Topology Design and Capacity Planning for Evolvable Information Networks Using Mutual Information

Submitted to

Graduate School of Information Science and Technology
Osaka University

January 2016

Lu CHEN

List of Publication

Journal papers

1. Lu Chen, Shin'ichi Arakawa, and Masayuki Murata,"Quantifying Network Heterogeneity by Using Mutual Information of the Remaining Degree Distribution, "*International Journal On Advances in Systems and Measurements*, vol. 6, no. 1 and 2, pp. 214-223, June 2013.

2. Lu Chen, Shin'ichi Arakawa, Hideyuki Koto, Nagao Ogino, Hidetoshi Yokota, and Masayuki Murata,"An Evolvable Network Design Approach with Topological Diversity, " to appear in *Computer Communications*, November 2015.

Refereed Conference Papers

1. Lu Chen, Shin'ichi Arakawa, and Masayuki Murata, "Analysis of Network Heterogeneity by Using Entropy of the Remaining Degree Distribution," in *Proceedings of International Conference on Advanced Communications and Computation (INFOCOMP 2012)*, pp. 84-89, October 2012.

2. Lu Chen, Shin'ichi Arakawa, Masayuki Murata, Hideyuki Koto, Nagao Ogino and Hidetoshi Yokota,"Designing an Evolvable Network with Topological Diversity, " in *Proceedings of IEEE International Workshop on Network Science for Communication Networks (in conjunction with IEEE Infocom 2014)*, pp. 789-794, April 2014.

Non-Refereed Technical Papers

1. Lu Chen, Shin'ichi Arakawa, and Masayuki Murata,"Analysis of Network Heterogeneity by Using Entropy of the Remaining Degree Distribution, "*Technical Report of IEICE (NS2012-13)*, vol. 112, no. 8, pp. 71-76, April 2012 (in Japanese).

2. Lu Chen, Shin'ichi Arakawa, and Masayuki Murata,"Sustainability Analysis of Topologies with Different Network Heterogeneity Described by Mutual Information of Remaining Degree, "*Technical Report of IEICE (NS2012-157)*, vol. 112, no. 392, pp. 93-98, January 2013 (in Japanese).

3. Lu Chen, Shin'ichi Arakawa, Hideyuki Koto, Nagao Ogino, Hidetoshi Yokota, and Masayuki Murata,"A Proposal of Evolvable Network Designing Approach with Topological Diversity, " *Technical Report of IEICE (PN2014-26)*, vol. 114, no. 334, pp. 31-36, November 2014.

Preface

As the Internet becomes a social infrastructure, it is important to design the Internet that has adaptability against environmental changes. However, operators of ISP networks usually add link capacity and routers in an ad-hoc way. That is, they enhance link capacity when link utilization exceeds a certain threshold, and they introduce new routers when existing routers become unable to accommodate traffic on enhanced links. Such an ad-hoc approach will not be useful when traffic volume is changing drastically and unpredictably. That is, when the nodes/links are added in an ad-hoc manner, those would not help accommodate another traffic increase or possibly taking place in the future. Even worse, those newly added nodes/links would be useless when the traffic is decreased around those equipment. Therefore, a new network design method which has a capability to adapt various kinds of environmental changes with less amount of equipment is necessary.

In this thesis, we propose a new network design method that is adaptive to dynamic environmental changes including traffic changes and node failures. Unlike a traditional design approach, our design approach tries to reduce a degree of specialization. That is, we do not simply increase network resources when facility expansion is found to be necessary. Instead, we tries to increase adaptability against future possible traffic changes as much as possible. In doing so, two problems arise for developing the new design method; how to quantify "a degree of specialization" and how to minimize or reduce "a degree of specialization."

We first introduce a mutual information to quantify the degree of specialization in a topological sense. Specifically, we define the topological diversity by the mutual information between degrees of two nodes that are connected by the direct link. We compare the mutual information for various topologies including router-level topologies and biological networks, and find that the mutual information of router-level topologies is higher than that of biological networks. The reason of high mutual information in the router-level topologies can be explained as follows. Since router-level topologies are designed under the physical and technological constraints such as the number of switching ports and/or maximum switching capacity of routers, there are some restrictions to construct the topology. Such the constraints lead to high correlation in degrees of two connected nodes. Further, and more importantly, progress of facility expansion leads to higher mutual information due to the specialization taken by environmental changes. It is often pointed out that the diversity is a source of keeping evolution in the biological networks. In our study, the diversity is quantified by the mutual information, and we can say that the router-level topologies are less diverse when compared with the biological networks. Actually, the network with low mutual information has a potential to evolve in various environments, that is, to be adaptive against the traffic changes and/or node failures.

We then propose a network design approach to enhance topological diversity by which the network can be easily adapted to deal with new environments without requiring a lot of additional equipment. Essentially, in our approach, a new node is connected to existing nodes to minimize the mutual information of the topology. We then evaluate the total cost, which is defined by the total amount of equipment, needed for accommodating traffic in two cases; in an ordinary situation where the traffic is increased gradually, and in the situation where a node failure takes place. Our results show that a thousand-node network evolved by our design approach reduces the total cost of the equipment by 15% comparing to a thousand-node network evolved by an ad-hoc design method.

Finally, we consider the diversity of link capacity in addition to the topological diversity. For

that purpose, we extend the definition of mutual information by considering load of the link and the available capacity after a failure of the link. Then, a network with low mutual information is obtained by repeatedly exchanging a small amount of capacity between links. Although the diversity of available capacity does not contribute for the case of single link failure, we expect the additional capacity will work for severe environmental changes and expect that less capacity is required as in the case of topological diversity. Using a 15-node topology with 28 links, we examine the effectiveness of the capacity planning with low mutual information. Our results show that the total amount of capacity is decreased when two or more links are failed simultaneously and is decreased by 20% at a seven-link failure.

Acknowledgments

This thesis could not have been accomplished without support from many people.

Prof. Masayuki Murata, who is my supervisor, introduced me to this research. The research started from his idea. He advised me when I lose my direction. I cannot think this research done without him. The greatest thanks go to him.

Members of the thesis committee, Prof. Teruo Higashino, Prof. Morito Matsuoka, Prof. Takashi Watanabe and Prof. Toru Hasegawa gave me helpful feedbacks. I would like to appreciate it.

Associate Prof. Shin'ichi Arakawa, who is my research group leader, supported me the most. Without his kind guidance, my research could not be finished. Special thanks go to him.

Dr. Hidetoshi Yokota, Dr. Nagao Ogino and Dr. Hideyuki Koto from KDDI laboratory gave me advices throughout the three years. I would like to appreciate their suggestion.

My laboratory staffs, Assistant Prof. Yuichi Ohsita and Assistant Prof. Daichi Kominami and past and present members in my laboratory gave me many worthful advices. I would like to thank all of them.

Last but not least, I would like to express my deepest gratitude to my parents for their many years of devoted support they provided during my undergraduate and graduate studies, which provided the foundation for this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The traffic in the Internet becomes more changeable and unpredictable. Actually, it is estimated that download and upload traffic by broadband subscribers is about 5.7 Tbps in Japan in 2015 [1], and is increased by 1.9 Tbps in the latest one year. The speed of traffic growth is 1.5 times per a year. Since this is only the current total traffic growth, the traffic around servers providing a new service which attracts many users is increasing even more. Furthermore, a Cisco report [3] forecasts that the total global IP traffic will grow about 1.9 times in the next four years as shown in Fig. 1.1. The increase of mobile traffic will be 1.9 times larger than that of traffic in fixed-line Internet. This infers that, not only traffic growth, but also changes in traffic pattern has a possibility to occur in the near future. According to an accident report [4], 54% of the accidents of telecommunication carriers in 2014 in Japan is caused by external factors. Among them, 86% of the accidents are caused by other carriers stopping their service, and the rest of them are caused by cable cut and power-outage. Comparing with the past report, the number of accidents caused by external factors is increasing.

Traditionally, ISPs design information networks with over-provisioning [5]. That is, ISPs put some extra link capacity and routers more than they actually needed. With the over-provisioning,

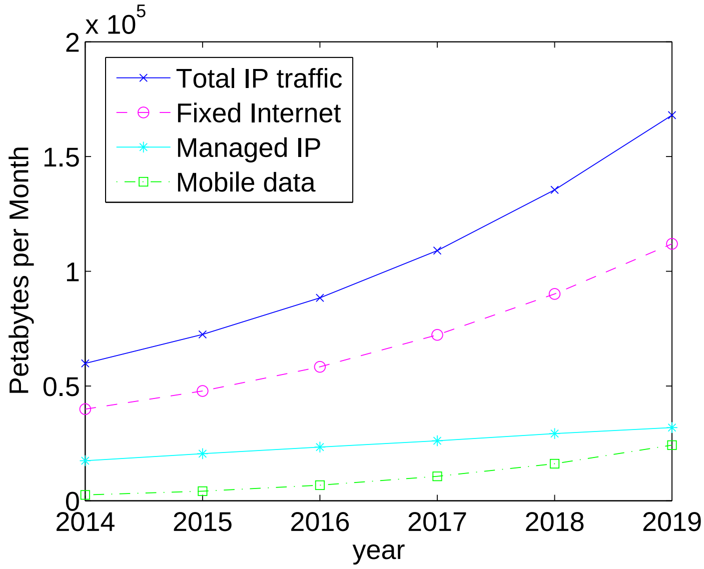


Figure 1.1: Traffic growth forecast (data from [1])

ISPs expect that a congestion will not occur when traffic changes. However, operators of ISP networks usually add link capacity and routers in an ad-hoc way. That is, they enhance link capacity when link utilization exceeds a certain threshold, and they introduce new routers when existing routers become unable to accommodate traffic on enhanced links. Such an ad-hoc approach will not be useful when traffic volume is changing drastically and unpredictably. That is, when the nodes/links are added in an ad-hoc manner, those would not help accommodate another traffic increase or possibly taking place in the future. Even worse, those newly added nodes/links would be useless when the traffic is decreased around those equipment. Therefore, a new network design method which has a capability of following various kinds of environmental changes with less amount of equipment is necessary. Throughout this thesis, we call a network having an adaptability against new environments without needs to increase/decrease its resources as "an evolvable network".

There are plenty of studies for network design methods. Many of studies formulate a network design problem as an optimization problem, and try to minimize, for example, operating cost by using given traffic demand, network size and equipment cost (see, e.g., [6]). However, when an environment surrounding networks changes rapidly, and the changes are hard to predict, a network designed through optimization does not guarantee its optimality under an unknown and/or inexperienced environment. Therefore, a new network design method which can follow any kind of environmental changes with less amount of equipment should be considered in facing with changeable and unpredictable environment.

To approach this problem, we can learn from a biological system which has a robustness to environmental changes. It is known that living things survive from environmental changes by repeatedly adapting and evolving themselves for a long while, and many researches in biological field investigate the mystery of mechanisms which let the biological system be robust [7, 8]. In a recent study, it has been found that the biological system has a characteristic where a component of the system is not intended to serve for a specific function, but is intended to serve for various kinds of functions [9, 10]. Thanks to the low degree of specialization of the components, a biological system obtain a robustness against failures of components in the system.

In this thesis, we propose a new network design method that can be adaptive to dynamic environmental changes including traffic changes and/or node failures. Unlike a traditional design approach, our design method tries to reduce the degree of *specialization*. That is, we do not simply increase network resources when facility expansion is found to be necessary. Instead, we tries to increase adaptability against future possible traffic changes as much as possible. In doing so, two problems arise for developing the new design method; how to quantify "a degree of specialization" and how to minimize or reduce "a degree of specialization."

We first introduce a mutual information to quantify the degree of specialization in a topological sense. Specifically, we define the topological diversity by the mutual information between degrees of two nodes that are connected by the direct link. Note that the concept of the mutual information is not a new measure. Solé et al. use the mutual information in [11] to analyze a topological characteristic. They calculate the mutual information of remaining degree distribution of biological networks and artificial networks such as software networks and electronic networks, and show that both of them have higher mutual information than randomly connected networks. In our study, we use the concept of mutual information to design an information network rather than to analyze the topological characteristic. We compare the mutual information for various topologies including router-level topologies and biological networks, and find that the mutual information of router-level topologies is higher than that of biological networks. The reason of high mutual information in the router-level topologies can be explained as follows. Since router-level topologies are designed under the physical and technological constraints such as the number of switching ports and/or maximum switching capacity of routers, there are some restrictions to construct the topology. Such constraints lead to high correlation in degrees of two connected nodes. Further, and more importantly, a progress of facility expansion leads to higher mutual information due to the specialization taken by environmental changes. It is often pointed out that the diversity is a source of keeping evolution in the biological networks [12]. In our study, the diversity is quantified by the mutual information, and we can say that the router-level topologies are less diverse when compared with the biological networks. Actually, the network with low mutual information has a potential to evolve in various environments, that is, to be adaptive against the traffic changes and/or node failures.

We then propose a network design approach to enhance topological diversity, by which the network is easily adaptable to deal with new environments without requiring a lot of additional equipment. Essentially, in our approach, a new node is connected to existing nodes to minimize the mutual information of the topology. We then evaluate the total cost, which is defined by the total amount of equipment needed for accommodating traffic in two cases; in an ordinary situation where the traffic is increased gradually, and in the situation where node failure takes place. Our results show that a thousand-node network evolved by our design approach reduces the total cost of equipment by 15% comparing to a thousand-node network evolved by an ad-hoc design method. Next, we evaluate the amount of additional capacity needed for accommodating traffic to deal with unpredicted environmental changes. Here, link capacities are assigned by assuming that single node failure occurs at networks, and we then consider a two-node failure for the unpredicted environment. Our results show that, when two-node failures occur, the network designed by our design approach can use existing capacity more than that by the ad-hoc design method. Note that a network evolved with our design approach requires more capacity on average to cover every pattern of single node failure than that evolved by the ad-hoc design method. This is because the topology constructed by our design approach is not specialized to a single node failure since our design approach does not consider physical length of links. However, such the unspecialized nature results in less requirements of link capacity to cover various kinds of environmental changes.

Finally, we consider the diversity of link capacity in addition to the topological diversity. For that purpose, we extend the definition of mutual information by considering load of the link and the available capacity after a failure of the link. Then, a network with low mutual information is obtained by repeatedly exchanging a small amount of capacity between links. Although the diversity of available capacity does not contribute in the case of single link failure, we expect the additional capacity will work for severe environmental changes and expect that less capacity is required as in the case of topological diversity. Using a 15-node topology with 28 links, we examine the effectiveness of the capacity planning with low mutual information. Our results show that the total amount of capacity is decreased when two or more links are failed simultaneously and is decreased by 20% at a seven-link failure.

1.2 Outline of Thesis

Quantifying Diversity in Network Using Mutual Information [13–16]

In Chap. 2, we first introduce a mutual information to quantify the degree of specialization in a topological sense. More specifically, we quantify the topological diversity by the mutual information between degrees of two nodes that are connected. From calculating the mutual information of router-level topologies and a BA topology, we find that the mutual information of router-level topologies are higher than that of the BA topology. The reason of high mutual information in the router-level topologies is partly explained by the fact that, since router-level topologies are designed under the physical and technological constraints such as the number of switching ports and/or maximum switching capacity of routers, there are some restrictions when constructing the network, leading to high correlation in degrees of two connected nodes. Then, we show some illustrative examples of topologies with different mutual information. We generate a mutual information maximized topology and minimized one whose number of nodes and links are same as that of a router-level topology. From the illustrative examples, we show that topology with high mutual information is less diverse, and have more regularity than the one with low mutual information.

Topology Design Approach Using Mutual Information for Evolvable Networks [17–19]

In Chap. 3, using the mutual information, we propose a network design approach, called EVN (EVolvable Network), to enhance topological diversity. In the approach, a new node is connected to existing nodes to minimize the mutual information of the topology. We compare the link capacity needed for accommodating traffic in an ordinary situation and single node failure occurring situations with an ad-hoc design method. We find that link capacity needed for a network grown with our design approach is less than that grown with the ad-hoc design method. We also find that a network designed by our design approach can reuse more link capacity than that by the ad-hoc design method. Next, we evaluate for additional capacity needed for accommodating traffic to deal with unpredicted environmental changes. As an unpredicted environment, we chose two node failures since this is not considered in the designing period. Our results show, when two node failures occur, the network designed by our design approach can use already placed capacity more than that by the ad-hoc design method. Note that a network grown with our design approach needs more capacity on average to cover one pattern of single node failure than that grown by the ad-hoc design method. This is because the topology generated by our design approach is not specialized to a single environment. However, such the unspecialized nature results in the necessity of less capacity to cover various kinds of environmental changes. Another disadvantage is that links in a topology designed by our design approach have large physical length since our design approach does not consider physical length of links. To see the disadvantage more clearly, we investigate the relationship between topological diversity and physical distance. We find that there is a trade-off relationship between them, and we can save a large amount of total link capacity with a little increase in physical distance.

Capacity Planning Using Mutual Information for Evolvable Networks

In Chap. 4, we present a capacity planning method that enhances the diversity in link capacity to make information networks more adaptive to large environmental changes with less amount of equipment. Our basic idea is to assign an additional capacity on each link such that the capacity can work for more severe environmental changes. More specifically, our method increases the diversity of available capacity on restoration path under single link failures. Although the diversity of available capacity does not contribute in the case of single link failure, we expect the additional capacity will work for severe environmental changes and expect that less capacity is required as in the case of topological diversity. To quantify the diversity of available capacity under the single link failures, we extend the definition of mutual information by considering load of the link and the available capacity after a failure of the link. Then, a network with low mutual information is obtained by repeatedly exchanging a small amount of capacity between links. Using a 15-node topology with 28 links, we examine the effectiveness of the capacity planning with low mutual information. Our results show that the total amount of capacity is decreased when two or more links are failed simultaneously and is decreased by 20% at a seven-link failure.

Chapter 2

Quantifying Diversity in Network Using Mutual Information

2.1 Introduction

As the Internet becomes the social infrastructure, it is important to design the Internet that has adaptability and sustainability against environmental changes [14, 20]. However, dynamic interactions of various network-related protocols make the Internet into a complicated system. For example, it is shown that interactions between routing at the network layer and overlay routing at the application layer degrade the network performance [20]. Therefore, a new network design method which has the adaptability against the failure of network equipment and has the sustainability against changes of traffic demand is becoming important. Since complex networks display heterogeneous structures that result from different mechanisms of evolution [11], one of the key properties to focus on is the network heterogeneity where, for example, the network is structured heterogeneous rather than homogeneous by some design principles of ISP networks.

Recent measurement studies on the Internet topology show that the degree distribution exhibits a power-law attribute [21]. That is, the probability P_x, that a node is connected to x other nodes, follows $P_x \propto x^{-\gamma}$, where γ is a constant value called scaling exponent. Generating methods of models that obey power-law degree distribution are studied widely, and Barabáshi-Albert (BA) model is one of it [22]. In BA model, nodes are added incrementally and links are placed based on the connectivity of topologies in order to form power-law degree distribution. The resulting topology has a large number of nodes connected with a few links, while a small number of nodes connected with numerous links. Topologies generated by BA model are used to evaluate various kinds of network performance [23, 24].

However, it is not enough to explain topological characteristics of router-level topologies by such models. It is because topological characteristics are hardly determined only by degree distribution [2, 25]. Li et al. [2] enumerate several different topologies with power-law, but identical degree distribution, and show the relation between their structural properties and performance. They point out that, even though topologies have a same degree distribution, the network throughput highly depends on the structure of a topology. The lessons from this work suggest us that the heterogeneity of the degree distribution is insufficient to discuss the topological characteristics and the network performance of router-level topologies.

In this chapter, we focus on the property of diversity, which is extensively studied in biological systems. Biological systems are systems that evolve robustly under many kinds of environmental changes. Similar studies have been done in complex system field [9,26–28]. Many of their networks also exhibit power-law attribute. A study of a key mechanism for adapting to environment changes in biological systems [9] explain that, because the system components can contribute to required traits diversely, the system can get traits required in a new environment by changing their contribution adaptively. Prokopenko et al. [12] consider changes of the diversity in a growing process of some complex systems. They say that an organized system, which we consider as a less diverse system here, has less configurations to evolve. They also said that a disorganized system, which is a diverse system, has more configurations to evolve, which makes the sytem be easily adapt to different environment. Therefore, the diversity is an important and applicable property to focus on in router-level topologies.

In [12], they used mutual information to measure the complexity, which we consider as diversity here. Inspired from their work, we investigate the topological diversity of router-level topologies by using mutual information. Mutual information yields the amount of information that can obtain about one random variable X by observing another variable Y. The topological diversity can be measured by considering Y as some random variable of a part of the topology and X as the rest of it. Solé et al. [11] study complex networks by using a remaining degree distribution as the random variable. They calculate the mutual information of remaining degree distribution of biological networks and artificial networks such as software networks and electronic networks, and show that both of them have higher mutual information than randomly connected networks. In this chapter, we evaluate the mutual information of some router-level topologies, and show that the mutual information represents the topological diversity as well.

Heterogeneity of structures have also been studied by Milo et al. [29]. They have introduced a concept called network motif. The basic idea is to find several simple sub graphs in complex networks. Arakawa et al. [25] shows the characteristic of router-level topologies by counting the number of each kind of sub graph which consists of four nodes respectively. They conclude that router-level topologies have more sub-graphs called “sector”, that is removing one link from four-node complete graph, than other networks. However, the network motif is expected to evaluate the frequency of appearance of simple structure in a topology, and is not suitable to measure the diversity of topology.

The rest of this chapter is organized as follows. The definition of remaining degree and mutual information is explained in Sec. 2.2. We investigate the topological characteristic and give some illustrative examples by changing the mutual information through a rewiring process in Sec. 2.3. In Sec. 2.4, mutual information of several router-level topologies is calculated, and is shown. Another topological characteristic, which is from the ISP network aspect, is shown in there too. Finally, we conclude this chapter in Sec. 2.5.

2.2 Definitions

Information theory was originally developed by Shannon for reliable information transmission from a source to a receiver. Mutual information measures the amount of information that can be obtained about one random variable by observing another. Solé et al. [11] use remaining degree distribution as the random variable to analyze complex networks. In this section, we explain the definitions of the mutual information of remaining degree.

Remaining degree k is defined as the number of edges leaving the vertex other than the one arriving along, so the remaining degree is one less than the degree of nodes. The example is shown in Fig. 2.1, where the remaining degree is set to two for the left node and three for the right node.

The distribution of remaining degree $q(k)$ is obtained from:

$$q(k) = \frac{(k+1)P_{k+1}}{\Sigma_k k P_k}, \tag{2.1}$$

where $P(P_1, \ldots, P_x, \ldots, P_K)$ is the degree distribution, and K is the maximum degree among nodes.

The mutual information of remaining degree distribution, $I(q)$, is

$$I(\text{q}) = H(\text{q}) - H_c(\text{q}|\text{q'}), \tag{2.2}$$

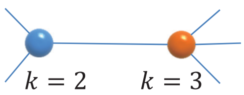


Figure 2.1: Example of remaining degree

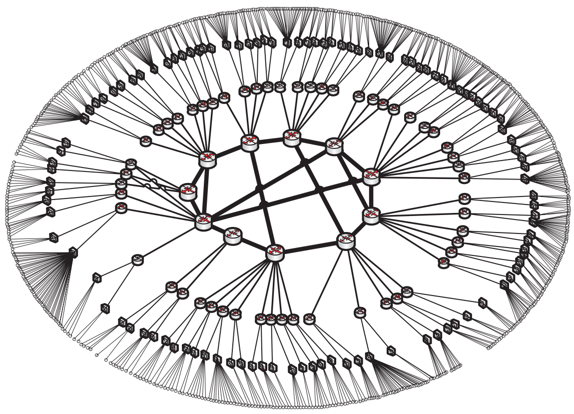


Figure 2.2: Abilene-inspired topology (referred from [2])

where q=($q(1)$, ... , $q(i)$, ... , $q(N)$) is the remaining degree distribution, and N is the number of nodes.

The first term $H(\mathrm{q})$ is entropy of remaining degree distribution:

$$H(\mathrm{q}) = -\sum_{k=1}^{N} q(k)\log(q(k)), \tag{2.3}$$

and $H(\mathrm{q})$ is greater or equal to 0. Within the context of complex networks, $H(\mathrm{q})$ provides an average measure of network's heterogeneity, since $H(\mathrm{q})$ measures the dispersion of the degree distribution of nodes attached to every link. $H(\mathrm{q})$ is 0 in homogeneous networks such as ring topologies. As a network become more heterogeneous, the entropy $H(\mathrm{q})$ gets higher. Abilene inspired topology [2], which is shown in Fig. 2.2, is heterogeneous in its degree distribution, as

Table 2.1: Mutual information of example topologies

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

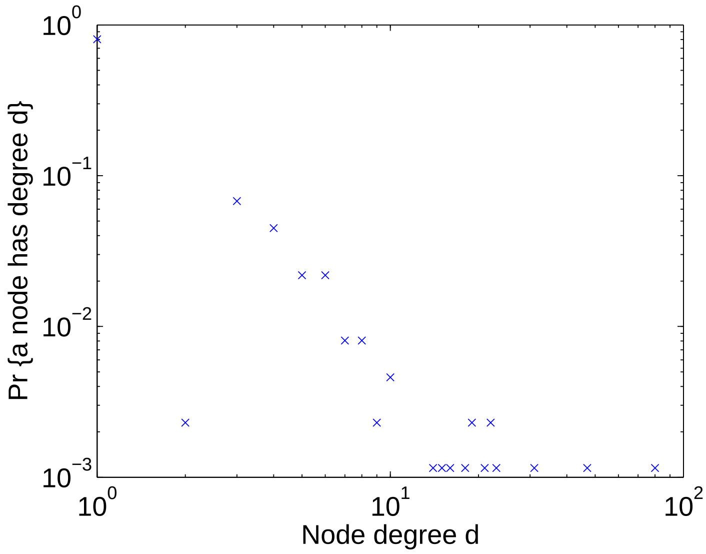


Figure 2.3: Degree distribution of Abilene-inspired topology

shown in Fig. 2.3. Therefore, the Abilene inspired topology has higher entropy as shown in Tab. 2.1.

The second term H_c(q|q') is the conditional entropy of the remaining degree distribution:

$$H_c(\text{q}|\text{q'}) = -\sum_{k=1}^{N}\sum_{k'=1}^{N} q(k')\pi(k|k')\log\pi(k|k'), \tag{2.4}$$

where $\pi(k|k')$ is conditional probability:

$$\pi(k|k') = \frac{q_c(k,k')}{q(k')}. \tag{2.5}$$

$\pi(k|k')$ gives the probability of observing a vertex with remaining degree k' provided that the vertex at the other end of the chosen edge has remaining degree k. Here, $q_c(k,k')$ is the joint probability, which gives the probability of existence of a link that connects a node with k edges and a node with k' edges, and $q_c(k,k')$ is normalized as:

$$\sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k,k') = 1. \tag{2.6}$$

The range of conditional entropy is $0 \leq H_c(\text{q}|\text{q'}) \leq H(\text{q})$. Ring topologies and star topologies have the lowest H_c, because, when the degree of one side of a link is known, the degree of the node on the other side is always determined. For Abilene inspired topology, because of its heterogeneous degree distribution, it is hard to determine the degree of the other side of a link than ring topologies or star topologies. Therefore, the conditional entropy $H_c(\text{q}|\text{q'})$ of the Abilene inspired topology is higher than ring topologies or star topologies. However, when we compare with a random topology that has almost the same $H(\text{q})$ as Abilene-inspired topology, we observe that the $H_c(\text{q}|\text{q'})$ of Abilene-inspired topology is lower than that of the random topology. This means the degree combination of a pair of nodes connected to a link is more biased in Abilene-inspired topology than in the random topology.

Using the distribution and probability explained above, the mutual information of the remaining degree distribution is expressed as follow:

$$I(\text{q}) = -\sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k,k') \log \frac{q_c(k,k')}{q(k)q(k')}. \tag{2.7}$$

The range of mutual information is $0 \leq I(\text{q}) \leq H(\text{q})$. The mutual information is higher in star topologies and Abilene-inspired topology since these topologies can get more information about the degree of a node by observing the other node connected by the direct link. $I(\text{q})$ of ring topologies and the random topology is low, but the reason for taking low mutual information is different. In ring topologies, because of the homogeneous degree distribution, no information can be obtained by observing degree of two nodes that are connected by a direct link. On the contrary, in the random topology, though the degree distribution is heterogeneous, because of the random connections, less information can be obtained. As we can see from these example topologies, $I(\text{q})$ is hard to discuss without considering about $H(\text{q})$. Hereafter in this chapter, we mainly use $H(\text{q})$ and $I(\text{q})$ to discuss the diversity of topologies.

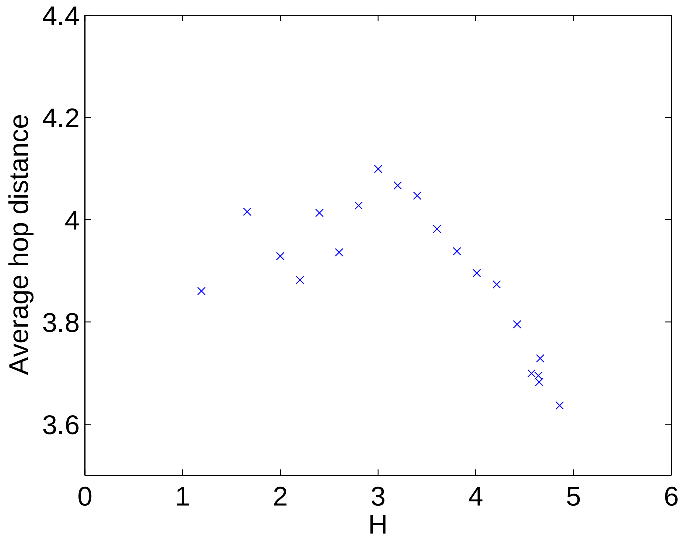


Figure 2.4: Average hop distance with topologies having different entropy

2.3 Mutual Information and the Characteristic of Topologies

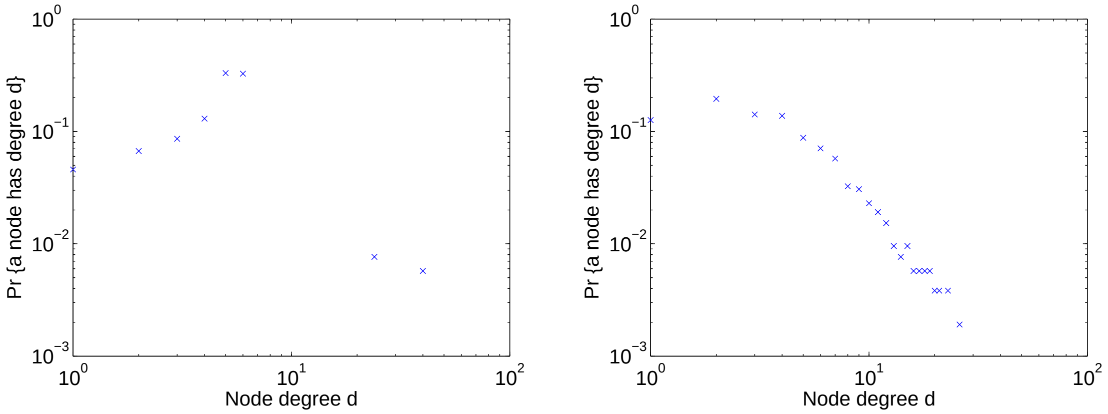
In this section, we explore the relationship between entropy and an average hop distance. Then, we show some illustrative examples of some topologies with different mutual information.

2.3.1 Entropy and Average Hop Distance

To show the relationship between entropy and the characteristic of topologies, we generate topologies having different entropy, and compare their average hop distance and degree distribution.

Topologies are generated by simulated annealing that looks for a candidate network that minimize the potential function $U(G)$. Here, the temperature is set to 0.01, and the cooling rate is set to 0.0001. The simulation searches 450000 steps. The initial topology is set to a topology obtained by BA model which has 523 nodes and 1304 links. Topologies are changed by random rewiring, and try to minimize the following potential function:

$$U(G) = \sqrt{(H - H(G))^2 + (H_c - H_c(G))^2}. \tag{2.8}$$



(a) Degree distribution ($H = H_c = 2.2$) (b) Degree distribution ($H = H_c = 4.2$)

Figure 2.5: Degree distribution with topologies having different entropy

Here, H and H_c are pre-specified value of entropy and conditional entropy, respectively. $H(G)$ and $H_c(G)$ are entropy and conditional entropy calculated from the topology G generated during in the search process. We generate topologies by setting H, H_c as $H = H_c$ from 1 to 5. Every time in the search process, $U(G)$ converge to approximately 0. Therefore, entropy and conditional entropy of the generated topologies are almost equal, and the mutual information of these topologies is approximately 0.

Figure 2.4 shows the average hop distance of topologies we generate. Degree distribution of a topology generated by setting $H = H_c = 2.2$ is shown in Fig. 2.5(a), $H = H_c = 4.2$ is shown in Fig. 2.5(b). Here, average hop distance is defined as the average of hop distance between every node pairs. We calculate the hop distance by assuming the minimum hop routing. From the result, we can see that, when H increases higher than 3, the average hop distance decreases. This is because, as H increases, the degree distribution become biased, and it gets close to power-law around $H = 4$.

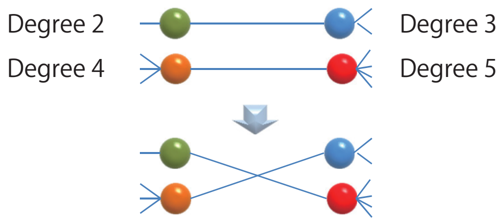


Figure 2.6: Rewiring method to leave the degree distribution unchanged

2.3.2 Mutual Information and Topological Diversity

Next, we show some illustrative examples of topologies with different mutual information. Because router-level topologies obey power-law, we compare topologies having high H.

Topologies are again generated by the simulated annealing. We set the same parameter and the same initial topology as we have use in the previous section. The different points are the way to rewire the topology and the potential function $U^I(G)$. For the first point, the topology is changed by a rewiring method [30] that leaves the degree distribution unchanged, i.e., by exchanging the nodes attached to any randomly selected two links (Fig. 2.6). For the second point, the potential function we used to minimize is $U^I(G)$ defined as,

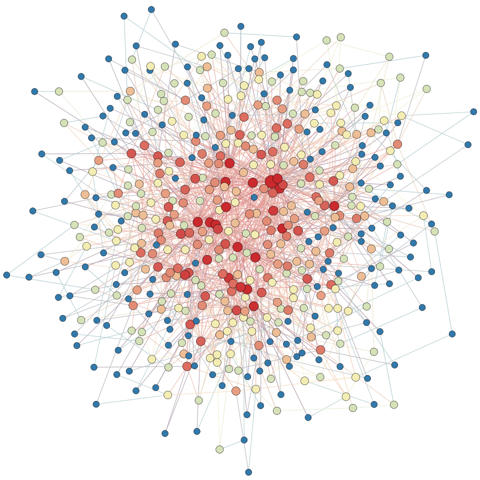
$$U^I(G) = |I - I(G)|, \tag{2.9}$$

where I is pre-specified mutual information, and $I(G)$ is mutual information calculated by the topology G generated in the search process. Note that looking for a pre-specified mutual information I is as the same as looking for a pre-specified conditional entropy H_c under the same entropy H. Because the entropy is same when the degree distribution unchanged, minimizing mutual entropy is identical to maximize conditional entropy.

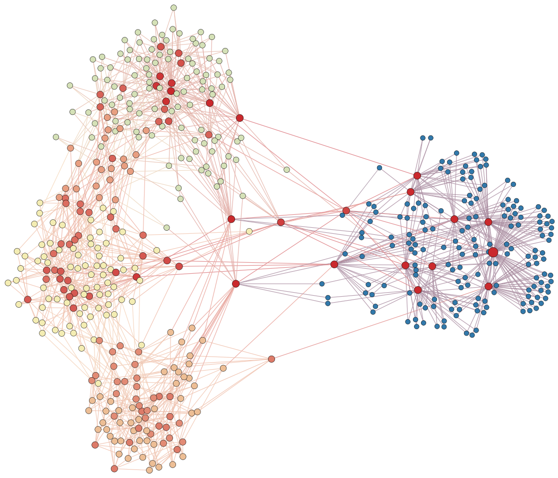
To show the relationship between mutual information and topological diversity, we use two

Table 2.2: Mutual information of topologies obtained by simulated annealing

Topology	Nodes	Links	$H(G)$	$H_c(G)$	$I(G)$
BA	523	1304	4.24	3.98	0.26
T_{Imin}	523	1304	4.24	4.13	0.12
T_{Imax}	523	1304	4.24	1.54	2.70



(a) T_{Imin} with minimum mutual information

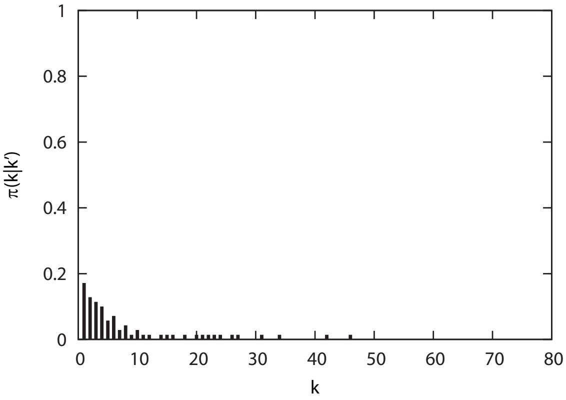


(b) T_{Imax} with maximum mutual information

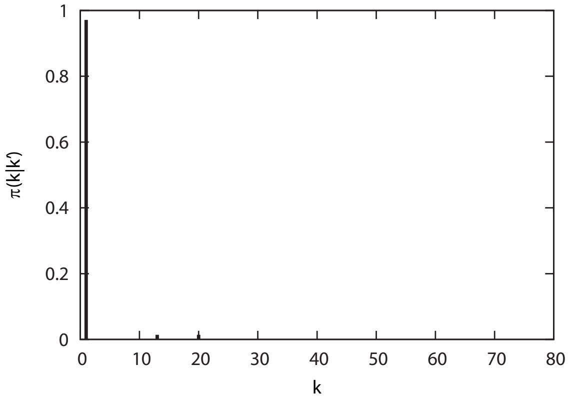
Figure 2.7: Illustrative examples with topologies having different mutual information

topologies: topology T_{Imin} with minimum mutual information and topology T_{Imax} with maximum mutual information. T_{Imin} is generated by setting $I = 0.0$ for simulated annealing, and the resulting mutual information is 0.12. The topology is shown in Fig. 2.7(a). T_{Imax} is generated by setting $I = 3.0$ for simulated annealing, and the resulting mutual information is 2.70. The topology is shown in Fig. 2.7(b). In both figures, colors represent the degree of nodes. Topological characteristics of the initial topology, T_{Imin} and T_{Imax} are summarized in Tab. 2.2.

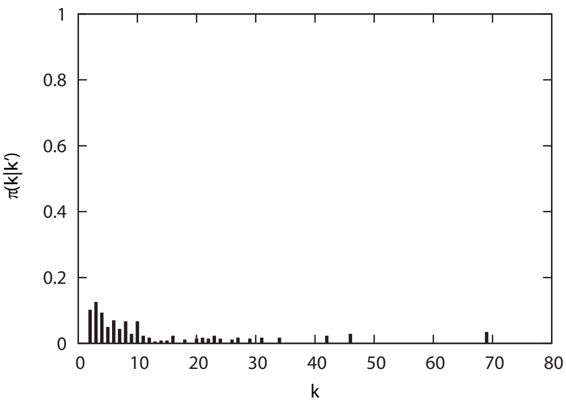
From Fig. 2.7(a) and Fig. 2.7(b), we can see that the topology with high mutual information is less diverse, and have more regularity than the one with low mutual information. From Fig. 2.8(a) to



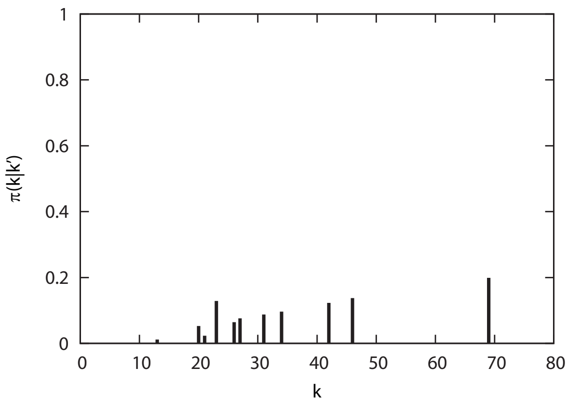
(a) $\pi(k|k')$ of nodes with the largest remaining degree in T_{Imin}



(b) $\pi(k|k')$ of nodes with the largest remaining degree in T_{Imax}



(c) $\pi(k|k')$ of nodes with the smallest remaining degree in T_{Imin}



(d) $\pi(k|k')$ of nodes with the smallest remaining degree in T_{Imax}

Figure 2.8: Degree distribution under different conditions

Fig. 2.8(d), we show $\pi(k|k')$ dependent on remaining degree k. $\pi(k|k')$ is defined as the probability of observing a vertex with k' leaving edges provided that the vertex at the other end of the chosen edge has k leaving edges. Fig. 2.8(a) and Fig. 2.8(c) show $\pi(k|k')$ of nodes with the largest remaining degree and nodes with the smallest remaining degree in T_{Imin}, respectively. Figure 2.8(b) and Fig. 2.8(d) show $\pi(k|k')$ of nodes with the largest remaining degree and nodes with the smallest

Table 2.3: Mutual information of router-level topologies

Topology	Nodes	Links	$H(G)$	$H_c(G)$	$I(G)$
Level3	623	5298	6.04	5.42	0.61
Verio	839	1885	4.65	4.32	0.33
ATT	523	1304	4.46	3.58	0.88
Sprint	467	1280	4.74	3.84	0.90
Telstra	329	615	4.24	3.11	1.13
BA	523	1304	4.24	3.98	0.26

remaining degree in T_{Imax}, respectively. We can see that $\pi(k|k')$ of T_{Imax} is more biased than that of T_{Imin}. This also represents that the topology with high mutual information is less diverse than the one with low mutual information.

2.4 Topological Diversity in Router-level Topologies

In this section, we calculate the mutual information for several router-level topologies. Based on the values of mutual information, we discuss the topological diversity of the router-level topologies. Then, we investigate the impact of mutual information on ISP networks from a link capacity perspective.

2.4.1 Mutual Information of Router-level Topologies

We calculate the mutual information for five router-level topologies: Level3, Verio, AT&T, Sprint and Telstra. The router-level topologies are measured by Rocketfuel tool [31]. For comparison purpose, we prepare a topology generated by BA model [22] which has the same number of nodes and links with AT&T. The results are summarized in Tab. 2.3 and Fig. 2.9.

From Tab. 2.3, we can see that all of the router-level topologies have high H, which means they have heterogeneous degree distribution. Level3 topology has highest H. This is because the Level3 topology includes many MPLS paths. These paths made the topology having high heterogeneity in degree distribution. Except Level3 topology, other router-level topologies shown in Tab. 2.3 have

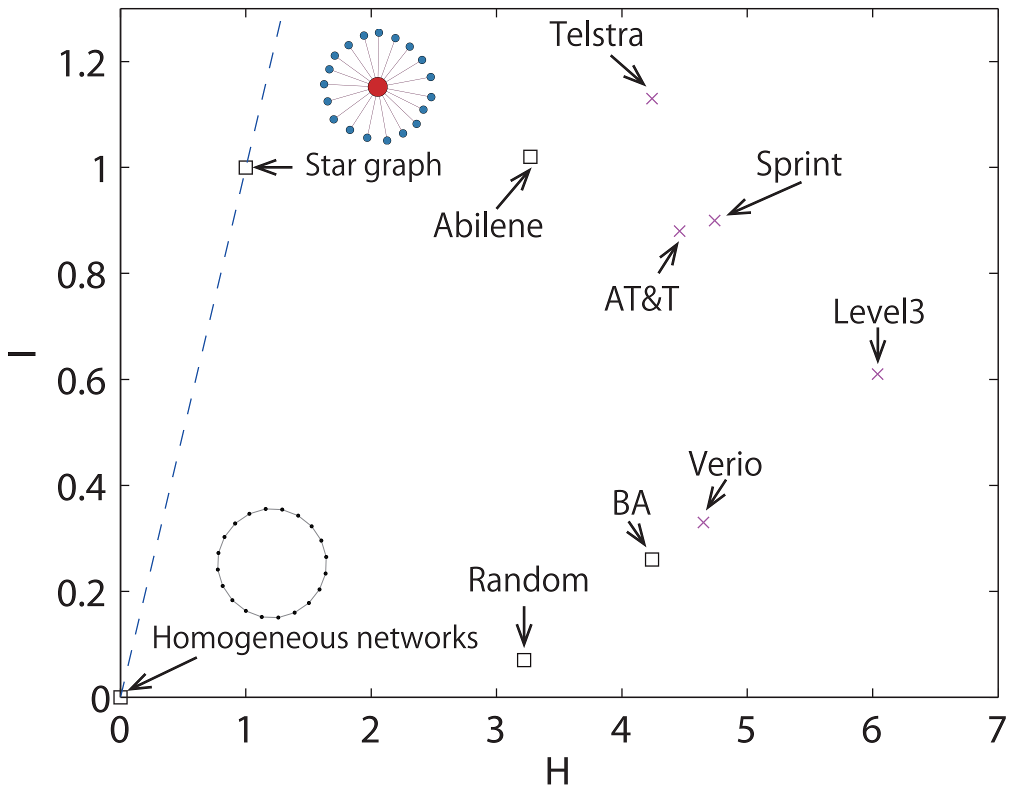


Figure 2.9: Entropy and mutual information of all the calculated topologies

almost the same H.

Comparing these topologies with BA topology, we can see that the mutual information of router-level topologies is higher than that of the model-based topology. This can be explained by a design principle of router-level topologies. Because router-level topologies are designed under the physical and technological constraints such as the number of switching ports and/or maximum switching capacity of routers, there are some restrictions and a kind of regulations on constructing the topologies, so they are less diverse. However, the mutual information of Verio topology is low. This can be explained by its growing history. Because Verio grows with a merge of small ISPs [32], the topology contains various kinds of design principles conducted in each ISP. Therefore, it is consided that

Table 2.4: Mutual information of topologies rewired from AT&T topology

Topology	AT&T$_{0.3}$	AT&T$_{0.4}$	AT&T$_{0.5}$	AT&T$_{0.6}$	AT&T$_{0.7}$	AT&T$_{0.8}$	AT&T
H	4.45583	4.45583	4.455834	4.45583	4.45583	4.45583	4.45583
H_c	4.17594	4.07697	3.97701	3.87589	3.77558	3.67903	3.57515
I	0.27989	0.37886	0.47882	0.57994	0.68025	0.77680	0.88068
Average hop	3.57439	3.56669	3.64005	3.74615	3.92027	4.18759	5.06338

Verio topology is more diverse than other router-level topologies.

2.4.2 Link Capacity Needed for Topologies with Different Mutual Information

In this section, we generate several topologies with different mutual information, but having the same entropy, and investigate the adaptability against environmental changes in ISP networks. It is preferable for ISP networks to have fewer changes in load on links even when node failures occur because the increase of load leads to high link usage, which possibly leads to increase delay, or high link capacity cost. Therefore, we evaluate changes in edge betweenness centrality after an environmental change. As for the environmental change, we consider the failure of nodes. We regard edge betweenness centrality as load on links, and evaluate the minimum link capacity needed to cover node failures. Note that the edge betweenness centrality does not reflect the actual load on links. Nevertheless, we use the edge betweenness centrality to characterize ISP networks because it gives a fundamental characteristic to identify the amount of traffic flow on topologies.

Topologies we use to compare are generated from the AT&T topology by rewiring links. The rewiring method leaves the degree distribution unchanged, which is as same as explained in Sec. 2.3.2. Because the topological diversity becomes lower as the rewiring proceeds, we calculate mutual information for every topology, and pick out topologies every time when the mutual information decreases 1.0. The entropy, conditional entropy and mutual information of selected topologies are summarized in Tab. 2.4. AT&T$_{0.3}$ is the topology with lowest mutual information obtained by our method with a long time of simulation. The average hop distance of each topology is also presented.

The failure we consider here is a single node failure. First, we evaluate the minimum link capacity needed to cover every pattern of single node failures. The link capacity $C(i)$ on link i is calculated as follow:

- Step 0: For all links i, set the initial edge betweenness centrality $E(i)$ as the link capacity $C(i)$:

$$C(i) = E(i). \tag{2.10}$$

- Step 1: When node j fails, calculate the new edge betweenness centrality $E_j(i)$ for every link. Renew the link capacity as (2.11) for every link:

$$\begin{cases} C(i) = E_j(i) & \text{if } (E_j(i) > C(i)) \\ C(i) = C(i) & \text{otherwise.} \end{cases} \tag{2.11}$$

- Step 2: Go back to Step1, select a new j until every node has been selected.

The total of edge betweenness centrality $\Sigma_i E(i)$ and the total of link capacity needed to cover every pattern of single node failure $\Sigma_i C(i)$ is shown in Fig. 2.10. Because $\Sigma_i E(i)$ is directly affected by average hop distance, the difference of $\Sigma_i E(i)$ in each topology is not important here. Instead, we focus on the amount of additional capacity needed to cover the node failures. We can see that $\Sigma_i C(i)$ tends to decrease as the mutual information of the topology decrease.

We next evaluate the changes of edge betweenness centrality on each link. The increment in edge betweenness centrality is also calculated for every pattern of single node failure. Let us denote the failured node as j. Then, we calculate $A_j(i)$ which represents the additional capacity of link i to cover the failure of node j as follows.

$$\begin{cases} A_j(i) = E_j(i) - E(i) & \text{if } (E_j(i) > E(i)) \\ A_i = 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

$A_j(i)$ for all the j sorted by link index i is shown in Fig. 2.11(a) and Fig. 2.11(b). Figure 2.11(a)

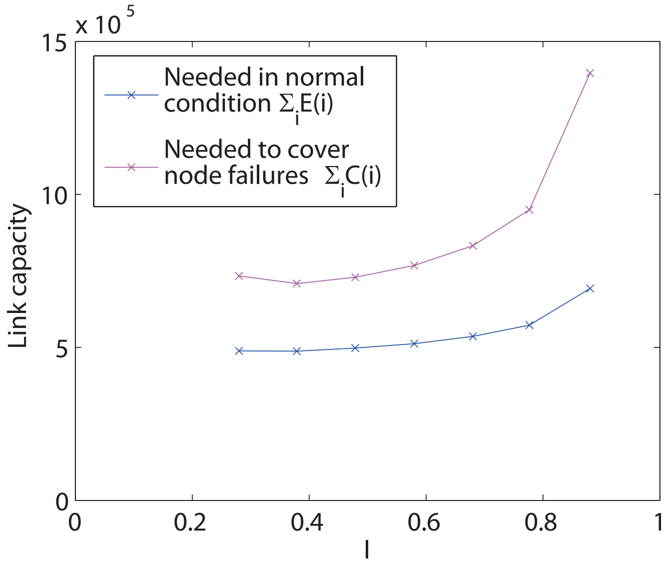


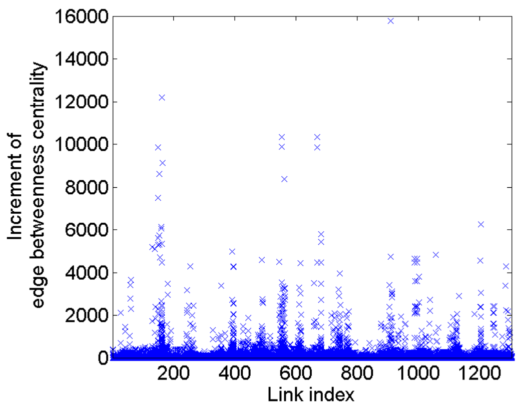
Figure 2.10: Link capacity with topologies having different mutual information

is calculated for original AT&T topology, and Fig. 2.11(b) is calculated for AT&T$_{0.3}$. We can see that, in the original AT&T topology, load at some of links are highly increased comparing with AT&T$_{0.3}$. This means many alternative paths tend to concentrate on some of the links when node failure occurs. In contrast, the increase of edge betweenness centrality on every link is small for AT&T$_{0.3}$. This is because that the alternative paths after the single node failure are balanced on many links, thanks to the nature of the diversity in AT&T$_{0.3}$.

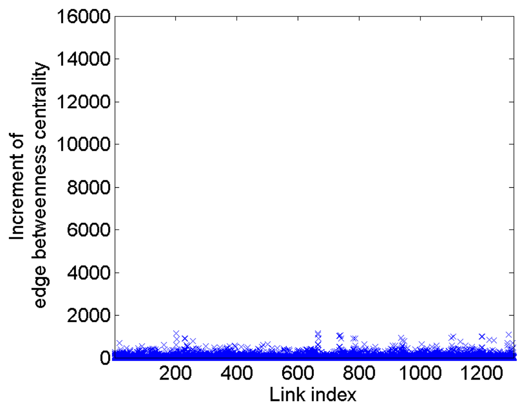
From these evaluation, we conclude that link capacity needed to deal with node failures decrease when the topology becomes diverse because alternative paths are less concentrated on links.

2.5 Summary

We investigated the network heterogeneity of router-level topologies by using mutual information. From calculating the mutual information of router-level topologies and a BA topology, we found that the mutual information of router-level topologies are higher than that of the BA topology. The



(a) Increament of edge betweenness centrality (AT&T)



(b) Increament of edge betweenness centrality (AT&T$_{0.3}$)

Figure 2.11: Increment of edge betweenness centrality with topologies having different mutual information

reason of high mutual information in the router-level topologies is explained by the fact that switching capacity of routers, there are some restrictions when constructing the network since router-level topologies are designed under the physical and technological constraints such as the number of switching ports and/or maximum Such constraints lead to high correlation in degrees of two connected nodes. Then, we showed some illustrative examples of topologies with different mutual information. We generated a mutual information maximized topology and minimized one whose number of nodes and links is same as that of a router-level topology. From the illustrative examples, we showed that topology with high mutual information is less diverse, and have more regularity than the one with low mutual information. From comparing the topology with different mutual information generated from AT&T, we found that link capacity needed to deal with node failures decrease when the topology becomes diverse because alternative paths less converge in the topology with high topological diversity.

Chapter 3

Topology Design Approach Using Mutual Information for Evolvable Networks

3.1 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 attracts many users, and there is no doubt that the environment surrounding the Internet will change even more in the future.

In spite of 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 chapter, we discuss whether this could be achieved by constructing a network that can easily adapt to deal with new environments. In ISP 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 [12], an ISP network 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 chapter, 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 the random network has topological diversity. However, it is not efficient to design an ISP network as a random network. A well-known disadvantage of the random network is that the average hop distance is larger than that in a scale free network. Because of this, the 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.

The rest of this chapter is organized as follows. Section 3.2 explains our proposed design approach. We explain the measure we use for design in Sec. 3.2.1. We then present our approach in

Sec. 3.2.2 and discuss characteristics of reliability against node failures in Sec. 3.2.3. In Sec. 3.3, 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 Sec. 3.3.3. Section 3.3.4 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 chapter in Sec. 3.4.

3.2 An Evolvable Network Design Approach

Evolution and evolvability have been studied for a long time in biology [8]. 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 [12]. 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 et al. [11] use mutual information to analyze topological characteristics of complex networks. The mutual information used in [11] is the difference between the heterogeneity in degree distribution and the degree-degree correlation, which is also known as assortativeness [33], appearing in the network's structure. It was shown in [13] that router-level topologies characterized by degree-degree correlation [2] lead to high mutual information. Following [13], we will minimize the information measure proposed in [11] to strengthen topological diversity. In Sec. 3.2.1, 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 Sec. 3.2.2.

3.2.1 Metric Used for Design

Solé et al. [11] use mutual information on the remaining degree distribution to analyze characteristics of complex networks. Following [11], 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 $\sum_k P(k)$ =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

$$q(z) = \frac{(z+1)P_{z+1}}{\Sigma_z z P_z}. \tag{3.1}$$

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

$$I(\boldsymbol{q}) = H(\boldsymbol{q}) - H_c(\boldsymbol{q}|\boldsymbol{q}'), \tag{3.2}$$

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 $\boldsymbol{q}'$ (=

$\{q(z')|1 \leq z' \leq N\}$ where z and z' are the remaining degrees of linked nodes). $H(\boldsymbol{q})$ is defined as

$$H(\boldsymbol{q}) = -\sum_{z=1}^{N} q(z) \log(q(z)), \tag{3.3}$$

and $H(\boldsymbol{q})$ always satisfies the inequality $H(\boldsymbol{q}) \geq 0$. 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 [2] is heterogeneous in the degree distribution, so it has higher entropy, as shown in Tab. 2.1. For reference, we also show $H(\boldsymbol{q})$ for a randomly generated topology. The topology was generated by Random 2 model [34] with 523 nodes and 1304 links, as in the AT&T topology.

The second term $H_c(\boldsymbol{q}|\boldsymbol{q}')$ of Eq. (3.3) is the conditional entropy of the remaining degree distribution:

$$H_c(\boldsymbol{q}|\boldsymbol{q}') = -\sum_{z=1}^{N}\sum_{z'=1}^{N} q(z')\pi(z|z') \log \pi(z|z'), \tag{3.4}$$

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

$$\pi(z|z') = \frac{q_c(z,z')}{q(z')}, \tag{3.5}$$

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,

$$\sum_{z=1}^{N}\sum_{z'=1}^{N} q_c(z,z') = 1. \tag{3.6}$$

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

$0 \leq H_c(\boldsymbol{q}|\boldsymbol{q}') \leq H(\boldsymbol{q})$. $H_c(\boldsymbol{q}|\boldsymbol{q}')$ 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 [2].

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

$$I(\boldsymbol{q}) = -\sum_{z=1}^{N}\sum_{z'=1}^{N} q_c(z,z') \log \frac{q_c(z,z')}{q(z)q(z')}. \tag{3.7}$$

$I(\boldsymbol{q})$ is high for the star and Abilene-inspired topologies (see the right-most column of Tab. 2.1), 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.

3.2.2 Design Approach

In this subsection, we describe our proposed design approach, which we call EVN (EVolvable Network). Fundamentally, EVN design approach reduces the mutual information on remaining degree, $I(\boldsymbol{q})$, so that the designed network has topological diversity. Note that 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 chapter, 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 chapter, 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 chapter 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 $H_0(\boldsymbol{q})$.

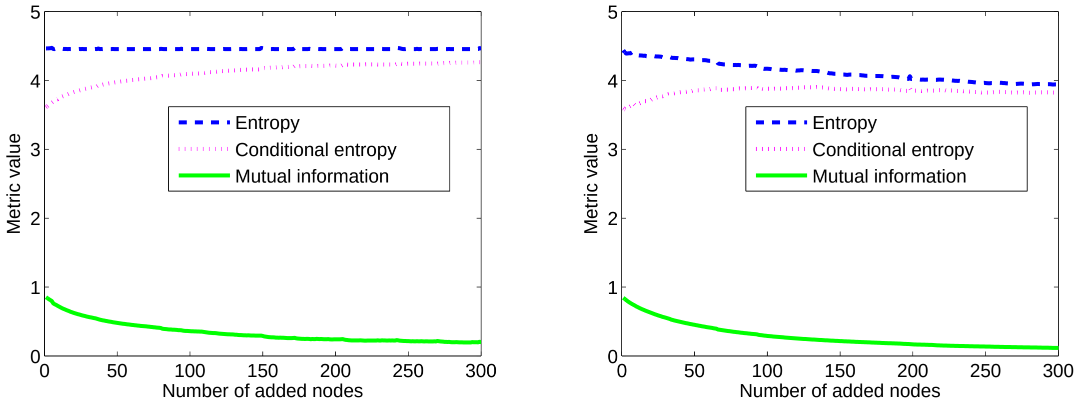
(b) Connect the node w and the m links, and obtain $G_k(V_k, E_k)$.

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 [35] 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. (3.1)), 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.

3.2.3 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 chapter, we only present the case of node failure, but we see similar results in the case of link failure.

Figure 3.1 shows the values of entropy, conditional entropy and mutual information of two networks: one is obtained by the EVN design approach (Fig. 3.1(a)) and the other is obtained by the EVN design approach without the entropy–restriction (Fig. 3.1(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



(a) EVN design approach

(b) EVN design approach without entropy–restriction

Figure 3.1: Entropy, conditional entropy, mutual information of topologies obtained by EVN design approach

result obtained by the Rocketfuel tool [31]; 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 not to 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, 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. 3.1(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. Figure 3.1(b) shows the case without an entropy–restriction. In this network, the entropy of remaining degree decreases as the network grows.

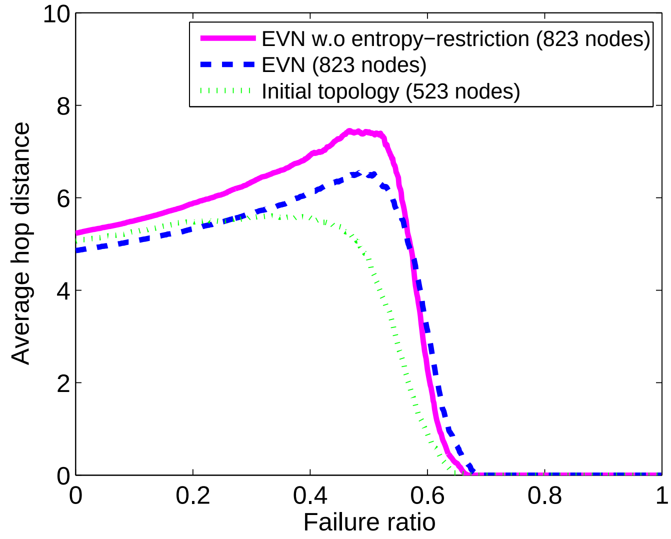


Figure 3.2: Average hop distance against random node failures

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. Figure 3.2 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.

3.3 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 [36], 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 ISP 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.

Network Design Method based on the FKP Model

The FKP model proposed by Fabrikant et al. [36] 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:

$$\alpha \cdot d(n_{new}, n_i) + h(n_i, n_0), \tag{3.8}$$

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 chapter, 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 [36] 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 modify 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. (3.8) for each existing node n_i.

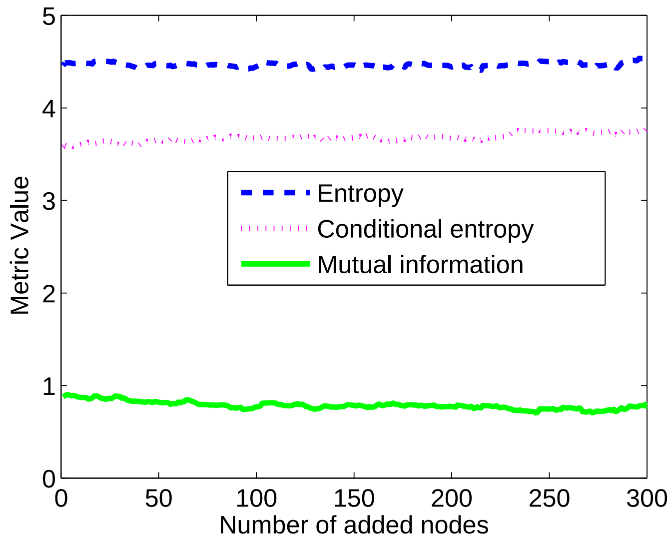


Figure 3.3: Entropy, conditional entropy, mutual information of topologies obtained by the FKP-based method

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 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 Subsections 3.3.1 and 3.3.2, the parameter α is 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.

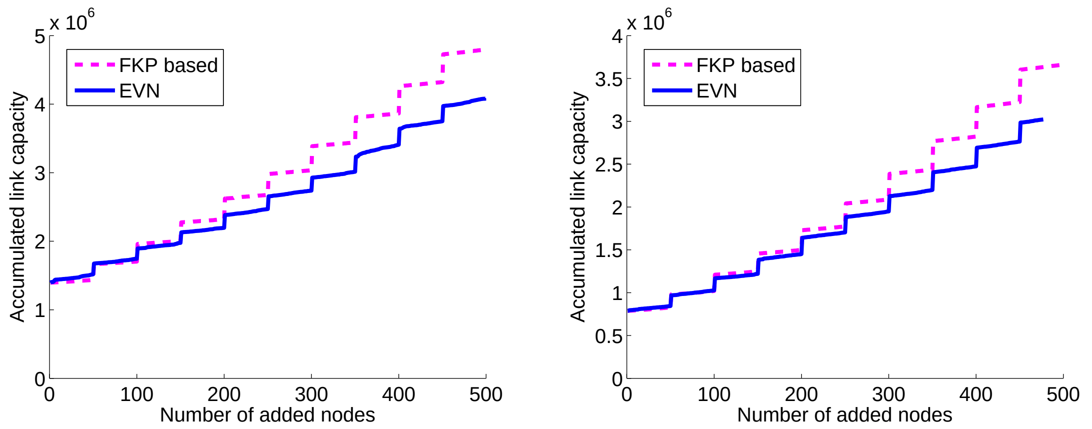
Figure 3.3 shows the entropy, conditional entropy and mutual information during network growth by the modified FKP-based design method. We use the AT&T topology as the initial topology, and set the number of added nodes to be n = 300 (i.e., the final topology is obtained after 300 steps) and the number of links for each step to be m = 2. The locations of nodes at the city-level are obtained from [31], 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. (3.8)). 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.

3.3.1 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

$$C_k^{EVN} = \sum_{e \in E} C_k^{EVN}(e), \tag{3.9}$$



(a) AT&T topology

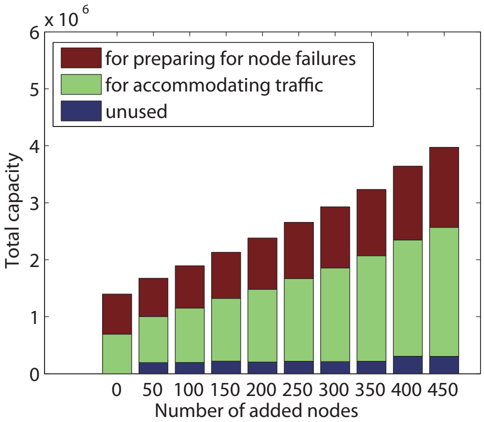
(b) Sprint topology

Figure 3.4: Accumulated link capacity with networks grown with different methods

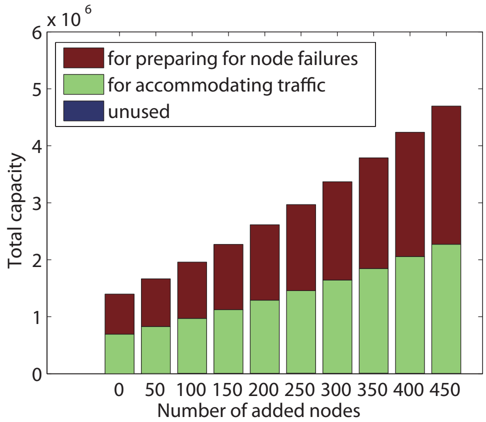
where $C_k^{EVN}(e)$ represents the capacity of link e. In the evaluation, the capacity of each link is chosen such that the link can accommodate the traffic arising from every pattern of single node failure in the topology G_k^{EVN}. The method of shortest path with equal hop path splitting [37] is 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}, is obtained in the same way.

Figure 3.4 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. 3.4(a) and to the Sprint topology (467 nodes and 1280 links) for Fig. 3.4(b). The Sprint topology is also a



(a) EVN design approach



(b) FKP-based design method

Figure 3.5: Capacity for preparing for node failures, capacity for accommodating traffic, unused capacity

measurement result obtained by the Rocketfuel tool [31]. Both figures indicate that our EVN design approach requires less link capacity than the FKP-based design method.

Table 3.1: Average of additional capacity needed to cover each single node failure

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

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. Figure 3.5(a) shows the results for the EVN design approach, and Fig. 3.5(b) shows the results for the FKP-based design method. Comparing Figs. 3.5(a) and 3.5(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 3.1 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 Tab. 3.1 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. 3.5 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 be used at the next (or a later) stage of evolution thanks to the unspecialized nature of the EVN design approach.

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 [38]. 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 Sec. 3.3.1. 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.

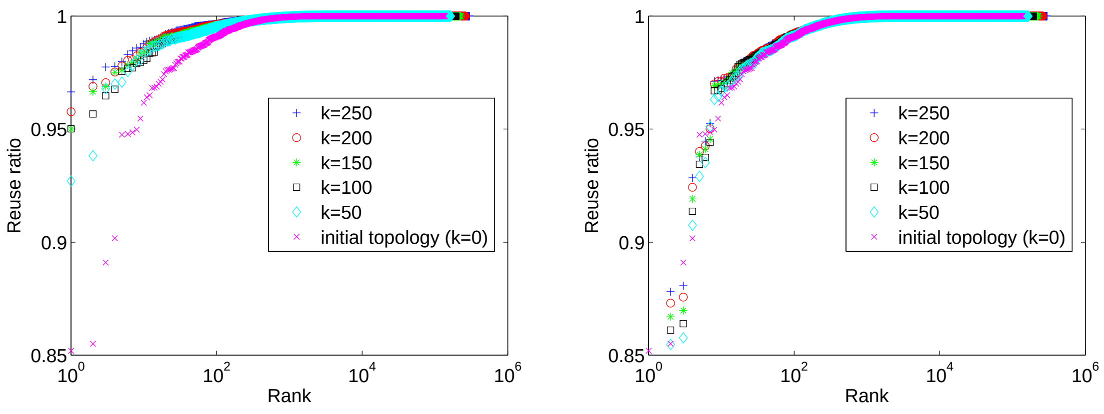
For evaluation, we introduce a *reuse ratio*, r_k, of a topology after k node additions defined by

$$r_k = \frac{F_k^{reused}}{F_k^{new}}, \tag{3.10}$$

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 k-th 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}.

Figure 3.6(a) shows $r_k^{EVN}(n_1, n_2)$ for all cases of two-node (n_1, n_2) failures and Fig. 3.6(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



(a) Reuse ratio for G_k^{EVN}

(b) Reuse ratio for G_k^{FKP}

Figure 3.6: Reuse ratio under two nodes failures

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.

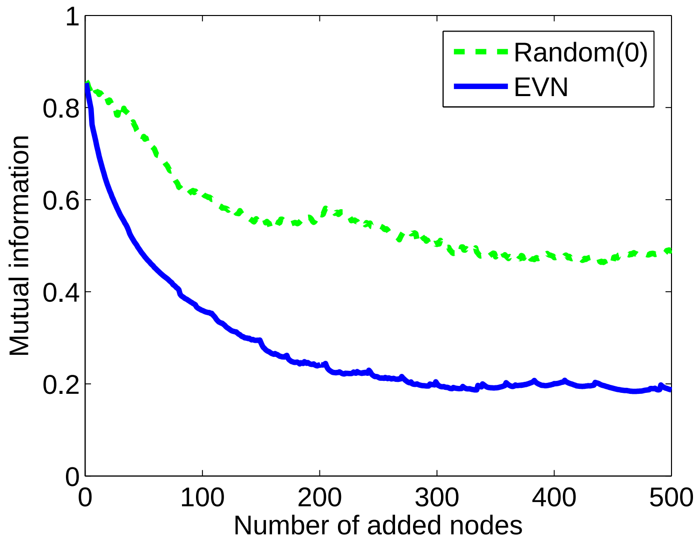


Figure 3.7: Entropy, conditional entropy, mutual information of a topology obtained by the random attachment method

3.3.3 Difference from Random Attachment Method

We showed that minimizing mutual information by the EVN design approach can lead to evolvable networks in Sec. 3.3.2 and Sec. 3.3.1. However, there are also other methods to lower mutual information. Since Solé et al. [11] show 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 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 add 2 links per node addition in the simulation. Hereafter we

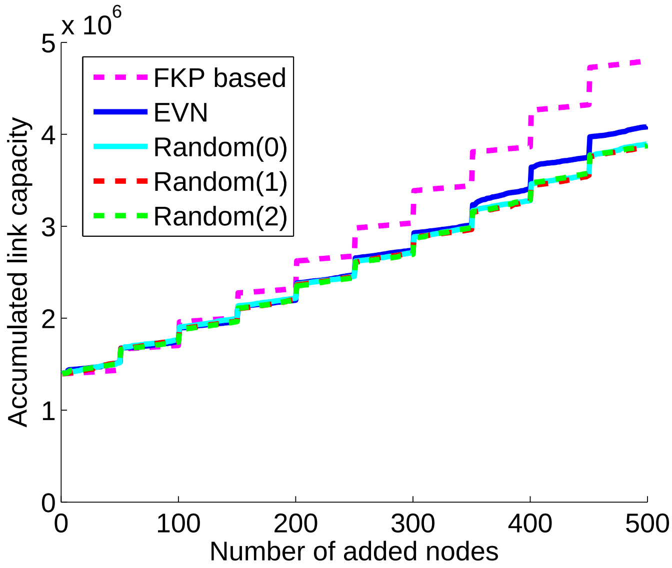
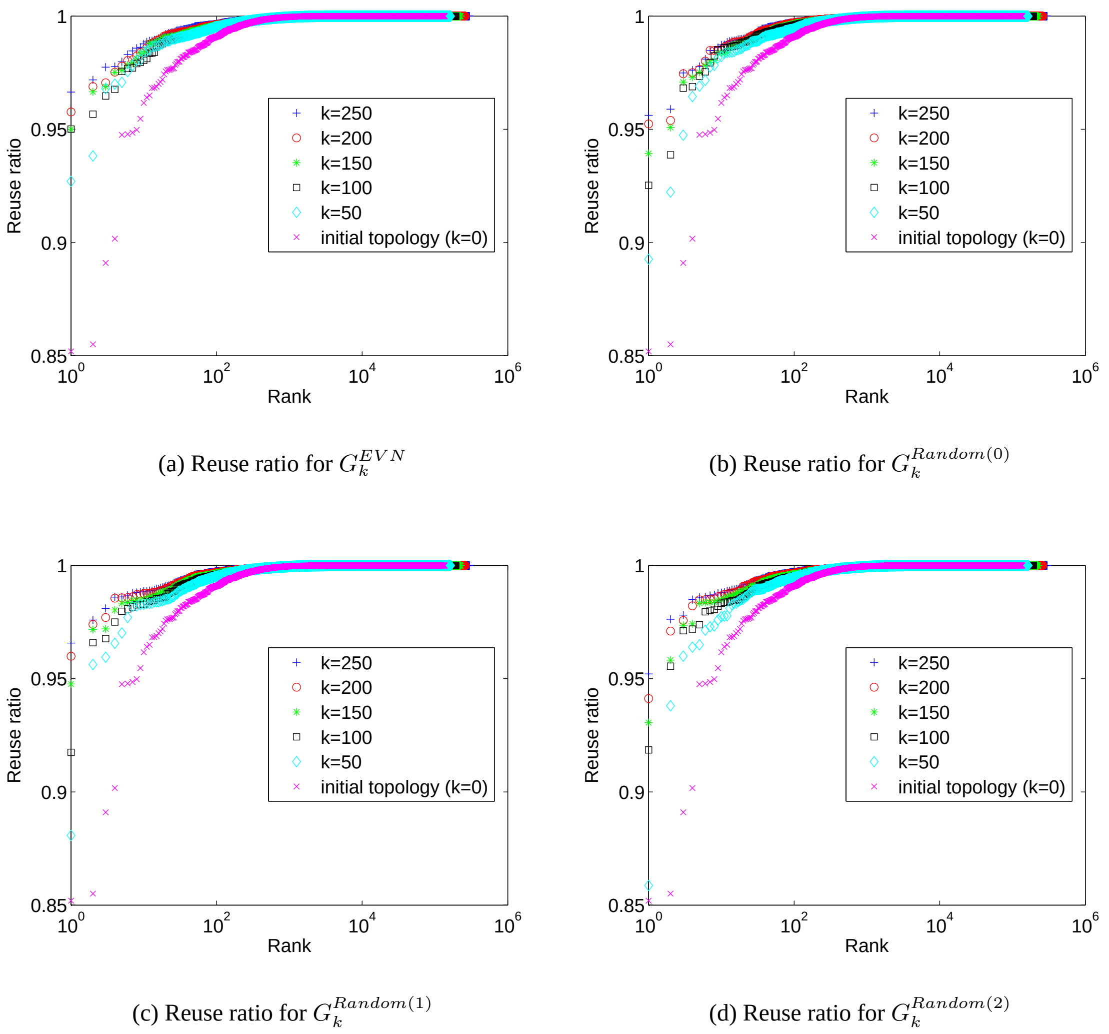


Figure 3.8: Accumulated link capacity of a network grown by the random attachment method

denote by G_k^{Random} the topology obtained by the random attachment method after the addition of k nodes. Figure 3.7 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 Tab. 2.1. 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



(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 $G_k^{Random(2)}$

Figure 3.9: Reuse ratio under two nodes failures of topologies grown by the random attachment method

in Sec. 3.3.1 and Sec. 3.3.2, 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. 3.8. 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. Figure 3.9 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.

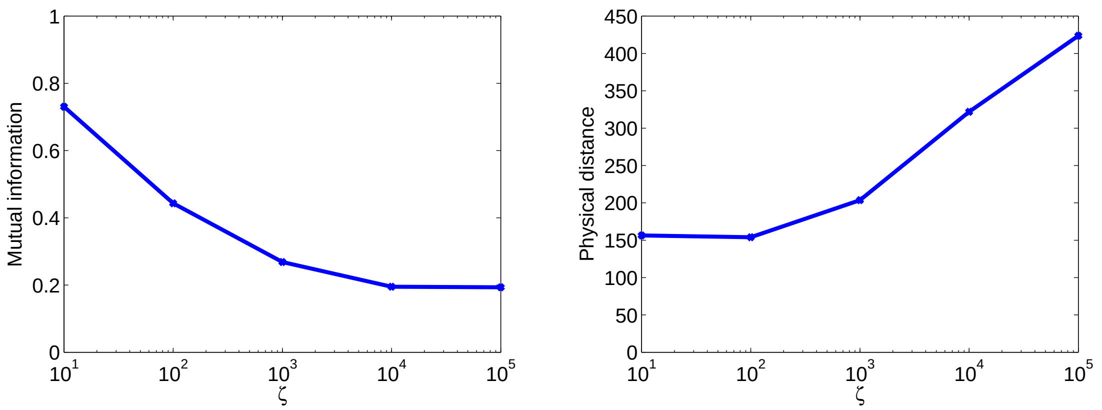
3.3.4 Trade-off between Topological Diversity and Physical Distance

In Sec. 3.3.2 and Sec. 3.3.1, 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 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:

$$\zeta \cdot I(q) + \sum_{i \in M} f(i), \tag{3.11}$$

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



(a) Mutual information

(b) Physical distance

Figure 3.10: Relationships between ζ and mutual information and physical distance

design method:

$$f(i) = \alpha \cdot d(n_{new}, n_i) + h(n_i, n_0) \tag{3.12}$$

See Eq. (3.8) in Sec.3.3 for details. M in Eq. (3.11) 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 $\sum_{i \in M} f(i)$ 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

$$E(\zeta) = (2/\pi) \cdot \arctan \zeta \cdot H_0(\boldsymbol{q}). \tag{3.13}$$

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

Table 3.2: Mutual information and physical distance of networks grown with different method

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

Accumulated capacity during network growth and reuse of facilities for unexpected environmental changes is evaluated for ζ equal to 0, 10, 100, 1000, $10\ 000$, $100\ 000$. For the details of the evaluation methods, please see Sec. 3.3.1 and Sec. 3.3.2, respectively. To compare with the EVN design approach and the FKP-based design method, in the simulation we add 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 change 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. 3.10(a) and in Fig. 3.10(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 Sec.3.3. 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 Tab. 3.2. 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. (3.12). 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}.

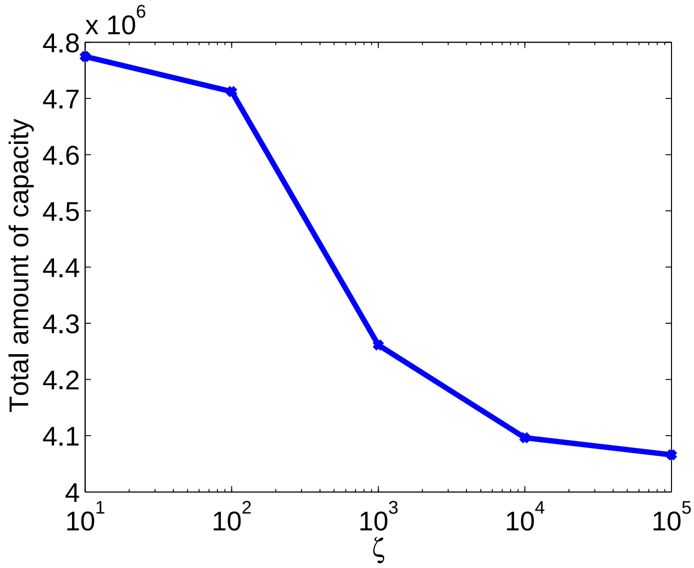


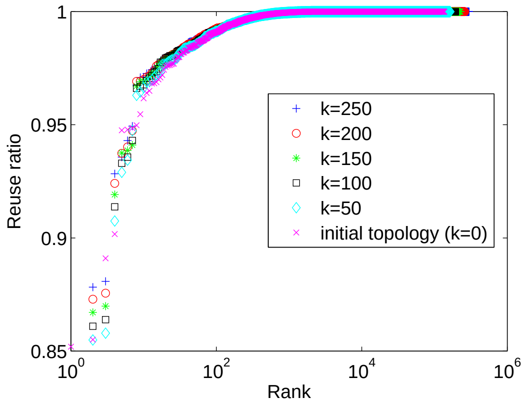
Figure 3.11: Relationship between ζ and total capacity

Figure 3.11 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 after the addition of 499 nodes. Moreover, much more capacity can be expected to be saved when the network grows even larger.

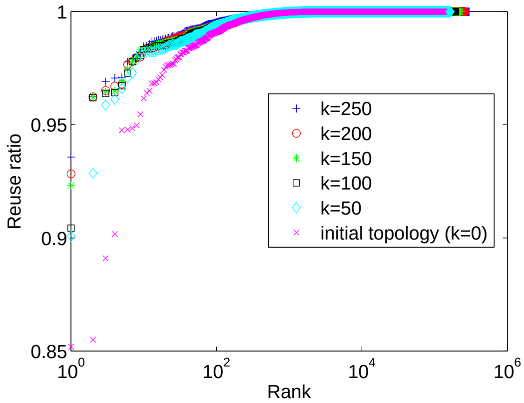
Figure 3.12 shows how the reuse of facilities for unexpected environmental changes depends on ζ. Results for ζ equal to 10, 1000 and $100\ 000$ are shown in Fig. 3.12(a), Fig. 3.12(b), Fig. 3.12(c), respectively. We can see that, though the worst reuse ratio of $G_{250}^{\zeta=1000}$ is worse than that of $G_{250}^{\zeta=100\ 000}$, it is better than that of $G_{250}^{\zeta=10}$.

3.4 Summary

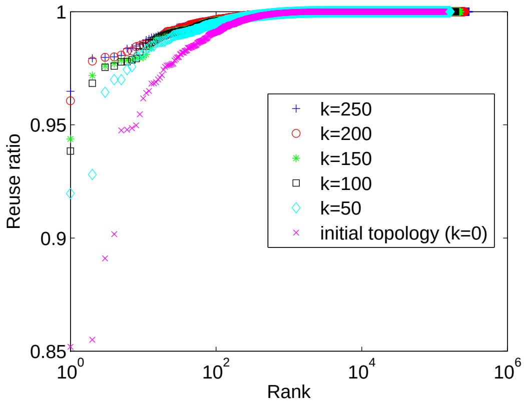
We proposed a network design approach, called EVN (EVolvable Network), to enhance topological diversity using the mutual information. In the approach, a new node is connected to existing nodes



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



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



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

Figure 3.12: Reuse ratio under two nodes failures of topologies grown by considering physical distance

to minimize the mutual information of the topology. We compared the link capacity needed for accommodating traffic in an ordinary situation and single node failure occurring situations with an ad-hoc design method. For the ad-hoc design method, we used FKP-based design method since FKP model includes primitive principles for designing an ISP network. We found that link capacity needed for a network grown with our design approach is less than that grown with the FKP-based design method. We also found that a network designed by our design approach can reuse more link capacity than that by the FKP-based design method. Next, we evaluated for additional capacity needed for accommodating traffic to deal with unpredicted environmental changes. As an unpredicted environment, we choose two node failures since this is not considered in the designing period. We found that, when severe two node failures occur, the network designed by our design approach can use more already placed capacity compared to that by the FKP-based design method.

A disadvantage of our approach is that a network grown with our network needs more capacity on average to cover one pattern of single node failure than that grown by the FKP-based design method. However, it needs less capacity to cover various environments. This is because the topology generated by our design approach is not specialized to a single environment. Another disadvantage is that, since our design approach does not consider physical length of links, links in a topology designed by our design approach have large physical length. However, from the discussion of trade-off between topological diversity and physical distance, 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 after the addition of 499 nodes. Moreover, much more capacity can be expected to be saved when the network grows even larger.

Chapter 4

Capacity Planning Using Mutual Information for Evolvable Networks

4.1 Introduction

In the previous chapter, we used the mutual information to quantify the diversity of topological structure, and developed a new design approach that enhances topological diversity using the mutual information. Because we focused on the topological diversity at the previous chapter, link capacity was assigned by a simple approach that assigns link capacity which is enough to accommodate traffic under any patterns of single node failure. In this chapter, we extend the concept of diversity to the link capacity, and investigate the effectiveness of capacity planning with high diversity for ISP networks.

Capacity planning for the ISP network is to obtain a capacity assignment that satisfies required quality such as accommodating traffic under traffic changes and/or equipment failures [39–42]. Since network traffic always changes or fluctuates, network operators usually put some extra capacity on links. The simplest capacity planning uses passive measurements of link utilization statistics and applies rules of thumb, such as upgrading links when they reach 50 percent average utilization, or some other target utilization. As a result, the links are always over-provisioned relative to the offered average load, but network operators expect that the capacity will be sufficiently over-provisioned for the peak load, and thus expect that congestion will not occur. A more sophisticated capacity planning approach has been investigated in many literatures. For example, Refs [41, 42] consider a spare capacity allocation under single link failures. Given information of topology and traffic demand matrix, the authors try to minimize capacity with consideration of single link failures.

Capacity designed by these approach can always satisfy required quality under the situation they considered. However, since, changes in traffic demand is becoming difficult to predict, there is a possibility that the quality largely deteriorate when an unexpected change occurs. Therefore, under the situation that changes in traffic demand becoming difficult to predict, a methodology to plan capacity with less deteriorated quality even when unexpected change occurs should be considered.

To approach to this problem, we learn from a mechanism of biological system which is robustness to environmental changes. It is known that living things survives from environmental changes by repeatedly adapting and/or evolving themselves for a long while, many researches in biological field investigate the mystery of mechanisms which let the biological system be robust. In a recent study, Whitacre et al. [9] point out that the biological system has a characteristic, called "degeneracy" where a component of the system is not intended to serve for a sole specific function, but serve for various kinds of functions. A diverse of components forms a function, and thereby makes a system be survivable against component failures. In an engineering system, components would be used for a specific function. What the optimal in engineering context means that roles of system resource are fully specialized by their governed objective function and constraints. Otherwise it will not be an optimal. In contrast, the components of biological system behave various kinds of different roles, which enable biological system survivable.

Inspired from the behavior of biological system as mentioned above, in this chapter, we develop a capacity planning methodology which can follow various kinds of environmental changes with less additional capacity. Our basic idea is to assign additional capacity on each link such that the capacity can work for more sever environmental changes as well as single link failures. More specifically, we increase the diversity of available capacity on restoration path under single link failures. Although the diversity of available capacity does not contribute for the case of single link failure, we expect the additional capacity would be needed to deal with severe environmental changes and expect that less additional capacity is required as like in the biological system. To quantify the diversity of available capacity under the single link failures, we define the mutual information between load on a link and available capacity along a restoration path for the failure of the link. The detail is explained in Sec. 4.2. In Sec. 4.3, we investigate network with its capacity on restoration paths assigned randomly, and show that the amount of additional capacity needed to cover environmental changes can be saved when capacity is assigned diversely. In Sec. 4.4, we investigate networks with different mutual information, and show that capacity with low mutual information can save more. Conclusion and future work are summarized in Sec. 4.5.

4.2 Mutual Information of Link Failure Preparation Ratio and Link Load

We use mutual information as a metric. Generally, mutual information yields the amount of information that can obtain about one random variable $\boldsymbol{X}$ by observing another variable $\boldsymbol{Y}$. It is used to quantify correlation between the two variables. In this chapter, we interpret $\boldsymbol{Y}$ as roles, and $\boldsymbol{X}$ as components. In detail, we use load on a link for $\boldsymbol{Y}$, and $\boldsymbol{X}$ is capacity placed on a restoration path when the link fails. In Sec. 4.2.1, we explain the definitions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ in detail. In Sec. 4.2.2, we show that process of discretization for $\boldsymbol{X}$ and $\boldsymbol{Y}$. Note that we do not use $\boldsymbol{X}$ and $\boldsymbol{Y}$ explained in Sec. 4.2.1 directly for calculating mutual information. This is because $\boldsymbol{X}$ and $\boldsymbol{Y}$ are continuous values. Instead of it, discretized value of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are used for calculation. Therefore, for simplicity, we discretize them. In Sec. 4.2.2, we show the process of discretization.

4.2.1 Definition of Variables Used for Mutual Information

As for the mutual information measurement, we use load on a link for $\boldsymbol{Y}$, and capacity placed on a restoration path when the link fails for $\boldsymbol{X}$. We could consider more simple definition for $\boldsymbol{X}$, such as considering $\boldsymbol{X}$ as capacity on a link. However, such definition is not sufficient to place the capacity diversely because the definition is used for accommodating traffic in an ordinary situation. Therefore, we use a spare capacity, which is assigned for accommodating traffic under equipment failure situation, for $\boldsymbol{X}$.

$\boldsymbol{X}$ is defined as

$$\boldsymbol{X} = \{\boldsymbol{X}^{u,v} | u, v \in N\}, \tag{4.1}$$

where $\boldsymbol{X}^{u,v}$ means spare capacity of end node pair (u, v), N describes the node set of the network.

$\boldsymbol{X}^{u,v}$ is defined as

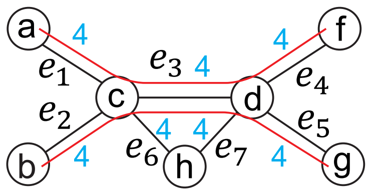
$$\boldsymbol{X}^{u,v} = \{z^{u,v}(e) | e \in \boldsymbol{R}^{u,v}\}, \tag{4.2}$$

where $z^{u,v}(e)$ is capacity on a restoration path when link e fails. $\boldsymbol{R}^{u,v}$ indicates a set of links where (u, v) uses under no link failure situation. Note that we use shortest path with equal hop path splitting [37] for calculating the links that (u, v) uses.

$z^{u,v}(e)$ is defined as

$$\frac{w^{u,v}(e)}{t(u,v)}, \tag{4.3}$$

where $w^{u,v}(e)$ indicates the capacity which is available for (u, v) after e fails. $t(u, v)$ indicates the traffic demand between (u, v). Hereafter, we call this value preparation ratio. We have two exceptions for defining $z^{u,v}(e)$. When (u, v) is unreachable after link e fails, we define $z^{u,v}(e)$ as 0. When $t(u, v)$ is 0 and if $w^{u,v}(e)$ is 0, we define $z^{u,v}(e)$ as 1. This is because the spare capacity



traffic demand (a, f) : 2 , (b,g) : 1

$$\boldsymbol{X}^{a,f} = \left\{\frac{w^{a,f}(e_1)}{t^{a,f}}, \frac{w^{a,f}(e_3)}{t^{a,f}}, \frac{w^{a,f}(e_4)}{t^{a,f}}\right\} = \left\{0, \frac{4-1}{2}, 0\right\}$$

$$\boldsymbol{X}^{b,g} = \left\{\frac{w^{b,g}(e_2)}{t^{b,g}}, \frac{w^{b,g}(e_3)}{t^{b,g}}, \frac{w^{b,g}(e_5)}{t^{b,g}}\right\} = \{0, \frac{4-2}{1}, 0\}$$

$\boldsymbol{X}^{a,b} = \{0,0\}$, $\boldsymbol{X}^{a,c} = \{0\}$, $\boldsymbol{X}^{a,d} = \{0, \infty\}$, $\boldsymbol{X}^{a,g} = \{0, \infty, 0\}$, $\boldsymbol{X}^{a,h} = \{0, \infty\}$, $\boldsymbol{X}^{b,c} = \{0\}$, $\boldsymbol{X}^{b,d} = \{0, \infty\}$, $\boldsymbol{X}^{b,f} = \{0, \infty, 0\}$, $\boldsymbol{X}^{b,h} = \{0,0\}$,$\boldsymbol{X}^{c,d} = \{\infty\}$, $\boldsymbol{X}^{\boldsymbol{c,f}} = \{\infty, 0\}$, $\boldsymbol{X}^{c,g} = \{\infty, 0\}$, $\boldsymbol{X}^{c,h} = \{\infty\}$, $\boldsymbol{X}^{d,f} = \{0\}$, $\boldsymbol{X}^{d,g} = \{0\}$, $\boldsymbol{X}^{d,h} = \{\infty\}$, $\boldsymbol{X}^{f,g} = \{0,0\}$,$\boldsymbol{X}^{f,h} = \{0, \infty\}$, $\boldsymbol{X}^{g,h} = \{0, \infty\}$

$$\boldsymbol{Y}^{a,f} = \{l(e_1), l(e_3), l(e_4)\} = \{2, 3, 2\}$$

$$\boldsymbol{Y}^{\boldsymbol{b,g}} = \{l(e_2), l(e_3), l(e_5)\} = \{1, 3, 1\}$$

$\boldsymbol{Y}^{a,b} = \{2,1\}$, $\boldsymbol{Y}^{a,c} = \{2\}$, $\boldsymbol{Y}^{a,d} = \{2,3\}$,$\boldsymbol{Y}^{a,g} = \{2,3,1\}$, $\boldsymbol{Y}^{a,h} = \{2,0\}$, $\boldsymbol{Y}^{b,c} = \{1\}$,$\boldsymbol{Y}^{b,d} = \{1,3\}$, $\boldsymbol{Y}^{b,f} = \{1,3,2\}$, $\boldsymbol{Y}^{b,h} = \{1,0\}$, $\boldsymbol{Y}^{c,d} = \{3\}$, $\boldsymbol{Y}^{\boldsymbol{c,f}} = \{3,2\}$,$\boldsymbol{Y}^{c,g} = \{3,1\}$, $\boldsymbol{Y}^{c,h} = \{0\}$, $\boldsymbol{Y}^{d,f} = \{2\}$, $\boldsymbol{Y}^{d,g} = \{1\}$, $\boldsymbol{Y}^{d,h} = \{0\}$, , $\boldsymbol{Y}^{f,g} = \{2,1\}$,$\boldsymbol{Y}^{f,h} = \{2,0\}$, $\boldsymbol{Y}^{g,h} = \{1,0\}$

$$\boldsymbol{X} = \{0, \frac{3}{2}, 0, 0, 2, 0,0,0,0,0, \infty, 0, \infty, 0,0, \infty, 0,0, \infty, 0, \infty, 0,0,0, \infty, \infty, 0, \infty, 0, \infty, 0,0, \infty, 0,0,0, \infty, 0, \infty\}$$

$$\boldsymbol{Y} = \{2,3,2, 1, 3, 1,2,1,2,2,3,2,3,1,2,0,1,1,3,1, 3,2,1,0,3,3,2,3,1,0,2,1,0,2,1,2,0,1,0, \}$$

Figure 4.1: Example of X and Y in the toy network

is as same as the traffic demand.

$\boldsymbol{Y}$ is defined as

$$\boldsymbol{Y} = \{\boldsymbol{Y}^{u,v} | u, v \in N\}. \tag{4.4}$$

where $\boldsymbol{Y}^{u,v}$ means a set of load of links on the path used by (u, v) for no link failure situation.

$\boldsymbol{Y}^{u,v}$ is defined as

$$\boldsymbol{Y}^{u,v} = \{l(e)|e \in \boldsymbol{R}^{u,v}\}, \tag{4.5}$$

where $l(e)$ indicates the load of link e.

Examples of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are shown in Fig.4.1. Numbers in the figure show link capacity. In this example, traffic is only occurring between (a, f) and (b, g), and the traffic demand between them is 2 and 1, respectively.

4.2.2 Discretization of Variables

In this section, we show how we discretized $\boldsymbol{X}$ and $\boldsymbol{Y}$. We express the former as $\bar{\boldsymbol{X}}$, and the latter as $\bar{\boldsymbol{Y}}$.

$z^{u,v,*}(e)$, an element of $\bar{\boldsymbol{X}}$, is defined as

$$z^{u,v,*}(e) = \lceil z^{u,v}(e)/\mathrm{CX}\rceil, \tag{4.6}$$

where CX is discretization interval of $\bar{\boldsymbol{X}}$. Note that, when $z^{u,v}(e)$ is ∞, then we regard it as a same value for calculation.

$l^*(e)$, an element of $\bar{\boldsymbol{Y}}$, is defined as

$$l^*(e) = \lceil l(e)/\mathrm{CY}\rceil, \tag{4.7}$$

where CY is discretization interval of $\bar{\boldsymbol{Y}}$. Through out this chapter, we set CX as 0.5, and set CY as 5 since the order of mutual information and entropy do not change when comparing two networks.

Using $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$, we define the mutual information measurement as:

$$I^c(\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}) = H^c(\bar{\boldsymbol{X}}) - H^c(\bar{\boldsymbol{X}}|\bar{\boldsymbol{Y}}), \tag{4.8}$$

where $H^c(\bar{\boldsymbol{X}})$ is entropy of $\bar{\boldsymbol{X}}$, and $H^c(\bar{\boldsymbol{X}}|\bar{\boldsymbol{Y}})$ is entropy of $\bar{\boldsymbol{X}}$ conditioned on $\bar{\boldsymbol{Y}}$. Note that the lower bound of $I^c(\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}})$ is 0, and the upper bound of $I^c(\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}})$ is $H^c(\bar{\boldsymbol{X}})$ or $H^c(\bar{\boldsymbol{Y}})$.

4.3 Entropy and Evolvability

As we explained in the previous section, the upper bound of mutual information depends on entropy. Therefore, before investigating the relationship between mutual information and additional capacity needed to deal with new environment, we investigate the relationship between entropy and additional capacity. To do this, we use networks having different entropy, and evaluate the additional capacity. Since $\boldsymbol{Y}$ do not related to capacity, we only consider entropy of $\boldsymbol{X}$. We explain the process of how we generated the networks in Sec. 4.3.1. The evaluation is explained in Sec. 4.3.2.

4.3.1 Networks Used for Evaluation

For comparison, we consider a highly engineered network. We call it N-ENG. N-ENG designs the capacity, enough to accommodate traffic when single link failures occur. Entropy of this network is low. Note that, however, it is not the lowest because elements of $\boldsymbol{X}$ are not all 1.0. This is because elements of $\boldsymbol{X}$ depends on topology. Let N-ENG be $N^I(G^I, C^I)$, where $N(G, C)$ indicates a network with topology G with its capacity designed as C. Also, let V indicates a set of nodes of G, and E indicates a set of links of G. For N-ENG, C^I is designed to accommodate traffic occurring under no link failure situation, and also under every single link failure situations when traffic demand is T^I.

To compare with N-ENG, we generate networks with higher entropy. Different from N-ENG, there is no generating method in engineered way for such networks. However, since the objective here is to show additional capacity needed to deal with environmental changes, we simply generate such networks by letting the elements of $\boldsymbol{X}$ be heterogeneous. To do this, we use N-ENG as an initial network, and add capacity randomly. Let $N_p(G_p, C_p)$ be the network before the capacity addition, and let $N_n(G_n, C_n)$ be the network after the capacity addition. With setting $N^I(G^I, C^I)$ as the first $N_p(G_p, C_p)$, we repeat the following process.

1. Select an edge e^* from G_p.
2. Select an end node pair (u^*, v^*) from end node pairs using e^* under no link failure situation, $\{(u, v)|e^* \in R^{u,v}\}$.
3. Calculate links E^* on restoration paths of (u^*, v^*) when link e^* fails.
4. Add one unit of capacity on every link in E^*, and let it be C_n.

As for $N^I(G^I, C^I)$, we use a topology with 15 nodes and 28 links as G^I. It is generated by BA model [22], then randomly rewired by keeping the degree distribution to be the same [30]. The reason why we generated it this way is because networks generated by such way can capture the properties of evolvable network topology explained in our previous work [18]. The properties are topological diversity and degree distribution heterogeneity. In the simulation, T^I was set to follow log-normal distribution $LN(1, 0.5^2)$.

We chose the network generated by repeating 100 times of such process. This network has enough capacity since the randomly added capacity is more than the capacity placed at original network. We call this network N-HH. To confirm the entropy of N-HH is high, we show the entropy of networks generated during the process in Fig. 4.2. X-axis is repeated times, and y-axis is the entropy. From the figure, we can see entropy rapidly becoming high at early stage. However, the increasing speed gets lower as entropy increases. Next, we show the mutual information of these networks in Fig. 4.3. X-axis is entropy, and y-axis is mutual information. We can see the mutual information is also becoming higher. For comparison purpose, we generate a network whose entropy is the middle of N-ENG and N-HH. We call this network N-MH. It is generated by repeating the process 17 times.

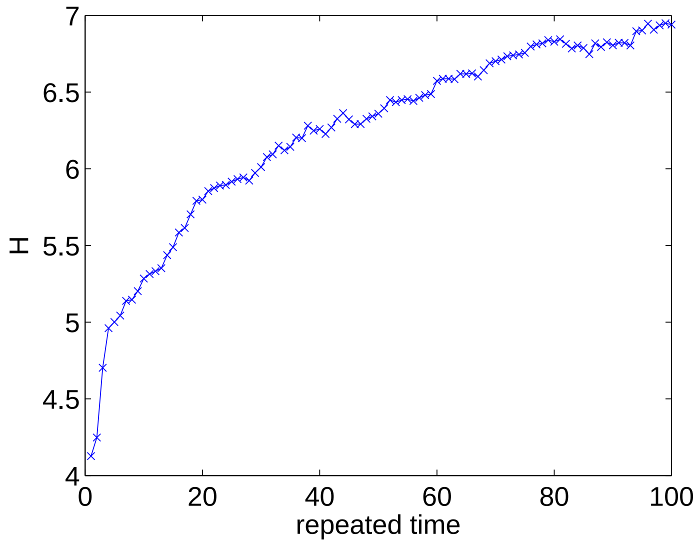


Figure 4.2: Change of entropy against capacity exchanging steps

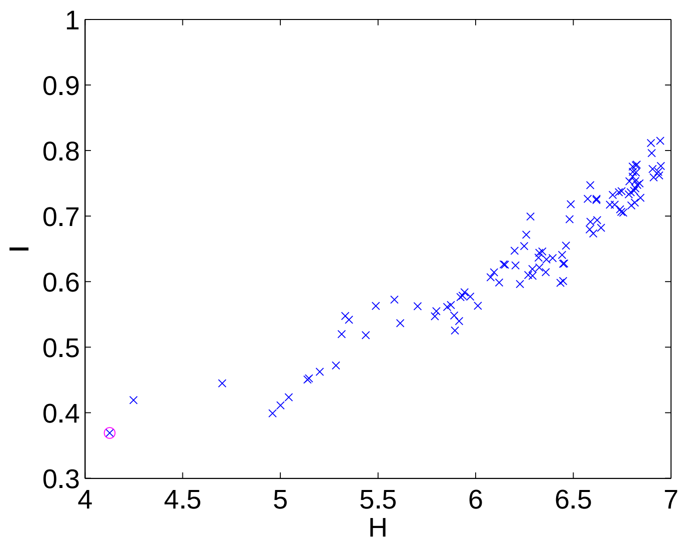


Figure 4.3: Entropy and mutual information of networks obtained during capacity exchanging process

Table 4.1 shows the value of entropy, conditional entropy, mutual information of the three networks with $H^c(\bar{\boldsymbol{Y}})$. We can see from the table that $H^c(\bar{\boldsymbol{Y}})$ in every networks are the same, and

Table 4.1: Entropy, conditional entropy and mutual information of selected networks

	$H^c(X)$	$H^c(X\|Y)$	$I(X;Y)$	$H^c(Y)$
N-ENG	4.12639	3.75713	0.369262	1.4875
N-MH	5.70286	5.14024	0.562614	1.4875
N-HH	6.91122	6.15229	0.75893	1.4875

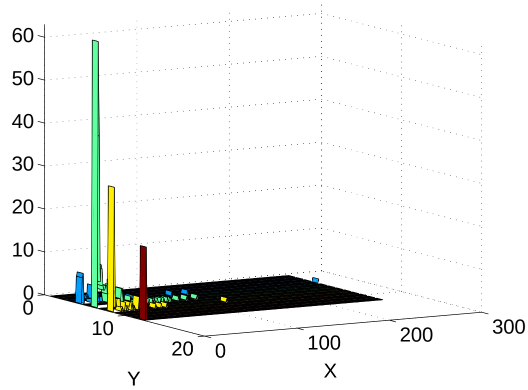
smaller than $H^c(\bar{\boldsymbol{X}})$.

We also show the distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ for each network. Since a value of entropy do not correspond to a single distribution, we clarify show the distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ for each of N-HH, N-MH and N-ENG in Fig.4.4. Note that the values of $\boldsymbol{X}$ and $\boldsymbol{Y}$ showed in the figure are after discretization, which are $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$. We can see from the figure that N-ENG has the most frequent combination of $(\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}})$. However, the frequency decreases in N-MH, and decreases even more in N-HH. The variation of $\bar{\boldsymbol{X}}$ becomes more in N-MH than in N-ENG, and also more in N-HH than in N-MH.

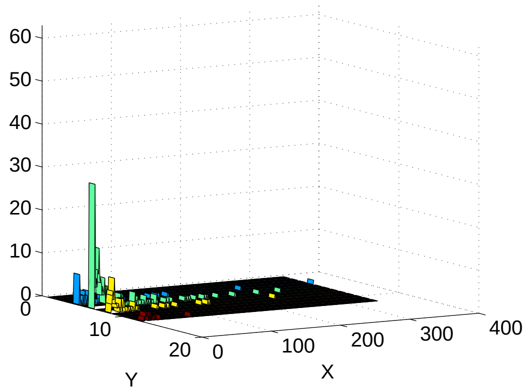
4.3.2 Evolvability Evaluation

We evaluate the difference of additional capacity with networks with different entropy. To evaluate under the same cost, we order the total capacity of all of networks. Then, we calculate the capacity needed to add to deal with environmental changes. The total amount of capacity is set to the maximum total capacity among N-HH, N-MH and N-ENG. We ordered the total capacity simply as follow: let $D(N)$ be the total amount of link capacity of an evaluation network N, and let $D_{\max}$ be the maximum total capacity among all the evaluation networks, then $c^*_{(N,e)}$, the new link capacity of a link e in network N, is

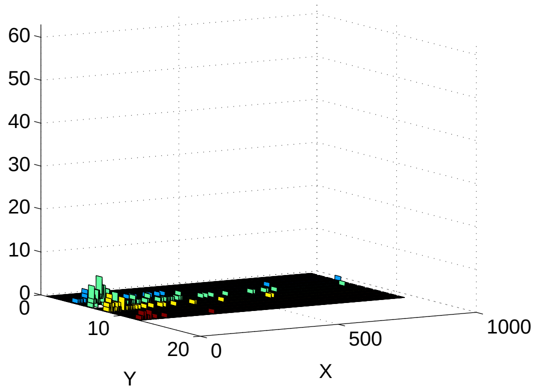
$$c^*_{(N,e)} = c_{(N,e)} * \frac{D_{\max}}{D(N)} \tag{4.9}$$



(a) N-ENG



(b) N-MH



(c) N-HH

Figure 4.4: Distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ in selected networks

where $c_{(N,e)}$ indicates the original link capacity of edge e in network N. We use N-HH^I, N-MH^I and N-ENG^I to indicate networks after their capacity ordered. Capacity added for ordering can be consider as simple provisioning for constant traffic growth. Therefore, comparing N-ENG^I and N-HH^I means comparing a network designed optimally with large provisioning for constant traffic growth and a network designed with high diversity but with less provisioning.

To show a network with its capacity assigned diversely can deal with environmental changes with less additional capacity, we evaluate additional capacity needed to deal with changes in traffic demand after single link failure. This is, for example, the case when traffic changes after a disaster.

We set $t^*(u,v)$, the changed traffic demand between an end node pair (u,v), as:

$$t^*(u,v) = t^I(u,v) + k, \tag{4.10}$$

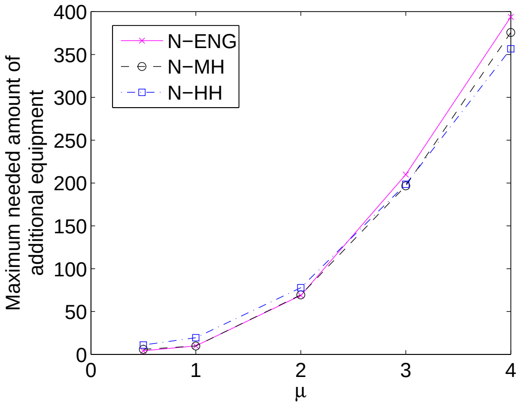
where $t^I(u,v)$ indicates the traffic demand between (u,v) in T^I, and k is a random number following log-normal distribution $LN(\mu, \sigma^2)$.

We explain the evaluation metric we used. With the changed traffic demand $T^* = \{t^*(u,v)\}$, we calculate additional capacity. $F^{e'}$, the capacity needed to add when link e' fails, is

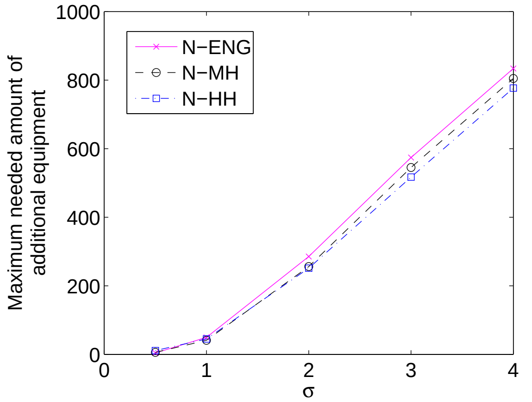
$$F^{e'} = \sum (c^{e'}_{(G,e)} - c^*_{(G,e)}), \tag{4.11}$$

where $c^{e'}_{(G,e)}$ is capacity needed on link e when link e' fails in G. In the simulation, we calculate $F^{e'}$ for every e' in G. We compare the maximum value among the obtained $F^{e'}$. This is because, as we explained in Sec. 4.1, our objective is to develop a capacity planning methodology which can follow various kinds of environmental changes with less additional capacity.

We evaluate the additional capacity against traffic demand growth by setting σ as 0.5 and μ as $0.5, 1, 2, 3, 4$. Figure 4.5(a) shows the results. Since the result depends on random number generator, we run the simulation for 100 times with different seeds. The results show the maximum value of $F^{e'}$ among 100 times for every e' in G. From the result, we can see that N-HH^I needs the largest additional capacity when μ is small, such as 0.5 and 1, while additional capacity of N-HH^I becomes lower compared to that of N-ENG^I when μ is large, such as 3 and 4. This indicates that additional capacity needed to cover traffic when traffic changes occur depends on the increase amount of traffic. N-ENG needs the fewest when traffic increases a little, while N-HH^I becomes the fewest as traffic increases.



(a) Amount of additional capacity when μ changes



(b) Amount of additional capacity when σ changes

Figure 4.5: Amount of additional capacity when traffic changes

Table 4.2: Networks needing less additional capacity with different scale of traffic changes

		μ				
		0.5	1	2	3	4
σ	0.5	N-ENG^I	N-MH^I	N-ENG^I	N-MH^I	N-HH^I
	1	N-MH^I	N-MH^I	N-MH^I	N-MH^I	N-HH^I
	2	N-HH^I	N-MH^I	N-MH^I	N-MH^I	N-HH^I
	3	N-HH^I	N-HH^I	N-HH^I	N-HH^I	N-HH^I
	4	N-HH^I	N-HH^I	N-HH^I	N-HH ^{I}	N-HH^I

We evaluate the additional capacity against traffic demand fluctuation by setting μ as 0.5 and σ as $0.5, 1, 2, 3, 4$. Figure 4.5(b) shows the results. The result also shows the maximum value of additional capacity among 100 times of simulation. From the result, we can see that N-HH^I needs the largest additional capacity when σ is small, such as 0.5 and 1, while additional capacity becomes lower compared to N-ENG^I when σ is large. This indicates that additional capacity needed to deal with traffic changes also depends on the size of traffic variation. N-ENG^I needs the fewest when traffic changes small, while N-HH^I becomes the fewest as traffic changes large.

To search for more patterns of traffic changes, we show the results when μ is $0.5, 1, 2, 3, 4$ and

σ is $0.5, 1, 2, 3, 4$ respectively. From Tab. 4.2, we can see that N-ENG^I requires smallest capacity when (μ, σ) is (0.5, 0.5) as expected. When μ or σ becomes larger, for example, when μ is 1 to 3 or when σ is 1 to 2, N-MH^I is the best. When μ and σ are large, N-HH^I becomes the best. This indicates that, when traffic demand does not change or changes a little, the network with low entropy requires the fewest. However, as traffic demand changes greatly, the network with high entropy requires the fewest capacity. Therefore, under the difficulty of traffic prediction in the future, capacity should be designed with high entropy.

4.4 Mutual Information and Evolvability

In this section, we show the relationship between mutual information and additional capacity needed to deal with environmental changes. To do this, we use networks with different capacity, and compare the additional capacity. We explain the process of how we generated the networks in Sec. 4.4.1. In Sec. 4.4.2, we show the additional capacity.

4.4.1 Networks Used for Evaluation

In Sec. 4.3, we showed that N-HH requires less capacity when environment changes greatly. Therefore, we generate networks with same entropy with N-HH, but with different mutual information for the evaluation.

For comparison purpose, we generate a network with higher mutual information than N-HH. The mutual information of a networks is high when whose elements of $\boldsymbol{X}$ with the same $\boldsymbol{Y}$ are less heterogeneous. We introduce a proportionality coefficient ζ in the generation method to generate a network having same entropy with N-HH. This is because, when we design to let each element of $\boldsymbol{X}$ proportional to the corresponding element of $\boldsymbol{Y}$, entropy of the network is too high. When calculated capacity cannot accommodate traffic when a single link failure occurs, we design network to accommodate the traffic. We generate the network as follow:

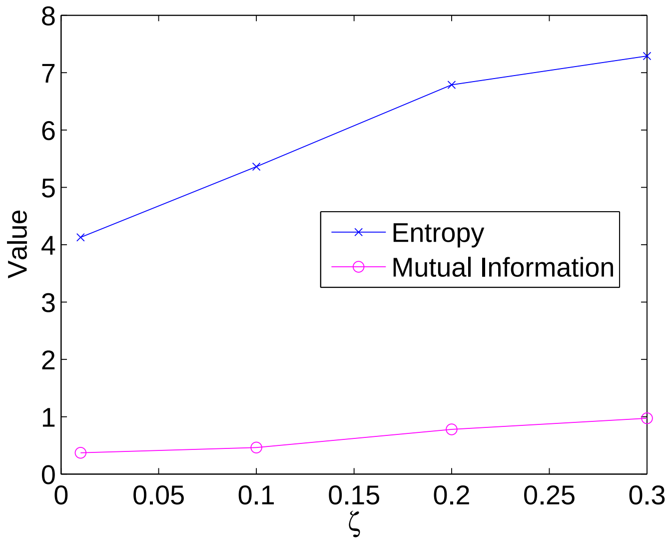


Figure 4.6: Mutual information and entropy against parameter ζ

1. Design capacity to accommodate traffic occurring under no link failure situation.

2. Design capacity to accommodate traffic occurring under link e failure situation.

   (a) Change traffic demand of end node pairs $\{(u,v)|e \in R^{u,v}\}$ as follow. Using $l(e)$ and $t^I(u,v)$, $t^{**}(u,v)$, the changed traffic demand between (u,v), is:

   $$t^{**}(u,v) = \zeta * l(e) * t^I(u,v), \tag{4.12}$$

   where ζ is a designing parameter as explained above. However, when $t^{**}(u,v)$ is lower than 1, network cannot accommodate traffic occurring in a single link failure situation. Because it is not realistic to designing capacity lower than what is expected in the designing period, we change $t^{**}(u,v)$ to 1 if it is lower than 1.

   (b) Using the changed traffic demand, design capacity for accommodating traffic under link e failure situation.

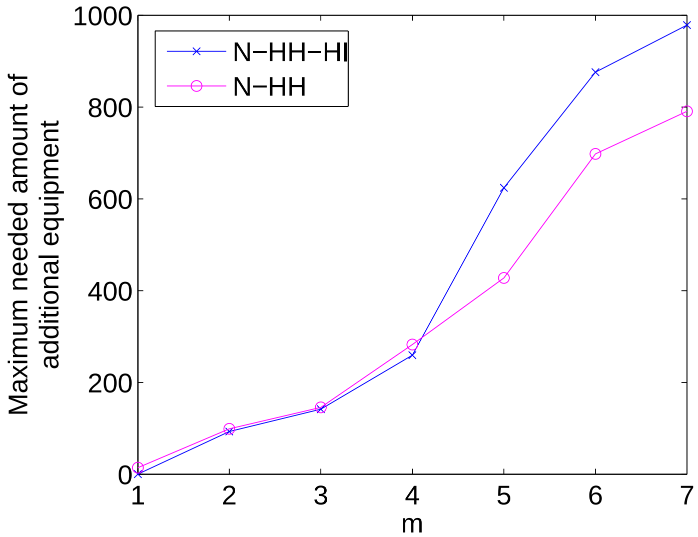


Figure 4.7: Amount of additional capacity against multi-link failure

3. Repeat (2) until capacity for every link failure is designed.

Figure 4.6 shows entropy and mutual information of networks with different ζ. A network generated by setting ζ to 0.001 is as same as N-ENG. This is because all the traffic demand is setting to 1 for every single link failures. From Fig. 4.6, we can see that, as ζ increases, the mutual information increases. Since, when ζ is 0.3, entropy of the network is as same as N-HH, we use this network for comparison. Hereafter, we call this network N-HH-HI. Note that mutual information of N-HH-HI is 0.21227 higher than that of N-HH.

4.4.2 Evolvability Evaluation

We evaluate additional capacity needed to deal with environmental changes with networks with different mutual information. To evaluate with the same cost, we order the total capacity of all the evaluation networks first. Then, we calculate the additional capacity.

To evaluate with the same cost, we order the total capacity N-HH and N-HH-HI to let them be even. The method we used is as same as what we used in Sec. 4.3.2. N-HH^{II} and N-HH-HI^{II} indicates networks after ordered respectively. Traffic demand is set to what we considered in the designing period. The evaluation metric we used is also same as Sec. 4.3.2.

Figure 4.7 shows the worst case among all the patterns of m link failure. We can see from the result that the worst additional capacity is 0 in N-HH-HI when under single link failure situation, while the order changes from five links failures. On the other hand, N-HH needs more additional capacity until four-link link failures. However, N-HH becomes less than N-HH-HI for more severe failure. This means when little number of links fails, a large amount of capacity is needed to add in network with high mutual information, while less amount of capacity is needed when large number of links fails.

4.5 Summary

We showed that a capacity planning methodology enhancing the diversity in capacity assignment can adapt to large environmental changes with less amount of equipment. Our basic approach is to assign additional capacity on each link such that the capacity can work for more severe environmental changes as well as single link failures. More specifically, we increased the diversity of available capacity on restoration path under single link failures. To quantify the diversity of available capacity under the single link failures, we first defined the mutual information between capacity on a restoration path of a link and load on the link which was failed. Then, we investigated network with its capacity on restoration paths assigned randomly, and showed that capacity assigned diversely can save additional capacity against environmental changes. We also investigated networks with different mutual information, and showed that capacity with low mutual information can decrease total amount of additional capacity against environmental changes.

Chapter 5

Conclusion

As the Internet becomes a social infrastructure, it is important to design the Internet that has adaptability against environmental changes. In this thesis, we investigated a network design method which has a capability of following various kinds of environmental changes with less amount of equipment is necessary.

Starting with our research background and overview of our studies in Chap. 1, we first introduced a mutual information to quantify the degree of specialization in a topological sense. In Chap. 2, we quantified the topological diversity by the mutual information between degrees of two nodes that are connected. From calculating the mutual information of router-level topologies and a BA topology, we found that the mutual information of router-level topologies are higher than that of the BA topology. The reason of high mutual information in the router-level topologies is partly explained by the following fact. Since router-level topologies are designed under the physical and technological constraints such as the number of switching ports and/or maximum switching capacity of routers, there are some restrictions when constructing the network, leading to high correlation in degrees of two connected nodes. We also enumerated topologies with different mutual information, and then showed that topology with high mutual information is less diverse, and have more regularity than the one with low mutual information.

In Chap. 3, we proposed a network design approach to enhance topological diversity, by which the network can easily adaptable to deal with new environments without requiring a lot of additional equipment. Essentially, in our approach, a new node is connected to existing nodes to minimize the mutual information of the topology. We then evaluated the total cost, which is defined by the total amount of equipment needed for accommodating traffic in two cases; in an ordinary situation where the traffic is increased gradually, and in the situation where node failure takes place. Our results showed that a thousand-node network evolved by our design approach reduces the total cost of equipment by 15% comparing to a thousand-node network evolved by an ad hoc design method. We also found that, when severe failures occur, the network designed by our design approach can effectively utilize already attached capacity compared to that by the ad-hoc design method. We revealed the trade-off relationship between topological diversity and physical distance, and revealed that our design method can suppress a large amount of total link capacity with a little increase in physical distance of links.

In Chap. 4, we next considered diversity of link capacity in addition to the topological diversity. For that purpose, we extended the definition of mutual information by considering load of the link and the available capacity after a failure of the link. Then, a network with low mutual information is obtained by repeatedly exchanging a small amount of capacity between links. Although the diversity of available capacity does not contribute for the case of single link failure, we expected the additional capacity will work for severe environmental changes and expected that less capacity is required as in the case of the topological diversity. Using a 15-node topology with 28 links, we examined the effectiveness of the capacity planning with low mutual information. Our results showed that the total amount of capacity is decreased when two or more links are failed simultaneously and is decreased by 20% at a seven-link failure.

Through the work presented in this thesis, we showed that diversities in information networks are key to adapt various kinds of environmental changes with less amount of equipment, and our design method that strengthens the topological diversity and diversity in capacity can adapt various environmental changes with less amount of equipments.

Several problems are left for future research. First, the calculation time of the proposed design approach should be improved. The calculation complexity of our design method is $O(n^2 * d^2)$, where n is the number of nodes and d is degree, there is a scalability problem of our approach. However, because the purpose is to higher topological diversity, strict minimization may be not needed. Approximate solution can be considered in the future work. Second, analytical investigation is required to provide more clear discussion of evolvability of our approach to several other unexpected environmental changes. Lastly, we have considered topological diversity and diversity of link capacity here, but diversity at the processing capacity of routers is another important factor to design an information network. We believe that a diversity in router's processing capability may help to obtain an evolvable network, which is remained for our future research topics.

Bibliography

[1] Ministry of Internal Affairs and Communications, “Aggregated results of traffic in Japan,” *Press Material*, Sept. 2015.

[2] 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.

[3] Cisco, “Cisco visual networking index: Forecast and methodology,” *White Paper*, May 2015.

[4] Ministry of Internal Affairs and Communications, “Accident occuring in telecommunication carrir,” *Press Material*, July 2015.

[5] Y. Huang and R. Guerin, “Does over-provisioning become more or less efficient as networks grow larger?,” in *Proceedings of IEEE International Conference on Network Protocols*, pp. 235–246, Nov. 2005.

[6] N. Katayama, M. Kubo, A. Tamura, and T. Matsui, *Network designing problem*. Asakura Publishing Co., Ltd., Oct. 2010.

[7] G. Tononi, O. Sporns, and G. Edelman, “Measures of degeneracy and redundancy in biological networks,” *Proceedings of the National Academy of Sciences*, vol. 96, pp. 3257–3262, Dec. 1998.

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

[9] J. Whitacre and A. Bender, "Degeneracy: A design principle for achieving robustness and evolvability," *Journal of Theoretical Biology*, vol. 263, pp. 143–153, Mar. 2010.

[10] J. Whitacre, "Degeneracy: A link between evolvability, robustness and complexity in biological systems," *Theoretical Biology and Medical Modelling*, vol. 7, p. 6, 2010.

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

[12] 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.

[13] 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, June 2013.

[14] L. Chen, S. Arakawa, and M. Murata, "Analysis of network heterogeneity by using entropy of the remaining degree distribution," in *Proceedings of International Conference on Advanced Communications and Computation*, pp. 84–89, Oct. 2012.

[15] L. Chen, S. Arakawa, and M. Murata, "Analysis of network heterogeneity by using entropy of the remaining degree distribution," *Technical Report of IEICE (NS2012-13)*, Apr. 2012.

[16] L. Chen, S. Arakawa, and M. Murata, "Sustainability analysis of topologies with different network heterogeneity described by mutual information of remaining degree," *Technical Report of IEICE (NS2012-157)*, Jan. 2013.

[17] L. Chen, S. Arakawa, H. Koto, N. Ogino, H. Yokota, and M. Murata, “An evolvable network design approach with topological diversity,” *Computer Communications*, Nov. 2015.

[18] L. Chen, S. Arakawa, M. Murata, H. Koto, N. Ogino, and H. Yokota, “Designing an evolvable network with topological diversity,” in *Proceedings of IEEE International Workshop on Network Science for Communication Networks*, pp. 789–794, Apr. 2014.

[19] L. Chen, S. Arakawa, H. Koto, N. Ogino, H. Yokota, and M. Murata, “A proposal of evolvable network designing approach with topological diversity,” *Technical Report of IEICE (PN2014-26)*, Nov. 2014.

[20] Y. Koizumi, T. Miyamura, S. Arakawa, E. Oki, K. Shiomoto, and M. Murata, “Stability of virtual network topology control for overlay routing services,” *OSA Journal of Optical Networking*, vol. 7, pp. 704–719, July 2008.

[21] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On power-law relationships of the Internet topology,” *ACM SIGCOMM Computer Communication Review*, vol. 29, pp. 251–262, Oct. 1999.

[22] A. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, pp. 509–512, Oct. 1999.

[23] R. Albert, H. Jeong, and A. Barabási, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, pp. 378–382, June 2000.

[24] K. Goh, B. Kahng, and D. Kim, “Universal behavior of load distribution in scale–free networks,” *Physical Review Letters*, vol. 87, Dec. 2001.

[25] S. Arakawa, T. Takine, and M. Murata, “Analyzing and modeling router-level Internet topology and application to routing control,” *Computer Communications*, vol. 35, pp. 980–992, May 2012.

[26] N. Wakamiya and M. Murata, “Bio-inspired analysis of symbiotic networks,” *Managing Traffic Performance in Converged Networks*, vol. 4516, pp. 204–213, June 2007.

[27] K. Leibnitz, N. Wakamiya, and M. Murata, “Biologically inspired networking,” *Cognitive Networks*, pp. 1–21, July 2007.

[28] Y. Koizumi, T. Miyamura, S. Arakawa, E. Oki, K. Shiomoto, and M. Murata, “Adaptive virtual network topology control based on attractor selection,” *IEEE/OSA Journal of Lightwave Technology*, vol. 28, pp. 1720–1731, June 2010.

[29] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, “Network motifs: Simple building blocks of complex networks,” *Science*, vol. 298, pp. 824–827, Oct. 2002.

[30] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, “Systematic topology analysis and generation using degree correlations,” *ACM SIGCOMM Computer Communication Review*, vol. 36, pp. 135–146, Oct. 2006.

[31] 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.

[32] M. Pentz, “Verio grows big with small clients,” *Business Journals*, Feb. 1999.

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

[34] J. Lloyd-Price, A. Gupta, and A. Ribeiro, “Robustness and information propagation in attractors of random boolean networks,” *PLOS ONE*, vol. 7, p. e42018, July 2012.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] Cisco, “Best practices in core network capacity planning,” *White Paper*, 2013.

[40] S. D. Borchgrave, “Capacity planning at Cisco,” *Cisco Blogs*, July 2010.

[41] Y. Liu and D. Tipper, “Spare capacity allocation for non-linear link cost and failure-dependent path restoration,” 2001.

[42] M. Patel, R. Chandrasekaran, and S. Venkatesan, “A comparative study of restoration schemes and spare capacity assignments in mesh networks,” in *Proceeding of International Conference on Computer Communications and Networks*, pp. 399–404, Oct. 2003.