SIP 终端中拓扑感知的 P2P 中继查找机制实现①

陈培培[1,2] 雷为民[1,2] 王 宁[1,2] 张秀武[2] (1.中国科学院沈阳计算技术研究所 辽宁 沈阳 110171; 2.中国科学技术大学 计算机科学与技术学院 安徽 合肥 230027)

摘 要: 由于 Internet 采用尽力而为的服务，VoIP 系统存在 QoS 问题。为改善 QoS,多数系统采用“应用层路由”方案，该方案关键是如何查找拓扑最优中继节点。结合 P2P 技术，提出一种拓扑感知的 P2P 中继查找机制。首先构建一个 Cluster Overlay 网络来模拟真实的 Internet 拓扑;在 Cluster Overlay 网络上设计一种拓扑感知的中继查找算法;最后在 SIP 终端中实现这种方案。在 NS2 平台上对 Cluster Overlay 网络和中继查找机制进行了仿真实验。结果表明:Cluster Overlay 网络和真实的 Internet 拓扑相似度很高，中继查找算法能查找到最优中继结点，从而改进了路径质量。

关键词: 应用层路由; P2P 中继; 拓扑感知; Cluster Overlay 网络; SIP 终端

Achieve of Topology-Aware P2P Relay Search Mechanism in SIP Terminal

CHEN Pei-Pei[1,2], LEI Wei-Min[1,2], WANG Ning[1,2], ZHANG Xiu-Wu[2]

(1. Shenyang Institute of Computer Technology, Chinese Academy of Sciences, Shenyang 110171, China; 2.University of Science and Technology of China, Heifei 230027, China)

Abstract: Because of the best-effort service of the Internet, the VoIP system has QoS problem. In order to improve QoS, most systems use application-layer routing mechanism. One of the key issues of the mechanism is how to search for relay node. Combining the P2P technology, this paper proposes a topology-aware based P2P relay search mechanism. First, in order to simulate the real Internet topology, a Cluster Overlay network was built. In the Cluster Overlay network, a topology-aware relay search algorithm was designed. Finally, this program was achieved in the SIP terminal. A simulation on the Cluster Overlay network and relay search mechanism was carried out on the NS2 platform. The results show that the similarity between the Cluster Overlay network and the real Internet topology is very high. The relay search algorithm can find the optimal relay node, thus to improve the quality of the path.

Keywords: application-layer routing; P2P relay; topology-aware; cluster overlay network; SIP terminal

1 引言

SIP[1]是当今主流多媒体通信应用层控制协议，由于其易于扩展、便于实现等诸多优点而得到了广泛应用。SIP 网络采用 C/S 网络架构，各用户之间通信信令和媒体流几乎都需服器来路由。P2P[2,3]网络采用的是分布式架构，它拥有高扩展性、健壮性和高容错性的特点，P2P 网络中各个节点都是对等的通信实体。Skype[4]是 P2P 结构 VoIP 的典型代表，自推出以来深受用户欢迎。

由于 SIP 网络的 C/S 架构，因此 SIP 系统存在“性能瓶颈”; 此外，Internet 的采用尽力而为的服务，使得 VoIP 系统存在 QoS 问题。为解决 Qos 问题，结合 SIP 系统和 P2P 系统特点，人们提出了 SIP P2P[3,5]架构。在这种架构中，提出了“应用层路由”解决方案。该方案在传统的 Internet 网络拓扑上构建一个 Overlay 网络[3]，在 Overlay 网络中查找一个或多个

① 收稿时间:2009-12-17;收到修改稿时间:2010-01-17

节点来充当中继节点，然后，通信的两个端点可以借助中继节点来转发他们的媒体流。目前应用层路由研究方案主要有：俄亥俄州立大学的 ASAP[6](an AS-Aware Peer-Relay Protocol for High Quality VoIP)，华盛顿大学的 SOSR[7](scalable one-hop source routing)，哈佛大学的 PPRR[8](path probing relay routing)，以及华中科技大学的 DPLM[9] scheme(A Novel Application-layer Routing Scheme for Low Delay VoIP)，这几种方案都提到中继节点查找机制。

2 相关工作

目前各种应用层路由研究方案中大多都提到用中继节点来提高网络服务质量。ASAP[6]方案是从 BGP 网关处获取主机信息，然后根据这些主机 IP 地址前缀将这些主机组织起来，IP 地址前缀相同的主机结合到一起组成一个小的主机集合，我们把这样的一个小的主机集合称为一个 Cluster。然后，从每个 Cluster 中随机选择一个节点作为本 Cluster 的代表，然后测出每对代表节点之间的时延。在 ASAP 系统中存在三种类型节点：①Boot-straps,这种节点主要是在系统刚启动时候，受理用户的登录和加入 Cluster 请求。②Cluster 代理节点,这些节点是本 Cluster 当中带宽很大，功能强大的节点。③普通的主机，他们是组成当前 Cluster 的主要成员。ASAP 通过组建 Cluster，并在各个 Cluster 当中指定代理节点，当 Cluster 中主机节点需要和其他主机节点通信时，首先向本 Cluster 代表发出请求，Cluster 代表根据目的端节点所在 Cluster 来选择合适的代理节点。ASAP 在组建 Cluster 时候是根据 IP 前缀来划分的，但是实际上很多主机具有相同的 IP 前缀，但是他们并不在同一个网络中，并且他们之间甚至有可能是不可达的，这将无法保证通过中继节点链路质量状况。

SOSR[7] 方案提供了一个简单可度量的从 Internet 路径失效中恢复过来的途径。它根据发往目的主机的端口的 TCP ACK 包以及目的主机相应的 TCP RST 包来判断通信的双方路径质量情况。该方案试图研究路径失效的特征，找出失效处的位置所在，然后选取一个或多个的主机作为中继节点，然后试图经过这些中继节点来恢复 Internet 路径失效，重新建立通信的源端和目的端的链路。该方案存在的问题是它并不能准确判断何时链路失效，而且在链路失效后只是随机选取一些节点充当中继，这无法保证链路的质量，有时通过这些选取的中继节点可能无法恢复失效的链路。

DPLM[9] scheme 构建 Overlay 网络方法是这样的，任何两个节点均探测出他们之间的时延，然后根据时延的大小将节点组织起来，时延比较小的这些节点会被组织到一起构成一个 Cluster，然后每个 Cluster 中选取几个节点充当中继节点，为 Cluster 中其他节点提供中继服务。这个方案存在的问题是：它只以时延来衡量两个端点之间的链路的好坏，忽略了丢包率及其它一些因素的影响，尽管有时两个端点的时延很小，但他们之间丢包率也会影响链路质量，这对于有些业务来说是不能接受的。

在综合考虑到各种因素和情况，我们提出：基于 SIP 的拓扑感知的 P2P 中继查找方案。在我们的方案中，依据 SIP 网络中的拓扑情况，根据探测出的每对节点之间的时延和丢包率，我们计算出每对节点之间的链路路径质量。路径质量计算公式为：

$$R=94.2-I_d-I_e \quad (1)$$

我们假设 SIP 终端中采用编解码(codec)为 G729a。其中：

$$I_e(G729a)=11+40.\times\ln(1+10e) \quad (2)$$

此处 e 为丢包率，包括网络丢包率和播放缓冲丢包率，我们暂不考虑播放缓冲引起的丢包率和时延。

$$I_d=-2.468\times10^{-14}\times d^6+5.062\times10^{-11}\times d^5-3.093\times10^{-8}\times d^4+1.344\times10^{-5}\times d^3-0.001802\times d^2+0.103d-0.1698 \quad (3)$$

此处：

$$d=D_{network}+D_{codec}+D_{playout} \quad (4)$$

我们暂不考虑 $D_{playout}$，对于 G729a，D_{codec} 为 35 ms。计算出每对节点之间的链路路径质量后。根据路径质量，我们将节点中路径质量比较好的那些节点组织到一个 Cluster 当中，最后这些节点都被组织到不同的 Cluster 当中。

3 总体方案设计

在这一部分介绍下我们组建的 Overlay 网络的总体的架构。在我们的系统中，通过构建 Overlay 网络，将各个节点组织起来，形成一个个 Cluster。轻量级

的 SIP 信令仍然通过 SIP 服务器来转发。在两个终端进行 SIP 信令交互时，双方一旦明确表示都支持 P2P 协议(P2P 协议是我们自己定义的)，那么接下来这两个终端将采用 P2P 协议进行交互。图 1 为我们的构建的 Overlay 网络的系统架构图。

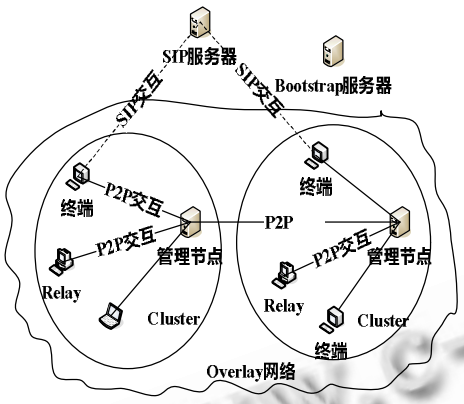

图 1 SIP P2P 网络系统架构图

下面介绍下系统中各个元素：

Relay 节点：主要是具备一定的能力（带宽，内存，CPU），可以为其它节点提供媒体中继服务的节点。

终端：运行 SIP 终端的计算机。该计算机在不满足条件时，SIP 终端中的 P2P 模块只开启 Client 子逻辑，该计算机成为 Overlay 网络中客户端节点；当满足条件时，则自举成为 Relay 节点。此时，该节点应同时运行 P2P 模块中的 Client 逻辑和 Relay Server 逻辑。此外，它还可以参照 Cluster 管理节点标准，申请 Cluster 管理节点功能，如果被批准成为 Cluster 管理节点，那么 P2P 模块除了开启 Client 逻辑和 Relay Server 逻辑外，它还要开启 Cluster Manager Server 逻辑。

管理节点：负责管理本 Cluster 当中其它终端节点信息，记录本 Cluster 终端节点和 Relay 节点信息。

Cluster：按照一定的标准将 SIP 网络中的运行 SIP 终端的主机节点划分到一个个的小的组织当中，这样的一个小的组织成为一个 Cluster，每个 Cluster 均有一个标识 ID，每个 Cluster 中的节点类型主要有：Relay 节点，终端节点，管理节点。

Bootstrap 服务器：它主要负责登记网络中划分的各个 Cluster 中的管理节点的信息，此外，它还负责为终端节点加入 Cluster 提供辅助信息。

SIP 服务器：它主要负责终端在起始阶段进行的 SIP 信令交互。

4 终端实现

本部分给出我们的 SIP 终端结构，如图 2 所示。SIP 终端主要组成部分有：用户接口，会话管理，SIP Stack，P2P 模块，MediaStack 等。其中 P2P 模块包括 3 个子逻辑：

①Client 逻辑：主要负责节点初始化，节点自我评价与自举，发出 Relay 请求，Relay 节点探测与收集。

②Relay Server 逻辑：主要负责媒体流处理，端口信息维护，处理中继服务请求。

③Cluster Manager Server 逻辑：维护本 Cluster 节点信息，处理本 Cluster 内各个节点请求，负责 Cluster Manager 选举，处理 Cluster 之间事务。

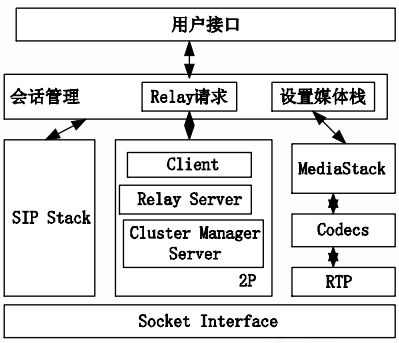

图 2 SIP 终端结构图

5 Overlay网络的组建

为更好的模拟现实世界当中 Internet 拓扑，我们采用 GT-ITM 生成的 Transit-Stub[10]形网络作为构建 Overlay 网络的原始拓扑。在我们的原始拓扑图中，不同的两个顶级域的节点之间：时延设置为满足正态分布 $N(80,20^2)$，丢包率满足正态分布 $N(2,1)$。同一个顶级域中 T-T 节点时延满足正态分布 $N(50,10^2)$，丢包率满足正态分布 $N(1,0.5^2)$。同一个顶级域中有边相连的 T-S 节点之间时延满足正态分布 $N(20,5^2)$，丢包率满足正态分布 $N(0.2,0.1^2)$。同一个顶级域中有边相连的 S-S 节点之间时延设置为满足正态分布 $N(10,2^2)$，丢包率为 0。

首先，需要确定终端节点到它即将加入的 Cluster 管理节点之间路径质量。如 图 3 所示。在同一个顶级域中，我们假定，Cluster 管理节点和在本 Cluster 中的终端节点之间最小路径质量为：关联在同一个 T 节点上的两个终端 A,B，其中终端 C,D 分别和终端 A,

B 相连，但终端 C，D 之间无边相连。我们把从 D 到 B，B 到 T1 以及 T1 到 A，A 到 C 的这条链路 C-A-T1-B-D 的路径质量定为终端节点和它即将加入的 Cluster 的管理节点之间最小路径质量。

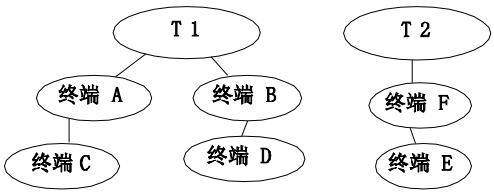

图 3　确定构建 Cluster 各个参数

其次，我们要确定两个 Cluster 成为邻居 Cluster 所应满足条件。如 图 3 所示。同一个顶级域中两个 Cluster 管理节点之间的路径质量定为：由终端 C 到终端 A，由终端 A 到 T1,T1 到 T2,T2 到终端 F，终端 F 到终端 E 的链路 C-A-T1-T2-F-E 的路径质量(注意：T1 和 T2 处于同一个顶级域中)。如果两个 Cluster 满足：它们的管理节点之间的链路路径质量不小于上述由 C 到 A，A 到 T1,T1 到 T2,T2 到 F，F 到 E 的链路 C-A-T1-T2-F-E 的路径质量(注意：T1 和 T2 处于同一个顶级域中)，我们就称这两个 Cluster 互为邻居 Cluster。

最后，确定不同的两个顶级域之间的两个节点之间的路径质量为 C-A-T1-T2-F-E 这条链路的路径质量（注意：此时 T1 和 T2 处于不同的顶级域中）。

接下来介绍 Overlay 网络组建过程。当一个新的终端要求加入 Overlay 网络时候，它首先探测距离自己最近的那个 Bootstrap 服务器，然后向 Bootstrap 服务器发出请求消息（消息中附带该终端的 IP 端口号码）。由于 Bootstrap 服务器保存着各个 Cluster 信息，当它收到加入请求时，将目前自己保存的各个 Cluster 信息发送给该终端。然后这个终端，向每个管理节点发送一组探测消息（共十个），测量出它到每个管理节点的网络时延和丢包率，根据这些测量参数计算它到各个管理节点的路径质量，然后选择一个和本终端路径质量最好的 Cluster，最后加入这个 Cluster。若这个节点发现，它到当前的各个 Cluster 的路径质量均达不到终端到管理节点的最低路径质量要求，或者当前还没有任何 Cluster 存在，那么它就会向 Bootstrap 申请建立一个新的 Cluster。

6　中继节点的查找

如图 5 所示。当某两个终端需要通信时，首先通过 SIP 消息进行交互，确认双方都支持 P2P 协议，然后获取对方的 IP 信息和 Cluster 信息，此处 Cluster 信息主要指双方所在的 Cluster 标识 ID 号。然后，发起端和目的端分别向自己的 Cluster 管理节点发出 Relay 请求,管理节点通过运行 Relay 查找算法，返回一个 Relay list。发起端和目的端在收到 Relay list 后，分别探测出二者到 Relay list 中各个节点时延和丢包率，然后目的端将探测出信息发给发起端，发起端综合考虑二者到 Relay list 中每一 Relay 时延和丢包率，计算出二者经由该 Relay 路径质量。最后选择一个 Relay 节点，使得发起端和目的端经由这个 Relay 的链路路径质量为最优。下面给出运行在管理节点上的进行最优 Relay 选择的算法。

```
Relay_search( s,d)
{  //s,d分别为通信发起端S 和目的端D 所在Cluster标识号
   //管理节点上保存信息有：1.本Cluster的中继节点信息列
   //表，每个Relay节点根据其当前负载状况及其本
   //身能力有一个评价值。
   //2.本Cluster的邻居Cluster的标识号和其管理节点IP地址
   //信息列表Neighbor list。
   初始化：Relay list为空；
   if    s等于d      then{
       评估各个Relay节点，返回一个最优Relay list；
     }
   else{
     if d在Neighbor list 中 then{
       向D所在Cluster管理节点M发出Relay请求；
       M回送一个评估后的Relay list1；
        将Relay list1并入Relay list；
       评估本Cluster Relay信息，形成Relay list2；
       将Relay list2并入Relay list;
        评估Relay list，形成最终的Relay list;
       返回最终的Relay list;
       }
      else {
         for 每个Cluster i∈Neighbor list do
          {
            向Cluster i管理节点M发出Relay请求；
             if d在i的Neighbor list中 then{
             M评估自己Relay节点，回送一个评估后的Relay
                          list i;
               将Relay list i并入Relay list；}
               else{ 回送d不在Cluster i的
                        Neighbor list中；}
          }
        评估Relay list,形成最终的Relay list;
        返回这个最终的 Relay list;
         }
       }
}
```

图 4　Relay 查找算法

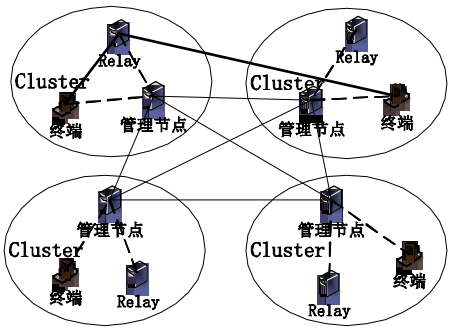

图 5　中继节点查找示意图

7　试验结果与分析

首先，评价我们构建的 Overlay 网络和原始网络

拓扑之间的相似度。衡量的标准是：每对节点在组建 Overlay 网络中的路径质量比上该对节点在原始网络拓扑中的路径质量。二者的相似度越高，说明我们组建的 Overlay 网络越能真实的反映原始的网络拓扑结构。

我们采用 GT-ITM 生成的 Transit-Stub 形原始网络拓扑中节点规模为 1616 个,顶级域 4 个，每个顶级域节点数目约为 400 个，T 节点数为 4 个。由原始拓扑构建的 Overlay 网络中：每个顶级域中 Cluster 大概有 4 个，每个 Cluster 中节点数目约为 100，总的 Relay 节点数为 100 个。然后随机找出 700 对节点，分别计算出它们在原始网络拓扑中的路径质量以及在构建好的 Overlay 网络中的路径质量，统计结果如 图 6 所示，然后将每对节点在构建的 Overlay 网络中的路径质量与原始网络拓扑中的路径质量之比绘制成曲线如图 7 所示。由 图 7 可以看出，对于绝大部分节点来说，二者的比值都大于 0.7。这说明我们构建的 Overlay 网络能够很好的反映原始网络拓扑的结构。

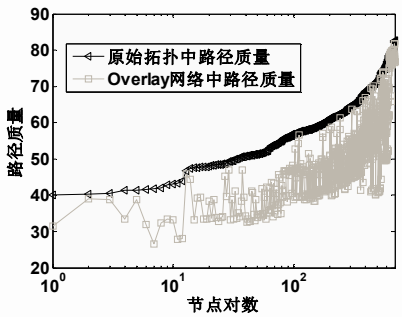

图 6 Overlay 网络与原始拓扑路径质量

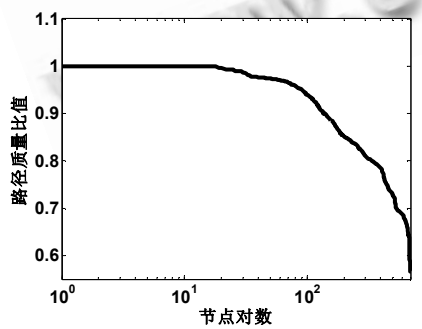

图 7 路径质量之比

接下来分析我们的方案中继服务方面的优势。为更好的反映我们构建的 Overlay 网络在为通信两个端点提供中继服务的性能，我们特地做了两次试验。第一个实验中组建的 Transit-Stub 形原始网络拓扑中节点规模为 1616 个,顶级域 4 个,每个顶级域节点数目约为 400 个，T 节点数为 4 个。由原始拓扑构建的 Overlay 网络中：每个顶级域中 Cluster 大概有 4 个，每个 Cluster 中节点数目约为 100，总的 Relay 节点数为 100 个。第二个实验节点规模为 9090,顶级域大概 9 个,每个顶级域节点数目为 1010 个，T 节点数为 10 个，每个顶级域 Cluster 大概 10 个。每个 Cluster 中节点数目为 101， 总的 Relay 节点数 400 个。然后我们随机找出 700 对节点，为每对节点提供中继服务，然后计算出每对节点经由查找出的中继节点的路径质量。为突出我们的中继查找方案的优点，对于每对节点在提供中继服务时，分别采用以下四种方案：

① 最优路径质量（optimization）：从所有 relay 节点中找出一个节点，这个节点使得通信两端经过这个节点的路径质量是经过所有 relay 节点中最好的。

② 随机(rand)：从当前 relay 节点集合中，随机选择一个节点作为 relay。

③ 单边最优(one-side opt):从 relay 节点集合中随机选择一个小的集合，然后从这个小的集合中选择一个与源节点路径质量最好的 relay 节点作为中继节点。

④ 我们的 Relay 查找方案(relay)：从通信的两个端点所在的 Cluster 中的 relay 节点中选出一个对这两个节点最优的一个节点作为它们的中继节点。

由图 8 图 9 可以看出，我们的 Relay 查找方案提供的中继服务中：中继节点查找的成功率约为 90%，通信两个端点经过我们的方案选出的中继节点的路径质量要比 rand 方案，one-side 方案都要优秀，有一部分甚至逼近最优方案查找出的中继节点的性能。

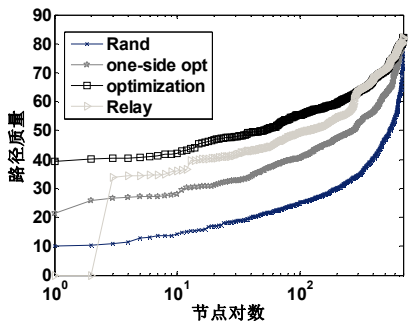

图 8 各种方案运行结果(1616 个节点)

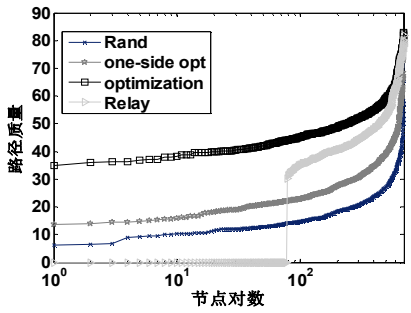

图 9 各种方案运行结果(9090 个节点)

8 结论与展望

本文主要讨论了拓扑感知的 P2P 中继查找方案，由上面的实验数据表明：我们的方案构建的 Overlay 网络能够很好的反映原始网络拓扑的结构，在我们构建的 Overlay 网络中提供的中继服务中找出的中继节点能够为通信的两个端点建立非常好的通信链路。

在将来，我们将会逐步完善和改进我们的 Cluster 组建算法以及中继查找算法。现实世界中的网络拓扑是极其复杂的，而我们用 GT-ITM 生成的 Transit-Stub 形原始网络拓扑则有些简单，和现实网络拓扑相比还有一些差距。我们在计算节点间路径质量时只考虑到了链路时延和丢包率，还有抖动和播放缓冲等因素没有考虑，这些将作为今后研究内容。

参考文献

1 Rosenberg J, Schulzrinne H, Johnston AR, Camarillo G, Peterson J, Sparks R, Handley M, Schooler E. SIP: session initiation protocol, RFC 3261. Internet Engineering Task Force. June 2002.

2 Garces-Erice L, Biersack E, Felber P, Ross K, Urvoy-Keller G. Hierarchical peer-to-peer systems. Lecture Notes in Computer Science. New York: Springer-Verlag, Proc. Euro-Par Parallel Processing. Jun. 2004. pp. 1230–1239.

3 Qi C, Dong ZJ, Principles and Applications of P2P SIP. ZTE Communications. 2007,13(6).

4 Guha S, Daswani N. An Experimental Study of the Skype Peer-to-Peer VoIP System. Proceedings of 2006 IPTPS. 2006.677－686.

5 Zhang XW, Lei WM, Yu B, Jia JY. Using P2P Overlay to Improve VoIP Quality in SIP+P2P System. In Processing of 2009 ICIE. 2009:255－259.

6 Ren S, Guo L, Zhang X. ASAP: An AS-Aware Peer-Relay Protocol for High Quality VoIP. Proceedings of 2006 IEEE ICDCS. 2006.70－79.

7 Gummadi KP, Madhyastha HV, Gribble SD, Levy HM, Wetherall D. Improving The Reliability of Internet Paths with One-hop Source Routing, Proceedings of 2004 USENIX. 2004.198－207.

8 Cheng C, Huang Y, Kung H, Wu CH. Path Probing Relay Routing for Achieving High End-to-end Performance. Proc. of 2004 IEEE GLOBECOM. 2004. 1359－1365.

9 Liao XF, Jin H, Zhu LL. DPLM: A Novel Application-layer Routing Scheme for Low Delay VoIP. ASIAN INTERNET NFERENCE (AINT-EC) .November 19, 2008 .

10 张宇,张宏莉,方滨兴. Internet 拓扑建模综述.软件学报, 2004,15(8):1220－1226.