

IPv6核心路由协议及其部署的研究

张云勇^{1,2} 刘韵洁² 张智江² 李国杰¹ 李忠诚¹

(中国科学院计算技术研究所 北京100080)¹

(中国联合通信有限公司技术部博士后科研工作站 北京100032)²

摘要 IPv6协议由于诸多优良特性,已成为未来互联网的标准。路由协议在所有协议中有着非常关键的作用,本文讨论、比较了IPv6的主要路由协议,并且为了使得网络具有可扩展性,满足网络的短期和长期的发展需求,还阐述了运营级网络核心路由协议的选择。

关键词 路由,开放最短路径优先,中间系统-中间系统,边界网关协议,多拓扑IS-IS

Research of IPv6 Core Routing Protocol and its Deployment

ZHANG Yun-Yong^{1,2} LIU Yun-Jie² ZHANG Zhi-Jiang² LI Guo-Jie¹ LI Zhong-Cheng¹

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)¹

(Postdoctoral Programme, Technology Department, China Unicom, Beijing 100032)²

Abstract IPv6 will be key protocols in future network due to its many unique features and in all IPv6 protocols, routing protocol acts a very important role. In this paper, main routing protocols of IPv6 are discussed and compared. Some suggestions of protocols selection are given out in order to let IPv6 network be scale and meet the network's short term and long term requirements.

Keywords Routing, Open shortest path first, IS-IS, Border gateway protocol, Multi-topology IS-IS

1 引言

随着第一代互联网取得的成功,以互联网为代表的信息网络已经渗透到社会生活的各个领域,成为现代信息社会最重要的基础设施之一。然而,随着网络规模的持续膨胀和新的网络应用需求不断增长,目前互联网发展面临着许多挑战:地址空间匮乏、带宽瓶颈、网络安全漏洞多、服务质量难以保证、互联互通监管难度大、不易开展新业务且难以实现创收的赢利模式、移动性支持有限等。正是在这一背景下,产生了以IPv6^[1,2]技术为基础的下一代互联网。

在应对挑战,给下一代网络服务搭建具有更高性能、更高质量、更加可靠、安全、经济与开放的平台的过程中,下一代互联网协议IPv6具有难以替代的作用。作为IPv6协议中的核心——路由协议起着非常重要的作用,它和IPv4路由协议有一定的区别^[3],为此,需专门加以研究。

2 IPv6路由协议

目前有以下路由协议可以支持IPv6:

2.1 静态路由

静态路由在IPv4中使用非常普遍,而且简单易用、不可缺少。在IPv6中也存在静态路由,只是把IPv4静态路由中的地址段增长到了128位。

2.2 RIPng

RIPng^[4]是在RIPv2的基础上开发的支持IPv6的RIP路由协议。RFC2080(RIPng for IPv6)、RFC2081(RIPng Proto-

col Applicability Statement)详细定义了实现细节。RIPng使用UDP协议521端口作为路由的收发。由于脱胎于RIPv2,因此实现方式基本一致:采用距离矢量作为计算路径的度量值;最长的网络路径受15跳的限制;使用计数到无穷、毒性逆转、水平分割、保持时间等方式来防止路由环。与RIPv2的区别是取消了路由认证(IPv6已内置安全)。

2.3 OSPFv3

OSPFv3^[5]是为IPv6开发的一套链路状态路由协议,大体与支持IPv4的OSPFv2版本相似。RFC2740(OSPF for IPv6)详细定义了实现细节。对比OSPF v2,在OSPF v3中有以下区别:

- 虽然OSPF v3是为IPv6设计的,但是OSPF的路由器ID、区域ID和LSA链路状态ID依然保持IPv4的32位的格式,而不是指定一个IPv6的地址,所以,即使运行OSPF v3也需要为路由器分配IPv4地址。

- 协议的运行是按照每一条链路(Per-link)进行的,而不是按照每个子网进行的(Per-subnet)。

- 把地址域从OSPF包和一些LSA数据包中去除掉,使得成为网络层协议独立的路由协议:①与OSPFv2不同,IPv6的地址不再出现在OSPF包中,而是会在链路状态更新数据包中作为LSA的负载出现;②路由器LSA和网络LSA也不再包含网络地址,而只是简单地表示拓扑信息;③邻居路由器的识别将一直使用Router ID,而不是像OSPF v2一样在某些使用端口会将端口地址作为标识。

- 附加洪泛范围。OSPF v3中定义了三种洪泛范围:链路

张云勇 博士后,主要研究方向为下一代网络与软交换、IPv6及中间件。刘韵洁 教授级高工,主要研究方向为计算机与通信网络。张智江 博士后,教授级高工,主要研究方向为计算机与通信网络、网络安全。李国杰 教授,博导,院士,主要研究方向为计算机系统结构、计算机网络。李忠诚 研究员,博士生导师,主要研究方向为计算机网络、测试、容错计算。

本地、区域和自治系统。

- 支持在一条链路上运行多个 OSPF 进程。对于 OSPF v2, 每一条链路只能运行一个 OSPF 进程, 在 OSPF v3 中进行了增强, 允许多个 OSPF 进程同时运行在一条逻辑链路上。

- 链路本地地址可以作为 OSPF 的转发地址。除了虚拟链路必须使用全局单播地址或者使用站点本地地址。

- 去掉了认证信息。在 OSPF v3 中不再有认证方面的信息。如果需要加密, 可以使用 IPv6 中定义的 IP 认证头来实现。

- OSPF 数据包格式发生了一些变化: ①OSPF 的版本号由 2 变成了 3; ②Hello 包和数据库描述包的选项域增加到 24 位; ③认证域去掉了; ④Hello 信息中不再包含地址信息; ⑤引入了两个新的选项: R 位和 V6 位; ⑥为实现单链路上多 OSPF 进程的实现在, 在 OSPF 包头中加入了 Instance ID 域; ⑦类型 LSA 3 名字改为: Inter-Area-Prefix-LSA, 类型 LSA 4 名字改为: Inter-Area-Router-LSA。

OSPF v2 和 OSPF v3 都使用最短路径优先算法, 在区域划分、链路类型、LSA 传播等方面基本一致。

2.4 中间系统-中间系统(IS-IS)

IS-IS 的发展较早, ISO10589 和 RFC1195 定义了如何使用 IS-IS 作为链路状态路由协议为 TCP/IP 服务。尔后的很多标准逐渐完善了 IS-IS 协议。而最新草案定义了如何使用 IS-IS 来支持 IPv6: draft-ietf-isis-IPv6-05.txt^[6] (Routing IPv6 with IS-IS)。

IS-IS 是一种扩展性很强的路由协议, 所以才能由支持 CLNS 的路由协议扩展到支持 IPv4 的。同样由于 IS-IS 的强健的扩展性使得通过草案 Routing IPv6 with IS-IS 使 IS-IS 支持了 IPv6。

草案 Routing IPv6 with IS-IS 定义了两个新的 TLV: IPv6 Reachability 和 IPv6 Interface Address。

- IPv6 Reachability TLV 该 TLV 的类型值为 236 (0xEC)。该 TLV 效果等同于 IPv4 中的两个 TLV: IP 内部可达 (IP Internal Reachability Information) 和 IP 外部可达 (IP External Reachability Information)。在 IPv6 Reachability TLV 中定义了 Up/Down 位和 External 位, 用来表示路由在 L2/L1 中相互重分发、是否是外部路由。

- IPv6 Interface Address TLV 该 TLV 的类型值为 232 (0xE8)。该 TLV 效果等同于 IPv4 中的 TLV: IP 端口地址 (IP Interface Address)。与原来的 TLV 相区别的是: 在新的 TLV 中地址段的长度由 128 位替换原先的 32 位。

单拓扑 IS-IS 只能运行单个 SPF 算法, 此时 IPv4 与 IPv6 对应的拓扑必须相同, 有一定的限制。而多拓扑 IS-IS 可运行多个 SPF 算法, IPv4 与 IPv6 的拓扑可以不相同, 带来了一定的灵活性。在多拓扑 IS-IS 中, 新增多拓扑 TLV (229)、MT Intermediate Systems TLV (222)、Multi-Topology Reachable IPv4 Prefixes TLV (125) 和 Multi-Topology Reachable IPv6 Prefixes TLV (237)。并且新增了如下道 MT ID 值:

- | | |
|------------|----------------------|
| 0: | 标准拓扑 |
| 1: | IPv4 带内 (in-band) 管理 |
| 2: | 为 IPv6 路由保留 |
| 3: | 为 IPv4 多播路由保留 |
| 4: | 为 IPv6 多播路由保留 |
| 5-3995: | 为 IETF 保留 |
| 3996-4095: | 为开发、试验和专用特性所保留。 |

2.5 BGP4

在 IPv4 环境中, BGP4 是一种广泛使用的用于自治系统

之间路由传播的路径矢量路由协议。在随后定义的一系列标准, 使得 BGP4 的功能更加强大, 可以用于承载多种协议, MPLS-VPN、多播等协议均是通过 BGP4^[7] 进行工作的。

在 RFC2545 (Use of BGP4 MultiProtocol Extensions for IPv6 Inter-Domain Routing) 中描述了如何使用 MP-REACH_NLRI 来传达 IPv6 的可达信息。

3 IPv6 网络路由协议的选择

路由协议的选取要有扩展性, 满足网络的短期和长期的发展。关于路由协议的选择如下:

3.1 网间路由协议

使用 BGP4 域间路由协议, 同时承载 IPv4 和 IPv6。

首先, 主干网的全部路由设备都处于同一个自治系统中, 核心层使用目前合法公共自治系统号码。这样, 在全网可以采用一致的路由策略对网络进行管理。同时, 在国际出口节点可以采用 BGP4 路由协议与国际 IPv6 网络交换路由信息。

另外, 由于两个出口路由器之间没有直连链路, 而且核心节点运行在一个大的 IGP 内, 因此内部自治系统存在 IBGP 非网状连接问题的情况, 需要通过路由反射技术 (RR) 将全部路由信息在核心路由器上同步。

3.2 网内路由协议

3.2.1 OSPFv3 与 IS-IS 的比较 对于支持 IPv6 的链路状态路由协议的选择, 目前包括 OSPFv3 与 IS-IS 两种, 在进行大型网络的内部路由协议的选用时, 不可避免地会产生 OSPF 与 IS-IS 路由协议的争论。从这两种路由协议本身来讲, 这两种路由协议都是被广泛应用的基于链路状态的路由协议。

下面就两种路由协议的优劣进行简要的比较。OSPF 用来交换 IPv6 路由信息的版本叫 OSPFv3, IS-IS 用来交换 IPv6 路由信息的版本叫 IS-ISv6。IS-IS 扩展了两个新的 TLV 来承载 IPv6 的路由信息。同 BGP 一样, IS-IS 可以同时承载 IPv4 和 IPv6 的路由信息。

OSPFv3 运行在 IPv6 网络中, 同 OSPFv2 并不兼容, 但同 OSPFv2 相比, OSPFv3 的机制和选路算法并没有本质的改变。OSPFv3 支持各种不同鉴别机制 (如简单口令验证、MD5 加密验证等), 并且允许各个系统或区域采用互不相同的鉴别机制; 提供负载均衡功能, 如果计算出到某个目的站点有若干条费用相同的路由, OSPFv3 路由器会把通信流量均匀地分配给这几条路由, 沿这几条路由把该分组发送出去。

下面对这两种路由协议进行比较:

- OSPFv3 所遵循的协议已经成为 RFC, IS-ISv6 还处于草案阶段, 将来有协议变化的可能。

- OSPFv3 只能用来交换 IPv6 路由信息, IS-ISv6 可以同时交换 IPv4 路由信息和 IPv6 路由信息。

- 目前, 只有少数厂家实现了 IS-ISv6, 而大多数厂家实现的是 OSPFv3, 如果要在网络中采用多厂商的设备, OSPFv3 更像是一个黏合剂, 可与不同类型的设备连接在一起。

3.2.2 网内路由协议的选择原则 OSPF 与 IS-IS 这两种路由协议没有本质区别, 在一个只有 IPv6 路由协议网络中, 从技术上说, 可以选择任何一种路由协议。

在一个同时存在 IPv4 路由协议和 IPv6 路由协议的网中, 其内部网关协议的选择有以下几种: a) OSPFv2 + OSPFv3; b) IS-ISv4 + OSPFv3; c) OSPFv2 + IS-ISv6; d) IS-ISv4/v6。

其中 c) 项很少在实际中使用, 而 b) 项由于要在同一个网

络中维护两种截然不同的路由协议,也很少使用。a)项需要在同一台路由器中运行两个互相独立的进程,这种机制称为“ships in the night”。而d)项只需要在一台路由器中运行一个进程。显然,a)项需要占用路由器更多资源,但由于a)项可以生成两个独立的IPv4和IPv6的路由表,IPv4和IPv6可以走不同的路由,从而在逻辑上可以形成不同的IPv4的网络拓扑和IPv6的网络拓扑。d)项占用路由器较少的计算资源,但由于IS-ISv6中IPv4和IPv6的拓扑计算是在同一个SPF计算进程中完成的,IPv4和IPv6的流量需要遵循同一个拓扑,在灵活性上不如a)项,同时在试验阶段,如果采用d)项,IPv6路由协议的崩溃将会导致IPv4的同时崩溃,而采用a)项则没有这种问题。如果IPv4的流量和IPv6的流量遵循同一个逻辑拓扑结构,此时采用d)项就会比较方便,如果采用a)项,在网络拓扑修改时,需要同时修改两个协议,容易造成拓扑的不一致,虽然目前有的厂商推出或正在研究IS-IS多拓扑计算,但还没有在国内外的IPv6网络上得到可靠性验证。

3.2.3 网内路由协议的具体选择 可使用IBGP和IS-IS链路状态协议,同时承载IPv4和IPv6,骨干网采用IBGP承载用户路由信息,IS-IS承载互联链路路由信息的路由政策。一是IS-IS路由协议网络可扩展性强,利于以后的扩容,IS-IS具有很好的分层、分域能力,适用于大型网络;二是IS-IS路由协议可只承载网络链路互联信息,不承担用户信息,路由表可以大大地简化和相对稳定。因此,网络链路稳定的情况下,在上百台路由设备的大型网络中应用,也可以满足要求。

对于用户路由信息的承载,选择IBGP可以提高全网路由的稳定性和层次化。因为BGP使用TCP会话保持对端之间的连接,所以具有面向连接的可靠性。另外,由于采用应用层数据报,因此,可以将大量的路由刷新在一次会话中完成,节约了网络带宽。另外,由于用户路由与链路路由的分离,

提高了网络的稳定性和层次化,便于网络维护。但由于BGP路由并不直接反映用户链路状态变化,因此,需要使用路由注入或手工配置的方式将用户可达信息在BGP路由表中传递,这会带来手工操作的工作量。另外,因为用户路由信息与网间互联信息都在BGP中承载,因此需要制订严格的维护规范,避免非法路由广播影响网络的稳定性。为了保证骨干网内路由条数尽量少,减少维护量,建议对地址进行规划,并且边缘路由器在广播IBGP路由时尽量广播完整的地址块。

小结 本文讨论了IPv6网络核心路由协议、选择原则及具体的选择。需注意,一般不推荐核心骨干网在隧道工作方式下运行OSPFv3或IS-ISv6,因为不论是OSPF还是ISIS不像BGP是通过TCP会话交换路由更新,在IPv4网络环境恶化的情况下,会引起SPF的频繁计算,造成IGP路由的不稳定。另外,实际的网络建设中,还需考虑网间国际出口BGP路由策略、网间国内出口BGP路由策略、网内IBGP路由策略、网内IGP路由策略、用户路由策略等,需对核心路由协议给予足够的重视。

致谢 感谢中国联合通信有限公司技术部技术开发处王明会博士、杨征先生与作者有益的讨论。

参考文献

- 1 Deering S, Hinden R. Internet Protocol. Version 6 (IPv6) Specification. RFC2460, 1998
- 2 Huitema C. 新因特网协议 IPv6. 陶文星等译. 北京:清华大学出版社, 1999
- 3 RFC2185. Routing Aspects of IPv6 Transition. 1997
- 4 RFC 2080. RIPng for IPv6, 1997
- 5 RFC2740. OSPF for IPv6, 1999
- 6 draft-ietf-isis-ipv6-02.txt, Routing IPv6 with IS-IS, 1999
- 7 Bates T, Chandra R, Katz D, Rekhter Y. Multiprotocol Extensions for BGP-4. RFC 2283, Feb. 1998

(上接第25页)

假设VFS的磁盘中存储200个媒体对象,则根据zipf分布的规律,用户请求的媒体对象位于VFS上的概率为 $\sum_{n=1}^{200} \frac{C}{n} = 0.7183$, (其中 $C = 1 / \sum_{i=1}^{2000} \frac{1}{i}$)。采用普通HSM系统时,有28.17%用户请求需要等待。

采用基于预置的数据放置策略后,在每小时100个用户请求的负载下(对应于 $\rho(4) = 0.33$),访问HSM的请求中有70.5%可以即时响应。因此,需要等待的用户点播请求占 $0.2813 * (1 - 0.705) = 0.083$ 。

若以增加VFS磁盘空间的方式使同样91.7%的请求得到即时响应,则由于 $\sum_{n=1}^{1015} \frac{C}{n} = 0.917$, (其中 $C = 1 / \sum_{i=1}^{2000} \frac{1}{i}$), VFS需要存放1015个媒体对象,相应的需要增加的磁盘空间为 $(1015 - 200) * 2G = 1.6T$ 。

总结 本文通过剖析三级存储设备的访问特性和媒体服务系统的特点,提出HSM系统中面向流媒体服务的一种数据放置策略。分析和仿真试验表明,采用这种策略,只需付出较小的存储代价,就可以使得HSM系统具有较高的即时响应性能,从而大大提高整个媒体服务系统的性能。

参考文献

- 1 Barnett S A, Anido G J. A Cost Comparison of Distributed and

- Centralized Approaches to Video-on-Demand. IEEE Journal on Selected Areas in Communications, 1996, 14(6): 1173~1183
- 2 Brubeck D W, Rowe L A. Hierarchical storage management in a distributed VOD system. IEEE Multimedia, 1996, 3(3): 37~47
- 3 Christodoulakis S, Triantafyllou P, Zioga F. Principles of optimally placing data in tertiary storage libraries. In: Proc. of Intl. Conf. on Very Large Data Bases, Athens, Greece, 1997. 236~245
- 4 Li J, Prabhakar S. Data Placement for Tertiary Storage. In: the 19th IEEE Symposium on Mass Storage Systems, Maryland, 2002
- 5 Ghandeharizadeh S, Dashti A, Shahabi C. A pipelining mechanism to minimize the latency time in hierarchical multimedia storage managers. Computer communications, 1995, 18(3): 170~184
- 6 Prabhakar S, Chari R. Minimizing Latency and Jitter for Large Scale Multimedia Repositories through Prefix Caching. International Journal on Image and Graphics (IJIG), 2003, 3(1): 95~117
- 7 Johnson T, Miller E L. Performance measurements of tertiary storage devices. In: Proc. of 24th Intl. Conf. on Very Large Data Bases, New York, 1998. 50~61
- 8 Yang Daoliang, Ren Xiaoxia, Chang Ming. Study on Data Replacement Algorithm in Continuous Media Server with Hierarchical Storage. In: 16th IFIP World Computer Congress, Beijing, Aug. 2000
- 9 Fishwick P A. Simpack: Getting Started with Simulation Programming in C and C++. In: 1992 Winter Simulation Conf. Arlington, VA, 1992. 154~162
- 10 SimPack Toolkit. <http://www.cise.ufl.edu/~fishwick/simpack/simpack.html>