Computer Communications 76 (2016) 101–110

Contents lists available at ScienceDirect

Computer Communications

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

An evolvable network design approach with topological diversity

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

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

ARTICLE INFO

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

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

ABSTRACT

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

1. Introduction

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

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

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

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

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

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

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

2. An evolvable network design approach

Evolution and evolvability have been studied for a long time in biology [2]. The core of evolution in living species is the presence of genetic diversity at the DNA-level and the adaptability of genetic diversity through natural selection in particular environments: individuals that are better adapted to their environment survive and pass on their genetic characteristics to the next generation. Various species exist today as a result of evolution over billions of years, under many kinds of environment.

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

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

2.1. Measure used for design

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

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

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

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

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

Table 1
Mutual information of example topologies.

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

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

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

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

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

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

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

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

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

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

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

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

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