Self transforming to power law topology for overlay networks

Suyong Eum*, Shin'ichi Arakawa*, and Masayuki Murata*
*Osaka University, Graduate School of Information Science and Technology
1-5 Yamadaoka, Suita, Osaka, 565-0871 Japan
Email: {suyong, arakawa, murata}@ist.osaka-u.ac.jp

***Abstract*— Designing the architecture of future network requires the understanding of how a topological structure influences the performance of a network. The study of the topology known as complex network theory has revealed the omnipresence of power law topology in many real networks. This power law topology is known to provide high efficiency in data exchange among individual nodes as well as robustness against random failures on them. For this reason, this topological structure has been adopted in several algorithms for the construction of an overlay network in the context of peer-to-peer (P2P) networks. All of them are based on a growing mechanism that requires continuous joining process of nodes.**

In this paper we propose a non-growing algorithm to construct a power law topology for overlay networks which have important implications for the evolution of the Internet to the next generation networks. The implementation of the algorithm is fully distributed that does not require a centrally dedicated server, thus, make them more secure against a single point failure of the system. Moreover, we demonstrate how the rewired topology takes advantage of its topological structure by analyzing its basic topological properties as well as robustness and efficiency.

***Index Terms*— Self-Organizing, Future networks, Power law topology, non-growing mechanism.**

I. Introduction

The number of users has increased from a few hundreds to more than a billion, and various types of new applications have been introduced. Aggressive user demands force the Internet either to evolve its fundamental design structure or to change its architecture totally based on "clean-slate" design principle for future networks.

An overlay network is one alternative architecture for future networks. It operates on top of a physical network that provides real network resource. For instance, Peer-to-Peer (P2P) overlay network makes use of the Internet as its physical network. Due to this separable structure, applications can be deployed virtually on top of the Internet by a growing number of network components without requiring modification to the basic Internet architecture. For this reason, an overlay network can accommodate various requirements of future applications or user demands, and these heterogenous requirements provide one justification for an overlay network in the future [1].

Topological structure of a network is believed to facilitate its performance. An overlay network is not exception. Thus, several studies have been carried out to construct an efficient topology for an overlay network in the context of P2P networks [2][3][4]. These approaches take advantage of a special topological structure called a power law topology or a scale free topology which is known to be a low diameter topology (efficiency in data exchange among individual nodes) as well as robustness against random failure of network components.

However, all these works are based on a network growing mechanism which involves the joining process of nodes to the existing network. In other words, they cannot be applied for the transformation of a topology, which is required in several cases in current or future network environment. For instance, disastrous multiple failures on peers may distort its structure severely so that inherent topological advantages may be removed, in such a case, some mechanisms are required to restore the distorted topological structure as well as the degraded performance. For this reason, in our previous paper [5] we proposed a rewiring method to transform the topological structure of P2P network to achieve this purpose.

In this paper, we improve the previous rewiring method [5] by introducing an energy function. The energy function indicates the heterogeneity level of a network and supports the rewiring process to transform a topology into a heterogenous topology. A heterogenous topology means that nodes with large number of links (large degree nodes) tend to join to other nodes with small number of nodes (small degree nodes). This topological structure is common in most biological networks which are believed to be efficient and robust.

In addition, we demonstrate how an overlay network can benefit from the topological structure emerged from the proposed rewiring process supported by the energy function. The proposed scheme can be also implemented in a self organizing manner that topological structure of a network is transformed simply by interaction among neighbor peers without help of centrally dedicated control units.

The rest of this paper is organized as follows. In Section II, we describe some basic properties that characterize a topology, namely degree distribution, clustering coefficient, and network distance. Section III presents the detail description of the proposed rewiring algorithm with assistance of heterogeneity energy function. This is followed by numerical evaluations of the proposed algorithm in Section IV. Finally, we conclude the paper in Section V.

II. Basic topological properties

Anything represented as a structure consisting of nodes and links can be analyzed using the theory of complex networks.

Since the proposed algorithm transforms the topological structure of a network, the complex network theory is a useful tool to investigate the performance of the proposed rewiring method. In this section we summarize some quantities and measures of complex networks that we use for characterizing the transformed topologies.

A. Degree distribution

A single node of a network can be characterized by its degree. The degree k_i is defined as the total number of links that are started from the node i. Degrees of all nodes in a network are characterized as a distribution function $P(k)$ that is the probability that a randomly chosen node has degree k. When the degree distribution of a network follows a power function shown in Equ. (1), the network is called a power law or a scale free network.

$$P(k) \sim k^{-\gamma} \tag{1}$$

A power law topology is known to have two interesting properties, small diameter and robustness, which are desirable topological properties for any efficient network.

B. Clustering coefficient

This property quantifies how well neighbor nodes of a given node are connected each other. The clustering coefficient C_i is defined as the fraction ratio between the existing links and possible number of total links among the neighbors of the node i. For instance, when a node i has k_i neighbors and there are E_i number of links among the neighbors, the clustering coefficient C_i is defined as follows,

$$C_i = \frac{2E_i}{k_i(k_i-1)} \tag{2}$$

The average clustering coefficient (ACC) of a topology simply averages the clustering coefficients of all nodes in the topology. Highly clustered topology is known to handle heavy traffic more efficiently [6], and this observation was applied to reduce utilization of nodes as well as to improve the reliability of a network against network failure [7].

C. Network distance

Network distance d_{ij} represents the number of links between two nodes i and j along the shortest path connecting them. This topological property has been used to measure the efficiency of the topological structure of a network due to the relation between this property and query transmission time among nodes. There are two popular ways to construct a topology with small distance. One is based on the rewiring process introduced by Watts et al [8] (It is called a small world model). The other method is to construct a power law topology. The difference between two approaches is that the former makes use of a rewiring mechanism and builds a homogeneous network (all nodes have approximately similar number of links) while the latter is based on a network growing mechanism and builds a heterogenous network. The average shortest path (ASP) of a topology is simply calculated by averaging the network distances of all node pairs in the network.

III. Proposed evolution model

The evolution model we propose here has two phases, namely rewiring phase and verification phase. The rewiring phase was proposed in our previous work [5], which is briefly introduced in Section III-A. The verification step is a selection process, which evolves a topology into a heterogeneous topology.

A. Rewiring step

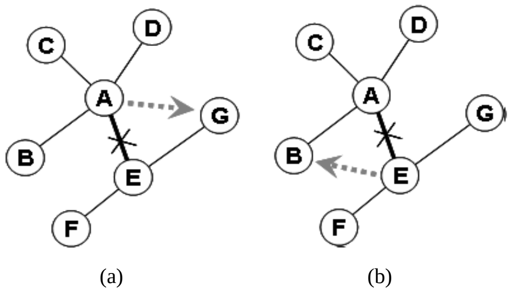

Fig. 1. Illustration of the rewiring phase. Either (a) or (b) is implemented each time with a probability of β and (1-β), respectively.

Fig. 1 illustrates the process of the rewiring phase. Firstly, a node is selected randomly from the existing network, and the selected node requests a neighbor to pass the identification (ID) of one of its neighbors.

A random peer can be chosen in various ways. Firstly, a dedicated hardware that maintains the identifications of existing peers in the networks can be used to select a random peer from the network. Although this approach provides the best performance, it may suffer from a single point failure or may not agree with one of the contributions in this paper which is the use of self organizing mechanism. The second choice can be to use some algorithms such as a random walk approach proposed by Vishnumurthy et al [9]. In this paper we do not consider the issue of random peer selection further since we assume that a random peer can be chosen by either one of the above methods.

In Fig. 1(a), let us assume that the peer A is chosen randomly and requests a neighbor peer E to pass the ID of one of the peer E's neighbors (F, G). Let's assume that the peer E passes the ID of peer G as a response, then the peer A disconnects the link to peer E and rewires to peer G. On the other hand, in Fig. 1(b) a randomly chosen peer A passes the ID of one of its neighbor peers to another neighbor. In this example, the peer A passes the ID of peer B to a randomly chosen neighbor E. Then, peer E disconnects the link to the peer A and rewires to peer B.

β is a parameter controlling the linear combination of the first and the second cases. The first case shown in Fig. 1(a) is

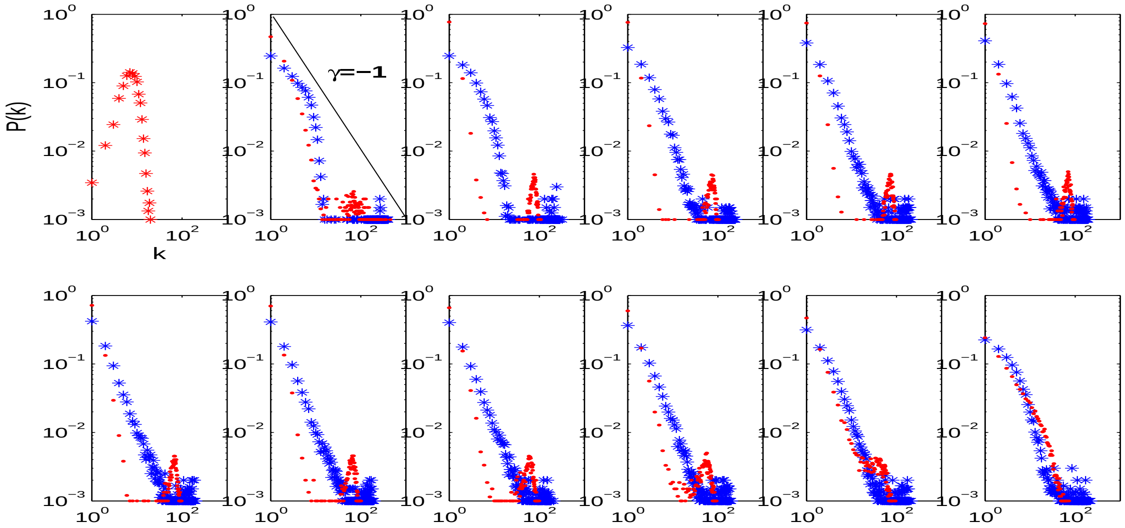

Fig. 2. The most top left figure shows the degree distribution of a random topology (N=1000, E=8000), and the rest of them show the degree distributions of the random topology after they are rewired N^2 times with different values of β (From top left to down right figures, β = 0.0, 0.1, . . . 1.0.). Each figure except the first one has two graphs. One with blue ($*$) is obtained after the random topology is rewired with the energy function in Equ. (4) (indicating the heterogeneity of the topology), and the other ones with red ($\cdot$) are obtained without the energy function. Each result is averaged over 10 realizations.

chosen with a probability β, and the second case in Fig. 1(b) is chosen with a probability $(1-\beta)$.

B. Verification step

In each rewiring event of the previous step, we accept the step selectively depending on the probability p in Equ. (3).

$$p = min\{1, e^{-[H(g^{'})-H(g)]}\} \quad (3)$$

where,

$$H(g) = \sum_{i,j \in N} -c_{ij} \left(1 - \frac{min\{k_i, k_j\}}{max\{k_i, k_j\}}\right) \quad (4)$$

Here, c_{ij} shows the existence of a link between ith and jth nodes. k_i is the degree of a node i. Thus, $H(g)$ in Equ. (4), we call it energy here, defines the heterogeneity level of a topology g. The probability shown in Equ. (3) is determined by the energy change due to the rewiring process ($g^{'}$ is the new topology after rewired). For instance, a new topology $g^{'}$ that emerges after one evolution (one rewiring) of the topology g is accepted with the probability p.

Heterogeneous topologies are common in most biological networks which are believed to be efficient and robust. Thus, any network adapting this topological structure can take advantage of this topological benefits. We discuss the benefits in the next section.

In addition, the whole rewiring process including this verification step can be implemented in a self organizing manner in such a way that each rewiring decision is made by individual nodes that just share degree information among their neighbor nodes. Due to this self organizing mechanism, this evolution algorithm has high level of scalability, in other words, a large size network can be treated as simple as a small size network. Also, the global properties of the network that may be used by hackers are hidden in a natural way.

IV. Evaluations

A. Emergence of a heterogeneous topology

For the evaluation of the proposed method, the emergence of a heterogenous topology from the method is plotted in Fig. 2. Initially we constructed 10 random topologies (The number of nodes (N):1000, and edges (E):8000) using Erdos and Renyi (ER) method [10] as original topologies. The averaged degree distribution, log values of $P(k)$ and k that were explained in Equ. (1), of the random topologies is plotted in the top left figure. The degree distribution follows a poisson distribution which represents a homogeneous network. The random topologies are rewired using different values of β from 0.0 to 1.0 increased by 0.1, and the degree distributions of the rewired topologies are plotted from the second figure to the last one. The individual figures have two graphs, ones with blue ($*$) and red ($\cdot$). They represent the topologies rewired with and without help of the heterogeneity function of Equ. (3), respectively. The degree distributions clearly demonstrate the emergence of heterogeneous topologies when the rewiring method is carried out selectively with the energy function. Especially, when β is between 0.3 and 0.8, clear power law topologies with exponent γ of 1.6 are observed.

B. ACC & ASP

Average clustering coefficient (ACC) and average shortest path length (ASP) described in Section. II play an important role to understand topological properties of a network.

A topology with high ACC is known to handle heavy traffic more efficiently [6]. For this reason, the authors in [7] increased this topological property of a network to improve its reliability against node failure as well as to reduce utilization of nodes.

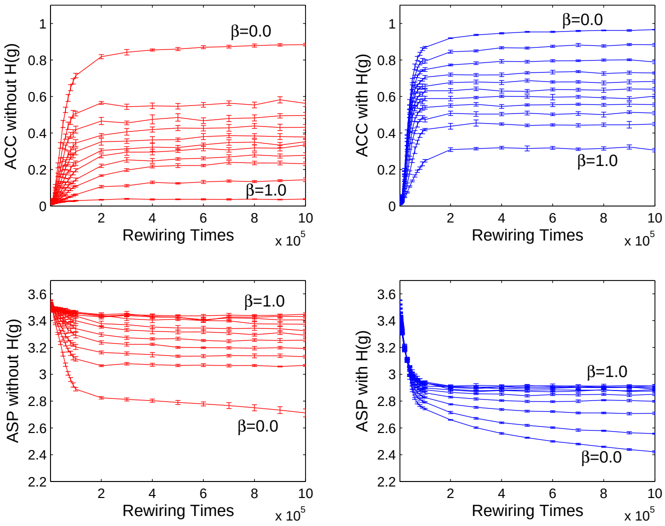

Fig. 3. Variation of basic topological properties (ACC: Average Clustering Coefficient, ASP: Average Shortest Path) of the rewired topologies with different values of β as the number of rewiring times increases with and without the energy function in Equ. (4). Regarding to ACC, nodes with at least two links are considered. 95% confidential intervals are plotted.

Moreover, ASP is closely related to the efficiency of a topology [11] since it represents the amount of network resources, e.g., the number of routers, that is required to serve one traffic flow. Thus, a topology with smaller ASP value is considered more efficient topology.

Fig. 3 plots the variations of ACC and ASP of the topologies being rewired as a function of rewiring times. The random ER topologies whose degree distributions are shown in the first figure of Fig. 2, are rewired using different values of β with and without support of the energy function in Equ. (4).

Firstly, the rewiring process increases ACC and reduces ASP of the random topologies regardless of the support of the energy function. Secondly, under the assumption that a topology with large ACC and small ASP represents a high performance topology, a rewired topology supported by the energy function shows higher performance.

The improvement of ASP through the rewiring process with the energy function can be expected from the degree distributions shown in Fig. 2 since it shows the formation of hub nodes. Emergence of hub nodes in a network is a sign that the network has small ASP value.

For more detail analysis of the clustering coefficients of the rewired topologies, Fig. 4 plots clustering coefficient distributions of the topologies whose degree distributions are shown in Fig. 2. One interesting observation here is that the clustering coefficient distributions of the rewired topologies supported by the energy function show power law distribution. It implies that small degree nodes are well clustered while high degree nodes are relatively less clustered. We discuss again this observation in terms of efficiency in Section IV-D.

C. Robustness

Robustness enables a system to withstand external and internal perturbations [12]. Since robustness is an indispensable property that guarantees certain performance of a network, it is interesting to investigate how the energy function influences the robustness of the rewired topologies.

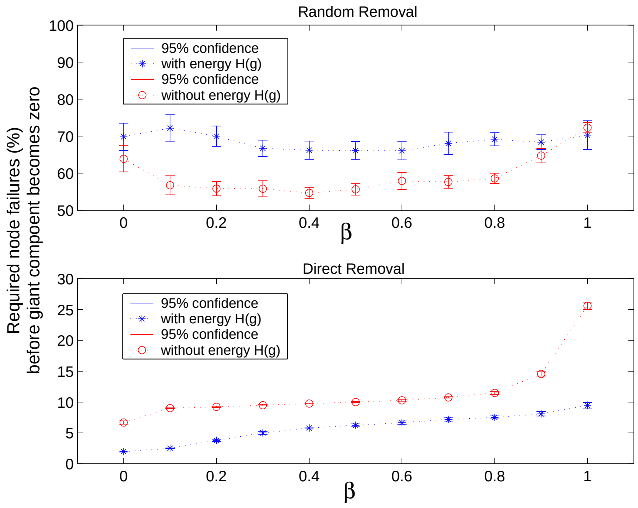

Fig. 5. Required number of node failures(%) before the size of giant component of individual topologies in Fig. 2 becomes zero.

For this reason, we adopted the classic simulation scenario proposed by Albert et al [13]. A node is kept removed randomly and directly (According to the degrees of individual nodes, e.g., a high degree node has high probability to be removed) from individual topologies whose degree distributions are shown in Fig. 2 until the size of giant component(A connected subgraph that contains a majority of the entire nodes) becomes zero, and then the numbers of removed nodes (%) are plotted in Fig. 5. Evolved topologies with the heterogeneity energy tend to be more robust against random failure, however, more vulnerable to direct attack than the ones evolved without the energy function. This outcome results from the power law degree distributions of the rewired topologies. The power law degree distribution implies that a few nodes have extremely large number of links, and large number of nodes have small degree. Thus, power law topologies are error tolerance - a randomly chosen node is likely to be a small degree node- and attack vulnerable - when high degree nodes are intentionally chosen [13].

D. Efficiency

The authors in [11][14] introduced a measure shown in Equ. (5) that indicates how efficiently data is exchanged

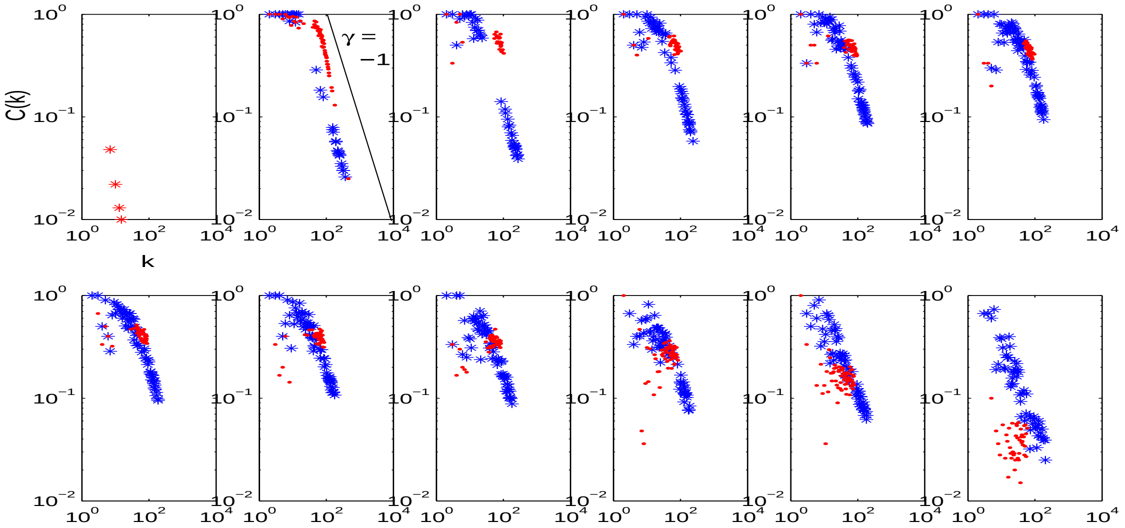

Fig. 4. Clustering coefficients $C(k)$ of topologies whose degree distributions are shown in Fig. 2 as a function of degree. Same as Fig. 2, from the second figure, the ones with blue ($*$) are obtained after the random topology is rewired with the energy function in Equ. (4) (indicating the heterogeneity of the topology), and the other ones with red ($\cdot$) are obtained without the energy function. Each result is averaged over 10 realizations.

over the network. We adopted this measure to evaluate the performances of the evolved topologies with and without the energy function.

$$AE = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{5}$$

where N and d_{ij} represent the total number of nodes in the network and the network distance between node i and j, respectively. When two nodes are disconnected, d_{ij} becomes infinite so that the metric (AE) is still able to quantify the efficiency of a non-connected topology.

As the authors [14] suggested, a node is kept removed randomly and directly (According to the degree of node) from individual topologies whose degree distributions are shown in Fig. 2 until the efficient AE becomes zero, and then the numbers of removed nodes (%) are plotted in Fig. 6.

Evolved topologies with the energy function tend to maintain higher efficiency against random failure as well as intended attack than the ones evolved without energy function except three cases (β=0.0, 0.9, and 1.0). This result is different from the one shown in Fig. 5, especially, in the scenario of the direct removal. Although, the rewired topologies with the energy function show less robustness than the ones without the energy function in the case of the direct removal, the former has higher efficiency than the latter. Since the analysis of efficiency takes into account of not only the giant component but also smaller components that are isolated from the giant component during the direct removal, high efficiency with low robustness of a network means that the network forms a clustered structure, and individual nodes inside subgraphs are generally well connected. We conjecture that this result is due to the clustering coefficient distributions shown in Fig. 4. As mentioned previously, small degree nodes in the evolved topologies with the energy function are well clustered so that it is hard to isolate individual nodes. That is why it maintains high efficiency although individual nodes continues to be removed.

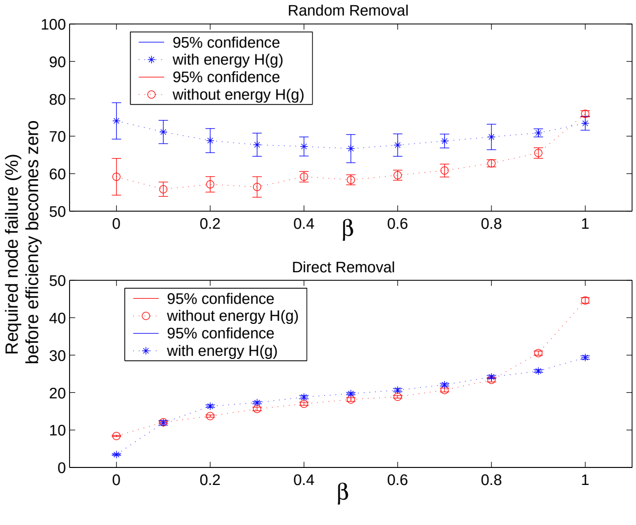

Fig. 6. Required number of node failures(%) before the efficiency of individual topologies in Fig. 2 becomes zero.

V. Conclusions

We demonstrated that a power law topology emerges through the proposed rewiring scheme with the energy function. The implementation of the approach can be carried out in a self organizing manner which enables the approach to inherit various advantages of a self organizing system such as scalability and security.

We analyzed the performance of the rewired topologies in terms of its basic properties namely clustering coefficient and network distance, and followed by more elaborated analysis in terms of robustness and efficiency. We observed two power law distributions, viz. one is in its degrees and the other is in the clustering coefficients of nodes. Firstly, due to the power law structure in its degree distribution, the rewired topology becomes a small diameter network that improves the efficiency in data exchange among individual nodes. Moreover, it inherits the robustness of a power law topology which is robust against random failure, however, vulnerable to intentional attack. Secondly, power law distribution in its clustering coefficient strengthens connections among nodes so that efficiency is better maintained in this topological structure.

Acknowledgments

This research was supported in part by the Global COE (Centers of Excellence) Program of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] D. Clark, B. Lehr, S. Bauer, P. Faratin, R. Sami, and J. Wroclawski, "Overlay Networks and the Future of the Internet," *COMMUNICATIONS & STRATEGIES*, vol. 63, no. 3, 2006.

[2] R. Wouhaybi and A. Campbell, "Phenix: supporting resilient low-diameter peer-to-peer topologies," *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 1, pp. 108–119, March 2004.

[3] M. Sasabe, N. Wakamiya, and M. Murata, "LLR: A construction scheme of a low-diameter, location-aware, and resilient p2p network," in *Proceedings of The First International Mobility, Collaborative Working, and Emerging Applications (MobCops 2006)*, Atlanta, USA, November 2006, pp. 1–8.

[4] S. Eum, S. Arakawa, and M. Murata, "Self-organizing scale free topology for peer-to-peer networks," in *2nd International Workshop on the Network of the Future (FutureNet II) in conjunction with IEEE GLOBECOM 2009*, Honolulu, USA, December 2009.

[5] ——, "Self organizing topology transformation for Peer-to-Peer(P2P) networks." *in IEICE Transactions on Communications*, vol. E93-B, no. 03, Mar, 2010.

[6] K. Hui, J. Lui, and D. Yau, "Small-world overlay P2P networks: construction, management and handling of dynamic flash crowds," *Computer Networks.*, vol. 50, no. 15, pp. 2727–2746, 2006.

[7] R. Fukumoto, S. Arakawa, T. Takine, and M. Murata, "Analyzing and Modeling Router-Level Internet Topology," *Lecture Notes in Computer Science*, vol. 52000, November 2008.

[8] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.

[9] V. Vishnumurthy and P. Francis, "On heterogeneous overlay construction and random node selection in unstructured p2p networks," in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, Barcelona, Spain, April 2006, pp. 1–12.

[10] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, pp. 17–61, 1960.

[11] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Efficiency of scale-free networks: error and attack tolerance," *Physica A: Statistical Mechanics and its Applications*, vol. 320, pp. 622–642, Mar 2003.

[12] H. Kitano, "Biological robustness." *Nat Rev Genet*, vol. 5, no. 11, pp. 826–837, November 2004.

[13] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, p. 378, 2000.

[14] V. Latora and M. Marchiori, "Efficient behavior of small-world networks." *Physical Review Letters*, vol. 87, p. 198701, 2001.