Self-Organizing scale free topology for Peer-to-Peer 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*— In operational P2P networks, an overlay topology needs to be constructed on the top of IP networks. Existing methods for the construction have limited self organizing behavior and some security problems.**

In this paper, we propose an algorithm of two parameters to construct a scale free topology for P2P networks in a self organizing manner. The emergence of scale free topologies from the algorithm is verified in both analytical and numerical ways. In addition, we demonstrate how P2P networks can benefit from the constructed topologies in terms of searching efficiency.

***Index Terms*— Self-Organizing, Scale free topology, Peer-to-Peer network.**

I. Introduction

There have been several research works on the construction of a low-diameter topology for unstructured P2P networks. The most popular way to construct a low-diameter topology is to build a power law network (Refer to Section II).

Since Barabási et al [1] introduced an algorithm to build a power law topology based on the preferential attachment mechanism, which implies a node with many links has a high probability to attract more links, several methods such as "Phenix" [2] and "LLR" [3] employed the mechanism to build a power law network for P2P networks.

The implementation of the preferential attachment requires a global view of the network or at least some information of the existing nodes such as their number of links. However, maintaining such a global view of the network is practically impossible as the size of network increases (scalability issue) [2], and moreover global or even small information of the existing nodes can be used for malicious purpose (security issue). For this reason, the implementation of the mechanism needs additional efforts to hide the identity of highly connected peer, which makes the implementation more complicated.

One possible approach for these scalability and security issues is to adopt the ability of self organization. The concept of self organization appears in many fields of science with different context. Here is one definition of self organization proposed by Prehofer et al [4].

> *A system is self-organized if it is organized without any external or central dedicated control entity. In other words, the individual entities interact directly with each other in a distributed peer-to-peer fashion. Interaction between the entities is usually localized.*

In the context of P2P overlay networks, it can be retranslated as individual peers autonomously interacting together to build a topology that possesses certain properties without any centrally dedicated control unit. Due to this distributed and localized operation among peers, a topology constructed in a self organizing manner has a high level of scalability which means a topology can be expanded at the same pace whether it is large or small, and also robustness against failure of some peers.

There have been various attempts to construct a power law topology in a self organizing manner in physics community. Firstly, Vazquez [5] introduced a method for the construction of a power law topology called random walk algorithm. It is a model of one parameter that initially a new node is attached to a randomly chosen node in the existing network and then it moves to a neighbor node[1] with a certain probability (e.g., parameter α). A link is created from the new node to the lastly chosen node from the random walk. They mainly showed numerically the emergence of various power law topologies. Saramaki et al [6] introduced one more parameter on the top of Vazquez's model which represents the length of random walk (e.g., parameter L), and demonstrated that, with any length even one-step, the random walk algorithm constructs a power law topology. Evans et al [7] investigated a wide set of parameters on the random walk algorithm of Saramaki et al. Both Saramaki and Evans et al's models provided limited analytical results to demonstrate the emergence of various power law topologies from their random walk algorithms. Smith et al [8] introduced one parameter model which is similar to Vazquez [5] model. They fully demonstrated, with a fairly complicated analytical method called the link space formalism, that various power law topologies can be constructed from their one parameter model.

In light of these observations, we propose in this paper a model of two parameters to construct various power law topologies in a self organizing manner for P2P networks. Comparing to previously introduced methods shown in physics literatures, the proposed model can construct more various power law topologies. This fine control over power law degree distribution can solve one known limitation of power law topologies, which is that high load is on very few number of peers. The emergence of various power law topologies from the proposed peer joining process is fully verified analytically and numerically.

Moreover, we evaluate the performance of the constructed topologies in terms of searching efficiency on them. From a

[1] That is why it is called "random walk".

network security point of view, local rules we employ in the algorithm are simple enough to hide some information of the existing network from attackers. In addition, in the proposed method a new peer attaches to the peer that is determined from the existing network. In other words, a new peer is not allowed to collect any information from the existing network to be attached. For this reason, the criticism that a power law topology is vulnerable to targeted attacks [9], can be mitigated through this naturally driven hiding mechanism.

The rest of this paper is organized as follows. In Section II, we describe a topological property called the degree distribution which is important to understand this paper. This is followed by a detailed description and a theoretical analysis of the proposed scheme in Section III. Section IV presents simulation results to evaluate the constructed topologies from the proposed method. Finally, we conclude the paper in Section V.

II. Degree distribution

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

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

A power law topology is known to have two contradictory properties, namely "robust yet fragile" [9] that means the high degree of error tolerance and attack vulnerability. It is because a power law degree distribution implies that a few nodes have extremely large degrees while most of them have small degrees. In other words, randomly chosen node is likely to be a small degree node - less damage against random removal, on the other hand, when a large degree node is removed intentionally, it damages the system severely.

III. Algorithm for power law topologies in a self organizing manner

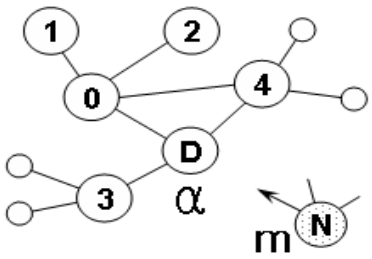

Fig. 1. An illustration of the proposed algorithm.

In this section, we describe the proposed algorithm that constructs various power law topologies. The algorithm involves two parameters, *viz* m and α as follows:

1) **m**: the number of links that a new peer has initially.
2) α: a peer in the existing topology is chosen for the endpoint of a new peer according to a combination of probability α for the random attachment and $(1-\alpha)$ for the preferential attachment.

In Fig. 1, assume that a new peer N appears with m links. Initially, a random peer is selected from the network.

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 the main contribution of this paper which is the use of the self organizing mechanism. The second choice can be to use some algorithms such as a random walk approach proposed by Vishnumurthy et al [10]. 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. Here, we just assume that there is a bootstrapping server to select a random peer from the network.

The peer N sends a request to a bootstrapping server to obtain IDs of randomly chosen m number of peers from the existing topology. Let the bootstrapping server select the peer D as one of randomly chosen m number of peers in the existing network. Then, the peer N initiates a request to peer D to ask ID of a peer to which the peer N attaches. Peer D passes its own ID with a probability α or passes one of neighbor peer's IDs $(0, 3, 4)$ with a probability $(1-\alpha)$. Finally, the peer N makes a connection to the peer whose ID is returned from the peer D.

We should note here that all decisions are actually made in the existing network. The new peer N makes a connection to the peer that is determined by the peer D. No peer needs to know more than which peers are its neighbor peers for the implementation of this algorithm. Because of this simplicity, the algorithm becomes robust and resilient against any targeted attack since any information about the existing topology is totally hidden from a new peer which can be an attacker.

A. Theoretical analysis

When a new node with degree m appears[2], each link of the new node is attached to a randomly chosen node from the existing topology with a probability α, and with a probability $(1-\alpha)$, the link is attached to a neighbor of the randomly chosen node which represents the preferential attachment.

Under the assumption that each new node appears every time unit, the total sum of the time elapse n is equivalent

[2]We assume that a m-node clique exists initially, and the initial clique is not included in this theoretical analysis.

to the total number of nodes in the network. Here, we define $N_k(n)$ as the number of nodes that have degree k after n time elapses (From this point, we use N_k instead of $N_k(n)$ for convenience). Fig. 2 illustrates the evolution of node degree

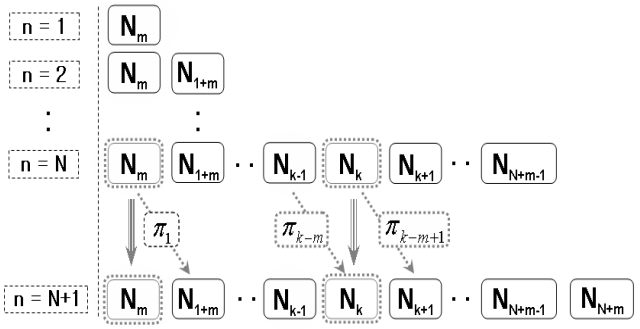

Fig. 2. Lattice diagram that shows the evolution of N_k as the number of peers n in the network increases.

counts N_k as the number of nodes in the network increases. At any time, it begins with N_m since there is always at least one node with degree m in the network.

Suppose that the constructed network has N number of nodes, and a new peer with degree m is about to join to the network. The probability that N_{k-1} evolves to N_k due to the attachment of the new node is defined as Π_{k-m} shown in Equation (2).

$$\Pi_{k-m} = \alpha\frac{mN_{k-1}}{n} + (1-\alpha)(k-1)\frac{N_{k-1}}{2n} \quad (2)$$

Here the first term of the right side of the equation represents the process that N_{k-1} evolves to N_k due to the random attachment governed by the parameter α. The probability that a randomly chosen node from the network has degree $k-1$ is equal to N_{k-1}/N, and the probability is proportional to the parameter m since each new peer makes m connections to the existing network.

The second term describes the process that N_{k-1} evolves to N_k due to the preferential attachment determined by the ratio $(1-\alpha)$. The probability that a neighbor of a randomly chosen node has degree $k-1$ is equivalent to the probability that a randomly chosen link belongs to a node with degree $k-1$. Thus, this probability can be obtained by dividing all links that are connected to nodes with degree $k-1$ by the total number of links in the network which becomes $(k-1)N_{k-1}/2mN$. This probability is also proportional to the parameter m since each new peer makes m connections to the existing network. Thus, cancelling out m in both the numerator and in the denominator so that there is not m in the second term.

In Fig. 2, the variation of N_k due to the attachment of a new peer can be defined as the difference between Π_{k-m} and Π_{k-m+1} as follows:

$$\begin{aligned}\frac{dN_k}{dn} &= \frac{N_{k-1}}{n}(\alpha m + (1-\alpha)(k-1)/2) \\ &\quad -\frac{N_k}{n}(\alpha m + (1-\alpha)k/2) \end{aligned} \quad (3)$$

Since the total elapse time n is equivalent to the total number of nodes in the network, N_k/n is equivalent to $P(k)$ which represents the probability that a randomly chosen node has degree k. In addition, dN_k/dn also becomes $P(k)$ in equilibrium state. Thus, by substituting N_k/n and dN_k/dn with $P(k)$, and N_{k-1}/n with $P(k-1)$, the Equation (3) becomes

$$P(k) = \frac{2\alpha m + (1-\alpha)(k-1)}{2 + 2\alpha m + (1-\alpha)k}P(k-1) \quad (k \geq m+1) \quad (4)$$

To complete this analysis, we define the initial condition of $P(m)$. In Fig. 2, the variation of N_m due to the attachment of a new peer can be obtained by calculating the probability that individual nodes with degree m maintain its degree without gaining a link. Thus,

$$\frac{dN_m}{dn} = 1 - \Pi_1 \quad (5)$$

In Equation (2), by substituting k with $m+1$, it becomes

$$\Pi_1 = \alpha\frac{mN_m}{n} + (1-\alpha)\frac{mN_m}{2n} \quad (6)$$

Hence, by substituting Equation (6) into Equation (5), the initial condition $P(m)$ becomes

$$P(m) = \frac{2}{\alpha m + m + 2} \quad (7)$$

Thus, we can derive the degree distribution of the constructed topology recursively from Equations (4) and (7) when the parameters of α and m are given.

IV. Experimental Results

In this section, we carry out various simulation studies to verify the theoretical analysis and to evaluate the performance of topologies that are constructed from the proposed algorithm. For the evaluation, we demonstrate that a topology with high searching efficiency can be constructed from the proposed algorithm.

A. Emergence of power law topologies from the proposed self organizing algorithm

We construct topologies numerically using the proposed algorithm with different sets of parameters, and then the degree distributions of the topologies are compared to analytical results. In Fig. 3, analytical results shown as solid white lines provide surprisingly good estimation for the simulation results shown as crosses. Especially, Fig. 3 shows that the number of large degree nodes decreases as the value of α increases (increasing the random attachment).

To observe how the parameters of α and m involve with the degree exponent γ defined in Equation (1), suppose that $m = \delta k$ $(0 < \delta \leq 1)$, substituting m with δk in Equation (4) gives

$$\begin{aligned}\frac{P(k)}{P(k-1)} &= 1 - \frac{3-\alpha}{(2\alpha\delta - \alpha + 1)k + 2} \\ &\approx 1 - \frac{3-\alpha}{(2\alpha\delta - \alpha + 1)}\frac{1}{k} \end{aligned} \quad (8)$$

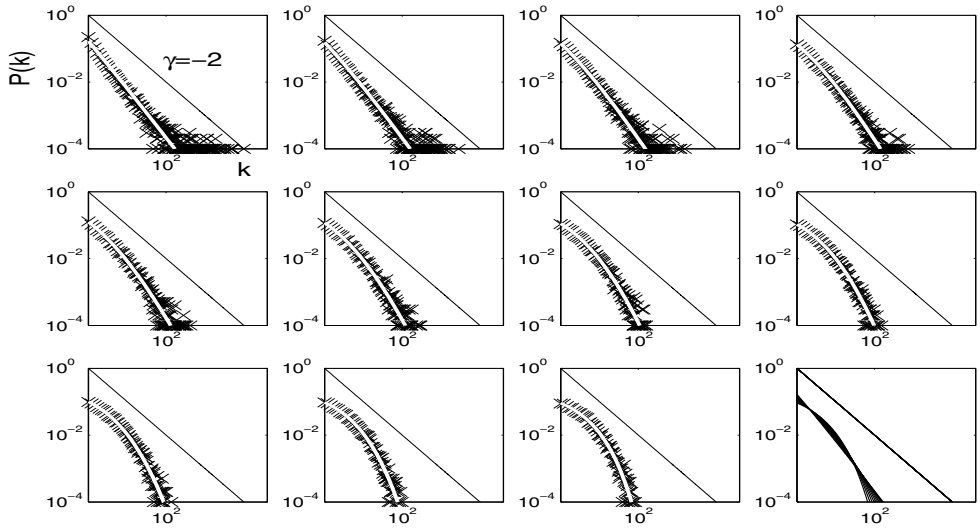

Fig. 3. The degree distributions from analytical (white solid lines) and numerical (crosses) results. For the simulation, a topology with 10^4 nodes is constructed with parameters $m = 5$, and α values that are 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, from top left to bottom right. The last figure plots the analytical results only by varying the value of parameter α.

In [11], it was shown that Equation (8) can be defined as a power law form shown in Equation (9) asymptotically.

$$P(k) = ck^{-\gamma} = ck^{-(3-\alpha)/(2\alpha\delta-\alpha+1)}$$
$$\Longrightarrow \gamma = \frac{3-\alpha}{(2\alpha\delta-\alpha+1)} \quad (9)$$

With the fixed value of m and large k which means ($\delta \to 0$), the exponent γ varies between 3 and ∞ according to the parameter α. Thus, as k increases in Fig. 3, the exponent γ asymptotically closes to 3 (first figure, α=0.0) and ∞ (second figure from the last, α=1.0). One problem of a power law

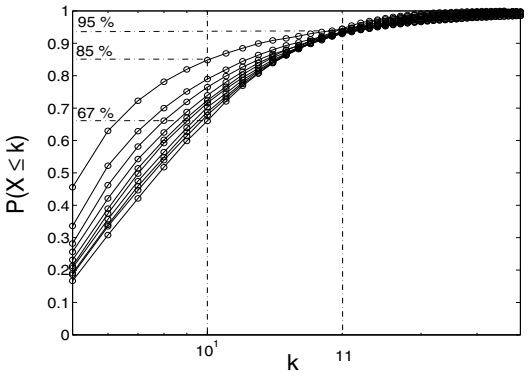

Fig. 4. The cumulative distribution of degrees shown in Fig. 3 with different values of α from bottom to top 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0.

topology is that high traffic is observed on a few of hub nodes. In unstructured P2P networks, this phenomenon is not desirable since peers are not willing to hold too much load. Thus, the number of degrees in a peer needs to be limited. Fig. 3 shows that the maximum number of degrees in the network can be controlled by varying the value of α.

The degree distributions shown in Fig. 3 are accumulated (the cumulative distribution of degree CDD) and plotted in Fig. 4. When α is equal to 1.0 and 0.0 respectively, around 85% and 67% of total nodes have less than or equal to 10 degrees. Moreover, 95% of total nodes have less or equal to 11 degrees.

B. Searching efficiency

In general unstructured P2P networks, there is not a central machine that assists peers to search desired files. Thus, searching a node that possesses a desired file is the responsibility of individual peers. Various searching algorithms have been proposed previously, however, these belong to two main categories, namely flooding and random walk searches.

1) Flooding Search (FDS): FDS is the most well known searching mechanism in unstructured P2P networks. A message is sent by a peer to its all adjacent nodes, and the nodes that receive the message resend it to its neighbor nodes excluding the source node. This process is replicated within a time frame called time-to-live (TTL). Since the average degree of a topology is proportional to m, a topology with large value of m (more links in a topology) shows high searching efficiency. When all topologies have the same number of nodes and links, a topology with small α achieves higher searching efficiency in right figure of Fig. 5. It is because a topology with small α has a small network diameter.

For the comparison, the coverage ratio of a random topology (its degree distribution is shown in the inset figure) is plotted also. All power law topologies from the proposed algorithm achieve higher searching efficiency than a random topology. For instance, when TTL is 3, FDS performs around 7 times better on a topology with ($\alpha = 0.0$) than on a random topology.

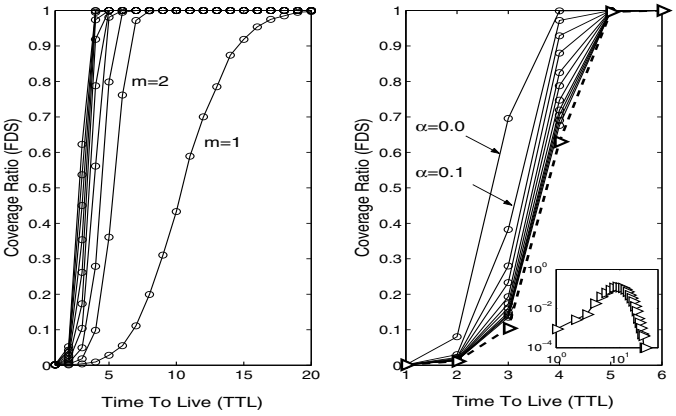

Fig. 5. Network coverage ratio by flooding search (FDS) on topologies that are generated by the proposed algorithm with different parameters (left figure with fixed (α = 0.5), right figure with fixed (m=5). For the comparison, the coverage ratio in a random topology (10^4 nodes and the average degree is 5. Its degree distribution is plotted in the inset figure.) is shown as a dot line with triangle marks in the right side.

2) Random Walk Search (RWS): RWS has been used as an alternative search mechanism since FDS generates very large amount of messaging traffic as well as poor granularity, i.e., one additional step (TTL) significantly increases the total messaging traffic in the network [12]. When a peer searches a desired file using RWS, it sends a message to one of its neighbors, and this process is repeated until the message is passed on to a neighbor of the target.

As observed previously, RWS also shows the best performance in a power law topology with a long tail (small α). Interpretation of the results shown in Fig. 6 is same as we did for the results of FDS shown in Fig. 5. In Fig. 7, the

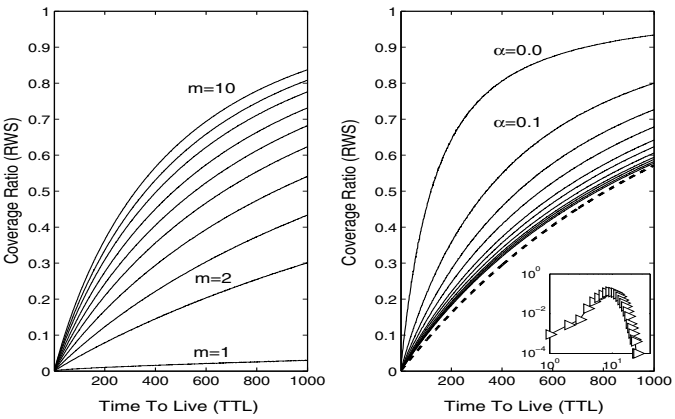

Fig. 6. Network coverage ratio by random walk search (RWS) in topologies that are generated by the proposed algorithm with different parameters (left figure with fixed (α = 0.5), right figure with fixed (m=5). For the comparison, the coverage ratio in a random topology (10^4 nodes and the average degree of a node is 5. Its degree distribution is plotted in the inset figure.) is shown as a dot line in the right side.

average peer to peer searching cost using RWS is plotted. It is interesting to observe that a topology with small number of total degrees (m) can benefit from the power law structure mostly in terms of searching efficiency. On the other hand, when a topology has large number of total degrees, it is hard to observe the improvement in searching efficiency as the value of α increases.

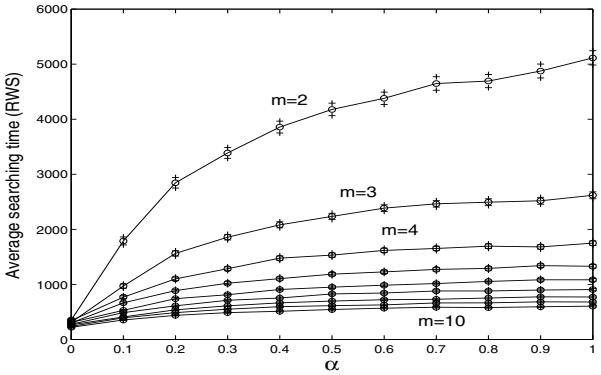

Fig. 7. The average searching time is calculated by averaging the number of hops between randomly chosen 10^4 pairs of nodes (random walk) with 95% confidential interval.

C. Peer leaving & recovering

Although the introduced algorithm ensures to construct various power law topologies, the overall structure keeps changing due to leaving peers. In this section, we investigate how leaving peers affect on topologies from the proposed algorithm and its performance.

Since leaving peers may disconnect the network, it is important to maintain the global connectivity using any mechanism. Thus, we introduce a recovering mechanism that rewires neighbors of a leaving peer in order to maintain the global connectivity and the original degree distribution as much as possible.

A peer can leave out of the network in two different ways. Firstly, it leaves forcefully, which means that the peer does not inform its leaving to its neighbor peers. In this case, neighbor peers can operate the proposed joining process again so it can keep its global connectivity. Secondly, it leaves gracefully, which means that the peer informs its leaving to its neighbor peers. In this case, the peer chooses a neighbor randomly and makes it a head peer. Then, the leaving peer passes IDs of its neighbors to the chosen head peer. The head peer makes a connection to each neighbor peer whose ID is obtained from the leaving peer. From this recovery process, a network can be protected from disconnection. Also, most neighbors maintain their number of degrees. Here we consider the second case further.

Let us define a parameter of μ that represents the mixture ratio between the number of joining peers and that of leaving peers. With a probability μ, a peer leaves the network, and with a probability $(1 - \mu)$ a peer joins the network.

In Fig. 8, we initially construct a topology of 1000 nodes with joining process only, and then leaving process is intro-

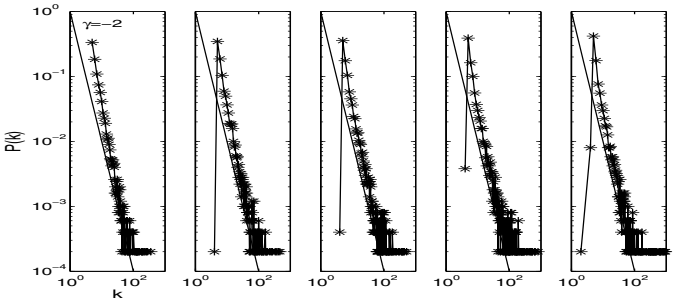

Fig. 8. The degree distributions of topologies that are constructed by the proposed algorithm ($\alpha = 0.1$, $m = 5$) with some leaving peers. From left to right, the values of μ are 0.0, 0.1, 0.2, 0.3, 0.4.

duced until the total number of nodes becomes 5000. With the proposed recovering mechanism, the global connectivity of the network can be maintained. However, we observe some distortion in the degree distributions of power law topologies. Especially, as the number of leaving peers increases, the number of large degree nodes increases also. It is because a head peer which is already a large degree peer[3] keeps gaining links. Interestingly, in spite of some distortion, the overall power law structure tends to be maintained under the proposed recovering process.

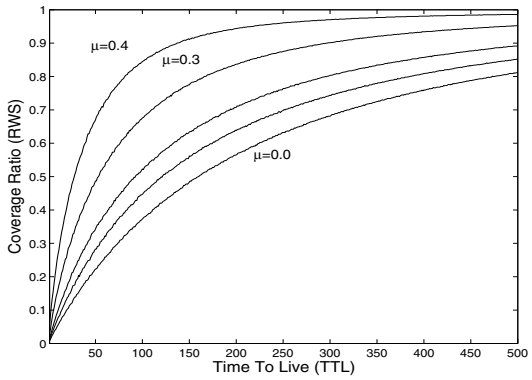

Fig. 9. Network coverage ratio by RWS in topologies shown in Fig. 8. From bottom to top, the values of μ are 0.0, 0.1, 0.2, 0.3, 0.4.

Another interesting observation is that a topology with large μ has higher searching efficiency as shown in Fig. 9. This is because the topology built with large μ has more high degree nodes than the one with small μ. Since higher degree nodes play as hub nodes, peers in the network can reach to other peers through the hub nodes within small number of hop counts so desired information can be found easily and quickly in such a small network.

V. Conclusions

A self organizing algorithm to construct various power law topologies has been proposed for P2P networks. We showed the emergence of various power law topologies analytically, and the accuracy of the analysis was confirmed with numerical results.

The proposed algorithm is simple. The only rule that each peer follows is to pass its own ID or ID of a neighbor peer on request. One important limitation of a power law topology, which is the existence of extremely high degree peers (high load), can be taken into account by varying the value of α in the proposed model.

We also demonstrated that, despite of its simplicity, the proposed algorithm is able to construct a topology that provides high searching efficiency that is highly required for P2P networks. Interestingly, the simple implementation of the algorithm enables individual peers not to share any information with other peers so that it makes attackers difficult to identify an important peer in the network. Thus, another criticism on a power law topology, which is vulnerable to targeted attacks, can be overcome.

[3]A neighbor of a randomly chosen peer tends to be a large degree peer as shown previously.

VI. Acknowledgment

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] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, October 1999.

[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. –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.

[4] C. Prehofer and C. Bettstetter, "Self-organization in communication networks: principles and design paradigms," *Communications Magazine, IEEE*, vol. 43, no. 7, pp. 78–85, July 2005.

[5] A. Vázquez, "Knowing a network by walking on it: emergance of scaling." [Online]. Available: arXiv:cond-mat/0006132

[6] J. Saramäki and K. Kaski, "Scale-free networks generated by random walkers," *Physica A Statistical Mechanics and its Applications*, vol. 341, pp. 80–86, Oct. 2004.

[7] T. S. Evans and J. P. Saramäki, "Scale-free networks from self-organization," *Physical Review E*, vol. 72, p. 026138, 2005.

[8] D. M. D. Smith, C. F. Lee, J.-P. Onnela, and N. F. Johnson, "Link-space formalism for network analysis," *Physical Review E*, vol. 77, p. 036112, 2008.

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

[10] 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.

[11] M. Mitzenmacher, "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics*, vol. 1, no. 2, pp. 226–251, 2004.

[12] H. Guclu and M. Yuksel, "Scale-free overlay topologies with hard cut-offs for unstructured peer-to-peer networks," in *Distributed Computing Systems, International Conference*, Los Alamitos, USA, June 2007, p. 32.