OSPFv3 路由协议在 FreeBSD 下的实现研究*)

孙庆南[1,2]　鲁士文[1]

(中国科学院计算技术研究所　北京 100080)[1]　(中国科学院研究生院　北京 100039)[2]

摘　要　随着 IPv6 协议在 Internet 中的广泛应用，人们更多地将注意力集中在路由器对 IPv6 协议的支持上。设计和开发基于 IPv6 的路由协议软件也更加重要。本文分析了 OSPFv2 路由协议与 OSPFv3 路由协议之间的不同，并且基于 RFC2740 设计开发了一个 OSPFv3 路由协议软件，该软件在实验网络环境中的运行取得了良好的效果。

关键词　IPv6，OSPFv2，OSPFv3，FreeBSD，路由协议软件

Research on Implemention of OSPFv3 Routing Protocol Software on FreeBSD Operating System

SUN Qing-Nan[1,2]　LU Shi-Wen[1]

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

(Graduate School of the Chinese Academy of Sciences, Beijing 100039)[2]

Abstract　With explosive expanding of IPv6 protocol used in Internet, people pay more attention on routers with IPv6 support. The area of designing and developing routing protocol based on IPv6 is more important. This paper analyzes the differences between OSPFv2 and OSPFv3, and designs an OSPFv3 routing software according to RFC2740. The software is excuted well in experimental network environment.

Keywords　IPv6, OSPFv2, OSPFv3, FreeBSD, Routing protocol software

1　前言

随着 Internet 技术在全球范围的飞速发展，OSPF 已成为目前 Internet 广域网和 Intranet 企业网采用最多、应用最广泛的路由协议之一。OSPF(Open Shortest Path First)路由协议由 IETF(Internet Engineering Task Force)IGP 工作小组提出，是一种基于 SPF(最短路径优先)算法的路由协议，在 IPv4 网络中，目前常用的 OSPF 协议版本号为 2[1]，随着 IPv6 网络在全球的部署，我国也于 2004 年建立了第一个 IPv6 主干网—CERNET2，OSPF 协议在 IPv6 网络上的实现及其行为研究也成为热点问题。目前 OSPF for IPv6 的版本为 OSPFv3[3]。本文针对 OSPFv3 路由协议和 OSPFv2 路由协议的不同点进行了分析，对其在 FreeBSD 操作系统上的实现方案进行了研究和设计，并编写了相应的软件。

2　OSPFV3 路由协议的新特性

OSPFv2 中的大部分特性在 OSPFv3 中都得以保留。然而 OSPFv3 对这些特性也进行了一些必要的改变，其中一部分改变是由于 IPv4 与 IPv6 协议的特性不同而引起的，另外一部分则是为了处理 IPv6 长地址而进行的简单改变。

2.1　基于链路的协议

OSPFv2 协议规范基于 IP 子网，而在 OSPFv3 中，对应 OSPFv2 协议规范中的“网络”、“子网”等词语都被“链路”所替换。在 IPv6 规范中这样定义术语“链路”：它是一种通讯设备或者介质，节点之间可以通过它在链路层相互通讯。在 OSPFv3 中，路由器接口连接到一条链路上，而不是连接到一个子网。

2.2　去除了地址语义

除了链路状态更新分组中的 LSA 载荷之外，IPv6 的地址不再出现在 OSPF 分组中。路由器 LSA 和网络 LSA 不再包含网络地址，但是包含简单的拓扑信息。OSPF 路由器 ID、区域 ID 和 LSA 链路状态 ID 保留为 IPv4 的 32 位大小，它们不能被赋值为 IPv6 地址。在 OSPFv3 中，邻接路由器由路由器 ID 区分，而不是 IPv4 中的按照广播的 IP 地址和 NBMA 网络区分。

2.3　LSA 洪泛范围增加

LSA 的洪泛范围已经被扩展为明确的三种独立的 LSA 洪泛范围，分别是本地链路范围、区域内部范围和自治域内部范围。

2.4　支持在一条链路上运行多个实例

可以通过 OSPFv3 分组头和接口结构中的实例 ID 在一条链路上运行多个 OSPFv3 实例，这样运营商可以直接在同一个物理网段上运行多个不同的 OSPF 域。

2.5　对本地链路地址的使用

IPv6 本地链路地址用在单独链路，邻居发现和自动配置等问题上。IPv6 路由器不转发那些有本地链路源地址的数据报。本地链路单播地址在 IPv6 地址范围 FF80/10 内选取。

2.6　验证的改变

在 OSPF for IPv6 中，将验证工作从 OSPF 本身移出，“AuType”字段和“Authentication”字段也从 OSPF 分组头中被去除。所有验证相关的字段都从 OSPF 分组格式和接口结构中去掉了。当在 IPv6 上运行时，OSPF 通过信赖 IP 分组的验证头以及 IP 封装的安全载荷来保证路由信息交换的完整性和机密性。

*)基金项目：国家高技术研究发展计划(863 计划)资助项目。孙庆南　博士研究生，主要研究方向为路由议和包转发技术；鲁士文　教授，博导，主要研究方向为网络协议和服务质量保证机制。

2.7 分组格式的改变

OSPF for IPv6 直接运行在 IPv6 上。除此之外，所有的地址语义都已经从 OSPF 分组头中去掉了，使得它成为网络协议无关的路由协议。

2.8 LSA 格式的改变

从 LSA 头和 Router-LSA，Network-LSA 中去掉了地址语义。这两个 LSA 描述了网络协议无关行为中的路由选择域的拓扑逻辑。加入了新的 LSA 来描述 IPv6 地址信息和下一跳解析所需的数据。此外还改变了一些 IPv4 的 LSA 名字以便彼此之间更加一致。

2.9 对未知 LSA 类型的处理

支持 IPv4 的 OSPFv2 协议简单地忽略了那些不支持的 LSA 类型。但是，为了使得运行于同一链路上的不同路由器之间相互兼容，在 OSPFv3 协议中，未知的 LSA 类型可以被路由器存储和洪泛，这样更大地扩展了路由器的兼容性。

2.10 Stub 区域的支持

在 OSPFv2 协议中，stub 区域的作用是最小化区域内路由器的链路状态数据库和路由表大小。这使得路由器可以用最小的资源来处理非常大的 OSPF 路由选择域。

在 OSPFv3 中保留了 stub 的概念。然而，不像在 OSPFv2 中，OSPFv3 允许有着不明 LS 类型的 LSA 被标记上“若类型不明，存储并且洪泛该 LSA”，当 LSA 的洪泛范围是区域内部或者本地链路时，或者 LSA 的 U-bit 置 0 时，一个不明 LS 类型的 LSA 可以在一个 stub 区域内洪泛。

2.11 通过路由器 ID 识别邻居

在 OSPFv3 中，一条给定链路上的邻接路由器总是通过它们的 OSPF 路由器 ID 相区分的。在 OSPFv2 中，在点到点网络和虚链路上的邻接路由器是通过它们的路由器 ID 区分的，而广播、NBMA 和点到多点链路上，是通过它们的 IPv4 接口地址区分的。

2.12 OSPFv3 的 IPv6 封装

OSPF 直接运行在 IPv6 的网络层上，因此，OSPF 分组直接被 IPv6 分组包装。OSPF 并没有定义一种拆分协议分组的方法，而只是当分组的大小超过链路 MTU 时依靠 IPv6 提供的拆分功能。OSPF 的 IPv6 封装有两个重要特征：

1)一些 OSPF 消息在网络上广播时使用了 IPv6 的组播方式。组播使用了两个特殊的 IP 组播地址。发往这两个地址的分组不被转发，也就是说这些分组只在网络上传输一跳的距离。因此这些分组的 IPv6 Hop Limit 应该被设为 1。这两个地址分别为：

A) AllSPFRouters

这个组播地址的值为 FF02::5。所有运行 OSPFv3 协议的路由器都要准备接收发往这个地址的数据分组。Hello 消息也总是发往这个地址。而且，在洪泛过程中，某些分组也要发往这个地址。

B) AllDRouters

这个组播地址值为 FF02::6。指定路由器和备份指定路由器都必须准备接收发往这个地址的数据分组。在洪泛过程中，某些分组也要发往这个地址。

2)在封装 OSPF 的 IPv6 头的 Next Header field 中设置为 89。

3 OSPFv3 路由协议软件的主要组成部分

OSPFv3 路由协议软件的主模块包括消息发送接收接口模块、消息接收处理模块、邻居路由器关系模块、路由器接口模块和链路状态数据库模块等。它们之间的相互调用关系如图 1 所示。

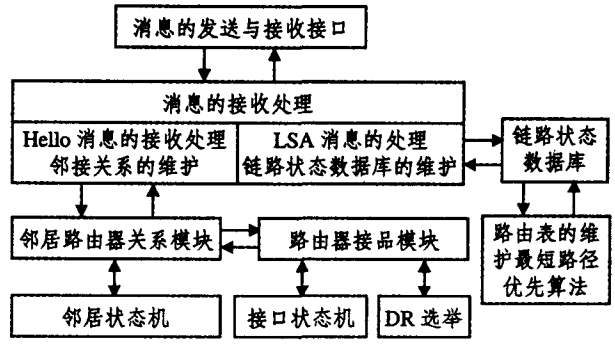

图 1 OSPFv3 路由协议软件主模块关系图

3.1 消息的发送与接收接口模块

消息的发送与接收接口模块提供了与下层网络协议的接口，用于实现 OSPFv3 协议消息的发送与接收，并且将接收到的消息分类交给消息接收处理模块进行进一步的操作。它提供的主要接口包括：

a) Hello 消息的发送与接收；

b)数据库描述分组的发送与接收；

c)链路状态请求分组的发送与接收；

d)链路状态更新分组的发送与接收。

3.2 消息的接收处理

消息接收处理模块主要用于处理 OSPFv3 路由协议的两大消息类型：Hello 消息和 LSA 消息。其中 Hello 消息用于维护邻居路由器的关系，LSA 消息用于对链路状态数据库的维护。邻居状态维护模块、接口状态维护模块和链路状态数据库维护模块对消息的接收处理模块提供了支持。

3.3 邻居的状态维护，邻居状态机

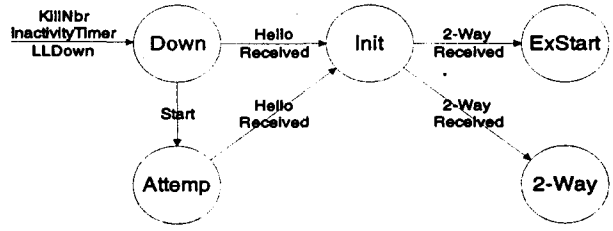

图 2 Hello 消息导致的邻居状态转换

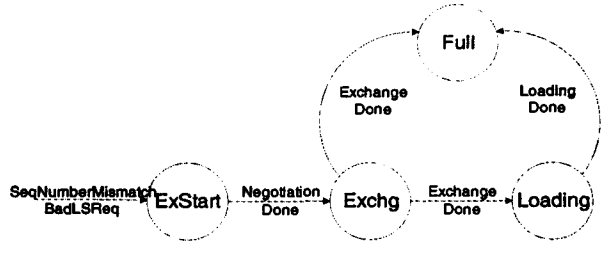

图 3 数据库描述消息导致的邻居状态转换

邻居状态机的状态转换图包括两个部分，一是由于接收到 Hello 消息导致的状态转换，一是由于接收到数据库描述消息导致的状态转换。根据邻居路由器的状态，在处理接收到的 LSA 消息时进行不同的处理。邻居状态机的状态转换图如图 2、图 3 所示。

3.4 接口的状态维护,接口状态机

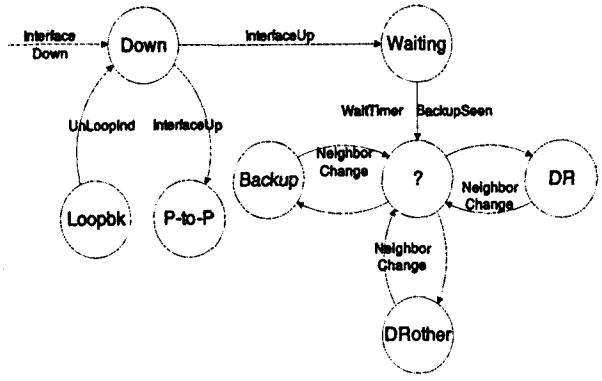

图 4 接口状态转换图

OSPFv3 路由协议中,定义接口为路由器与链路之间的连接,所有路由器产生的路由协议分组都要标记接口的 Area ID。在接口的状态转换中,完成区域中指定路由器的选举。接口的状态转换图如图 4 所示。

3.5 链路状态数据库的维护、路由表的计算与最短路径优先算法

链路状态数据库维护模块对链路状态数据库的添加、删除、查找和更新等操作提供了接口。路由表的计算也要使用最短路径优先算法对链路状态数据库进行操作。作为 OSPFv3 路由协议的核心,最短路径优先算法的流程图如图 5 所示。

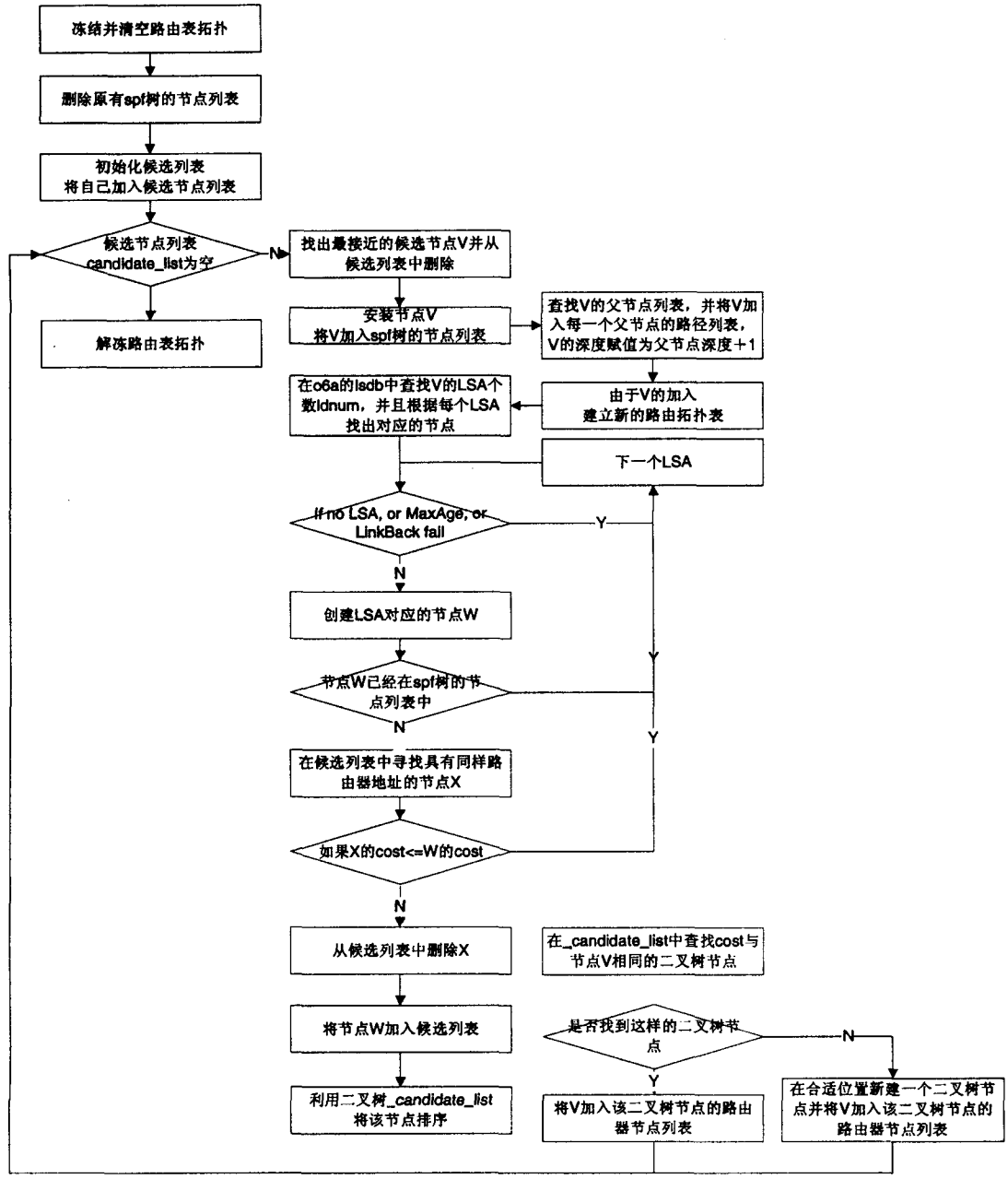

图 5 最短路径优先算法流程图

OSPF 的路由表是通过它的链路状态报告 LSA 来完成的。路由表的路由可以分类三类:区域内的路由、AS 内部区域与区域之间的路由和通往其他 AS 的路由。路由器在生成路由表的时候,首先在域内要生成一个最短路径树。路由器

首先找到由本地路由器生成的路由器 LSA 或网络 LSA。如果本地路由器是网络指定路由器，检查在这些 LSA 中每个条目所描述的对象，如果这个对象不在树上，则生成一条新的路由到这个对象的路由条目，否则就将其与之进行比较，保留代价较小的除去旧的代价更大的路由条目，并将以前树中的那一项删除掉。同时，如果对象描述的是一个路由器，就将其放入一个队列中。处理完本地路由器生成的域内 LSA 的所有条目后，算法再从队列中取出一个路由器，同样也在本地的数据库中找到它生成的域内 LSA，重新上面的步骤。在处理域内路由器及其生成的域内 LSA 后，此时就得到了全部的域内路由表。

然后依次计算本地路由器数据库中的网络汇总 LSA、ASBR 汇总 LSA 和 AS 外部 LSA 得到整个的路由表。

4 OSPFV3 路由协议软件及实验环境

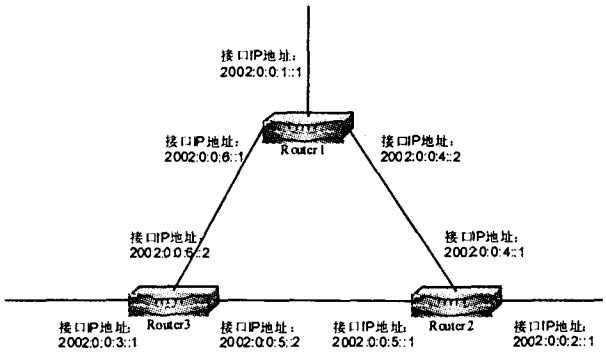

图 6 OSPFv3 路由协议软件实验环境拓扑图

基于 RFC2740 和 RFC2328 编写的 OSPFv3 路由协议软件运行于 FreeBSD 操作系统之上，所使用的验证系统实验平台拓扑图如图 6 所示。其中，实验环境的路由器系统由三台 PC 机组成，它们的硬件环境和软件环境分别如表 1 所示，各路由器之间以及路由器与终端主机之间通过 100M 以太网连接。

表 1 实验环境路由器软硬件配置

路由器编号	CPU	内存	操作系统
Router1	Pentium4 2.4G	512M DDR	FreeBSD 4.7-RELEASE
Router2	Pentium3 800M	512M SDRAM	FreeBSD 4.7-RELEASE
Router3	Pentium4 1.4G	512M SDRAM	FreeBSD 4.7-RELEASE

经过多次实验，当路由软件启动后，可以有效地建立路由器内部的路由转发规则。当一条转发通路上的某条链路不可用时，平均经过 47.1 秒的时间，路由器可以重建路由表，选择通过其它路由器转发数据分组，实现了路由选择协议的主要功能。

小结 随着 IPv6 协议日益广泛的应用，路由器作为网络互连的核心设备，它对于 IPv6 的支持程度成为人们所关心的问题。本文在对比分析 OSPFv3 路由协议的基础上，给出了一种 OSPFv3 路由协议软件的实现方案，并且通过实验验证取得了良好的效果，这对于进一步研发具有自主知识产权的 IPv6 路由器有着重要的意义。

参 考 文 献

1 Moy J T. OSPF Version 2. Request for Comments 2328, April 1998
2 Coltun R. The OSPF Opaque LSA Option. RFC2370, July 1998
3 Coltun R, Ferguson D, Moy J. OSPF for IPv6. Request for Comments 2740, Dec. 1999
4 Shaikh A, Goyal M, Greenberg A, Rajan R, RamakrishnanK K. An OPSF Topology Server: Design and Evaluation. IEEE J. Selected Areas in Communications, 2002,20(4)

(上接第 10 页)

4.4 其他方法

通过在算法更新或是重建算法数据结构的预处理过程中进行预计算和添加相应指针，可以在一定程度上避免报文查找过程中的回溯；算法的数据结构建立后，通过结点合并、路径压缩、提取公共子集消除冗余措施等预处理，可以对算法进一步改善，也有助于提高算法处理的空间和时间效率。但是这种方法通常会使更新复杂化，因此较适用于规则更新不频繁的环境中，或是用于优化原本就已需要在更新时重建整个数据结构的算法。

总结 Internet 的带宽和速度的增长要求网络单元要能在单位时间内能够处理更多数目的报文。而随着各种网络应用的发展，未来的网络必须为用户提供更多的服务类型和更好的服务质量，因此高性能 IP 报文分类算法的研究十分必要。最近几年以来，各种不同的 IP 路报文分类算法被提出，本文对 IP 报文分类算法进行了概述和总结，并在分类速度、更新速度、存储耗费、适用范围、对分类维度和规则数目的可扩展性、规则表示的灵活性等方面进行比较，并对 IP 报文分类算法研究的方法和趋势进行了分析和总结。IP 报文分类算法是众多网络上层服务和功能的基石，对它的回顾和总结将帮助我们加深 IP 报文分类问题的理解，也有助于对此问题的进一步研究。

参 考 文 献

1 Buddhikot M M, Suri S, Waldvogel M. Space decomposition techniques for fast layer-4 switching. In: Proc. of Conf. on Protocols for High Speed Networks, August 1999. 25～41
2 Knuth D E. The art of computer programming. vol3: sorting and searching, Addison-wesley, 3rd editon. 1998
3 Tsuchiya P. A search algorithm for table enTries with non-contiguous wildcarding: [unpublished report]. Bellcore
4 Srinivasan V, et al. Fast and scalable layer 4 switching. In: Proc. of ACM Sigcomm'98, september 1998
5 Baboescu F, Singh S, Varghese G. Packet Classification for Core Routers: Is there an alternative to CAMs? in INFOCOM, 2003
6 Feldman A, Muthukrishnan S. Tradeoffs for packet classification, In: Proc. of Infocom, March 2000,3:1193～202
7 Gupta P, McKeown N. Packet Classification using Hierarchical Intelligent Cuttings. In: Proc. Hot Interconnects VII, August 99, Stanford. This paper is also available in IEEE Micro, January/ February 2000,20(1):34～41
8 Singh S, Baboescu F, Varghese G, Wang J. Packet Classification Using Multidimensional Cutting. In: Proc. of ACM SIGCOMM, Karlsruhe, Germany, August 2003
9 Gupta P, McKeown N. Packet Classification on Multiple Fields. In: Proc. Sigcomm, Computer Communication Review, Sept. 1999,29(4):147～60
10 Srinivasan V, Suri S, Varghese G. Packet classification using tuple space search. The ACM Sigcomm'99, 1999
11 Baboescu F, Varghese G. Scalable packet classification in Proc of ACM Sigcomm'01, september 2001
12 Lakshman T V, Stidialis D. High speed policy-based packet forwarding using efficient multi-dimensional range matching. In: Proc. of ACM Sigcomm '98, sept. 1998
13 Kounavis M E, Kumar A, Vin H, Yavatkar R, Campbell A T. Directions in Packet Classification for Network Processors. Second Workshop on Network Processors (NP2), Anaheim, California, February, 2003. 8～9