S. Kamei
NTT Communications
T. Momose
Cisco Systems
T. Inoue
T. Nishitani
NTT Communications
July 15, 2012

ALTO-Like Activities and Experiments in P2P Network Experiment Council
draft-kamei-p2p-experiments-japan-08

Abstract

This document introduces experiments to clarify how ALTO-like approach was effective to reduce network traffic made by a Council in Japan to harmonize P2P technology with the infrastructure. And this also provides some suggestions that might be useful for ALTO architecture learned through our experiments.

Table of Contents

1. Introduction

An overlay network, which is used by P2P and other applications, offers the advantage of allowing flexible provision of services while hiding the lower layer network. The downside is that inefficient routes are often taken in the lower IP network, thereby increasing the network load. Several proposals have been made to build an overlay network that takes account of the information about the lower layer network. [1] [2] Since the management of the Internet is highly distributed, it is difficult to implement such proposals and thus optimize a network without the cooperation of network providers.

Recently, the controversy between the overlay network and the network providers has been rekindled. Under these circumstances, some researchers have studied overlay network control technology that takes account of the network topology information obtained from network providers.

One of the activities concerning this issue has been made by the P2P Network Experiment Council in Japan. This document reports on the issues addressed and experiments being made by the council, focusing on the experiments made from 2007 to 2008.

2. Background in Japan

2.1. P2P traffic

As of 2008, the world most popular P2P file sharing application, Bittorrent, isn't widely deployed in Japan. Instead, other Japan specific file sharing P2P applications such as Winny [3], Share [4], and so on, still occupy 40% of the Internet traffic in Japan even though many those P2P users were arrested for sharing illegal files with these P2P apps.

Each P2P file sharing application has a unique protocol and none of them have a large market share therefore making it hard to effectively control.

2.2. Impact on network infrastructure

One of the advantage of using P2P technology for content delivery is that peers exchange content directly among themselves. This reduces the load on servers. Also, P2P applications can reduce upstream traffic from an original content server. This is significant that the charge for upstream traffic is usually delay-sensitive for content delivery services, and it is not negligible.

It is also known that server cost could be reduced with P2P technology. However, the story is quite different for network providers. From the viewpoint of network providers, the traffic that content servers generate has shifted to the edge network and the amount of traffic has not necessarily been reduced with using P2P technology for reducing server cost. Another problem for network providers that an extremely inefficient routing may be selected has been raised. It is because overlay network systems are configured without any regard to the structure of the lower layer network or network geometry.

In some cases, the total amount of traffic on the Internet used to be limited by the capacity of servers. For those cases, P2P technology can improve the scalability of servers , however it may exhaust network resources. Moreover, using P2P applications increases the volume of traffic per user remarkably.

Faced with increase in the load on network infrastructure, network providers are compelled to take actions to overcome the sudden increase in facilities' cost. Representative actions include placing content in internet exchanges or data centers, introducing bandwidth control, and raising the access fees [5].

In the future, video posting sites, which has been delivered using client-server applications, may adopt P2P system. The increase in traffic arising from such a shift will be a great threat to the network.

2.3. The object of P2P Network Experiment Council

In order to reduce Internet traffic and encourage legitimate use of P2P technologies, the Japanese government led to establish a new council called P2P Network Experiment Council conjunction with commercial P2P application vendors and ISPs in 2006.

Then the council had started to develop regulations that include several guidelines such like an advance notice to restrict bandwidth to heavy traffic users. In accordance with the regulations, some ISPs introduced solutions that reduce traffic caused by P2P file sharing applications.

Besides this activity, the council also looked for new ways to control traffic by commercial P2P applications with ISPs, carriers, contents providers and P2P system vendors. In this work, the council had experiments that introduced ALTO-like system and observed how the traffic was reduced by redirecting to proper peers on the real Internet in Japan.

In our experiment, the council settled hint servers, which are described in section 4. Hint servers have a protocol offering network distance to peers, which is disclosed to P2P application vendors.

Using hint server, P2P application vendors can introduce ALTO concepts easily to their P2P distribution systems. Because the protocol provided of hint servers is independent on specific P2P application vendors like Bittorrent, the council defines the protocol to be able to use any P2P application vendors. It needs to gather network information from ISPs to offer network distance to peers, however many ISPs dislike to disclose such information to others. Therefore, hint servers are designed to offer little information about ISPs' network architecture to P2P application vendors.

To monitor traffic of peers, the council also settled dummy node, which are described in section 3.1.

This memo describes the overview of the experiments.

3. The details of the experiments

The council has already learned that the server cost could be reduced with using P2P technology for contents delivering by investigating data offered by the members of the council. For example, the data brought by the vendors shows as follows:

> 90% of traffic was reduced with UG Live by Utagoe Inc [6].
>
> The costs of delivering to tens of thousand subscribers was reduced to 1/5 with BBbroadcast with TV Bank Corp. [7]

On the other hand, these reduced server costs may affect network load. One of the goals of our experiments is to visualize the impacts and propose an architecture to reduce network load caused by these new technologies.

To satisfy the above goals, the framework to be proposed should be well generalized as possible that doesn't rely specific P2P application behaviors because multi P2P application vendors join these experiments. In addition, the traffic should be captured beyond multi ISPs.

3.1. Dummy Node

As mentioned before, while the effect of delivery using P2P technology on reducing the traffic and the load on servers is well

known, traffic behavior in the inter-ISP is not known. In Japan, there is a backbone traffic report cooperated with ISPs and IXes [8]. However, this measurement requires to capture packets on subscribers line to know end user's activity. It is not realistic to measure the behavior of P2P applications at user terminals connected to the Internet because that would require a large-scale arrangement for measurement, such as using Deep Packet Inspection (DPI) on aggregated lines.

To solve these problems, we put several nodes called 'dummy nodes' in the ISP's networks. The dummy nodes emulate an end user's PC and P2P applications are running on the nodes. Every P2P node provided by participating vendors in the experiment was configured so it always contacted the hint server.

By introducing dummy nodes, we can observe and evaluate how much P2P applications have affected networks by measuring the traffic on dummy nodes. Since this method can't measure every subscriber's traffic, the accuracy would be less than other methods. But this make it possible to adapt to situations many different P2P applications coexist on a network. We can say this is suitable for these experiments.

A dummy node consists of Intel PC server, Linux(CentOS), VMWare and Windows XP works on VMWare. With this configuration, all packets can be captured without any impacts to the network, nodes and application behaviors. And it enable us to use different P2P applications for windows and evaluate them generally.

To see behaviors of the node, incoming and outgoing packets are captured on Linux because every packets are transmitted through it. In these experiments, we captured source/destination address, port number, amount of traffic and start/end time to see flow information.

60 Dummy nodes are put on access networks that are closest subscriber as possible in different 40 networks.

```
+----------------------+
|+--------------------+|
||+------------------+||
||| P2P Application  |||
|||    WindowsXP     |||
|||        +--+      |||
||+--------|N |------+||
||  VMware |e |       ||
|+---------|t |-------+|
|   Linux  |IF| capture|
+----------|  |--------+
           +--+
```

Dummy nodes

Figure 1

4. Hint Server ('08)

In Japan, bottleneck in IP networks have been shifting from access networks to backbone networks and equipments, such as bandwidth between ISPs and capacity in IXs, since FTTH has rapidly spread all over Japan. Under these circumstances, the Council proposed a less restrictive and more flexible cooperation between ISPs than existent P4P experiments [9]. The proposed method consists of the following elements: (1) P2P clients, (2) P2P control servers, and (3) a hint server: a peer selection hint server. (1) and (2) are existing systems but whether (2) exists depends on each application. (3) is a server that provides a hint as to the selection of a peer, and plays a role equivalent to that of ALTO Server. Note that this proposal was based on results of experiments using dummy nodes. The results showed that it was possible to reduce unnecessary traffic that flows across the boundaries of geographical districts or ISPs through providing information about the physical network to P2P applications.

When a peer joins the network, it registers its location information (IP address) and supplementary information (line speed, etc.) with&# 12288;the hint server. The hint server calculate network distance between peers (P2P client) based on network topology information obtained from the ISP and generates a priority table for peer selection. The hint server returns the table to the peer.

If all information can be made publicly, the above procedure can produce a result which is close to overall optimization. However, some information held by ISPs can often be confidential. Besides, in some cases, the volume of calculation required to process all information can be excessive. To avoid these problems, it is planned

to conduct experiments with a limited set of functions, analyze experiments results, and gradually expand the scope of optimization.

A control mechanism that makes use of all possible information is difficult not only technically but also difficulties to achieve coordination among providers. In consideration of these difficulties, the council has been limiting the implementation and experiments to the following scope since 2006.

Figure 2 shows an outline of the hint server.

```
+---------+   GetLocation    +-------------GeoIP DB Server---------+
|         |  +-----------+   |   +----------+      +-----------+   |
|         |--|IP Address |-->|   | GeoIP DB |      |Quagga etc |   |
|         |  +-----------+   |   +----------+      +-----------+   |
|         |                  | +-------------+  +----------------+ |
|         |  +-----------+   | |  District   |  |    Routing     | |
|         |--|AS Code:   |---| | information |  |information(DGP)| |
|         |  |Regional   |   | |             |  |                | |
|P2P Peers|  |Information|   | |   Range of  |  |AS Code(origin) | |
|   or    |  +-----------+   | | IP address  |  |                | |
| Contro| |                  | +-------------+  +----------------+ |
| Server  |                  +-----------------------------------+
|         |                                |      ^
|         |  PeerSelection                 v      |
|         |  +-----------+   +-------------------------------------+
|         |--|IP Address |-->| +--Priority Node Selection System--+ |
|         |  |    List   |   | |                                 | |
|         |  +-----------+   | |     Peer candidate ranking      | |
|         |  +-----------+   | |                                 | |
|         |--|  Ranking  |-->| +---------------------------------+ |
|         |  +-----------+   +-------------------------------------+
+---------+
```

Peer selection hint server

Figure 2

The network information used by the hint server is not information solicited from individual ISPs but the AS number and district information, which are more or less already public. Routing tables are not generated. Instead, peers within the same ISP or the same district are selected with higher priority in order to confine traffic to within the same ISP or the same district.

When the hint server receives an IP address, it returns its attribute information, to achieve the above. A peer can select a peer based on

the returned information. This operation is called GetLocation. However, in preparation for the time when it becomes necessary to hide topology information, an interface is provided through which a priority order is returned in response to an input of a list of candidate peers. This operation is called PeerSelection.

Although the target node is selected based on the criterion that it is within the same ISP or the same district, this type of selection is not very effective if the number of participating peers is small. Table 1 shows ratio of peers within the same AS or the same prefecture calculated from the distribution of ASs and prefectures in the IP address space from one-day data on a Winny network.

Conditions	ratio
AS matches	6.70%
Prefecture matches	12.76%
Both match	2.09%
Neither match	78.45%

Table 1: AS and prefecture distributions

Since, in addition to the above, the presence/absence of content affects the result, the control of selecting a peer within the same district may be inadequate. Therefore, it is necessary to introduce the weight of a continuous quantity that reflects the physical distance or the AS path length as an indicator of the proximity of the areas involved.

In consideration of the above, the following two measures are used for the evaluation of proximity between peers in a hint server.

- AS path length (distance between ISPs)

  AS path length calculated from BGP full routes. Since a full routing table retrieved at an ISP can show only a best path, it may not get an accurate length if the AS hop of both ISPs is too large. To avoid this, we use multiple BGP information received from different ISPs and combine them. Based on this concept, we used BGP routing information's offered by three ISPs operated by big telecommunication couriers and made a topology tree. Then it enables to calculate the shortest path between given two ASes.

- Geographical distance

Distances between peers are measured using physical distance of prefectural capitals that target peers belong to. The distance between prefectural capitals is used to calculate physical distance. Distances between prefectural capitals are sorted into ascending order, and then into bands, with weights 1 to 15 assigned to them so that there are a more or less equal number of "capital pairs" in each band. If either of their location is indefinite, distance is equal to 15 and, if they are in the same prefecture, distance is equal to 0.

Evaluation of distances between peers showed that the distribution of distances was almost uniform when distances between peers are normalized. This result suggests that using normalized distances expands the area where the control by a Hint Server is effective. The geographical distance is only used when the AS path length is same.

An example of the request and the response

o Request

```
POST /PeerSelection HTTP/1.1
Host: ServerName
User-Agent: ClientName
Content-Type: text/plain; charset=utf-8

v=Version number
[application=Application identifier]
ip=IP address of physical interface
port=Port number of physical interface
[nat={no|upnp|unknown}]
[nat_ip=Global IP address using UPnP]
[nat_port= Global port number using UPnP]
[trans_id=transcation ID]
[pt=Flag of port type]
[ub=upload bandwidth]
[db=download bandwidth]
```

o Response

```
HTTP/1.1 200 OK
Date: Timestamp
Content-Type: text/plain; charset=utf-8
Cache-control: max-age=max age
Connection: close

v=Version number
ttl=ttl
server=hint server name
...
trans_id=transaction ID
pt=Flag of port type
client_ip=Peer IP address observed from server
client_port=Peer port number observed from server
numpeers=number of respond peer
n=[src address] dst address / cost / option
```

5. High-Level Trial Results

5.1. Peer Selection with P2P

Table 2 shows the result of the analysis of communication in a node of an ISP installed in Tokyo, as an example of measurement results.

In these two experiments we evaluate different P2P applications, in 1st experiment, the P2P topology is generated by tree algorithm, and in 2nd experiment, it is generated by mesh algorithm. Both of them result in similar performance.

Conditions	Experiment 1	Experiment 2
*Peers selected within the same ISP	22%	29%
*Peers selected within the same district	19%	23%
*Peers selected within the same district and the same ISP	5%	7%

Table 2: Percentage of communication within the same ISP

The table shows that the probability of communication with peers in

the same ISP is proportional to the number of population and the share of the ISP in each district. The data show that peers were selected at random. Note that the vendor of a P2P application used in these experiments explained that the mechanism of selection a peer using network information can be implemented. However, peer selection is normally based on past information because users often cannot actually perceive the effect of using network information.

5.2. Peer Selection with the Hint Server

The main objective of these experiments was to verify the operations of the hint server and P2P applications. The distances between a dummy node and a peer were obtained from data on the dummy nodes. An examination of the distances between a dummy node and a peer revealed that mean value of distance after the hint server was introduced was reduced by 10% and that 95% value of that was reduced by 5%. The results show introducing hint server can reduce network loads by P2P applications.

6. Considerations

We clarified followings throughout our experiments.

1. Dispersed dummy nodes can figure out the behavior of peers and traffic between inter-ISP networks, which peers are selected by each peer. Therefore it proves that the importance of peer selection control mechanism proposed in ALTO.

2. Using our peer selection control mechanism, called hint server, could achieve significant differences. Our hint server can lead each peer to select nearer peer.

In the experimental result of peer selection control, it is smaller in intra-ISP traffic than other experiments [10] We think that it is because there are smaller peers in each area of traffic control. When there are many peers in one ISP, it is easy to select peers in the same ISP. However, when there are small peers in one ISP, it is difficult to select peers in the same ISP. In the situation of our experiments, there are many ISPs of peers belonging, and there are relatively smaller peers exist in same ISP.

Moreover, we didn't force P2P vendors to limit their implementation policy, therefore we can observe differences how each implementations weigh the information from the hint servers. Especially, in tree overlay topology P2P applications, such mechanism is very effective, on the other hand, in mesh overlay system, less effective.

6.1. Next steps

The experiments are on going as of 2011. Current experiments in 2011, we've changed the communication protocol to hint servers to ALTO based because it is nearly standardized. In our implementation, PIDs and the value of cost are mapped to ISP subnets, and ISP distance respectively. We also implement services for compatibility required by ALTO such as Service Capability and Map Services. But the Endpoint Cost Service is mainly used because of backward compatibility of our experiments.

We also study hierarchical hint server structure, in order to control in coarse inter-ISPs and in detail intra-ISP. It is also effective for limiting the area of information disclose.

6.2. Feedback to ALTO WG

This section describes what the authors learned with these experiments would be useful for the ALTO WG.

6.2.1. Hierarchical architecture for ALTO servers

In our experiments, we present the possibility of traffic control among multi-ISPs and multi-P2P applications using ALTO mechanism. On the other hand, we found several problems in ISP operations to adapt the mechanism. One is the granularity of network information. Among inter-ISP area, it is relatively easy to treat information for public purpose using BGP full route. On the other hand, among intra-ISP area, it may be difficult to disclose private information of each ISP. [11] propose some modification for ALTO protocol in order to hide ISP information. We propose hierarchical structures. From the viewpoint of cooperation between ISPs, fine-grained information is not necessarily required and moreover it is difficult to exchange the fine-grained information between ISPs. Considering this situation, the authors use only coarse-grained information to control backbone traffic in the experiments this year, though demand of controlling traffic within an ISP using fine-grained information will arise in the near future. Therefore it led us that introducing hierarchical structure into ALTO is necessary to cope with both situations. Actually, the authors plan to adapt a hierarchical control mechanism in the next steps, which include the following two steps.

- o In the first step, coarse-grained information about whole the network is used to select ISPs.

- o Next, fine-grained information within the ISP is used to select a peer.

6.2.2. Measurement mechanism

In the experiments, there were two difficulties as follows:

- o Evaluating effect of introducing a hint server was difficult, since P2P applications had their own measurement mechanisms.

- o How to treat priority orders of peers suggested by a hint server could not be predetermined for P2P applications.

From these experiences, the authors consider that clarifying requirements about measurement mechanisms for P2P applications are necessary also in ALTO.

7. Security Considerations

This document does not propose any kind of protocol, practice or standard.

8. IANA Considerations

No need to describe any request regarding number assignment.

9. Acknowledgments

Thanks to strong support by MIC (Ministry of Internal Affairs and Communications of Japanese government), the council was established. These experiments were performed under cooperation among P2P Network Experiment Council members, and DREAMBOAT co.,ltd., Bitmedia Inc., Utagoe. Inc. and Toyama IX have especially supported analyses of the experiments. The authors appreciate Tohru Asami, Hiroshi Esaki and Tatsuya Yamshita for their constructive comments.

10. Informative References

[1] "On the Quality of Triangle Inequality Violation Aware Routing Overlay Architecture", INFOCOM 2009: 2761-2765.

[2] "QRON: QoS-aware routing in overlay networks", IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 22, NO. 1, JANUARY 2004.

[3] "Winny on Wikipedia", <http://en.wikipedia.org/wiki/Winny>.

[4] "Share on Wikipedia",

<http://en.wikipedia.org/wiki/Share_(P2P)>.

[5] Taniwaki, "Broadband Competition Policy in Japan", 2008, <http://www.smartireland.jp/en/forum/may-2009/>.

[6] Utagoe Inc., "UGLive technology introduction", http://www.utagoe.com/en/technology/grid/live/index.html, March, 2011.

[7] TVBank, "Live Delivery using `BB Broadcast'Achieving 96% Saving in Traffic!", http:.wwww.tv-bank.com/jp/20081031.html, 2008 (in Japanese).

[8] Cho, Fukuda, Esaki, and Kato, "The Impact and Implications of the Growth in Residential User-to-User Traffic", SIGCOMM2006, pp207-218, Pisa, Italy, September 2006.

[9] Open P4P, "P4P Field Tests: Yale-Pando-Verizon", http://www.openp4p.net/front/, 2009.

[10] "RFC5632: Comcast's ISP Experiences in a Proactive Network Provider Participation for P2P (P4P) Technical Trial", September 2009.

[11] "ALTO H12,draft-kiesel-alto-h12-02 (work in progress)", March 2010.

Authors' Addresses

Satoshi Kamei
NTT Communications Corporation
Granpark Tower 17F, 3-4-1 Shibaura
Minato-ku, Tokyo 108-8118
JP

Phone: +81-50-3812-4697
Email: skame@nttv6.jp

Tsuyoshi Momose
Cisco Systems G.K.
9-7-1 Akasaka
Minato-ku, Tokyo 107-6227
JP

Phone: +81-3-6738-5154
Email: tmomose@cisco.com

Takeshi Inoue
NTT Communications
3-4-1, Shibaura
Minato-ku, Tokyo 108-8118
JP

Phone: +81-3-6733-7177
Email: inoue@jp.ntt.net

Tomohiro Nishitani
NTT Communications
1-1-6, Uchisaiwaicho
Chiyodaku, Tokyo 100-8019
JP

Phone: +81-50-3812-4742
Email: tomohiro.nishitani@ntt.com