

文章编号:1001-9081(2010)04-0888-04

## 基于网络测量的 P2P 跨域流量优化机制

郭涛<sup>1</sup>, 周旭<sup>1</sup>, 王治平<sup>2</sup>, 唐晖<sup>1</sup>

(1. 中国科学院声学研究所 高性能网络实验室, 北京 100080; 2. 中兴通讯股份有限公司, 南京 210012)

(guot@hpl.ac.cn)

**摘要:** P2P 技术的普及优化了用户的体验, 但对带宽的过度消耗也带给网络运营商巨大的压力。据此提出基于网络测量的、业务相关 P2P 跨域流量优化机制, 该机制对底层网络建立模型, 通过综合考量底层的网络信息和具体 P2P 业务的特殊性来优化节点互联。实验结果表明, 该处理机制明显减少了跨域流量, 优化了 P2P 用户的体验。

**关键词:** 对等网; 节点选择; 网络测量; 流量优化

**中图分类号:** TP393 **文献标志码:** A

## Mechanism for optimizing P2P traffic between networks based on network measurement

GUO Tao<sup>1</sup>, ZHOU Xu<sup>1</sup>, WANG Zhi-ping<sup>2</sup>, TANG Hui<sup>1</sup>

(1. High Performance Network Laboratory, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China;

2. Zhongxing Telecommunication Equipment Company, Nanjing Jiangsu 210012, China)

**Abstract:** The popularization for P2P technique has optimized the Internet experience; however, it brings big pressure to the Internet Service Provider (ISP) for it overly consumes the network bandwidth. This paper proposed a mechanism for optimizing P2P traffic based on the network measurement between networks related to different services, which set up the base layer network model and can optimize the connection between peers through comprehensively considering the base layer network information and particularity of services. Performance analysis and simulation show that this mechanism can not only reduce the traffic between networks but also optimize user experience.

**Key words:** Peer-to-Peer (P2P); peer selection; network measurement; traffic optimization

### 0 引言

对等网 (Peer-to-Peer, P2P) 流量对网络带宽的掠夺性占用给运营商的网络带来了巨大的压力, 在中国运营商的长途骨干网上, P2P 流量占一半以上, 严重影响了互联网正常业务的开展和其他传统主流内容的传播效果。但是 P2P 技术代表了互联网先进生产力的发展方向, 单纯的封堵和限制, 意味着技术的倒退, 也不符合用户的需求。这就迫切需要制定一个互联网运营商和 P2P 厂商都能够接受的合作方案, 来达到双赢的局面。

美国耶鲁大学网络系统实验室提出了 Proactive network Provider Participation for P2P (简称 P4P) 的解决方案<sup>[1]</sup>, 提出在运营商 (Internet Service Provider, ISP) 和 P2P 之间进行通信和协作来完成 P2P 流量的管理优化。德国的 Deutsche 电信实验室提出了 Oracle<sup>[2]</sup> 技术, 认为流量的本地化是优化 P2P 流量的关键。文献[3]提出了一种基于片段融合度的节点选择算法, 以减少网间流量。文献[4]中提出视频直播系统通过对 LandMark 值进行特殊标识, 以固定标识国家、身份、地区和城市来就近选择节点。

以上研究工作从不同的视角出发, 旨在减少网间流量, 但大多针对特定的 P2P 应用协议提出, 无法兼容多种不同的 P2P 协议。这些方案虽然减少了跨域流量, 但对不同的 P2P

业务的特殊性却不能区分以待。为此本文在支持自治域发现与优化的 P2P 管理协议 (Domain Discovering Protocol, DDP)<sup>[5]</sup> 基础上, 对其节点选择算法加以改进, 使其减少跨域流量的同时, 兼顾 P2P 业务的多样性, 优化用户体验。

### 1 DDP 系统简介

DDP 是由中科院声学研究所高性能网络实验室提出的一种 P2P 流量优化<sup>[6]</sup> 的管理方案。该方案在不改变原有 P2P 协议的基础上, 能综合协调多种 P2P 应用协议。在此系统上, 运营商可以根据需求和成本, 灵活调整部署方案, 从而实现引导、代理、缓存等多种流量优化功能, 同时支持自治域的分级域管理, 协助 P2P 厂商分级别、分层次管理 Peer 的互联行为。DDP 是一个能实现运营商与 P2P 厂商双方双赢的解决方案。

DDP 系统通常包括: P2P 重定向服务器 (P2P Redirect Server, PPR)、P2P 缓存服务器 (P2P Cache, PPC)、域名解析服务器 (Domain Name System server, DNS) 三个实体。

1) PPR 是一个部署在运营商自治域的网络实体。它通过开放的标准协议与域内的 P2P 客户端通信, 为本域的 P2P 用户服务, 一旦收到了 P2P 客户端发送的 Peer 列表的请求, 就会根据网络拓扑信息和策略信息选择 Peer 列表返回。

2) DNS 为客户端提供域名解析服务, 客户端可以向其查

收稿日期: 2009-09-30; 修回日期: 2009-11-28。 基金项目: 中兴高校合作基金资助项目。

**作者简介:** 郭涛 (1984-), 男, 湖南沅江人, 硕士研究生, 主要研究方向: P2P 应用技术; 周旭 (1976-), 男, 四川成都人, 副研究员, 博士, 主要研究方向: 分布式网络、P2P 应用技术; 王治平 (1977-), 男, 安徽安庆人, 博士, 主要研究方向: 分布式计算、数据挖掘; 唐晖 (1971-), 男, 山东潍坊人, 研究员, 博士, 主要研究方向: 网络通信、下一代互联网。

询域内 PPR 服务器的实际地址。

3) PPC 部署在域内,为域内 P2P 用户提供代理和缓存功能。

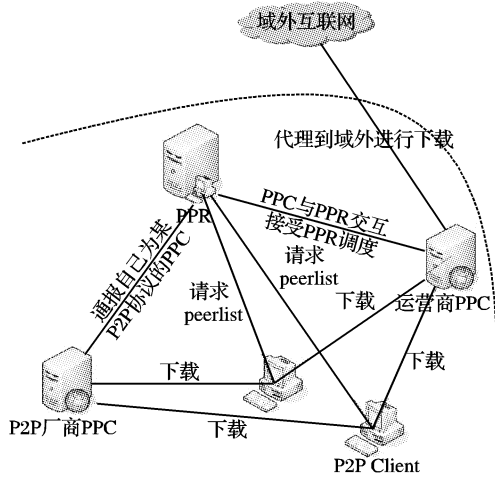


图1 DDP系统部署图

P2P 客户端采用标准的协议与 PPR 通信,当客户端要访问或者请求某个文件时,它会向 PPR 发送请求报文,其中包含了请求文件的 infohash、所使用的 P2P 协议类型以及 PeerID 等信息。PPR 收到客户请求后会查看是否有其他 Peer 使用相同的 P2P 协议访问着同样的文件,若找到则将其一定数量的 Peer 信息返回给客户端。此过程中,PPC 相当于一个超级 Peer,若本域内部署了 PPC 服务器,PPR 会优先将 PPC 服务器地址一起返回给客户端。客户端收到返回信息后就会与 PPC 及返回的 Peer 连接。如图 2 所示。

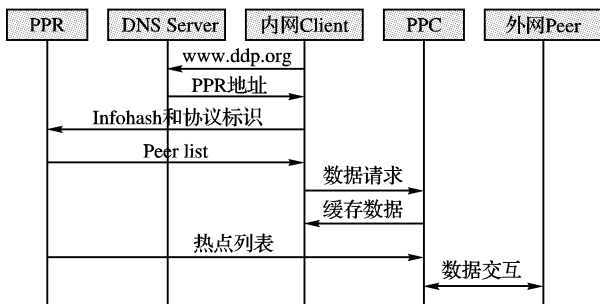


图2 P2P流量管理时序图

## 2 基于网络测量的P2P跨域流量优化机制

从图2可以看出,当内网的Client向本域的PPR发送节点请求时,PPR会根据内网Client所发出的请求内容予以识别,再返回节点列表。如果将整个互联网上的节点视为一个全集,在PPR的服务进程中就需要设计一种节点选择机制,从全集中选出一个子集返回给请求客户端。当发送请求的Client与该子集内节点连接时,不但能产生较少的跨域流量,而且符合该节点所请求的业务对于带宽、时延等要求的特殊性。具体过程阐述如下。

### 2.1 网络拓扑图的转换

PPR 部署在运营商内部,可以得到网络实际的拓扑信息并予以抽象化。根据网络的实际情况和网络管理的粒度,将网络划分为不同的域,每个域下可能存在多层次的子网。运营商将其拓扑信息以特定的格式存入到数据库中。

现将整个网络抽象化成一张有向连通图,节点表示网络,

有向边表示网络之间的链路,如图3所示。

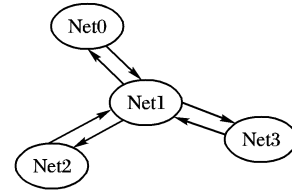


图3 网络拓扑抽象

为了度量不同网络间和同网络的不同节点之间的参数,根据参数来优化节点的选择,本文对该图进行数据抽象化描述,创建以下三张表。

第一张是 Net 表,用以描述域的属性。基本表项为  $\langle \text{NetID}, \text{ParentID}, \text{Type}, \text{AddrStart}, \text{AddrEnd}, \text{Mask}, \text{Capacity} \rangle$ 。网络彼此之间的层次关系可以由  $\langle \text{NetID}, \text{ParentID} \rangle$  两个参数来决定;Type 用来表述网络的类型以及接入方式,比如 LAN, ADSL; AddrStart 和 AddrEnd 用来表示该网络内拥有节点的 IP 范围;Mask 表示子网掩码;Capacity 用来表示网络内的节点数目。

第二张是 Peer 表,用以描述域内节点的基本属性。基本表项为  $\langle \text{PeerID}, \text{IP}, \text{PeerUplink}, \text{PeerDownlink}, \text{Option} \rangle$ 。PeerID 和 IP 分别表示该节点的逻辑标识和网络层标识;PeerUplink 和 PeerDownlink 分别表示该节点的上传和下载带宽;Option 用来做扩展选项,可以用它来表示计算机的负载和计算资源等情况,做深入优化。

第三张表是 NetEdge 表,用于表示各个网络之间网络链路的参数信息。基本表项为  $\langle \text{NetA}, \text{NetB}, \text{BandWidth}, \text{Delay}, \text{Lost}, \text{Distance}, \text{Option} \rangle$ 。NetA 和 NetB 为网络 ID 号,表示该链路连接哪两个网络;BandWidth 表示网络的出口带宽;Delay 表示链路的时延;Lost 表示链路的丢包率和可靠性;Distance 表示该链路跨越了几个路由器。Options 为扩展选项。

需要特别指出的是:该图有向图。 $\langle \text{NetA}, \text{NetB} \rangle$  和  $\langle \text{NetB}, \text{NetA} \rangle$  分别表示同一网络链路的两个方向,因为同一链路的两个方向其出口带宽和时延等参数可能不同。

### 2.2 算法描述

以上各个表格的底层信息参数由 PPR 收录。考虑到运营商和 P2P 应用需要针对各个参数的不同侧重点来进行节点的优化选择,本文以各个参数为自变量,综合考虑各个参数的权重,计算出不同节点之间链路的总开销值  $Cost$ 。在进行节点选择的时候就可以依据不同 P2P 业务的  $Cost$  值来选择网段和节点,公式如下:

$$Cost = f(\text{BandWidth}, \text{Delay}, \text{Lost}, \text{Distance}, \text{Option})$$

(1)

根据式(1)中的参数,本文参考传统路由器的选择协议,以借鉴它们处理底层信息参数的方法。比如:OSPF (Open Shortest Path First) 协议对于路由距离的度量考虑了带宽和延迟等因素;Cisco 私有的 EIGRP<sup>[7]</sup> (Enhanced Interior Gateway Routing Protocol) 协议也综合考虑了带宽、时延、可靠性、负载传输单元大小等因素(但通常也缺省地忽略了可靠性、负载和传输单元三个因素)。这些参数都能影响底层的路由选择,但又具有不同的物理单位,而 OSPF 和 EIGRP 协议均是把这些参数经权衡之后等价考量。本文借鉴这种处理方式,并在综合考虑这些因素的前提下,提出同时面向 ISP 和 P2P 应

用的综合开销计算公式:

$$Cost = \left[ \frac{K_1}{BandWidth} \times \mu_1 + \left( \sum Delay \right) \times K_2 \times \mu_2 + \right. \\ \left. Lost \times K_3 \times \mu_3 \right] + Distance \times K_0 \times \mu_0 \quad (2)$$

式(2)中,  $K_i \times \mu_i$  分别为带宽、时延、丢包率、路由跳数的权衡系数,其中  $\mu_0, \mu_1, \mu_2, \mu_3$  之和为 100%,  $K_i$  为  $\mu_i$  的归一化参数。

在以上各权衡系数中,  $\mu_0$  主要面向运营商,用以调节  $Distance$  权重的大小。 $Distance$  为某条路由选择的路径上所跨越的路由跳数。运营商希望所选节点的  $Distance$  值尽量小,从而减少跨域流量,这就需要在设置  $Cost$  值计算公式时提升系数  $\mu_0$  的大小。 $\mu_0$  并不是固定值,而是针对不同 P2P 应用有不同的权重值,但具体的设置仍需 P2P 厂商和运营商之间做更加深入具体的协商。

$\mu_1, \mu_2$  和  $\mu_3$  的调节主要面向 P2P 应用,分别用来表示带宽、时延和可靠性系数(丢包率)的大小。式(2)中的时延需要综合考虑链路本身的时延和网络设备的时延。不同的 P2P 业务需要选择不同的  $\langle \mu_1, \mu_2, \mu_3 \rangle$  矢量,对于类似 Skype 语音等时延敏感的业务,需要加大  $\mu_2$  系数,对于 Bittorrent 等下载业务,就需要加大带宽的选择系数  $\mu_1$ ; 对于视频等流媒体业务,则需要综合考虑带宽和时延等因素的影响。

各具体的 P2P 应用在第一次向 PPR 注册时,需要提交本身的业务属性,包括  $\langle \mu_1, \mu_2, \mu_3 \rangle$  矢量值。当执行该业务的客户端再次向 PPR 请求节点时,基于不同的业务,PPR 会根据之前注册的矢量值来计算待选节点所处网络的  $Cost$  值,从而优化节点选择。然后在设置好计算综合开销值所需的权重系数后,PPR 采用传统的 Dijkstra 算法生成每个网络到其他所有网络的  $Cost$  序列,升序排列后最终得到的计算结果。

对于其计算结果,以下面矩阵的方式呈现:

$$\begin{bmatrix} P2PType_1 \\ P2PType_2 \\ P2PType_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} RNetID_{11} & RNetID_{12} & RNetID_{13} & \cdots \\ RNetID_{21} & RNetID_{22} & RNetID_{23} & \cdots \\ RNetID_{31} & RNetID_{32} & RNetID_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (3)$$

其中:  $NetID_i$  表示 ID 号为  $i$  的网络;  $RNetID_{mn}$  表示 P2P 类型为  $m$ , 身处在 ID 为  $n$  的网络中。当发送节点选择请求的时候,PPR 根据  $Cost$  计算出来的网络优先级序列,表示为:

$$\begin{bmatrix} RNetID_{11} \\ RNetID_{12} \\ RNetID_{13} \\ \vdots \end{bmatrix} = \begin{bmatrix} n_1IDr_1 & n_1IDr_2 & n_1IDr_3 & \cdots \\ n_2IDr_1 & n_2IDr_2 & n_2IDr_3 & \cdots \\ n_3IDr_1 & n_3IDr_2 & n_3IDr_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (4)$$

其中  $n_aIDr_b$  表示某节点处在 ID 为  $a$  的网络中,当它发出请求时,PPR 为其选择的优先级为  $b$  的网络 ID 号。

### 3 效率分析和实验设计

实验设备: PPR 服务器(OS: Linux ES5)一台, PC 机一台(OS: Linux FEDORA 9)。首先构建网络拓扑图。

为了降低复杂度,实验中认为同网络的节点都具有同等的上下行带宽和计算资源,同时忽略网络链路的有向性,即认

为两个网络之间的链路  $\langle NetA, NetB \rangle$  和  $\langle NetB, NetA \rangle$  具有相同的属性。网络拓扑如图 4 所示。

图 4 中网络 2 和 3 具有相同的 ParentID, 且子网掩码均为 25 位, 说明它们同是 192.168.1.0/24 网段的子网。因为在实验中忽略了网段内节点的差异性, 所以将 Peer 表简化为图 4 中的 PeerUplink 项, 表示网段内节点的上行链路大小。

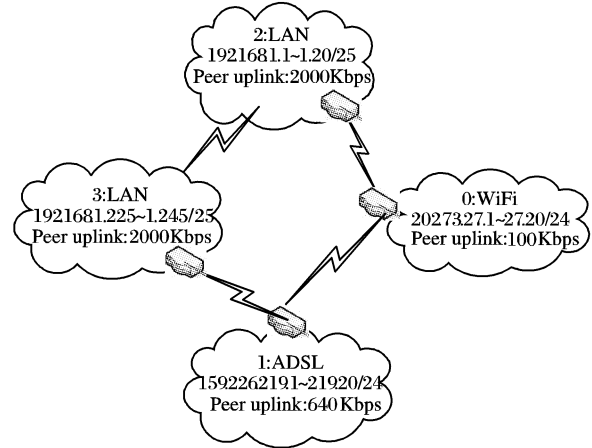


图4 DDP系统网络拓扑测试图

实验中 ADSL 的上行速率大小设置为 640 Kbps; WiFi 客户端实际上行速率受网络实际环境影响很大(比如多人共享), 实验中设置为 100 Kbps; LAN 内部节点之间互传, 100 MB 链路上行最大可达约 8 Mbps, 假设受客户端本身限制的影响, 实验中将 Peer uplink 约束为 2000 Kbps。

各个网段之间的链路属性如表 1 所示。

表1 实验拓扑图网间链路属性

NetA	NetB	Bandwidth / Mbps	Delay / $\mu s$	Lost / %	Distance	Option
0	1	10	20000	10.00	2	NULL
0	2	50	10000	2.00	2	NULL
1	2	100	4000	0.50	2	NULL
2	3	100	2000	0.05	1	NULL

从表 1 可以看出, 实验环境中设定 LAN 之间以及 LAN 与 ADSL 之间均用 100 Mbps 链路连接; WiFi 标准是 802.11g/b, 其中 802.11g 是 54 Mbps, 802.11b 是 11 Mbps。考虑到实际情况, 设定 WiFi 与 LAN 之间链路带宽为 50 Mbps, 与 ADSL 之间为 10 Mbps; 相邻网络之间需要经过 2 跳路由, 但由于网络 2 和 3 有相同的上层网络, 路由跳数设为 1; Option 选择暂不考虑, 设为空。

为了验证节点选择算法的优化性, 本文设计实验对 Bittorrent 和 PPCDN (PPCDN 是中科院声学所高性能网络实验室自主开发的流媒体点播系统) 协议进行测试。假设初始化各  $\mu_i$  值为 25%, 平均条件下网络链路的属性为  $\langle Bandwidth = 50 \text{ Mbps}, Delay = 6400 \mu s, Lost = 1\%, Distance = 1.5 \rangle$ , 且各个参数在  $Cost$  权重中的数值均为 12800, 由此得到归一化参数矢量  $\langle K_1, K_2, K_3, K_0 \rangle$  值为  $\langle 256 \times 10^7, 8, 512 \times 10^2, 35 \times 10^3 \rangle$ 。实验中以此归一化参数矢量和表 2 中设定的  $\mu_i$  计算  $Cost$ 。

Bittorrent 和 PPCDN 协议具体的各  $Cost$  计算参数  $\mu_i$  值如表 2 所示。表 2 中 Bittorrent 协议与 PPCDN 流媒体协议关于  $Cost$  计算参数设置的不同, 是为了体现 P2P 业务的多样性以及特定的业务开展时对于底层信息的特殊要求。对于类似

Bittorrent 的文件下载协议,人们不仅关心其传输文件的带宽大小,更希望能减少它在传输过程中的跨域流量,考虑到这两个因素的重要性,实验中将 *Distance* 和 *Bandwidth* 的权重值设置为 60% 和 30%;对于类似 PPCDN 等流媒体协议,除了 *Distance* 和 *Bandwidth* 之外,客户更期望观看过程中节目画面的正确性和流畅性,因此在满足其基本视频速率的前提下(目前大多数频道速率都在 800 Kbps 以下),实验中加大了对时延和链路可靠性的考虑,分别设置为 15%。

表 2 Bittorrent 与 PPCDN 的 Cost 计算参数

P2P Type	$\mu_0/\%$	$\mu_1/\%$	$\mu_2/\%$	$\mu_3/\%$
Bittorrent	60	30	5	5
PPCDN	50	20	15	15

实验时,首先在服务器端启动 PPR 进程,之后在客户端启动进程(客户端进程由 Python 语言编写)对 PPR 发送节点请求,并显示返回的节点列表。所得实验结果如图 5~8。

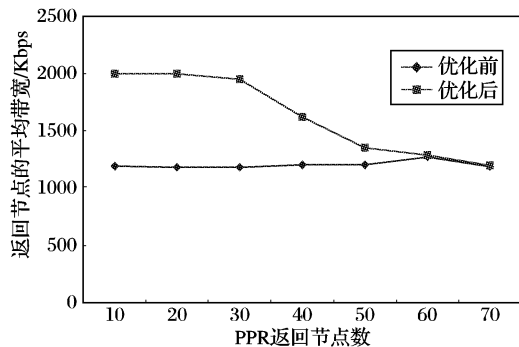


图 5 优化前后 Bittorrent 协议返回节点的平均带宽对比

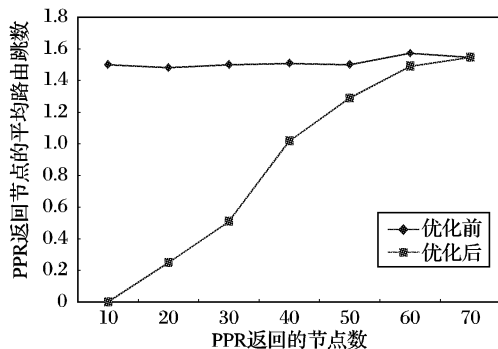


图 6 优化前后 Bittorrent 协议返回节点的平均路由跳数对比

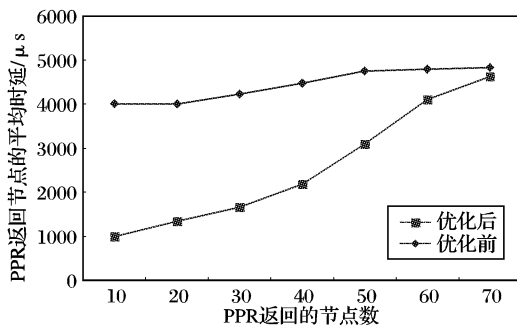


图 7 优化前后 PPCDN 返回节点平均时延对比

从图 5、6 可以看出,Bittorrent 协议优化后,PPR 返回节点的路由跳数不仅明显减少,即减少了跨域的流量,而且返回节点的平均带宽也有明显的增加。从图 7、8 可以看出 PPCDN 协议优化后,PPR 返回节点的平均时延和丢包率均明显减少,大大提高了用户体验质量。

此外,当选取节点数明显小于总节点数时,优化效果明

显。而当选取节点数接近总节点数时,优化前后效果趋同,这是由于实验中设定的总节点数有限所致。而实际网络环境下的节点数非常大,相比而言,P2P 业务开展时选取的节点数非常小,由实验结果可知,本文所述算法在实际网络环境中可以收到很好的优化效果。

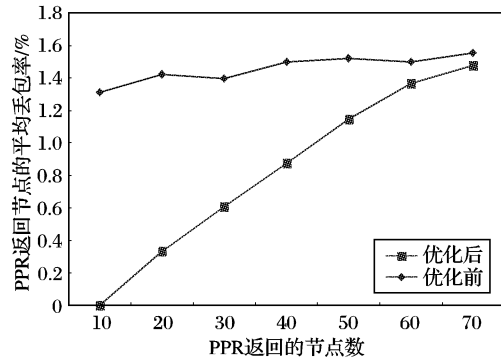


图 8 优化前后 PPCDN 返回节点的平均丢包率对比

## 4 结语

目前,P2P 流量对于带宽的过度消耗使得网络运营商不堪重负,P2P 与网络运营商之间的合作能促使彼此双方达到双赢的局面,既能减少跨域流量,又能优化用户体验。

基于 P2P 业务的多样性和复杂性,单一的优化策略不能满足多种 P2P 应用的要求。本文在中科院声学所开发的 DDP 系统的基础上,针对 P2P 业务的不同特性和要求对节点选择机制加以改进,针对具体的底层参数信息,提出了一种在节点选择过程中总开销的权衡机制,并以此为依据进行节点的筛选。实验结果表明,该处理机制明显减少了跨域流量,也优化了不同 P2P 用户的体验,在考虑了运营商利益的同时,也最大限度地兼顾了各类 P2P 应用的需求。但是本文进行节点选择时屏蔽了同网内节点的差异性,而且没有针对互联网上所有 P2P 应用展开研究,所以如何对同网络内节点的差异性进行量化和标准化(比如链路负载、计算负载等),如何满足更多 P2P 业务的需求,仍需要今后进一步研究。

## 参考文献:

- [1] XIE HAIYONG, KRISHNAMURTHY A, SILBERSCHATZ A, et al. P4P: Explicit communications for cooperative control between P2P and network providers [R]. P4PWG, 2007.
- [2] PETSCHICK M. The P2P oracle [EB/OL]. [2009-09-01]. [http://www.net.t-labs.tu-berlin.de/teaching/.../IR.../petschick\\_slides.pdf](http://www.net.t-labs.tu-berlin.de/teaching/.../IR.../petschick_slides.pdf).
- [3] 欧阳荣, 苗卉, 雷振明. 一种减少网间 P2P 流量的 Peer 选择算法 [J]. 计算机工程, 2008, 34(8): 108-110.
- [4] LIAO X F, JIN H, LIU Y H, et al. Any see: Peer-to-peer live streaming [C] // 25th IEEE International Conference on Computer Communications. Barcelona: IEEE Press, 2006: 1-10.
- [5] ZHOU X, TANG H, QIN W, et al. DDP: A novel P2P traffic management and optimization protocol [C] // Third International Conference on Communications and Networking in China. Hangzhou: IEEE Press, 2008: 208-212.
- [6] 蒋卓明. P2P 网络管理优化若干关键问题的研究 [D]. 北京: 中国科学院高能物理研究所, 2009.
- [7] CATHERINE P. 组建可扩展的 CISCO 互连网络 [M]. 陈宇, 袁国忠, 译. 北京: 人民邮电出版社, 2003: 103-137.