# Autonomous Topology Improvement for Unstructured Peer-to-Peer Networks and Its Evaluation

Haoyi Wan, Norihiro Ishikawa, Johan Hjelm †
NTT DoCoMo Inc. † Ericsson Research
{man, ishikawanor}@nttdocomo.co.jp † johan.hjelm@ericsson.com

**Abstract**

The topology of a peer-to-peer overlay network is usually constructed randomly without considering the characteristics of underlying physical links. Thus searching and routing between peer nodes is often inefficient. This paper proposes a mechanism that a peer node autonomously improves the topology of a peer-to-peer network based on the metrics of underlying networks. The simulation results show that this mechanism improves the topology of a peer-to-peer network and significantly decreases the average hop count between any pair of peer nodes.

## 非構造ピアツーピアネットワークのためのトポロジー自律改善と評価

万 皓毅　石川 憲洋　ヨハン・ジェレム †
(株) NTT ドコモ † エリクソンリサーチ
{man, ishikawanor}@nttdocomo.co.jp † johan.hjelm@ericsson.com

**概 要**

本研究では、ピアツーピアネットワークトポロジーの自律的改善方式について検討を行った。一般的にピアツーピアオーバレイネットワークのトポロジーはランダムに構成され、下位層の物理リンクの状況を考慮していないため、ピアノード間の情報検索及びルーチングは必ずしも効率的ではない。本研究では、下位層ネットワークのメトリックスを基に、ピアノードが自律的にネットワークトポロジーを改善する手法について提案した。更に、ピアツーピアシミュレータを用いた性能評価により、提案方式は任意のノード間の平均ホップ数を大幅に減少させる効果があることを示した。

## 1. Introducation

Peer-to-Peer (P2P) systems have attracted much attention and many similar systems such as Gnutella [1], Freenet [2] and JXTA [3] have been proposed since Napster music file exchanging service was widely used. P2P systems can be classified into structured and unstructured systems [4]. Unstructured P2P systems, such as Gnutella, are popularly used in the Internet, because they require no centralized directories and no control over network topology. In unstructured P2P networks, all the peer nodes form a P2P overlay network over a physical network. When a new node wants to join a P2P network, it tries to connect with the peer nodes by using the IP addresses provided by a bootstrapping node. A mismatching problem occurs between the P2P overlay network topology and the underlying physical network topology when a new node joins the network randomly according to the process described above. And it generates excessive traffic load in the Internet infrastructure and influences the performance of a P2P systems when searching information or routing in the P2P networks. The effectiveness of information searching and routing among peer nodes is one of the most important indicators to evaluate a P2P system. In this paper we focus on unstructured P2P systems and try to improve searching and routing effectiveness by solving the topology mismatching problem.

To solve the topology mismatching problem, improving P2P network topology is necessary. It can be realized by calculating metrics between every pair of peer nodes. In a large scale P2P network, due to the frequent join, leave and failure of peer nodes, the topology of the network changes time by time. It is not realistic to obtain the state of the network and calculate metrics of every pair of peer nodes. Hence, a heuristic mechanism is required to improve the network topology. In this paper, we propose an autonomous topology improvement mechanism. With this mechanism, a peer node autonomously reconfigures the topology based on the neighbor nodes' information within the range of 2 hops. Not only the virtual links of P2P level, but also the hop counts of physical links between peers are considered in the proposed autonomous topology improvement mechanism.

We organize this paper as follows. Section 2 summarizes the requirements on the network improvement mechanism. Section 3 depicts the details of the proposed mechanisms. Section 4 shows the performance evaluation. We show the related works in section 5 and conclude this paper in Section 6.

## 2. Requirements

Requirements on the autonomous topology improvement mechanism are listed as follows.

*Autonomous Topology Improvement based on local information*

Peer nodes should be able to improve the network topology autonomously based on the local information because it is not realistic to obtain the information of the whole P2P network in terms of traffic load.

*Scalability*

A mechanism should be able to be deployed in a large scale P2P networks which is constituted by thousands to tens of thousands peer nodes.

*Applicability to heterogeneous networks*

A mechanism should be able to be deployed across heterogeneous networks, including the Internet, ad-hoc networks, home networks and so on.

*Adaptation to dynamic change of network topology*

A mechanism should be rapidly able to be adapted to the dynamic change of the topology of P2P networks due to frequent join and leave of peer nodes.

## 3. Network Topology Improvement

Based on the above requirements, we propose a network topology improvement mechanism. We describe the basic design concept and its detailed mechanisms in this section.

### 3.1 Design Concept

#### 3.1.1 Exchanging of Local Topology Information

A complete understanding of the whole network topology is meaningful to improve the network topology. However, in P2P networks, exchanging of the topology information of the whole network results in heavy traffic load, since the topology of a P2P network changes very frequently. Hence, for P2P networks, heuristic topology improvement based on local topology information is feasible. To this end, local topology information around a node should be available. In our mechanism, each node periodically broadcasts its own information to other nodes within the range of 2 hops. As a result, each node holds local topology information within the range of 2 hops in its cache memory. Fig. 1 shows an example of this design concept. Exchanging of topology information is controlled in a limited area so that such traffic will not increase the load of the overall network. When the local topology information broadcasted to a peer node changed comparing with the old information cached in the peer node, the peer node starts the proposed network improvement mechanism explained in Section 3.2.

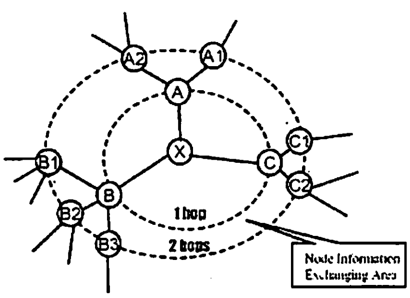


Fig. 1 The node information exchanging area

#### 3.1.2 Criteria of Topology Improvement

Considering the physical link characteristics for topology improvement of P2P networks, metrics is introduced into our proposal. Metrics is defined as a cost between two adjacent nodes. The metrics can be hop count, bandwidth or delay of the physical link between any pair of nodes. Which metrics to be used depends on the user's choice when designing a P2P network. While the proposed mechanism is independent of any particular metrics, we use the hop count of the physical link between any pair of nodes as metrics in this paper. The hop count between any pair of adjacent nodes can be measured by "Traceroute" in the case of the Internet.

Based on this consideration, we propose a mechanism which takes underlying layer characteristics into account to improve the network topology. The hop count from one node to its neighbor nodes within the range of 2 hops is used in our mechanism.

It is assumed that all of the nodes can adjust their connections with adjacent nodes. Since each node can start improvement independently, collision may occur when any pair of two adjacent nodes tries to improve their local topologies simultaneously. To avoid it, a directed graph is used in our mechanism. This directed graph is created according to the sequence of the node participation in the network. A parent-child relationship is built between any two adjacent nodes. Each node becomes a child node when it connects to a node and joins a P2P network. The connected node becomes a parent node. Each node only improves the topology along its parent's direction. A directed topology graph converted from Fig. 1 is shown in Fig. 2. Node *A* is the parent of Node *X* that is the parent of Node *B* and Node *C*. Node *X* only considers the topology improvement along its connection with Node *A* and improvements of other connections are conducted by Node *B* and Node *C*.

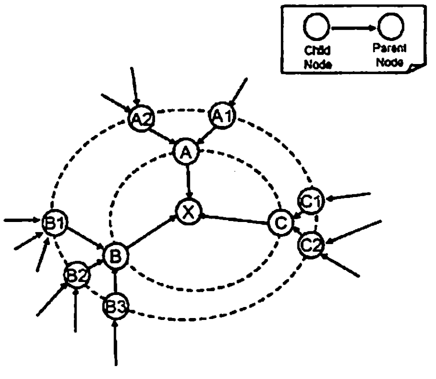


Fig. 2 Directed graph

### 3.1.3 Node Information for Broadcasting

In this subsection, we will explain the details on what kind of node information should be broadcasted and how the information is used in the topology improvement mechanism.

Table 1 shows basic node information which will be used for broadcasting to other adjacent nodes. The table includes six elements. Node ID is the identifier of a node in a P2P network. It is a global unique ID. Node Address is the physical address in a network such as IP address. The next element is maximum connections a node can hold. It is a parameter that shows the ability of a node. This value of the parameter can be statistically defined by users. The following three elements are related to an adjacent node of the node. Neighbor ID is adjacent node's Node ID. Relation depicts the relationship with an adjacent node. In this example, Node A is the parent of Node X while Node B and Node C is the child of Node X. We use hop count as metrics in this paper.

| Node ID | Node Address | Maximum Connection | Neighbor ID | Relation | Metrics |
|---|---|---|---|---|---|
| X | xxxx | 5 | A | Parent | 3 |
| | | | B | Child | 4 |
| | | | C | Child | 2 |

Table 1 Basic node information

The basic node information shown in Table 1 is broadcasted within the range of 2 hops periodically. Hence, each node will make up a table shown in Table 2 which is a description of local topology information within the range of 2 hops. The grey part of this table is the information of the node itself (Node X).

| Node ID | Node Address | Max Connection | Neighbor ID | Relation | Metrics |
|---|---|---|---|---|---|
| X | xxxx | 5 | A | Parent | 3 |
| | | | B | Child | 4 |
| | | | C | Child | 2 |
| A | xxxx | 4 | A1 | Child | 3 |
| | | | A2 | Child | 4 |
| | | | X | Child | 3 |
| A1 | ... | ... | ... | ... | ... |
| A2 | ... | ... | ... | ... | ... |
| B | xxxx | 3 | X | Parent | 4 |
| | | | B1 | Child | 4 |
| | | | B2 | Child | 2 |
| | | | B3 | Child | 1 |
| B1 | ... | ... | ... | ... | ... |
| B2 | ... | ... | ... | ... | ... |
| B3 | ... | ... | ... | ... | ... |
| C | xxxx | 7 | X | Parent | 2 |
| | | | C1 | Child | 4 |
| | | | C2 | Child | 4 |
| C1 | ... | ... | ... | ... | ... |
| C2 | ... | ... | ... | ... | ... |

Table 2 Neighbor nodes' information within the range of 2 hops

## 3.2 Network Improvement Mechanism

Based on the consideration in section 3.1, the following formula is used for network improvement processes.

$$M(x,k) = \frac{\sum_{i=0}^{n-1} (M_{ki} \times N_i)}{\sum_{i=0}^{n-1} N_i}$$

*M(x, k)* is the evaluation formula if Node *x* releases the connection to its parent node and reconnect to Node *k*. This connection to Node *k* is called as a virtual connection. *Mki* means the metrics from Node *x* to Node *i* via Node *k*. *Ni* is the number of nodes existing in Node *i* area, which are connected to Node *i*. The grey part shown in Fig. 4 is Node *i* area. Node *i* and the nodes outside the range of 2 hops from Node *X* which connected directly to Node *i* are in this area. *n* is the number of nodes within the range of 2 hops from node *x*.

Fig. 3 is an example of our topology improvement mechanism. Node *A* is the parent node of Node *X*. The metrics between Node *A* and its neighbors is obtained by node information broadcasting process which is described in section 3.1. When Node *X* starts its improvement process, it follows the next three steps to improve its local network topology.

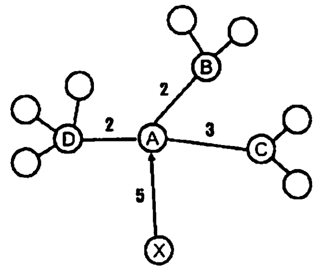


Fig. 3 Example of a node connection

*Step 1: Calculate the metrics from node x to every potential adjacent node in the P2P network as Fig. 4 shows.*

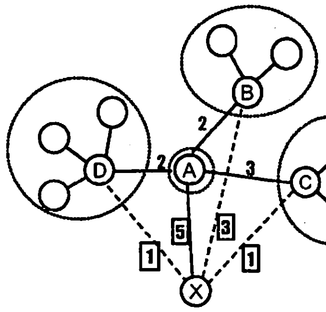


Fig. 4 Virtual connection metric calculating

In this step, virtual connections are built by Node X to calculate the metrics to its potential adjacent nodes along its parent node direction. "Traceroute" is used to calculate the value of this parameter as we mentioned in section 3.1.2. Then the process is moved to step 2 to build a set of tables for calculating the evaluation formula *M(x, k)* based on the assumption that Node *X* is connected to Node *K*.

*Step 2: Connection table is built for further computation*

In step 2, using the metrics with potential adjacent nodes a table is built for each potential adjacent node.

For Node *K*, Table k is built. *Mki* is set to a cell at the "Area i" row and "Cost Value" column. *Ni* is set to a cell at "Area i" row and "No. of Nodes" column.

Virtual connection table for Node *X* (See Table 3) is built using the P2P network shown in Fig. 3. The calculation result *M(x, k)* according to Table 3 is as follows:

$M(X, A) = 78/11, M(X, B) = 66/11,$

$M(X, C) = 49/11, M(X, D) = 40/11$

When using the virtual connection with Node *D*, the evaluation formula *M(x, k)* is minimum.

Connection to Node A: (5+21+24+28)/11=78/11

| Area | Cost | No. of Nodes | Metrics |
|---|---|---|---|
| A | 5 | 1 | 5 |
| B | 5+2 | 3 | 21 |
| C | 5+3 | 3 | 24 |
| D | 5+2 | 4 | 28 |

Connection to Node B: (5+9+24+28)/11=6

| Area | Cost | No. of Nodes | Metrics |
|---|---|---|---|
| A | 3+2 | 1 | 5 |
| B | 3 | 3 | 9 |
| C | 3+2+3 | 3 | 24 |
| D | 3+2+2 | 4 | 28 |

Connection to Node C: (4+18+3+24)/11=49/11

| Area | Cost | No. of Nodes | Metrics |
|---|---|---|---|
| A | 1+3 | 1 | 4 |
| B | 1+3+2 | 3 | 18 |
| C | 1 | 3 | 3 |
| D | 1+3+2 | 4 | 24 |

Connection to Node D: (3+15+18+4)/11=40/11

| Area | Cost | No. of Nodes | Metrics |
|---|---|---|---|
| A | 1+2 | 1 | 3 |
| B | 1+2+2 | 3 | 15 |
| C | 1+2+3 | 3 | 18 |
| D | 1 | 4 | 4 |

Table 3 Virtual connection table for Node X

*Step 3: The Node with the minimum value is selected and the connection is set up to this node.*

As a result, Node *X* releases its connection with Node A and reconnects to Node *D*. Fig. 5 shows the reconnection of Node *X*.

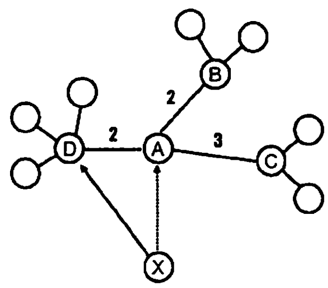


Fig. 5 Reconnection of Node X

## 4. Performance Evaluation

In this section, we present some simulation results of our proposed autonomous topology improvement mechanism.

### 4.1 P2P Network Simulator

We developed a P2P network simulator to generate overlay network topologies and simulate behaviors of peer nodes. Our P2P network simulator allows realistic Internet-like topologies to be used in simulations. The underlying network topologies are generated by BRITE [5] – a universal topology generation tool. BRITE is used by researchers to generate Internet topologies constituted by Autonomous Systems and Routers. BRITE generates a network topology as the next four-step process:

1. Placing the nodes in the plane.
2. Interconnecting the nodes.
3. Assigning attributes to topological components (delay and bandwidth for links, AS id for nodes, etc.)
4. Generating a topology using the specific format.

Our P2P network simulator uses the network topology generated by BRITE as the underlying network topology. P2P overlay networks built by peer nodes at the application-level are generated by our simulator. Our simulator can adapt to the dynamic change of P2P network topology due to join, leave or failure of peer nodes

## 4.2 Simulation Procedures

We do the simulations on a 3.2 GHz Pentium 4 PC with 3.0 GB of RAM running Windows XP as the following procedures:

1. We use BRITE to generate a random network topology constituted by 100 routers using the Waxman's probability model for interconnecting the nodes as the underlying network topology in our simulations. The Waxman's probability model is given by:

$$P(u,v) = \alpha e^{-d/(\beta L)}$$

   where $0 < \alpha, \beta \leq 1$, $d$ is the Euclidean distance from node $u$ to node $v$, and $L$ is the maximum distance between any two nodes.

2. We use the P2P network simulator to generate overlay network topologies constituted by 100 to 3000 peer nodes over the underlying network generated by BRITE. Before the topology generations start, some parameters are set as follows: number of first peer nodes = 3, maximum number of peer nodes each node can be connected to = 6, node information broadcasting interval = 10 simulation steps, node information broadcasting range = 2 hops.
3. We edit the simulation scenarios where 10% of the whole simulation steps are node join, 8% are node leave and 2% are node failure. The whole simulation steps are set to 20% of the node number in the P2P networks.
4. We use the simulation scenarios to measure the change of the average hop count between any pair of peer nodes in different scale networks during the same time interval.
5. We change the node information exchanging range to measure the number of messages transferred to improve the network topology in the whole simulation process.
6. We output the simulation results to spread sheets and show the figures in section 4.3.

## 4.3 Simulation Results

In Fig. 6, we compared the average hop count between any pair of peer nodes in the networks constituted by various number of peer nodes.

When the number of nodes is less than 1000, the average hop count between any pair of peer nodes in the network without the network improvement mechanism is almost similar to that in the network with the improvement mechanism. When the number of nodes is larger than 1000, the network with the improvement mechanism keeps the average hop count between any pair of peer nodes around 9 hops and is significantly small compared with the average hop count measured in the network without the improvement mechanism. At the same probabilities of node join, leave and failure, our proposed mechanism is effective in the network constituted by over 1000 nodes.

We changed the simulation time to make the probabilities of node join, leave and failure be the same in different scale networks. Because the improvement process is executed every 10 simulation steps, the larger the network is, the more the network improvement process is executed. This may lead to the result shown in Fig. 6. To eliminate the time factor of the improvement process, we measured the average hop count of any pair of peer nodes in different scale networks in the same time interval. The result is shown in Fig. 7.

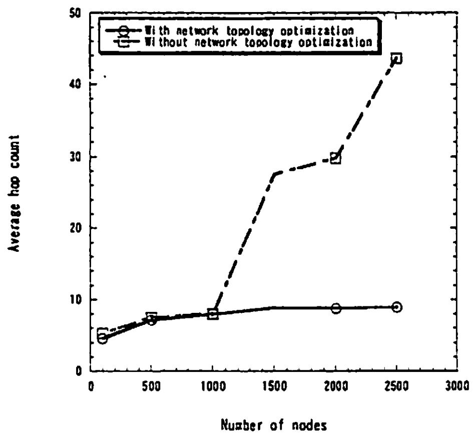


Fig. 6 Average hop count comparison at the same probabilities of node join, leave and failure

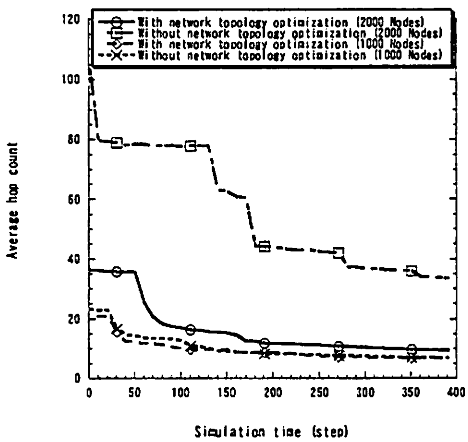


Fig. 7 Average hop count comparison in the same simulation interval

From Fig. 7, we can see that the average hop count between any pair of peer nodes in different scale networks is decreased by introducing our proposed network topology improvement mechanism. And the ratio of the average hop count decreased in the network with 2000 nodes is larger than that in the network with 1000 nodes. This explains that our proposed mechanism is more effective for large scale networks.

In Fig. 8, we show the relationship between the node information broadcasting range and the number of messages sent in the networks constituted by various number of peer nodes. "Message" indicates the messages transferred by nodes to broadcast node information during the network improvement process. In Fig. 8, 2 hops mean that each node broadcasts its own node information to other nodes within the range of 2 hops.

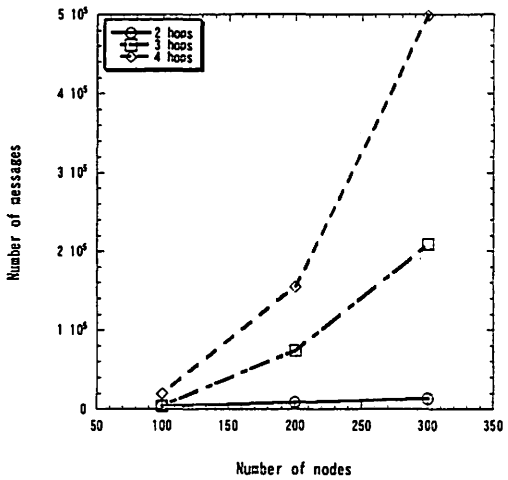


Fig. 8 Relationship between node information broadcasting range and number of messages

When the node information broadcasting range becomes more than 2 hops, the number of messages sent totally in the network rapidly increases in proportion to the increase of the number of nodes.

## 4.4 Summary

According to the performance evaluation shown above, we summarize the results as follows:

- At the same probabilities of node join, leave and failure, our proposed mechanism is effective in the network constituted by over 1000 nodes.
- Our mechanism is effective to decrease the average hop count in any scale network. In particular, our mechanism is quite effective for large-scale networks.
- Node information broadcasting within 2 hops range is effective to reduce the number of messages used by network topology improvement processes.

## 5. Related works

There have been some researches regarding the topology improvement of a P2P network. [6] proposed a distributed flow control and topology construction algorithm that (1) restricts the flow of queries into each node, so they don't become overloaded and (2) dynamically evolves the overlay topology, so that queries flow towards the nodes that have sufficient capacity to handle them.

In [7], each node independently defines a level of satisfaction. This quantity between 0 and 1 represents how satisfied a node is with its current set of neighbors. As long as a node is not fully satisfied, the topology adaptation continues to search for appropriate neighbors to improve the satisfaction level. Finally, high capacity nodes are indeed the ones with high degree and low capacity nodes are placed within the short reach of high capacity ones.

However, both [6] and [7] did not consider the characteristics of underlying physical links (e.g. hop count and delay) while adapting the overlay topology.

## 6. Conclusions

This paper proposes the network topology improvement mechanism for unstructured P2P networks. We showed the requirements of this mechanism. Based on the requirements, we propose the autonomous topology improvement mechanism using local topology information, considering the characteristics of underlying physical links. Finally, we presented the performance evaluation results of our proposed mechanism using the P2P simulator. The simulation results show that our proposed network topology improvement mechanism is quite effective to decrease the average hop count between any pair of peer nodes for various scale P2P networks.

## References


[1] Matei Ripeanu, "Peer-to-Peer Architecture Case Study: Gnutella Network", Technical Report, University of Chicago, 2001

[2] http://freenetproject.org/

[3] http://www.jxta.org/

[4] Q. Lv, P. Cao, E. Cohen, K. Li and S. Shenker. "Search and replication in unstructured peer-to-peer networks", Proceedings of the 16th ACM International Conference on Supercomputing, June 2002

[5] A. Medina, A. Lakhina, I. Matta and J. Byers. BRITE: Boston University Representative Internet Topology Generator. *http://www.cs.bu.edu/brite/*

[6] Qin Lv, Sylvia Ratnasamy and Scott Shenker, "Can Heterogeneity Make Gnutella Scalable?", In Proceedings of the 1st International Workshop on Peer-to-Peer Systems, March 2002

[7] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham and S. Shenker, "Making Gnutella-like P2P Systems Scalable", Proceedings of ACM SIGCOMM 2003