Scientific
Research

A Topology-Aware Relay Lookup Scheme for P2P VoIP System

Xiuwu ZHANG[1], Ruifeng GUO[1,2], Weimin LEI[1,2], Wei ZHANG[1]

[1]*School of Computer Science and Technology, University of Science and Technology of China, Hefei, China*
[2]*Shenyang Institute of Computer Technology, Shenyang, China*
Email: *zhangxiuwu@sict.ac.cn,* grf@sict.ac.cn, leiwm@sict.ac.cn, zhangwei@sict.ac.cn

Abstract

Because of the best-effort service in Internet, direct routing path of Internet may not always meet the VoIP quality requirements. Thus, many researches proposed Peer-to-Peer VoIP systems such as SIP+P2P system, which uses relay node to relay RTP stream from the source node to the destination node and uses application-layer routing scheme to lookup the best relay nodes. The key of those systems is how to lookup the appropriate relay nodes, which we call relay lookup problem. This paper presents a novel peer relay lookup scheme based on SIP+P2P system. The main ideas are to organize the P2P network using a Cluster overlay and to use topology-aware to optimize relay selection. We introduce the mechanism in detail, and then evaluate this mechanism in NS2 network simulation environment. The results show that our scheme is scalable and can get high relay hit ratio, which confirm the feasibility of a real system. We also make comparison with traditional schemes and the results show that our scheme has good path quality.

1. Introduction

Unfortunately, the current Internet provides a Best-effort service, which cannot guarantee VoIP service qualities on the IP-layer sometimes. That is, sometimes direct routing path in the Internet may not always provide good path quality. For the good quality of service, it is important to select alternate paths between the source node and the destination node in the application layer. Thus, it has been popular to build overlay routing for VoIP application based on Peer-to-Peer (P2P) networks in the Internet [1–3].

Many previous research works [2,4,5] show that overlay routing can effectively improve VoIP session's performance by avoiding IP-layer's path failure. In application-layer routing scheme for VoIP systems, we should first lookup appropriate relay nodes in the application layer to bring an intermediate routing scheme. We call it the relay lookup problem. The definition of relay node is one which can not only give relay services for local users, but also serving as a bridge for remote nodes in the topology. The role of relay node is not only to support two/-

multiple node's communication (traverse a NAT or firewall), but also to avoid congestion or failure in IP-layer to improve the quality of communication capacity. Many researches [2,6] show that relaying RTP packet is an effective way to improve the quality of VoIP communication.

There are many application-layer routing projects which propose relay lookup scheme, such as MIT's resilient overlay network (RON) [4,5], Ohio state University's AS-Aware Peer-Relay Protocol (ASAP) [1], which all are to solve the Internet path failure through one-hop overlay routing. Besides these systems, there is one commercial software which is Skype [3,7,8]. As we know, Skype successfully use P2P technology and supports millions of online users, and the most important technique is the relay mechanism. Skype profits from the relay mechanism, but the fact is that, Skype encrypts its packets so that the exact routing methods are unknown. These defects hinder its further growth.

In our initial work, SIP+P2P system [2], we propose a P2P relay mechanism, which is to make those nodes with good network situation and machine performance become relay nodes, and to use a P2P overlay to organize those nodes. In the original system, we use a simple relay lookup scheme which we call one-side optimal selection

---

An early work of this paper will be presented in part at the 5th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2009, September, Beijing, China.

scheme. In one-side optimal selection, client nodes cache a list of relay nodes and keep heartbeat detecting. When need relays, the caller client node selects several nearest relay nodes in cache by means of RTT and negotiates with peer client to find proper relays. But this mechanism is a unilateral optimal method, not suitable for bilateral optimal requirement in VoIP communication.

Thus, how to design an effective scheme in SIP+P2P system to find the appropriate nodes is really a field worth researching. In this paper, we propose a novel peer relay lookup scheme based on cluster participation, which can be used in SIP+P2P system. The main idea is to use topology-aware to optimize the peer relay selection. We divide the whole P2P relay network into many clusters, and organize the whole P2P network using Cluster overlay. When one node joins the system, it will automatically upgrade its identity to be a landmark node or just be an ordinary node according to its own network situation and relative location. If it is upgraded to be a landmark node, it will be in charge of building clusters in the overlay network, and other ordinary nodes will join in the cluster. The landmark node is the responsible one of the cluster, and it will record the relay nodes in its cluster which have the ability to relay with other clusters. When the node in the cluster needs relay nodes during its communication, the landmark will give out a relay node candidate list, and the communication node chooses the right relay node from the list after testing and negotiating with peer node. In the aspect of cluster topology constructing, we use path quality as the measurement metric, which is quite different from ASAP. ASAP use AS information from BGP routing table to build cluster topology, which is unrealistic in large scale network and mobile network. Instead, our measurement metric of path quality also reflects the proximity relationship in topology distance, and is easy to realize and suitable for large scale network.

2. Related Work

The resilient overlay network (RON) [4,5] is a solution for general application-layer routing problems, which aims to provide optimized packet routing performance through routing packets in an overlay network. Using link state routing protocol, RON server detects all the links in regular cycle, thus it gets the update link state of the whole network, including the delay between any paired nodes. When there is a failure in the direct path, the nodes ask the server to find relay nodes. The server will check all the links around the source node and the destination node, and then return a proper node list which is near to both nodes. RON is only efficient for small network application in which the nodes number is below 50. Besides, RON is not specifically designed for VoIP system, thus, is not suitable for relay service in VoIP system.

Many researchers attempt to gain insights into the Sk-ype relay selection algorithm in a black-box manner [3,7, 8]. They capture the packets to study the protocol in Skype and analyze the key techniques according to these traces. From the perspective of specific technologies, Skype's application layer routing has several advantages, including quick routing, lower operating costs and network overhead. Then the system can greatly reduce the burden of the central server. Nevertheless, on the one hand, the relay selections do not take topology into consideration, so it can not always find proper relays. On the other hand, due to its private signal and encrypting packet transmission, we still do not know the detailed design of the system up to now.

Our previous work, SIP+P2P system [2], proposed a P2P relay mechanism to improve the VoIP quality. We select nodes which have good network situation and machine performance to become relays and provide relay service for VoIP sessions. Then we use a P2P overlay to organize those nodes. In our original system, we used a simple peer relay lookup scheme: all client nodes cache a list of relay nodes and keep alive; when a node needs relays to communicate, it will select several nearest relays and negotiate with the peer node; after detecting and negotiating, the two nodes find appropriate relay nodes for their communication. In order to optimize the selection procedure, client node evaluates relay nodes in its cache by means of *ping/pong* RTT value, and sorts them according to the evaluation result. We also classify all relay nodes into different ISP sets to enhance path diversity and reduce the cache space. For all that, this mechanism can not always find the near relays for both sides yet, because it is a unilateral optimal selection. But VoIP systems often require that the selected relays are near to both the source node and the destination node. Thus this relay lookup mechanism needs to be further improved.

ASAP (AS-Aware Peer-Relay Protocol) [1] is specially designed for VoIP systems and as it experienced, they got a good result for VoIP communication nodes. The main idea is to divide the whole network into AS domains. According to IP prefix, the system classifies all nodes that have joined the system into different ASs, which are called clusters. There are Bootstrap servers in the system in charge of the IP prefix of clusters and the overall AS graph. There are responsible nodes, also called proxy nodes, in every cluster. These proxy nodes are chosen from the clusters with better network state. They get the location of AS and the IP prefix table from the Bootstrap server, and then make a record of nodes in the cluster. When nodes join the system, they can find their cluster nodes based on its own IP, and let the cluster record. Also proxy node will detect the surround AS and record the near ones. When two communicating nodes need relays, they will ask proxy nodes for the near clusters' information and the two clusters' information will make an intersection and the nodes in the set will be the relay candidates. The two nodes will then detect the relay

candidates to find the appropriate relays. The key of this system is to find the IP prefix to give nodes basic judgment to join the right cluster. But getting the IP prefix from the BGP routing table is not easy to bring out for large scale network, and the real-time update one is even more difficult and needs lots of work, so the actual deployment of system in large scale network is unrealistic. But the system itself is great reference for relay lookup mechanism.

As introduced above, both Skype and SIP+P2P system have some limits: 1) their peer selection is suboptimal; 2) there are a large number of unnecessary probes. Those systems can be further improved by using topology-aware technology. ASAP uses AS-Aware to optimize relay selection, but it still has some limitations. Thus, we propose a novel peer relay lookup based on SIP+P2P system, which still uses a Peer-to-Peer relay mechanism. We organize the P2P relay network using Cluster overlay, and use topology-aware to optimize relay lookup. When build cluster overlay and select relay, we use path quality as its measurement metric, which is easier to realize in large scale network.

3. Design of Peer Relay Lookup Scheme

Our design still confirms to the framework of SIP+P2P system, in which the SIP signal is client/server based, but the media stream is transmitted in P2P relay network. We still adopt one-hop relay scheme, and use topology-aware technology to help selecting the proper relays. We divide the whole relay network into different clusters according to the path quality, which denotes the distance between two nodes. Thus, nodes which have good quality path (e.g. MOS>3.5) to each other are near in topology and will join in the same cluster. We use delay and packet loss to evaluate the path quality, because it is generally thought that the voice quality mainly relates to the delay and the packet loss among all network factors. When the one-way delay is lower than 200ms and the packet loss less than 2%, the voice quality is usually acceptable. In this cluster system, there are three kinds of nodes:

- Landmark node: Landmarks are chosen from ordinary nodes when they have good hardware condition and network surrounding. The landmark is the management center of the cluster.
- Relay node: Relay nodes provide relay service for other nodes. They usually have good network situation. When a node becomes a relay, it will record its address on landmark.
- Ordinary node: Ordinary nodes will join the cluster by being recorded on the landmark and maintain the states with periodic heartbeat method.

Relay nodes and ordinary nodes are all cluster nodes. Sometimes, a landmark node may also be a relay node. When communication nodes need relaying, they will ask

the recording landmark for proper relay nodes candidates, and then detect the relay to see if it can be the relay node.

3.1. System Architecture

As shown in Figure 1 and Figure 2, the system is composed of some Bootstrap Servers, a cluster layer and a logical relay layer. The Bootstrap Server is only server-like component in the system which needs to be deployed previously. It is responsible for node login, landmark management, and maintains a bootstrap's data structure, including: a cluster graph, a cluster landmark IP table. Cluster layer is used to organize nodes and maintain the real topology. Cluster layer is formed with several clusters. In each cluster, nodes are tagged as landmarks or cluster nodes. Each cluster node is attached to only one landmark. In normal cases, the quality of path between a node and its landmark is very good. The landmark node maintains its near clusters to form the topology, as show in Figure 1. Relay logical layer is used to help nodes in cluster to find appropriate relay nodes. Relay layer is a virtual layer with all nodes mapped from cluster layer; the connection relationship is kept the same as the one in cluster layer, simply adding some relay related information on each node to help find relay.
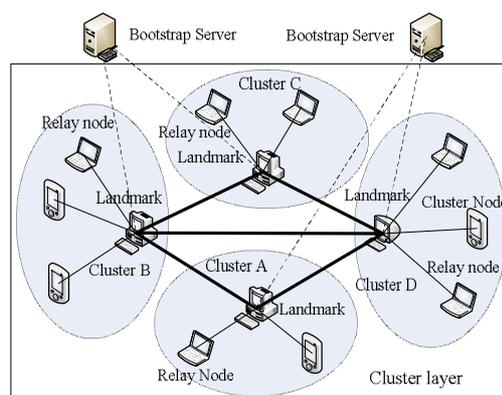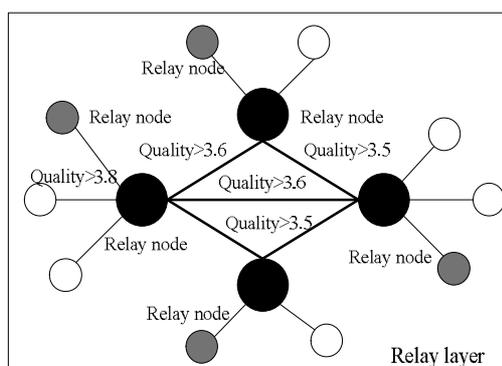

Figure 1. Deployment of system architecture.

Figure 2. A relay logical layer mapped from the cluster-layer.

3.2. Module Design on the Peer

Figure 3 shows the module design on one peer. There are two main modules, including cluster management module and relay module. There are two parts in cluster management module: the landmark logic and the cluster node logic. And relay module also consists of the relay server logic and the relay client logic.

Cluster management module: Its main function is to maintain the cluster topology. The landmark logic maintains a neighbor landmark set, a cluster node set and a relay node set. Landmark logic is opened when this node becomes a landmark. The cluster node logic maintains its landmark set and a standby landmark set. There will be one candidate landmark which is decided by landmark competing. All nodes keep heartbeat communications with other nodes, and will help them to find the relay nodes candidates. The node will change the landmark set according to the dynamic network states.

Relay module: The main function is to provide relay service for upper SIP sessions. Relay server logic is an independent function. When a cluster node can become a relay, it will open this function and record its address on the landmark. Relay client logic mainly helps to find candidate relay nodes by detection. At the same time, it will cache a local relay list which has been found to be the proper relays. But the relay client will first ask the cluster management module to find proper relay nodes; if it is failed, it will then use the local relays.

3.3. Node Operations

Node joining: There is a fixed address list of the bootstrap servers stored in every client. The cluster node will send a *RTT* detection request to the bootstrap server and receive the response when it logins. Then according to the *RTT*, the normal node requests a list of landmarks from the nearest bootstrap server. After got a list of landmarks, this node detects them and finds out the nearest landmark according to the path quality. We set a threshold when judging whether a node belongs to a cluster. If

there is no corresponding cluster landmark, this node will be automatically promoted to be a landmark to form a new cluster, and then will get a cluster ID from the bootstrap server which is used to represent the new cluster.

The maintenance of topology: After a node joins a cluster, it will request near landmark list from its cluster landmark and detect them. If there is a landmark that is nearer than the current one, this node will join the new cluster by attaching to that landmark. This situation is common in mobile network because of node migration. In addition, each landmark stores a list of nearby ones. At the same time these landmarks will exchange their stored nearby landmark lists with each other so that the new joined landmark can be broadcasted in the network in time. Meanwhile, the landmark will also notify messages of new landmarks to its cluster nodes simultaneously. Then the nodes in the cluster will check their own condition and do some adjustments.

The maintenance of cluster node: the landmark maintains heartbeats with its cluster nodes and also with the relay nodes it stores.

Collecting information of relay: when a normal node joins the network, it will check out whether it could be a relay or not. If a relay node knows more than two landmarks whose path qualities (from itself to one landmark) are greater than a constant Q_0 value, it will notify the landmark. The landmark will also send back the list of previous landmarks addresses. The structure is called relay-cluster data structure in the system.

The initial communication: when two nodes communicate and need relay, they will first test the quality of the direct connection. And then, they will begin new inquiry for relays. Our SIP+P2P system supports multiple paths in one session, and uses optimizing routing [2,6] to improve QoS performance.

Inquiry for Relay: a normal node will ask landmark for candidate relay nodes by notifying the destination node's cluster, otherwise it will find candidate relay nodes locally. The candidate relay nodes are found in Relay cluster date structure. The landmark first searches the relay-cluster data structure for available relays of the corresponding destination cluster's relay nodes set and then sends the result to the source peer for detecting. If there are no proper relay nodes at last, the source node will ask for the near cluster set (which is got from the neighbor landmark set) and make an intersection with the destination node's near cluster set. Then it asks the cluster's landmark in the cluster set for relay nodes, makes them for the initial candidate relay nodes, and begins the detecting. If both of the two methods fail, the source node will randomly search nodes in its neighbors set for candidate relay nodes. Of course, this is the worst condition.

Building the communication: whether the source node gets the candidate relay list from local or from the landmark, these relays must be negotiated with the destination node, and then be detected. After the negotiation and
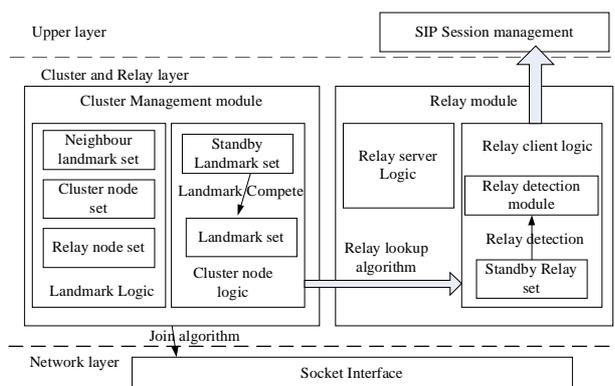

Figure 3. Module design.

detection, those suitable relay nodes will provide the application layer routing service.

3.4. Summary

Our system is a typical distributed system. As depicted above, the bootstrap server in our method is similar to a central server, and the landmark is similar to local server in CDN network. The cluster node should register and log off at the landmark and maintain a certain link with the landmark through heartbeat packages. All the landmarks make up a coverage network through the network topology, and maintain the cluster topology by heartbeat detecting. The landmark node and the cluster node compose a relation of tree. Landmarks exchange information with each other so that they could keep a close watch on the change of the network. And the landmark also maintains a list of relay of its autonomy cluster. The whole cluster overlay is similar to the Internet topology. The communication nodes request relays from its landmark and do the detection. If it fails, it will get the intersection of landmarks of the source node and the destination node. These landmarks and the relays in their clusters will be the candidate relays. This will reduce the load of the
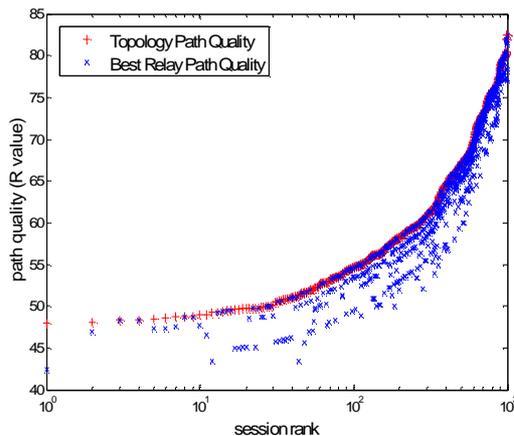

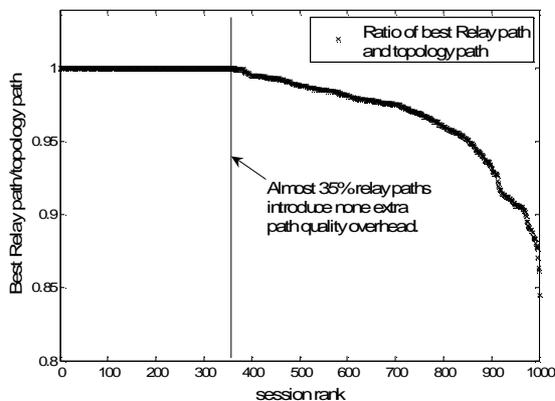**Figure 4. Topology path quality vs best relay path quality.**

Figure 5. Distribution of relative quality overhead.

bootstrap server greatly and make the network very scalable. The node detection and relay detection are finished when building the coverage network topology, so the cost is shared with the process of building the coverage network topology. And this mechanism has a high hit ratio in the simulation system.

4. Experiments and Analyses

In this section, we will first introduce our experiment environments, including the simulation platform and the topology generator. Then we will introduce our evaluate methods. We mainly consider the performance of cluster overlay and the efficiency of relay lookup algorithm.

We evaluate our cluster overlay mainly from two aspects:

1) Firstly, we evaluate the similarity degree of topology between our cluster overlay and the Internet.

2) Then, we build cluster overlay in different network scales to evaluate its scalability.

The simulation results show that our cluster overlay topology is similar to Internet and has high scalability.

When evaluate the performance of relay lookup scheme, we mainly consider two evaluations below:

1) Hit ratio in different network scales.

2) Comparing path quality of our scheme with original SIP+P2P system and a simple random selection.

The simulation results show that our scheme has higher hit ratio and performs better than other schemes.

4.1. Experiment Environments

We simulate our design in the NS2 simulation platform. There are three main steps in our experiment:

1) Firstly, we use a topology generator tool to generate network topology which is as same as Internet.

2) Secondly, we build our cluster overlay based on this network topology.

3) Finally, we run relay lookup algorithms on our cluster overlay, and evaluate its performance.

In the NS2 environment, there is a tool to build network topology called GT-ITM [9]. We adopt the network model of random connecting of Transit AS with Stub AS (Transit-Stub model), which is most similar to the real network. With different sets of delay and packet loss between nodes (read from a static configure file), we guarantee that the topology is the same in different experiments for a given network scale. We build cluster overlay networks in different network scales to evaluate the scalability, and compare node pair's quality in cluster overlay with quality in original topology. After building cluster overlay, we run different relay lookup scheme to evaluate the performance of our topology aware algorithm.

4.2. Similarity Degree of Topology

We define relative quality ratio as the quality ratio of best relay path and topology path. When the relative qua-

lity ratio is higher, our cluster overlay topology is similar to Internet. We select 1000 node pairs and compute there topology path qualities and the best relay path qualities, which are plot in Figure 4.

Then, we compute the relative quality ratio, and plot its distribution below. As we can see in Figure 5, almost 35% relay paths introduce none extra quality overhead and 90% of relative quality ratio are above 0.9. These results prove our cluster overlay is very similar with real Internet topology.

4.3. Scalability of Cluster Topology

The scalability of this system requires that when the network scale increases greatly, the cluster scale does not increase greatly. The scalability is very important to our cluster overlay, because when the cluster overlay has high scalability, the maintenance cost of cluster topology does not increase greatly according to the network scale. As show in Figure 6, if the network scale increases, the number of landmarks and cluster nodes is increasing as well. But, the number of cluster nodes is increasing greatly and the number of landmark has no clear changes. This result proves that our scheme has good scalability.

4.4. Hit Ratio of the Best Relay Nodes

The object of this scheme is to find the best relays in P2P relay network which are near to both the source node and the destination node. This proportion of best relay found in this scheme is called hit ratio. In this experiment, we compare the relay nodes resulted from simulation with those resulted from calculation according to topology graph, and get a hit ratio. As show in Figure 7, we compare the hit ratio in different network scales for SIP+P2P and our new scheme. The results show that our scheme has better hit ratio than original SIP+P2P system. The reason is that our new scheme uses topology-aware to help look up the best relay.

4.5. Path Quality of Peer Relay Routing

In this section, we evaluate the QoS performance of our new scheme. We choose a network with 3000 AS nodes for simulation, and randomly select 10^4 pairs of nodes sessions need relay. We only select those node pairs which are in different ASs, and whose path qualities are bad. Then we use different scheme to select relays and compare their path quality (Random selection, One-side optimal selection, Topology Aware, and Optimal search calculated according to topology graph):

As we can see in Figure 8, our new scheme performs better than One-side optimal and random selection, and is very close to optimal search. However, when using topology-aware lookup algorithm, there are 8% sessions can not find relay. As mentioned above, when a client lookup relay failed, it will use a relay in local cache.
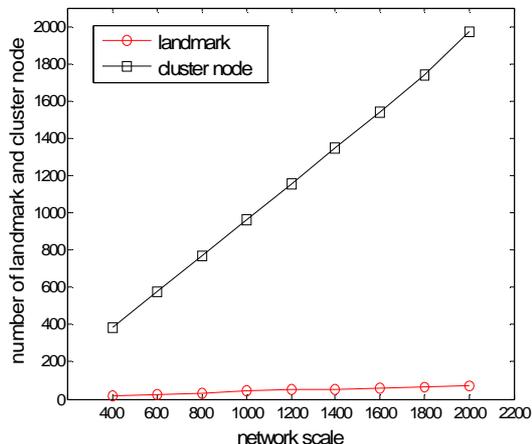

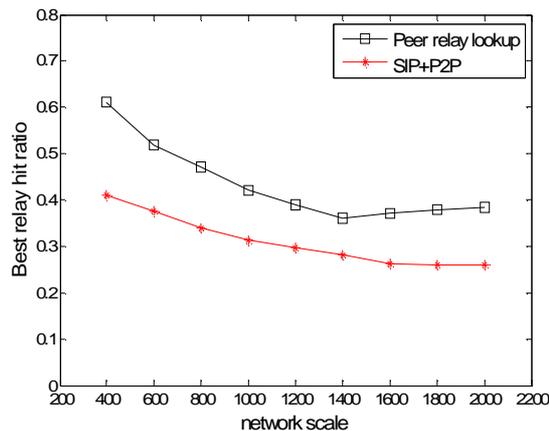**Figure 6. The number of cluster node and landmark.**

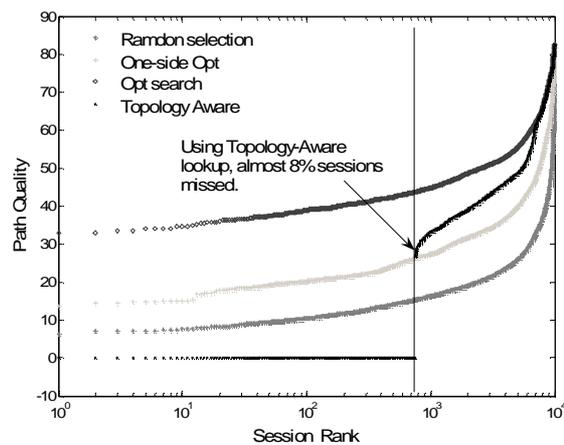**Figure 7. Hit ratio of the best relay.**

Figure 8. Path quality of peer relay routing

5. Conclusions

In this paper, we propose a novel peer relay lookup scheme based on SIP+P2P system to improve its relay selection. This approach uses topology-aware technology to optimize the relay selection. First, we divide the whole

P2P relay network into different clusters. When building the cluster topology, we use path quality between two nodes as the measurement metric. Then, when communicating node needs relay, it asks its landmark for proper relays and the cluster landmark will choose candidate relays according to the network topology. This new scheme can guarantee that the selected relay is near to both the source node and the destination node, which is better than the original SIP+P2P system.

We introduce the whole design in detail, including the system architecture, the peer design, and the node operation procedure. Then we evaluate our design in NS2 simulation, and the results show that our scheme is scalable. We also make comparison with traditional schemes and the results show that our scheme can get high best relay hit ratio and better path quality. Those experimental results confirm the feasibility of a real system.

Our future work is to realize this scheme in our SIP+P2P UA [2], and to evaluate it in a real network environment. Moreover, to modify this scheme to make it suitable for mobile network is another future work.

6. References

[1] S. Ren, L. Guo, and X. Zhang, "ASAP: An as-aware peer-relay protocol for high quality VoIP," Proceedings of 2006 IEEE ICDCS, pp. 70–79, 2006.

[2] X. W. Zhang, W. M. Lei, and W. Zhang, "Using P2P network to transmit media stream in SIP-based system," Proceedings of ICYCS 2008, pp. 362–367, November 2008.

[3] S. Guha and N. Daswani, "An experimental study of the skype peer-to-peer VoIP system," Proceedings of 2006 IPTPS, pp. 677–686, 2006.

[4] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "Resilient overlay networks," ACM SIGCOMM Computer Communication Review, Vol. 32, No. 1, pp. 75–83, 2001.

[5] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "The case for resilient overlay networks," Proceedings of 2001 Workshop on Hot Topics in Operating Systems, pp. 131–143, 2001.

[6] S. Tao, K. Xu, A. Estepa, T. Fei, *et al.*, "Improving VoIP quality through path switching," Proceedings of 2005 IEEE INFOCOM, pp. 515–523, 2005.

[7] K. Suh, F. Daniel, K. Jim, and D. Towsley, "Characterizing and detecting skype-relayed traffic," Proceedings of 2006 IEEE INFOCOM, pp. 292–301, 2006.

[8] H. Xie and Y. Yang, "A measurementbased study of the skype peer-to-peer VoIP performance," Proceedings of 2007 IPTPS, 2007.

[9] K. Calvert and E. W. Zegura, "GT-ITM: Georgia tech internetwork topology models," In http://www.cc.gatech.edu/projects/gtitm/, 1997.