\boldsymbol{q})$ is low, that is, little information can be obtained, because nodes are randomly connected. In the ring topology, $I(\boldsymbol{q})$ is 0 because of the homogeneous degree distribution.

2.2. Design approach

In this subsection, we describe our proposed design approach, which we call EVN (EVolvable Network) design. Fundamentally, EVN design reduces the mutual information on remaining degree, $I(\boldsymbol{q})$, so that the designed network has topological diversity. Note that an EVN is not designed to satisfy particular design constraints, for example, performance constrains or budget constraints. Therefore, networks designed by this design approach may not be as optimal as highly "engineered" networks that are specialized to meet particular design constraints. Instead, as we will see later in this paper, a network with topological diversity designed by our approach is evolvable, that is, it can easily be adapted to deal with new environments without requiring a lot of additional equipment.

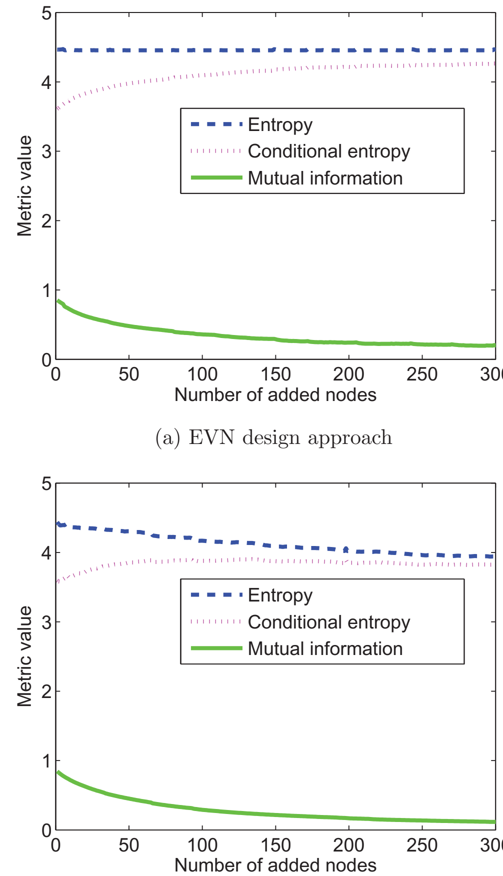
When designing a network, we should consider various design constraints such as network performance or budget constraints. In this paper, we do not explicitly consider the validity or effectiveness of a particular design constraint; instead, we consider whether networks produced using our design approach are 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 for real network maintenance, but we expect that the characteristics obtained by our approach are not much different from those of realistic cases. Furthermore, for simplicity, we assume for most of this paper that topology is the only information we use to decide where to attach a new node.

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 by m. Also, let $G_k(V_k, E_k)$ be the topology obtained by the kth step of the algorithm, then it has k additional nodes and km additional links compared with the initial topology, that is, $|V_k| = |V_0| + k$ and $|E_k| = |E_0| + km$. Note that, because our aim is to show the potential of a design method based on minimizing mutual information, we use an exhaustive search for deciding on the appropriate node to connect.

1. Calculate the entropy $H_{k-1}(\boldsymbol{q})$ 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) Choose m different nodes for to connect to the new node w by m links.
      - For this purpose, first enumerate all of the topologies for all the possible cases of m additional links, and calculate the entropy $H(\boldsymbol{q})$ and the mutual information $I(\boldsymbol{q})$ for each topology. Note that we simply use notation $\boldsymbol{q}$ here, but formally, it should depend on the topology including the new node and links.
      - Choose m nodes that minimize mutual information while making the entropy greater than or equal to the entropy $H_0(\boldsymbol{q})$.

   (b) Connect the node w and the m links, and obtain $G_k(V_k, E_k)$.

In each node addition, we add m links such that the entropy $H_k(\boldsymbol{q})$ of the new topology is greater than or equal to the initial $H_0(\boldsymbol{q})$. The reason why this entropy-restriction is included is that the reliability of a network is improved by increasing the entropy of the degree distribution, as Wang et al. [8] have shown 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 $H(\boldsymbol{q})$ measures the heterogeneity of the remaining degree distribution, the distribution is derived from the degree distribution (Eq. (1)), so the entropy of the remaining degree distribution should not be decreased after the node addition. In the next subsection, we will illustrate this by showing network growth with and without the entropy constraint.



(a) EVN design approach

(b) EVN design approach without entropyrestriction

Fig. 1. Values of entropy, conditional entropy and mutual information.

2.3. Improvement in robustness

In this section, we show the difference in network robustness against equipment failure between two growing networks with and without the entropy-restriction. Note that in this paper, we only present the case of node failure, but we see similar results in the case of link failure.

Fig. 1 shows the values of entropy, conditional entropy and mutual information of two networks: one is obtained by the EVN design approach (Fig. 1(a)) and the other is obtained by the EVN design approach without the entropy-restriction (Fig. 1(b)). For both networks, 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 the Rocketfuel tool [9]; it has 523 nodes and 1304 links. Then, we apply both design approaches with $n = 300$ added nodes, that is, we iterate 300 steps of our design approach. Also, we set $m = 2$, i.e., we add two links in each step of node addition. The reason why two links are added in each step is to not let the average degree of the designed networks become 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 node addition in reality, we just assume here that the average degree will not change greatly in the near future. In the figure,

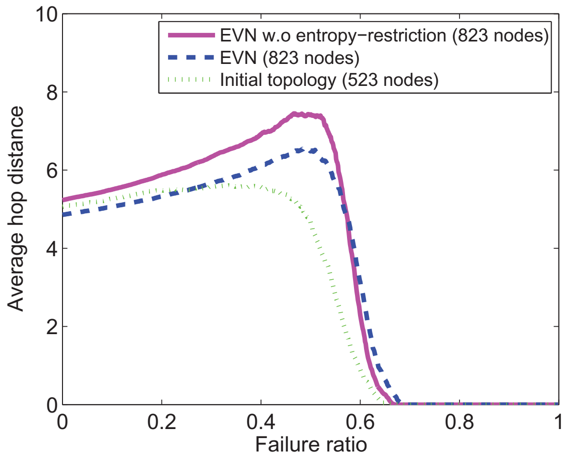


Fig. 2. Average hop distance against random node failures: Comparison between EVN design approaches with and without entropy-restriction.

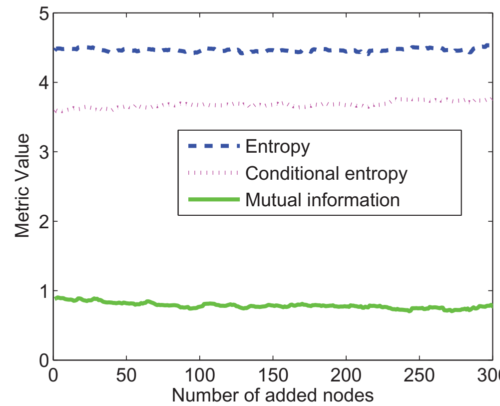


Fig. 3. Values of entropy, conditional entropy and mutual information obtained by the modified FKP-based network design method.

the horizontal axis represents the number of added nodes and the vertical axis represents the value of entropy, conditional entropy and mutual information for the topology. We can see from Fig. 1(a) that the 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 the remaining degree distribution is kept high by our algorithm, as expected. Fig. 1(b) shows the case without an entropy-restriction. In this network, the entropy of remaining degree decreases as the network grows.

We now compare the robustness of the two networks just after 300 nodes have been added. The measure of network robustness investigated here is the change of average hop distance when node failures occur. The shortest path routing is used for calculating the hop distance. Fig. 2 shows how the average hop distance changes as nodes are removed one by one in a random order. The horizontal axis is the failure ratio which is defined as the number of failed nodes over the number of initial nodes. The vertical axis is the average hop-count distance for the most connected component after the node failures. In the figure, we observe that the average hop distance of the network designed with the entropy-restriction is lower than that of the network designed without the entropy-restriction. Comparing with the results for the initial topology (AT&T topology), when the failure ratio is low, the average hop distance of the network designed with the entropy-restriction is lower, while that of the network designed without the entropy-restriction is higher. From this figure, it can be seen that a network designed with the entropy restriction achieves better performance even with node failure, and a robust network is built. This is the reason why we consider the entropy-restriction in our EVN design approach. Note that we will evaluate the "evolvability" of our design approach against equipment failure in more depth in the next section.

3. Evaluation

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 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 methods can be considered, we will consider the FKP model [10], in which nodes and links are incrementally added such that a new link connected to the new node is added to keep minimizing the weighted sum of physical distances and hop distances. The reason why we consider the FKP model is that it includes primitive principles for designing an information network. Therefore, a result that shows better performance than the FKP-based method indicates that our approach has features that would be useful in real-life networks. Hereafter, we call the topology growth method based on the FKP model, the FKP-based design method. Please see Appendix A for details of FKP-based design method.

Fig. 3 shows the entropy, conditional entropy and mutual information during 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 to be $n = 300$ (i.e., the final topology is obtained after 300 steps) and the number of links for each step to be $m = 2$. The locations of nodes at the city-level are obtained from [9], and we rescale 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 results that entropy, conditional entropy and mutual information are unchanged during network growth. This is because a principle of growth in the FKP model is to minimize the distance metric (Eq. (A.1)). 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 contrary, that of a network grown by the EVN design approach is low, which means topological diversity is kept high.

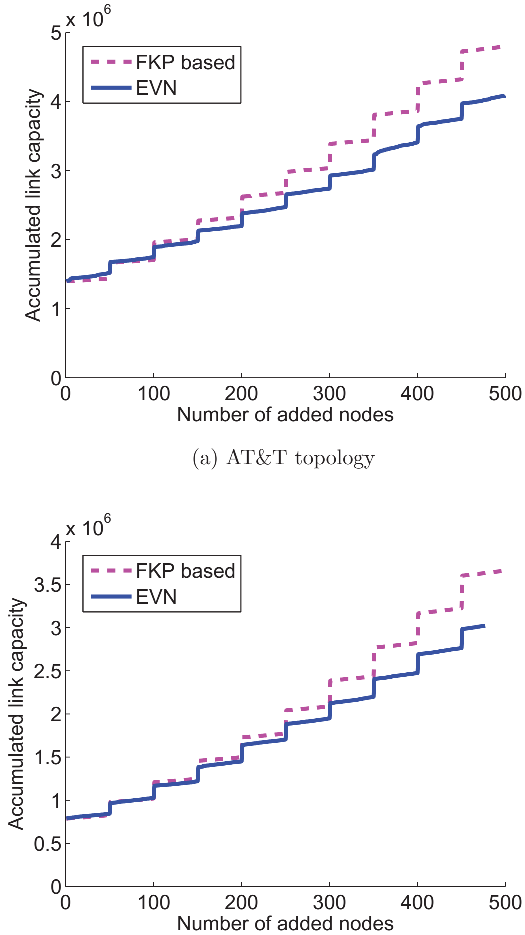
3.1. Evaluation of accumulated capacity

We, first, evaluate equipment accumulated during network growth. In the design process, we assume that there is an enhancement of equipment needed to cope with single node failure. The reason for considering this enhancement is to see how designed networks absorb surges of traffic arising from node failure. The equipment we consider here is the total capacity of links for the same number of added nodes and links in the EVN design approach and in the FKP-based design method.

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

$$C_k^{EVN} = \sum_{e \in E} C_k^{EVN}(e), \tag{8}$$

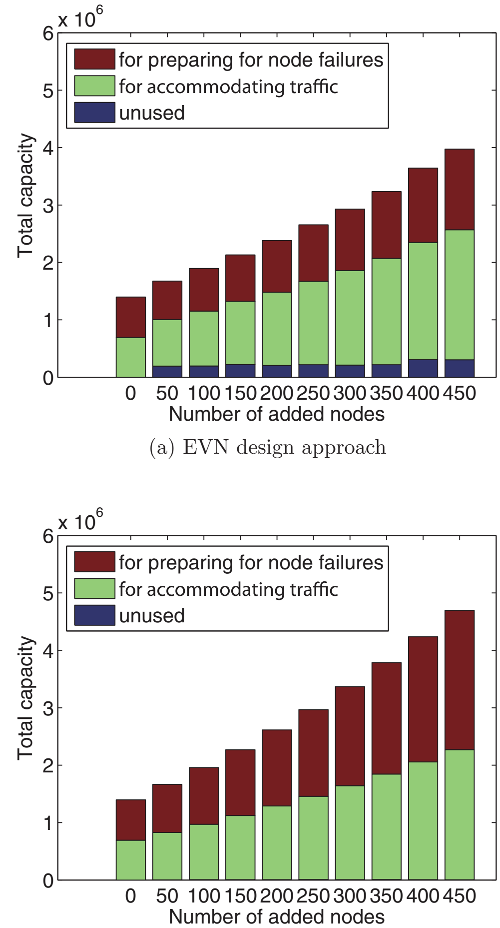
where $C_k^{EVN}(e)$ represents the capacity of link e. In the evaluation, the capacity of each link is chosen such that the link can accommodate



(a) AT&T topology

(b) Sprint topology

Fig. 4. Accumulated link capacity.



(a) EVN design approach

(b) FKP-based design method

Fig. 5. Capacity for preparing for node failures, capacity for accommodating traffic, unused capacity.

the traffic arising from every pattern of single node failure in the topology G_k^{EVN}. The method of shortest path with equal hop path splitting [11] is applied for calculating the capacity. The traffic demand is set to one unit between all node pairs in G_k^{EVN} for simplicity.

The link capacity is re-designed to cope with an increase of traffic at every node addition and to cope with single node failures at every 50-node addition. The link capacity is incremental, i.e., if 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 failure. $C_k^{FKP}(e)$, the total capacity of G_k^{FKP}, was obtained in the same way.

Fig. 4 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 the AT&T topology (523 nodes and 1304 links) for Fig. 4(a) and to the Sprint topology (467 nodes and 1280 links) for Fig. 4(b). The Sprint topology is also a measurement result obtained by the Rocketfuel tool [9]. Both figures indicate that our EVN design approach requires less link capacity than the FKP-based design method.

To see in more detail how a network with topological diversity can scale up with less equipment, we consider three kinds of link capacity: capacity for preparing for node failures, capacity for accommodating traffic, and unused capacity based on the difference of link capacity between before and after the addition of 50 nodes. Fig. 5(a) shows the results for the EVN design approach, and Fig. 5(b) shows the results for the FKP-based design method. Comparing Fig. 5(a) and (b), we can clearly see that the FKP-based design method requires more capacity for preparing for node failures, while capacity for accommodating traffic is almost the same as for the EVN method. This is caused by the overlap in equipment placement in each single node failure. Table 2 shows the average 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 2 that G_{450}^{EVN} needs more capacity on average to cover one pattern of single node failure than G_{450}^{FKP}. However, it needs less capacity to cover every pattern of single node failure. This is because the topology generated by the EVN design approach is not specialized to a single environment. Therefore, it can efficiently use the network equipment placed for one single node failure to cover another single node failure. We observe from Fig. 5 that the unused capacity in the EVN design approach is larger than that in the FKP-based design method. This means that the EVN design approach may under-utilize the capacity at a given stage of evolution. However, this unused capacity will used at the next (or a later) stage of evolution thanks to the unspecialized nature of the EVN design approach.

Table 2
Average additional capacity needed to cover each single node failure.

	EVN	FKP
Capacity	6.0535×10^3	5.8868×10^3

3.2. Reuse of facilities for unexpected environmental changes

In the previous subsection, we showed that a network with topological diversity requires less capacity during network growth. Thanks to the unspecialized design of the topology, most link capacity is reused in the new environment. However, that evaluation only assumed that link capacity is designed to protect against single node failure. This subsection evaluates evolvability for cases other than single node failure. However, since unpredicted environmental change is hard to define, we use a scenario of unpredicted environmental changes following the evaluation presented in [12]. 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 on a large scale. 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 Section 3.1. Although actual traffic demand will be different, our intention here is to show how the designed network reuses existing capacity in response to unexpected environmental changes. Thus, we use unit traffic demand for simplicity.

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

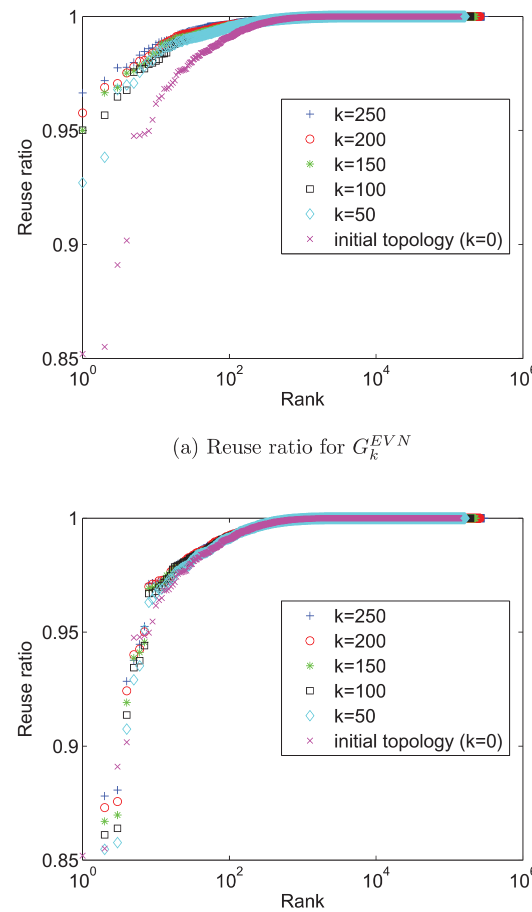
$$r_k = \frac{F_k^{reused}}{F_k^{new}}, \tag{9}$$

where F_k^{reused} represents the capacity that can be reused from the capacity that has already been deployed, and F_k^{new} represents the capacity that *was* required to deal with unpredicted environmental changes for the kth network, that is, the network with k nodes added. The ratio r_k ranges from 0 to 1.0. For r_k close to 1.0, capacity that is already in place can be reused for unpredicted environmental change. However, more capacity is required to deal with unpredicted environmental change as r_k decreases.

We evaluate the reuse ratio for the case of 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 the reuse ratios as $r_k^{EVN}(n_1, n_2)$ for G_k^{EVN} and $r_k^{FKP}(n_1, n_2)$ for G_k^{FKP}.

Fig. 6(a) shows $r_k^{EVN}(n_1, n_2)$ for all cases of two-node (n_1, n_2) failures and Fig. 6(b) shows $r_k^{FKP}(n_1, n_2)$. Note that we again use the AT&T topology as the initial topology. In these figures, the horizontal axis represents the rank of reuse ratio in ascending order, and we show the change of reuse ratio as a result of changing k. Looking at reuse ratios for ranks from 1 to 200, those obtained by the EVN design approach are higher than those of the FKP-based design method, and this tendency becomes clearer as k increases. This is due to the increase of topological diversity. Because alternative paths for a single node failure would be less likely to be biased toward 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. However, when the topology is less diverse, paths would be likely to be biased toward 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 in place that have less capacity than the biased links, which leads to lower values of reuse ratio.

We can also observe the non-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 for which $r_{250}^{FKP}(n_1, n_2)$ is less than 1 is 7557. This 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 a slight increase in the over-provisioning of links.



(a) Reuse ratio for G_k^{EVN}

(b) Reuse ratio for G_k^{FKP}

Fig. 6. Evaluation results of Scenario A: failure of two nodes.

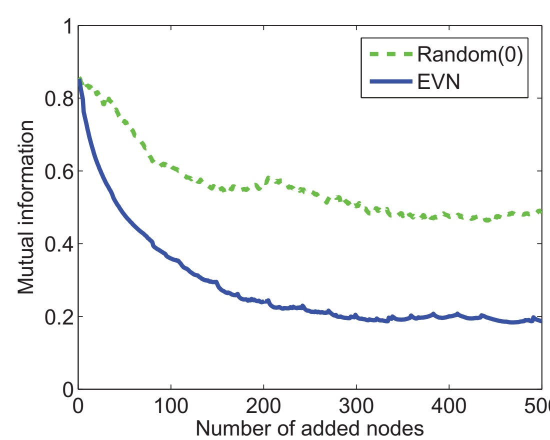


Fig. 7. Variation of mutual information.

4. Comparison with random attachment method

We showed that minimizing mutual information by the EVN design approach can lead to evolvable networks in Section 3. However, there are also other methods to lower mutual information. Since Solé et al. [3] showed that mutual information of a random graph can be approximately 0, a simple method could be to attach new nodes to randomly selected existing nodes. Though the computation time of that method is faster than the EVN design approach we show in this section that the randomly attachment method could not increase the

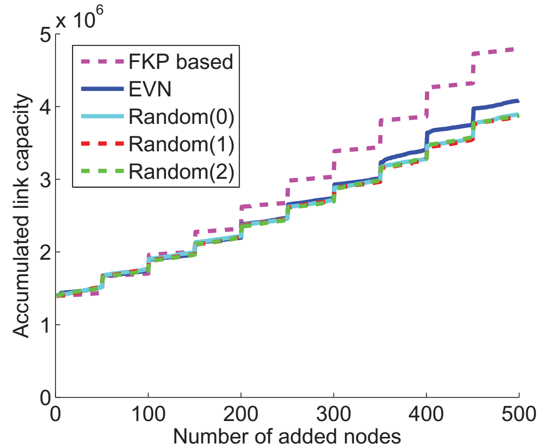


Fig. 8. Accumulated link capacity for the random attachment method.

topological diversity, and could have bad performance in terms of evolvability compared to the EVN design approach.

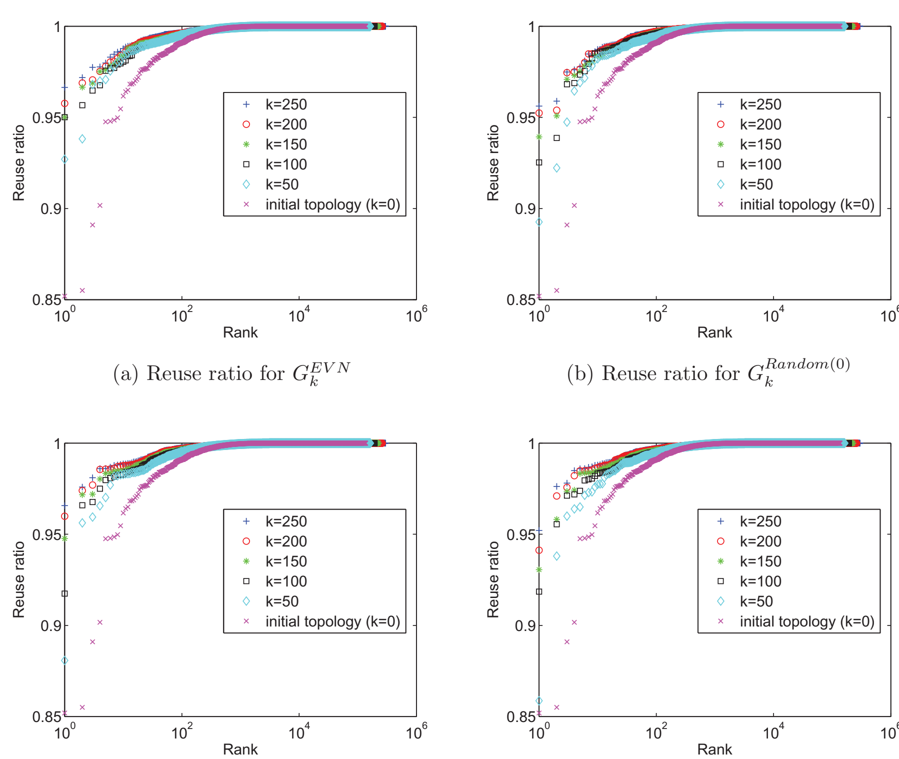
The random attachment method is to attach new nodes to randomly selected existing nodes. To maintain reliability we also add the entropy restriction to this method. Note that without the restriction, the evolvability, especially the accumulated capacity, would be much worse. To compare with the EVN design approach, , we added 2 links per node addition in the simulation. Hereafter we denote by G_k^{Random} the topology obtained by the random attachment method after the addition of k nodes. Fig. 7 shows the variation of mutual information during network growth. We can see that the mutual information of G_{499}^{Random} is approximately 0.5, while that of a random topology is approximately 0 as we stated in Table 1. We suppose that the mutual information of G_{499}^{Random} is influenced by the initial (AT&T) topology. Since the AT&T topology has a power-law behavior in its degree distribution, it has many nodes with degree 2. Therefore, when selecting nodes randomly from the AT&T topology, nodes with degree 2 have a high probability to be selected. Because nodes with degree 2 are mostly at the edge of the topology, newly added nodes will be attached to edge nodes with a high probability. Hence, it is difficult to increase the diversity of the core part of the topology, and we think this is the reason why the mutual information remains high as the network grows.

To see the difference in evolvability, we evaluate accumulated capacity during network growth and reuse of facilities for unexpected environmental changes. Details of these evaluations are given in Sections 3.1 and 3.2, respectively. We used three different random seeds for the simulation. Hereafter, $G_k^{Random(v)}$ denotes a topology generated by a seed v.

The accumulated capacity is shown in Fig. 8. Total amount of facilities of $G_{499}^{Random(0)}$ is lower than that of G_{499}^{EVN}. This means the random attachment method can save facilities when compared with the EVN design approach. However, this is only in the environment which is expected. In an unexpected environment, the network produced by the EVN design approach performs better than that produced by the random attachment method. Fig. 9 shows reuse ratios under unexpected environmental changes. We can see that the worst reuse ratio of $G_k^{Random(0)}$ is lower for every k than that of G_k^{EVN}. We suppose this is caused by the lower diversity of the core part of $G_k^{Random(0)}$ as we explained above. Though we only used seed zero in the explanation, topologies generated with other seeds also have the same tendency.

5. Trade-off with physical distance

In Section 3, we showed that our EVN design approach is better than the FKP-based design method in terms of evolvability. However, links in a topology designed by the EVN design approach have large



(a) Reuse ratio for G_k^{EVN} (b) Reuse ratio for $G_k^{Random(0)}$

(c) Reuse ratio for $G_k^{Random(1)}$ (d) Reuse ratio for $G_k^{Random(2)}$

Fig. 9. Results for the random attachment method in the unexpected environmental change scenario.

Table 3
Values of the two measures for G_{499}^{EVN} and G_{499}^{FKP}.

	G_{499}^{EVN}	G_{499}^{FKP}
Mutual information	0.186689	0.791514
Physical distance	477.95	157.855
Total facilities	$4.0870{*}10^6$	$4.7990{*}10^6$

physical lengths. In more detail, the total physical length of all links of G_{499}^{EVN} is about 3 times greater than that of G_{499}^{FKP}. In this section, we are going to show that, even considering physical distance, an approach that increases topological diversity can lead to an evolvable network. Here, we use an objective function that considers both physical distance and topological diversity to generate networks, and discuss whether a network with topological diversity is evolvable even taking physical distance into account.

The objective function we used is a weighted sum of mutual information and physical distance:

$$\zeta \cdot I(q) + \sum_{i \in M} f(i), \tag{10}$$

The first term consists of a weight ζ and the mutual information of remaining degree $I(q)$. When ζ approaches infinity, the topology generating process is almost as same as the EVN design approach. The second term is the summation of $f(i)$ which is the objective function used in the FKP-based design method:

$$f(i) = \alpha \cdot d(n_{new}, n_i) + h(n_i, n_0) \tag{11}$$

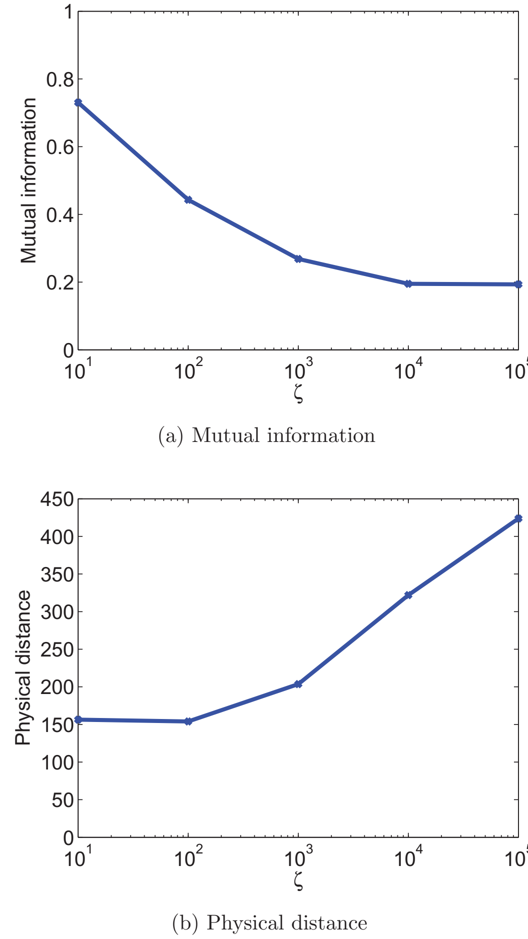
See Eq. (A.1) in Appendix A for details. M in Eq. (10) represents a set of candidate nodes connected with a newly added node n_{new}. If m links are added to n_{new} in each step, the number of elements in M will be m. When ζ is 0, the topology generating process will search for an M that minimizes $\Sigma_{i \in M} f(i)$ at each step. This is as same as choosing m nodes having small $f(i)$ in an ascending order. Therefore, this is the same as the FKP-based design method.

The entropy restriction is also changed. Since the entropy restriction should be active when ζ approaches infinity, we set the entropy restriction to be

$$E(\zeta) = (2/\pi) \cdot \arctan \zeta \cdot H_0(\boldsymbol{q}). \tag{12}$$

Therefore, when ζ approaches infinity, $E(\zeta)$ is $H_0(\boldsymbol{q})$, and when ζ is 0, $E(\zeta)$ is 0.

Accumulated capacity during network growth and reuse of facilities for unexpected environmental changes are evaluated for ζ equal to 0, 10, 100, 1000, 10 000, 100 000. For the details of the evaluation methods, please see Sections 3.1 and 3.2, respectively. To compare with the EVN design approach and the FKP-based design method, in the simulation we added 2 links for every node addition. Hereafter, the topology obtained by adding k nodes for $\zeta = p$ will be denoted by $G_k^{\zeta=p}$. To investigate how mutual information and physical distance changes with ζ, we show values for $G_{499}^{\zeta=0}$, $G_{499}^{\zeta=10}$, $G_{499}^{\zeta=100}$, $G_{499}^{\zeta=1000}$, $G_{499}^{\zeta=10\,000}$ and $G_{499}^{\zeta=100\,000}$ in Fig. 10(a) and in Fig. 10(b), respectively. Note that physical distance here indicates the total physical length of all links when nodes are placed in $[0, 1]^2$. Details are given in Appendix A. For any ζ, a node newly added at step k is placed in the same physical position that the FKP-based design method would place a new node at step k. For comparison, the mutual information and physical distance of G_{499}^{EVN} and G_{499}^{FKP} are shown in Table 3. When ζ is 0, 10 or 100, mutual information is close to that of G_{499}^{FKP}. The reason why the mutual information of G_{499}^{FKP} differs from that of $G_{499}^{\zeta=0}$ is that, in some steps, different nodes are chosen to attach to a new node when there are more than 3 nodes that all minimize Eq. (11). The mutual information for $G_{499}^{\zeta=10\,000}$ or $G_{499}^{\zeta=100\,000}$ is close to that of G_{499}^{EVN}, and that of $G_{499}^{\zeta=1000}$ is just 0.075 higher than that of G_{499}^{EVN}. When ζ is 0, 10 or 100, the physical distance is almost as same as that of G_{499}^{FKP}. The physical distance for $G_{499}^{\zeta=1000}$ is only 1.3 times larger than that of G_{499}^{EVN}; while those of $G_{499}^{\zeta=10\,000}$ and $G_{499}^{\zeta=100\,000}$ are more than 2 times larger than that of G_{499}^{EVN}.



(a) Mutual information

(b) Physical distance

Fig. 10. Relationships between ζ and the two measures.

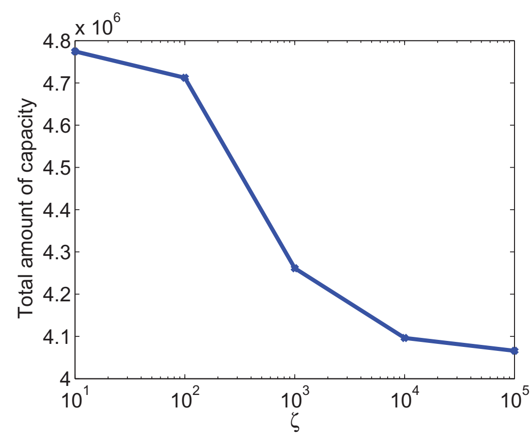
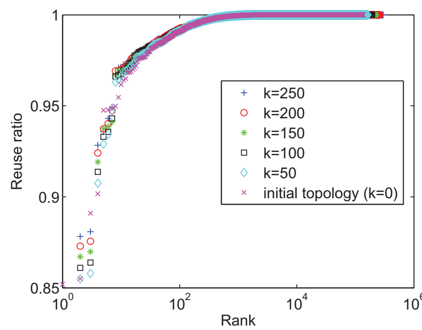
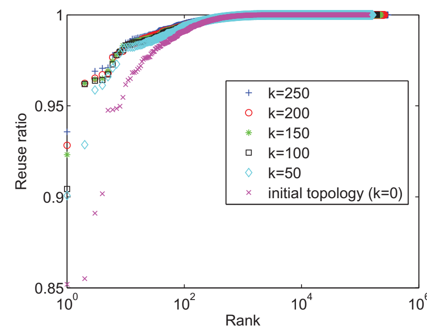


Fig. 11. Relationship between ζ and total capacity.

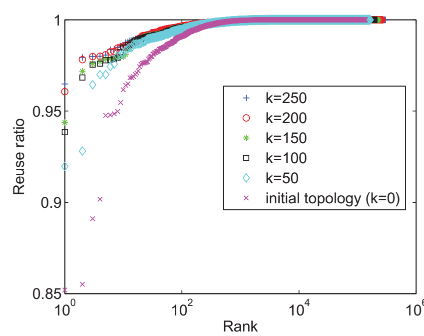
Fig. 11 shows how total capacity decreases as ζ increases. There is a large difference between $G_{499}^{\zeta=100}$ and $G_{499}^{\zeta=1000}$, while there is only a slight difference in capacity needed for $G_{499}^{\zeta=10\,000}$ and $G_{499}^{\zeta=100\,000}$. If one allows 1.3 times more physical distance than that used in the FKP-based design method, than one can save 11% of total link capacity



(a) Reuse ratio for $G_k^{\zeta=10}$



(b) Reuse ratio for $G_k^{\zeta=1000}$



(c) Reuse ratio for $G_k^{\zeta=100\ 000}$

Fig. 12. Results for the approach considering physical distance in the unexpected environmental change scenario.

after the addition of 499 nodes. Moreover, much more capacity can be expected to be saved when the network grows even larger.

Fig. 12 shows how the reuse of facilities for unexpected environmental changes depends on ζ. Results for ζ equal to 10, 1000 and 100 000 are shown in Fig. 12(a), Fig. 12(b), Fig. 12(c), respectively. We can see that, though the worst reuse ratio of $G_{250}^{\zeta=1000}$ is worse than that of $G_{250}^{\zeta=100\,000}$, it is better than that of $G_{250}^{\zeta=10}$.

6. Conclusion and future work

In this paper, we have proposed a design approach (the EVN design approach), based on minimizing mutual information, to strengthen topological diversity and make an evolvable network. We have shown that a network grown using our design approach can grow with less capacity than 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, so a network grown using our design approach shows an overlap between equipment placement in an old environment and that in a new one. We also showed that the random attachment method could not increase the topological diversity compared to the EVN design approach, so that a topology designed by the EVN design approach could have a high reuse ratio under unexpected environmental change scenarios. Although the EVN design approach did not consider the physical lengths of links, we showed that, even considering the physical lengths of links, an approach that increases topological diversity can lead to an evolvable network.

Several problems are left for future research. First, further evaluation of the EVN design approach is needed. In the simulations in this paper, we only add two links for each additional node in order to keep the average degree similar to that of the initial topology. However, there are also other cases in practice. Although we believe that the topology will also be diverse and evolvable when adding three or more links for a node, this should be investigated by further simulation. Second, because the order of the EVN design approach is $O(n^2 \cdot d^2)$, where n is the number of nodes and d is the degree, there is a scalability problem. However, because the purpose is to enhance topological diversity, strict minimization may not be needed. Approximate solutions can be considered in future work. Third, analytical investigation is required to provide a clearer discussion of the evolvability of networks designed using our approach in response to several other unexpected environmental changes. Lastly, we have considered topological diversity here, but diversity at a higher-level, such as the diversity of link capacity distribution or processing capacity of nodes may help to improve evolvability.

Acknowledgment

This work was partly supported by Grant-in-Aid for Scientific Research (A) 24240010 from the Japan Society for the Promotion of Science (JSPS).

Appendix A. Network design method based on the FKP model

The FKP model proposed by Fabrikant et al. [10] incrementally adds nodes and connects existing nodes such that physical distance and hop distance metrics are minimized.

In the original FKP model, the first node n_0 is set to be 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:

$$\alpha \cdot d(n_{new}, n_i) + h(n_i, n_0), \tag{A.1}$$

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 the 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-neighbor nodes. A topology with high α can be shown to behave like a random topology. A power-law degree distribution appears at moderate values of α. The power-law attribute here is used to determine moderate α. Though the power-law degree distribution found in [10] is said to be different from those given by other Internet models, we think this point is not important here.

For comparing with our method, we modified the FKP model as follows. Given a topology $G_0(V_0, E_0)$ and the physical locations of nodes, our modified version of the FKP model adds a node and m links for each node addition in the k-th step according to 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 the node existence ratio in each area. The node existence ratio of an area is defined as the number of nodes 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 existence ratio.
4. Calculate the distance metric defined by Eq. (A.1) for each existing node n_i.
5. Select m nodes in ascending order of their value of the 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 node is determined with a probability proportional to the node existing ratio (Step (ii) above). The reason for this is that, because routers are often added to areas where traffic demand is large, an area attracts more traffic as more routers exist in the area. Second, we add multiple links per node addition so that the average degree of the designed networks can be controlled (Step (v)).

In the evaluations in Sections 3.1 and 3.2, the parameter α was set to 10.0, where the average hop distance is lowest under the condition that the entropy $H(\boldsymbol{q})$ is moderate, so as not to obtain a star-like topology.

References

[1] M. Prokopenko, F. Boschetti, A. Ryan, An information-theoretic primer on complexity, self-organization, and emergence, Complexity 15 (1) (2009) 11–28.

[2] C. Dovrolis, J.T. Streelman, Evolvable network architectures: what can we learn from biology? ACM SIGCOMM Comput. Commun. Rev. 40 (2) (2010) 72–77.

[3] R. Solé, S. Valverde, Information theory of complex networks: on evolution and architectural constraints, Complex Netw. 650 (2004) 189–207.

[4] M.E. Newman, Assortative mixing in networks, Physical Rev. Lett. 89 (20) (2002) 208701.

[5] L. Chen, S. Arakawa, M. Murata, Quantifying network heterogeneity by using mutual information of the remaining degree distribution, Int. J. Adv. Syst. Meas. 6 (1 and 2) (2013) 214–223.

[6] L. Li, D. Alderson, W. Willinger, J. Doyle, A first-principles approach to understanding the Internet's router-level topology, ACM SIGCOMM Comput. Commun. Rev. 34 (2004) 3–14.

[7] J. Lloyd-Price, A. Gupta, A.S. Ribeiro, Robustness and information propagation in attractors of random boolean networks, PloS One 7 (7) (2012) e42018.

[8] B. Wang, H. Tang, C. Guo, Z. Xiu, Entropy optimization of scale-free networks robustness to random failures, Physica A: Stat. Mech. Appl. 363 (2) (2006) 591–596.

[9] N. Spring, R. Mahajan, D. Wetherall, T. Anderson, Measuring ISP topologies with rocketfuel, IEEE/ACM Trans. Netw. 12 (1) (2004) 2–16.

[10] A. Fabrikant, E. Koutsoupias, C. Papadimitriou, Heuristically optimized trade-offs: a new paradigm for power laws in the Internet, Autom., Lang. Program. 2380 (2002) 110–122.

[11] A. Nucci, S. Bhattacharyya, N. Taft, C. Diot, IGP link weight assignment for operational Tier-1 backbones, IEEE/ACM Trans. Netw. 15 (4) (2007) 789–802.

[12] J. Whitacre, P. Rohlfshagen, A. Bender, X. Yao, Evolutionary mechanics: new engineering principles for the emergence of flexibility in a dynamic and uncertain world, Nat. Comput. 11 (2012) 431–448.