

# 基于 P2P 的 IPv6 虚拟网络构造

秦丰林<sup>1,2</sup>, 葛连升<sup>1</sup>, 刘 璐<sup>2</sup>

(1. 山东大学网络中心, 济南 250100; 2. 山东大学信息科学与工程学院, 济南 250100)

**摘要:** IPv4 向 IPv6 转换是 Internet 发展的趋势。针对双栈和隧道过渡技术存在的局限性, 提出一种基于 P2P 的 IPv6 虚拟网络的设计方法。该方法采用双层无结构拓扑实现 IPv4/IPv6 路由表的构造和查找, 通过动态隧道管理机制自动设置和维护 IPv6 节点间的隧道, 提高隧道管理效率。分析结果表明, 该方法能够满足系统可扩展性和可行性的需求。

**关键词:** IPv6 虚拟网络; 对等网络; 动态隧道管理

## Construction of P2P-based IPv6 Virtual Network

QIN Feng-lin<sup>1,2</sup>, GE Lian-sheng<sup>1</sup>, LIU Ju<sup>2</sup>

(1. Network Center, Shandong University, Jinan 250100; 2. School of Information Science and Engineering, Shandong University, Jinan 250100)

**【Abstract】** Transition to IPv6 from IPv4 is recognized as the trend of future Internet. This paper proposes a P2P-based IPv6 virtual network for the defects of dual stacking and tunnel transition techniques. This system uses a two-layered unstructured topology to construct and lookup the IPv4/IPv6 routing table, and adopts a Dynamic Tunnel Management(DTM) mechanism to automatically setup and maintain the tunnels between IPv6 nodes, which can improve the efficiency of tunnel management. Analysis results show that the system is scalable and realizable.

**【Key words】** IPv6 virtual network; P2P network; Dynamic Tunnel Management(DTM)

### 1 概述

IPv6 具有 128 位的地址空间、灵活的地址配置机制、增强的安全性和移动性等优点, 因而 IPv4 向 IPv6 转换是 Internet 发展的趋势。双栈和隧道是 2 种基本的 IPv4/IPv6 过渡技术, 主要存在以下问题:

(1) 双栈技术需要 ISP 升级设备来支持 IPv6, 升级成本高, 维护难度大, 过渡时间长;

(2) 隧道技术需要手工配置, 可管理性和可扩展性差(虽然 6to4 是一种自动隧道技术, 但面临管理和安全问题<sup>[1]</sup>)。

隧道技术可抽象为  $N-1$  和  $N-N$  2 种模型, 如图 1 所示。其中,  $N-1$  模型的管理复杂度是  $O(N)$ ,  $N-N$  模型是  $O(N^2)$ , 因而隧道的管理工作随  $N$  增大而大量增加。

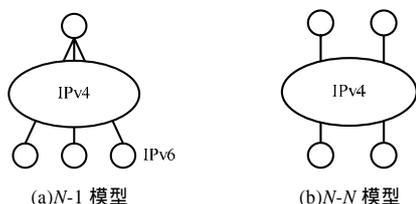


图 1 隧道技术模型

6Bone 是一种建立在 IPv4 网络上的 IPv6 覆盖网络, 用户主要通过隧道技术接入 IPv6, 隧道配置和管理工作复杂。目前, 相对于 ISP 缓慢升级设备来支持 IPv6, 端用户主机大多已支持 IPv4/IPv6 双协议栈。基于这种考虑, 文献[2]提出利用 P2P 技术实现 IPv6 虚拟网络, 通过 DHT 技术建立和维护 IPv4/IPv6 路由表。然而, DHT 的优点是数据精确查找, 不适用于基于最长匹配的路由查找, 而且文献[2]也未涉及 IPv6 节点间隧道的设置和管理问题。

在此基础上, 本文提出构造一种基于 P2P 的 IPv6 虚拟网

络(P2P-6Bone), 采用双层无结构拓扑构造和维护路由表, 实现路由的最长匹配查询。针对大规模隧道配置的复杂性问题, 提出一种动态隧道管理(Dynamic Tunnel Management, DTM)机制, 实现 IPv6 节点间隧道的自动设置和维护, 提高隧道管理效率。

### 2 系统结构

P2P-6Bone 从逻辑上可划分为核心网络(Core Network, CN)和边缘网络(Edge Network, EN)2 个部分, 如图 2 所示。CN 是指 IPv4 骨干网络, 内部节点称为外部网关(External Gateway, EG), 负责边缘网络的接入。EN 是指 IPv6 用户网络, 主要包括普通 IPv6 主机(Host)和作为网关的内部网关(Internal Gateway, IG)2 类节点。

图 3 显示了 EG(或 IG)节点的结构, 主要包括 P2P 模块、隧道模块和地址配置模块 3 个部分。地址配置模块负责为 EN 中 IPv6 主机自动分配地址, 具体实现可参考 Radvd, 本文不作详细介绍。

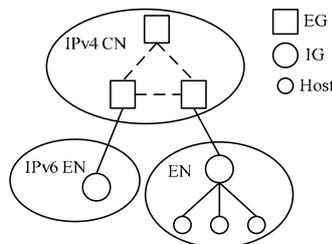


图 2 系统结构

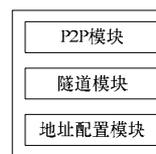


图 3 节点组成结构

**基金项目:** 山东省自然科学基金资助项目(Y2007G42)

**作者简介:** 秦丰林(1978 -), 男, 工程师、博士研究生, 主研方向: 网络管理与应用, 对等网络; 葛连升, 副教授; 刘 璐, 教授

**收稿日期:** 2009-04-20 **E-mail:** qfl@sdu.edu.cn

P2P 模块采用类似 KaZaA<sup>[3]</sup> 的双层无结构拓扑, 其中 EG 作为超级节点, IG 作为普通节点。主要因为: (1) 充分考虑 P2P 网络的异构性; (2) 提高系统的可扩展性; (3) 实现路由的最长匹配查找; (4) 解决存在 NAT 设备时 IG 节点间通信问题。路由查询完成后, 隧道模块通过 DTM 机制在 IG 间自动建立和维护隧道, 实现 IPv6 主机间的通信。

### 3 关键技术

本节将介绍 P2P-6Bone 中的 P2P 路由协议和 DTM 机制, 并讨论系统的其他方面, 如安全性和 NAT 等问题。

#### 3.1 P2P 路由协议

P2P-6Bone 中 EG 和 IG 共同组成一个双层无结构 P2P 网络, 其中 EG 是超级节点, IG 是普通节点。EG 的选举方式与 KaZaA 协议相同, 由具有公网连接、一定的计算能力和存储空间节点来承担。

每个 IG 维护一个 EG 节点列表, 在启动时首先选择延时最小的 EG 建立连接, 并定期向其发送通告报文, 包括自己的 IPv4 地址和负责的 IPv6 地址前缀。EG 负责管理与 IG 和其他 EG 的连接, 并构造和维护一个 IPv4/IPv6 路由表。EG 在接收到 IG 的通告报文后, 将其中的 IPv4 和 IPv6 地址映射添加到路由表中, 最后 EG 路由表中含有所有 IG 邻居的 IPv4/IPv6 地址映射, 如表 1 所示。 $T_{expire}$  表示路由表项的过期时间, 在  $T_{expire}$  时间超时后将该表项删除, 保证路由的实时性和有效性。

表 1 P2P-6Bone 路由表

IPv4 地址	IPv6 地址	$T_{expire}/\text{min}$
128.190.102.1	3001:1200:3001:0:100::/80	30
202.111.222.1	3002:f25a:9:237:fd::9d3a:1	30
112.121.201.1	3000:1002:9:256:e2::fc8a:1	30

IPv6 主机在与其他 EN 中 IPv6 主机通信时, 其内部网关 IG 首先向 EG 发出请求获得对端 IG 的 IPv4 地址, 以建立到该 IG 的隧道连接。EG 在接收到查询请求(IG Query)后, 首先检查路由表中是否有对应的表项, 若存在则返回最长匹配记录(IG Reply); 若不存在, 则向其他 EG 邻居洪泛路由查询请求(EG Query), 最后 EG 从接收到的路由查询应答(EG Reply)中选择最长匹配记录, 并返回 IG Reply。P2P 路由查询的典型流程如图 4 所示。

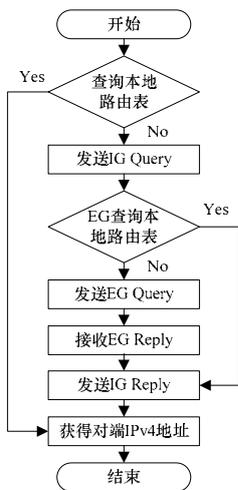


图 4 P2P 路由

为提高路由查询效率, P2P-6Bone 采取了 2 种优化方法: (1) IG 缓存一个 IPv4/IPv6 路由表, 在路由查询时首先检查本地路由表; (2) EG 采用“一跳复制”策略<sup>[4]</sup>, 即首先从 EG 列

表中随机选择  $d$  个邻居建立连接并交换路由信息, 以后只需周期性地增量更新(Incremental Update)来实现路由同步; 同时, EG 使用“心跳算法”检查 EG 邻居的连通性, 在某个 EG 邻居离开或失效时, 向其他邻居洪泛节点失效报文, 再随机选择一个 EG 建立连接, 保持节点度  $d$  不变, 这也有助于提高 P2P-6Bone 的鲁棒性, 一般设  $d = 5$ 。

为实现路由实时性和通信开销的折衷, 需合理设置路由刷新时间  $T_{refresh}$ 。设 IG 在获得新 IPv6 地址后, 原地址在时间  $T_{transmission}$  内仍然有效, 则路由刷新时间需满足  $T_{refresh} < T_{transmission}$ , 一般设  $T_{transmission} = 24 \text{ h}$ ,  $T_{refresh} = 15 \text{ min}$ ,  $T_{expire} = 30 \text{ min}$ 。在  $T_{refresh}$  时间超时后, IG 向 EG 发送通告报文更新路由表项。

#### 3.2 DTM 机制

IG 节点(表示为  $IG_i$ ,  $i$  为节点编号, 下同)通过 P2P 路由获得对端 IG( $IG_j$ )的 IPv4 地址后, 使用 DTM 机制在两者之间自动建立隧道以进行 IPv6 通信。隧道建立流程如图 5 所示。

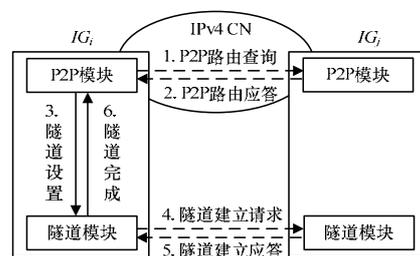


图 5 DTM 机制

隧道建立的步骤如下:

**步骤 1~步骤 2**  $IG_i$  通过 P2P 路由查询得到  $IG_j$  的 IPv4 地址。

**步骤 3**  $IG_i$  的 P2P 模块向隧道模块发送隧道设置消息, 包含  $IG_j$  的 IPv4/IPv6 地址对等。

**步骤 4**  $IG_i$  的隧道模块向  $IG_j$  的隧道模块发送隧道建立请求消息, 包含本端的 IPv4/IPv6 地址对、隧道类型(如 IPv6 in IPv4 或 GRE)等。

**步骤 5**  $IG_j$  的隧道模块从隧道建立请求消息中取出 IPv4/IPv6 地址对和隧道类型等信息, 在成功建立隧道后, 向  $IG_i$  返回隧道建立应答消息。

**步骤 6**  $IG_i$  的隧道模块建立隧道, 向 P2P 模块发送隧道完成消息, 最后  $IG_i$  和  $IG_j$  开始 IPv6 通信。

隧道建立成功后, 隧道模块之间通过“心跳算法”检测隧道的连通性。在检测到隧道中断时, 隧道模块向 P2P 模块发送隧道中断消息, P2P 模块间隔  $T_{interval}$  时间后重新发起隧道建立过程。同时, 为提高通信效率, 隧道模块可维护  $N_{max}$  个隧道连接。在接收到隧道设置请求时, 如果到目的地址的隧道不存在且当前隧道数目  $N_{current} < N_{max}$ , 则隧道模块直接向  $IG_j$  发送隧道建立请求; 否则, 首先根据“LRU 算法”拆除隧道, 再发送隧道建立请求。

#### 3.3 安全性和 NAT

P2P-6Bone 易受到冒名攻击, 例如 IG 向 EG 通告自己的 IPv4 地址和其他 IG 负责的 IPv6 地址前缀, 造成路由劫持问题。这可以通过公钥密码技术(PKI)来解决<sup>[2]</sup>。在 EG 的路由表项中增加 2 个证书字段: 地址所有证书和地址映射证书。地址所有证书可由 IPv6 地址分配机构(如 IANA)向 IG 颁发, 含有 IPv6 地址前缀和公钥, 对应的私钥则仅由该 IG 保存; 地址映射证书则由 IG 自己生成, 含有 IPv4 和 IPv6 地址的映

射。IG 在通告报文中包含这 2 个证书,在接收到通告报文时,EG 首先通过这 2 个证书来检查地址映射的真实性。

实际网络中 IG 大多处于 NAT 设备后面,因此,NAT 是 P2P-6Bone 面临的重要问题。双层无结构 P2P 拓扑可通过“连接反转”的方法实现单节点 NAT 穿越,对于双节点 NAT 等复杂的穿越问题,可参考文献[5]来解决。

#### 4 性能分析

与文献[2]中数据和路由都通过中间节点转发相比,P2P-6Bone 实现了数据平面与控制平面的分离。IG 通过隧道直接传送数据,路由信息则由 EG 转发,从而提高数据转发效率,降低 P2P 网络开销。

可扩展性是 P2P-6Bone 的一个重要指标。由于 IG 是普通节点,不负责路由的查找和维护,因此主要分析 EG 的可扩展性。假设 EG 生存时间服从参数为  $\lambda$  的指数分布,则在 1 s 时间内节点失效概率为  $1-e^{-\lambda}$ 。设网络中 EG 数量为  $N$ ,则失效节点数为  $N(1-e^{-\lambda})$ ,发出的节点失效报文长度  $S=S_0N(1-e^{-\lambda})$ , $S_0$  是单个节点失效报文长度。节点度至少为  $d$  的第  $n_i$  个 EG 节点的平均邻居数可表示为

$$E(n_i) = \begin{cases} d + \sum_{l=d+2}^N \frac{d}{l-1} & 0 < i < d+1 \\ d + \sum_{l=i+1}^N \frac{d}{l-1} & d+1 < i < N \end{cases} \quad (1)$$

根据式(1)可以得出前  $d+1$  个 EG 具有最大的平均邻居数。因此,考虑最差情况,即前  $d+1$  个节点中某一节点同时从邻居接收到节点失效报文,并向其邻居转发,其带宽开销  $C$  为

$$C = 2NS \max_{0 < i < N} (E(n_i)) = 2dNS_0(1-e^{-\lambda}) \left(1 + \sum_{l=d+2}^N \frac{1}{l-1}\right) \quad (2)$$

假设 EG 只将节点带宽  $B$  的 10% 用于传输控制信息,则需满足

$$C \leq B \times 10\% \quad (3)$$

就节点带宽  $B$  和生存时间  $t$  对 P2P-6Bone 可扩展性的影响进行了分析,如图 6 所示。其中, $B$  的变化范围是 128 Kb/s~1 024 Kb/s; $t$  的范围是 0.5 h~4 h;其他参数  $d=5$ , $S_0=256$  bit 是固定值。可以看出,随着节点带宽和生存时间增加,系统支持的节点数也随之增加,特别的,在  $B=1$  Mb/s, $t=2.5$  h 时节点数约为  $2.4 \times 10^4$ 。文献[3]对 KaZaA 网络的测量结果显示,超级节点平均生存时间约为 2.5 h,节点数在  $2.0 \times 10^4 \sim 4.0 \times 10^4$  之间,符合上面的分析结果,也充分说明 P2P-6Bone 能够满足系统可扩展性和可行性的要求。

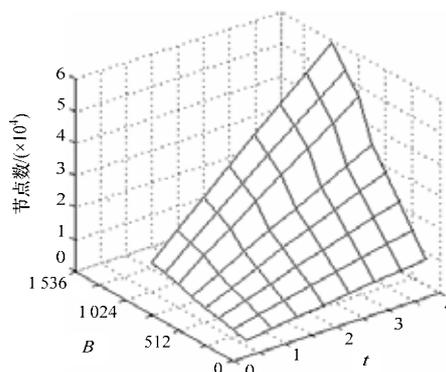


图 6 节点带宽和生存时间对节点数的影响

#### 5 结束语

IPv4 向 IPv6 转换是 Internet 发展的趋势,然而目前 IPv6 并未取得预期的广泛部署和应用。本文中的 P2P-6Bone 具有以下特点:(1)可扩展性。双层无结构 P2P 协议可保证系统的可扩展性;(2)灵活性。可根据需求在 Internet 或局域网中进行增量部署;(3)可管理性。DTM 机制能实现大规模隧道的自动设置和维护,提高隧道管理效率;(4)低成本。无需改变现有设备,仅利用端用户主机来实现 IPv6 虚拟网络。P2P-6Bone 既可作为 IPv6 网络的访问和测试平台,也可作为 IPv6 下的一种 P2P 资源共享应用,对促进 IPv6 的普及具有一定的积极意义。下一步工作主要是 P2P-6Bone 的实际部署与应用,并对网络的运行结果进行测量与分析。

#### 参考文献

- [1] Savola P, Patel C. Security Considerations for 6 to 4[S]. RFC 3964, 2004.
- [2] Zhou Lidong, Robbert V. P6P: A Peer-to-Peer Approach to Internet Infrastructure[C]//Proc. of the 3rd International Workshop on Peer-to-Peer Systems. [S. l.]: ACM Press, 2004.
- [3] Liang Jian, Kumar R, Ross K W. The KazaA Overlay: A Measurement Study[J]. Computer Networks, 2005, 50(6): 842-858.
- [4] Chawathe Y, Ratnasamy S, Breslau L, et al. Making Gnutella-like P2P Systems Scalable[C]//Proc. of ACM SIGCOMM'03. [S. l.]: ACM Press, 2003: 407-418.
- [5] 张国印, 叶在伟, 曲丽君. 一种 UDP 穿越 NAT 的新方案[J]. 计算机工程, 2008, 34(12): 112-113, 131.

编辑 顾逸斐

(上接第 96 页)

#### 参考文献

- [1] 朱玉贤, 李毅. 现代分子生物学[M]. 北京: 高等教育出版社, 2002.
- [2] Thakurta D G. Computational Identification of Transcriptional Regulatory Elements in DNA Sequence[J]. Nucleic Acids Research, 2006, 34(12), 3585-3598.
- [3] Martin T, Li Nan, Timothy L B, et al. Assessing Computational Tools for the Discovery of Transcription Factor Binding Sites[J]. Nature Biotechnology, 2005, 23(1): 137-144.
- [4] Obayshi T, Kinoshita K, Nakai K, et al. ATTED-II: A Database of Co-expressed Genes and CIS Elements for Identifying Co-regulated Gene Groups in Arabidopsis[J]. Nucleic Acids Research, 2007, 35(Database Issue): 863-869.
- [5] Timothy L B, Nadya W, Chris M, et al. MEME: Discovering and Analyzing DNA and Protein Sequence Motifs[J]. Nucleic Acids Research, 2006, 34(Web Server Issue): 369-373.

编辑 金胡考