[19] 中华人民共和国国家知识产权局

[12] 发明专利申请公布说明书

[21] 申请号 200580034130.6

[51] Int. Cl.
H04L 12/56 (2006.01)
H04J 3/22 (2006.01)
H04L 12/28 (2006.01)
G06F 15/173 (2006.01)
H04J 3/16 (2006.01)
G06F 15/16 (2006.01)

[43] 公开日 2007 年 9 月 12 日

[11] 公开号 CN 101036355A

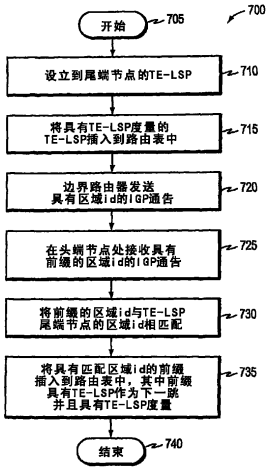
[22] 申请日 2005.11.17
[21] 申请号 200580034130.6
[30] 优先权
[32] 2004.12.1 [33] US [31] 11/001,459
[86] 国际申请 PCT/US2005/041797 2005.11.17
[87] 国际公布 WO2006/060184 英 2006.6.8
[85] 进入国家阶段日期 2007.4.6
[71] 申请人 思科技术公司
地址 美国加利福尼亚州
[72] 发明人 让 - 菲利普·瓦瑟尔
斯特凡诺·B·普雷维蒂
[74] 专利代理机构 北京东方亿思知识产权代理有限责任公司
代理人 王 怡
权利要求书 4 页 说明书 20 页 附图 9 页

[54] 发明名称

具有 IGP 扩展的域间 TE - LSP

[57] 摘要

本发明公开了一种用于在计算机网络中将流量工程(TE)标签交换路径(LSP)的尾端节点的可达性信息传播到 TE - LSP 的头端节点的技术。 TE - LSP 优选地跨越网络的多个域，从而使尾端节点驻留在与头端节点的域不同(远离头端节点的域)的某一域中。 域间信息传播技术采用内部网关协议(IGP)来将可达性信息从驻留在与尾端节点相同的域中的目标节点发送到头端节点。 头端节点使用远程信息来计算可以从尾端节点到达的路由(即，地址前缀和关联属性)，以插入到其路由表中。



1. 一种用于在计算机网络中将流量工程（TE）标签交换路径（LSP）的尾端节点的可达性信息传播到所述 TE-LSP 的头端节点的方法，所述 TE-LSP 跨越所述网络的多个域，从而使所述尾端节点驻留在远离所述头端节点的域的某一域中，所述方法包括：

经由内部网关协议（IGP）通告将所述可达性信息从驻留在所述尾端节点的域中的目标节点发送到所述头端节点；

在所述头端节点处接收所发送的可达性信息；

在所述头端节点处利用所接收的信息计算可以从所述尾端节点到达的路由；以及

将所计算出的路由插入到所述头端节点的路由表中。

2. 如权利要求 1 所述的方法，还包括：利用区域标识符（区域 id）数据结构通告从所述目标节点可达的地址前缀，所述区域 id 数据结构标识了每个前缀所属的域。

3. 如权利要求 2 所述的方法，还包括：在所述 IGP 通告内发送扩展可达性对象数据结构，所述扩展可达性对象数据结构包含所述区域 id 数据结构。

4. 如权利要求 1 所述的方法，还包括：利用节点 id 数据结构通告所述尾端节点的域中下游的可达地址前缀，所述节点 id 数据结构将所述尾端节点标识为发起所述前缀。

5. 如权利要求 4 所述的方法，还包括：跨域地在所述 IGP 通告内发送扩展可达性对象数据结构，所述扩展可达性对象数据结构包含所述节点 id 数据结构。

6. 如权利要求 5 所述的方法，还包括：利用前缀成本数据结构通告从所述尾端节点可到达的每个地址前缀的度量，所述前缀成本数据结构包含了由不发起所述前缀的节点计算的度量。

7. 如权利要求 6 所述的方法，还包括：跨域地在所述 IGP 通告内发送所述扩展可达性对象数据结构，所述扩展可达性对象数据结构包含所述前

缀成本数据结构。

8. 如权利要求 1 所述的方法，还包括：在计算路由之前将所返回的请求信息存储在所述头端节点处的影子路由表中。

9. 如权利要求 1 所述的方法，其中所述 TE-LSP 跨越多个区域。

10. 一种用于在计算机网络中将流量工程（TE）标签交换路径（LSP）的尾端节点的可达性信息传播到所述 TE-LSP 的头端节点的系统，所述 TE-LSP 跨越所述网络的多个域，从而使所述尾端节点驻留在远离所述头端节点的域的某一域中，所述系统包括：

适合于将所述可达性信息从驻留在所述尾端节点的域中的目标节点传送到所述头端节点的 IGP 通告；

被配置为在所述头端节点处存储所发送的可达性信息的影子表；

在所述头端节点上执行的路由信息库（RIB），其被配置为利用所接收的信息计算可以从所述尾端节点到达的路由；以及

被配置为存储所计算的路由的所述头端节点的路由表。

11. 如权利要求 10 所述的系统，还包括：用于通告从所述目标节点可达的地址前缀的区域标识符（区域 id）数据结构，其中所述区域 id 数据结构标识了每个前缀所属的域。

12. 如权利要求 11 所述的系统，还包括：所述 IGP 通告内用于发送所述区域 id 数据结构的扩展可达性对象数据结构。

13. 如权利要求 10 所述的系统，还包括：用于通告所述尾端节点的域中下游的可达地址前缀的节点 id 数据结构，其中所述节点 id 数据结构将所述尾端节点标识为发起所述前缀。

14. 如权利要求 13 所述的系统，还包括：所述 IGP 通告内用于跨域发送所述节点 id 数据结构的扩展可达性对象数据结构。

15. 如权利要求 10 所述的系统，还包括：通告从所述尾端节点可达的每个地址前缀的度量的前缀成本数据结构，其中所述前缀成本数据结构包含由不发起所述前缀的节点计算的度量。

16. 如权利要求 15 所述的系统，还包括：所述 IGP 通告内用于跨域发送所述前缀成本数据结构的扩展可达性对象数据结构。

17. 如权利要求 10 所述的系统，还包括：在所述头端节点处用于在计算路由之前存储所返回的请求信息的影子路由表。

18. 一种用于在计算机网络中将流量工程（TE）标签交换路径（LSP）的尾端节点的可达性信息传播到所述 TE-LSP 的头端节点的装置，所述 TE-LSP 跨越所述网络的多个域，从而使所述尾端节点驻留在远离所述头端节点的域的某一域中，所述装置包括：

用于经由内部网关协议（IGP）通告将所述可达性信息从驻留在所述尾端节点的域中的目标节点发送到所述头端节点的装置；

用于在所述头端节点处接收所发送的可达性信息的装置；

用于在所述头端节点处利用所接收的信息计算可以从所述尾端节点到达的路由的装置；以及

用于将所计算出的路由插入到所述头端节点的路由表中的装置。

19. 一种包含可执行程序指令的计算机可读介质，所述程序指令用于在计算机网络中将流量工程（TE）标签交换路径（LSP）的尾端节点的可达性信息传播到所述 TE-LSP 的头端节点，所述 TE-LSP 跨越所述网络的多个域，从而使所述尾端节点驻留在远离所述头端节点的域的某一域中，所述可执行程序指令包括用于执行以下步骤的程序指令：

经由内部网关协议（IGP）通告将所述可达性信息从驻留在所述尾端节点的域中的目标节点发送到所述头端节点；

在所述头端节点处接收所发送的可达性信息；

在所述头端节点处利用所接收的信息计算可以从所述尾端节点到达的路由；以及

将所计算出的路由插入到所述头端节点的路由表中。

20. 一种用于在计算机网络中将流量工程（TE）标签交换路径（LSP）的尾端节点的可达性信息传播到所述 TE-LSP 的头端节点的扩展可达性对象数据结构，所述 TE-LSP 跨越所述网络的多个域，从而使所述尾端节点驻留在远离所述头端节点的域的某一域中，所述数据结构包括：

用于通告从所述目标节点可达的地址前缀的区域标识符（区域 id）数据结构字段，其中所述区域 id 数据结构标识了每个前缀所属的域；

用于通告所述尾端节点的域中下游的可达地址前缀的节点 id 数据结构字段，其中所述节点 id 数据结构将所述尾端节点标识为发起所述前缀；以及

用于通告从所述尾端节点可达的每个地址前缀的度量的前缀成本数据结构，其中所述前缀成本数据结构包含由不发起所述前缀的节点计算的度量。

具有 IGP 扩展的域间 TE-LSP

技术领域

本发明涉及计算机网络，更具体而言，涉及跨计算机网络的域取得可达性信息。

背景技术

计算机网络是由通信链路和网段互连的地理上分布的节点的集合，通信链路和网段用于在诸如个人计算机和工作站之类的末端节点之间传输数据。许多类型的网络是可用的，其类型范围从局域网（LAN）到广域网（WAN）不等。LAN 一般通过位于同一个总的物理位置（例如建筑物或校园）中的专用私有通信链路连接节点。另一方面，WAN 一般通过长距离通信链路连接地理上分散的节点，所述长距离通信链路例如是公共载波电话线、光路径、同步光网络（SONET）或同步数字层级体系（SDH）链路。因特网是连接全世界的不同网络的 WAN 的一个示例，其提供了各种网络上的节点之间的全球通信。节点一般通过根据预定义协议交换离散数据分组或帧来经由网络通信，所述预定义协议例如是传输控制协议/因特网协议（TCP/IP）。在这种上下文中，协议由定义节点彼此之间如何交互的一组规则构成。计算机网络还可以由诸如路由器之类的中间网络节点互连，以扩展每个网络的有效“大小”。

由于可以证明互连的计算机网络的管理是很繁重的，因此较小的计算机网络群组可以被维护作为路由域或自治系统。自治系统（AS）内的网络一般通过传统的被配置为执行域内路由协议的“域内”路由器耦合在一起，并且通常受制于一个共同的权力机构。为了提高路由可缩放性，服务提供商（例如 ISP）可以将 AS 划分为多个“区域”。然而，可能希望增大能够交换数据的节点的数目；在这种情况下，执行域间路由协议的域间路由器被用于互连各种 AS 的节点。而且，可能希望互连工作在不同管理

域下的各种 AS。这里所用的 AS 或（更具体而言）区域通常被称为“域（domain）”，并且将不同域互连在一起的路由器通常被称为“边界路由器”。

域间路由协议的示例是边界网关协议版本 4（BGP），其通过在系统的邻居域间路由器之间交换路由和可达性信息来执行域（AS）间的路由。相邻（adjacency）是一种在选定的邻居（对等）路由器之间形成的关系，其目的是交换路由信息消息并提取网络拓扑。由 BGP 对等路由器交换的路由信息一般包括目的地地址前缀，即，目的地地址中被路由协议用来进行路由（“下一跳”）判决的部分。这种目的地地址的示例包括 IP 版本 4（IPv4）和版本 6（IPv6）地址。BGP 通常工作在诸如 TCP 之类的可靠的传输协议上，以建立 TCP 连接/会话。BGP 协议是公知的，并且在 1995 年 3 月公布的题为“*A Border Gateway Protocol 4 (BGP-4)*”的注释请求（RFC）1771 中有一般描述。

域内路由协议或内部网关协议（IGP）的示例是开放最短路径优先（OSPF）路由协议和中间系统到中间系统（ISIS）路由协议。OSPF 和 ISIS 协议是基于链路状态技术的，因此被统称为链路状态路由协议。链路状态协议定义了路由信息和网络拓扑信息在域中被交换和处理的方式。该信息通常涉及域内路由器的本地状态（例如，路由器的可用接口和可达邻居或相邻情况）。OSPF 协议在 1998 年 4 月的题为“*OSPF Version 2*”的 RFC 2328 中有所描述，用在 IP 上下文中的 ISIS 协议在 1990 年 12 月的题为“*Use of OSI ISIS for routing in TCP/IP and Dual Environments*”的 RFC 1195 中有所描述，这两篇 RFC 都通过引用结合于此。

中间网络节点经常将其路由信息存储在由路由信息库（RIB）维护和管理的路由表中。路由表是一种可搜索数据结构，其中网络地址被映射到其关联路由信息。然而，本领域技术人员将理解路由表不需要被组织为表，而是也可以是另一类型的可搜索数据结构。尽管中间网络节点的路由表可以被配置具有预定的一组路由信息，但是该节点也可以在其发送和接收数据分组时动态获取（“学习”）网络路由信息。当在中间网络节点处接收到分组时，分组的目的地地址可以用于标识包含与接收的分组相关联

的路由信息的路由表条目。分组的路由信息指示分组的下一跳地址等等。

为了确保其路由表包含最新路由信息，中间网络节点可以与其他中间节点协作以散布代表当前网络拓扑的路由信息。例如，假定中间网络节点检测到其邻居节点（即，相邻网络节点）之一变得不可用，例如由于链路故障或邻居节点“离线”等等。在这种情形下，中间网络节点可以更新存储在其路由表中的路由信息，以确保数据分组不被路由到不可用的网络节点。此外，中间节点还可以将该网络拓扑的变化传输给其他中间网络节点以使得其可以更新它们的本地路由表并绕过不可用节点。以这种方式，每个中间网络节点“知道”拓扑的变化。

一般来说，路由信息是根据预定的网络通信协议，例如链路状态协议（例如 IS-IS 或 OSPF）在中间网络节点之间散布的。传统的链路状态协议使用链路状态分组（或“IGP 通告”）来在互连的中间网络节点（IGP 节点）之间交换路由信息。这里所用的 IGP 通告通常描述 IGP 路由协议用来在互连的 IGP 节点（即，路由器和交换机）之间传输路由信息的任何消息。在工作中，第一 IGP 节点可以生成 IGP 通告，并通过其每个耦合到其他 IGP 节点的网络接口“流播（flood）”（即，发送）分组。其后，第二 IGP 节点可以接收被流播的 IGP 通告，并基于包含在接收的 IGP 通告中的路由信息更新其路由表。随后，第二 IGP 节点可以通过其每个网络接口流播接收的 IGP 通告，但是接收到 IGP 通告的接口除外。该流播过程可以重复，直到每个互连的 IGP 节点都已接收到 IGP 通告并更新其本地路由表为止。

在实际中，每个 IGP 节点一般生成并散布这样的 IGP 通告，其路由信息包括中间节点的邻居网络节点的列表以及与每个邻居相关联的一个或多个“成本”值。这里所用的与邻居节点相关联的成本值是用于确定与该节点进行通信的相对难易程度的任意度量。例如，成本值可以按照到达邻居节点所需的跳数、分组到达邻居节点的平均时间、耦合到邻居节点的网络链路上的可用带宽或网络流量的量等等方面来量度。

如上所述 IGP 通告通常被流播，直到每个中间网络 IGP 节点已经从每个其他互连的中间节点接收到 IGP 通告为止。然后，每个 IGP 节点可以通

过聚集所接收的邻居节点列表和成本值来构造相同的网络拓扑“视图”。为此，每个 IGP 节点可以将该接收到的路由信息输入到“最短路径优先”（SPF）计算中，该计算确定将中间节点与每个其他网络节点相耦合的最低成本网络路径。例如，Dijkstra 算法是一种用于执行这种 SPF 计算的传统技术，在 1999 年 9 月公布的 Radia Perlman 的课本“*Interconnections Second Edition*”的 12.2.4 部分中有更详细的描述，该课本通过引用全文结合于此。每个 IPG 节点基于其 SPF 计算的结果更新存储在其本地路由表中的路由信息。更具体而言，RIB 更新路由表以将目的地节点与关联到最低成本路径的下一跳接口相关以到达这些节点，如 SPF 计算所确定的。

多协议标签交换（MPLS）流量工程已被开发出来以满足数据联网需求，例如有保证可用带宽或快速恢复。MPLS 流量工程采用了现代标签交换技术来构建通过标签交换路由器（LSR）的 IP/MPLS 网络的有保证带宽端到端隧道。这些隧道是一类标签交换路径（LSP），因而通常被称为 MPLS 流量工程（TE）LSP。MPLS TE 的示例可以在 2001 年 12 月的题为“*RSVP-TE: Extensions to RSVP for LSP Tunnels*”的 RFC 3209、2004 年 6 月的题为“*Intermediate-System-to-Intermediate-System (IS-IS) Extensions for Traffic Engineering (TE)*”的 RFC 3784 以及 2003 年 9 月的题为“*Traffic Engineering (TE) Extensions to OSPF Version 2*”的 RFC 3630 中找到，这些 RFC 的内容通过引用整体结合于此。

从头端 LSR 到尾端 LSR 的 MPLS TE-LSP 的建立涉及计算通过 LSR 的网络的路径。最优情况下，所计算的路径是在某种度量下测得的“最短”路径，其满足所有相关的 LSP 流量工程约束，例如所需带宽、用于包括在路径中的每条链路和节点的备用旁路隧道的可用性等等。路径计算或者可以由头端 LSR 执行，或者可以由充当路径计算元件（PCE）的某个其他实体执行。头端 LSR（或 PCE）使用其关于网络拓扑和每条链路上的可用资源的知识来根据 LSP 流量工程约束执行路径计算。可以使用各种路径计算方法，包括 CSPF（约束最短路径优先）。MPLS TE-LSP 可以被配置在单个域（例如 IGP 区域或级别）内，或者也可以跨越多个域（例如多个 IGP 区域或级别）。

在跨过域边界时出现的一个困难是在头端 LSR 处的路径计算需要跨头端和尾端 LSR 之间的整个网络的网络拓扑和资源的知识。然而服务提供商一般不跨域边界彼此共享该信息。具体而言，网络拓扑和资源信息通常并不流过域边界，即使单个服务提供商可以操作所有区域也是如此。头端 LSR 或任何单个 PCE 都没有足够的知识来计算路径。因此，要计算域间 TE-LSP 需要 MPLS 流量工程路径计算技术。

PCE 的使用已经适合于创建分布式 PCE 体系结构，以扩展 MPLS TE-LSP 跨域边界。这种分布式体系结构的示例在 2003 年 9 月 18 日由 Vasscur 等人提交的题为“COMPUTING INTER-AUTONOMOUS SYSTEM MPLS TRAFFIC ENGINEERING LSP PATHS”的共同拥有、共同未决的美国专利申请 No. 10/767,574 中有所描述，该申请的内容通过引用整体结合于此。在分布式 PCE 体系结构中，计算路径所需的可视性被在相邻域之间扩展以使得 PCE 可以通过在保留跨域的机密性（例如当适用于 AS 时）的同时交换虚拟最短路径树（VSPF）来协作计算跨多个域的路径。

某些应用可能结合了单向数据流，该单向数据流被配置为根据某一“服务质量”（QoS）将时间敏感流量从计算机网络中的源（发送者）传送到网络中的目的地（接收者）。这里，网络资源可以被预留用于单向流以确保维持与数据流相关联的 QoS。资源预留协议（RSVP）是一种使应用能够预留资源以为其数据流获得特殊 QoS 的网络控制协议。RSVP 与路由协议结合工作，以例如为计算机网络中的数据流预留资源从而建立数据流所需的 QoS 水平。RSVP 在 R. Braden 等人的“*Resource ReSerVation Protocol (RSVP)*”的 RFC 2205 中定义。在流量工程应用的情况下，RSVP 信令被用于建立 TE-LSP，并沿着遵循这组所需约束的 TE-LSP（其路径可能已通过各种方式计算）向诸如边界路由器之类的路由器传送各种 TE-LSP 属性。

偶尔，头端 LSR 或节点会具有到其自身的域外部的（即，远程的）特定域（例如区域或级别）中的多个 TE-LSP。这些域间 TE-LSP 可以在单个尾端 LSR 或远程域的节点处终止，或者可以在同一远程域内的不同尾端节点处终止，这取决于其初始设置。已知的这种域间 TE-LSP 的限制在于当

试图到达包含在尾端节点的域内的节点或前缀时不能自动将流量引导到这种 TE-LSP 上。该限制主要是由于头端节点可用的有限网络拓扑信息。当前，这种可达性信息的缺乏已经要求使用静态的或基于策略的路由，这通常要求由系统管理员利用网络拓扑的先验知识进行手工配置。这种替换方案可能是麻烦的且应用性有限，并且在某些情况下（例如，误配置情况下）可能是网络故障的起因。

发明内容

本发明涉及一种用于在计算机网络中将流量工程（TE）标签交换路径（LSP）的尾端节点的可达性信息传播到 TE-LSP 的头端节点的技术。TE-LSP 优选地跨越网络的多个域，从而使尾端节点驻留在与头端节点的域不同（远离头端节点的域）的某一域中。域间信息传播技术采用内部网关协议（IGP）来将可达性信息从驻留在与尾端节点相同的域中的目标节点发送到头端节点。头端节点使用远程信息来计算可以从尾端节点到达的路由（即，地址前缀和关联属性），以插入到其路由表中。

在这里描述的示例性实施例中，TE-LSP 是利用资源预留协议（RSVP）TE 信令消息建立的。头端节点将对所建立的 TE-LSP 的索引（标签）插入到其路由表中，该标签被指派有 RSVP TE 信令消息所发现的度量。目标节点利用跨域（例如 IGP 区域/级别）发送的 IGP 消息传播（“通告”）与 TE-LSP 的尾端节点有关的远程可达性信息。目标节点被示例性地实现为区域边界路由器或第 1 级/第 2 级路由器，并且 IGP 消息被实现为开放最短路径优先（OSPF）或中间系统到中间系统（IS-IS）通告消息（“IGP 通告”）。注意，IGP 通告包括用于传送远程可达性信息的类型/长度/值（TLV）编码格式。

根据本发明的一方面，目标节点利用新的区域标识符（区域 id）子 TLV 通告其域中所有可达的地址前缀，区域 id 子 TLV 标识了每个前缀所属的域。区域 id 子 TLV 被包含在被跨域发送到 IGP 节点的 IGP 通告的扩展可达性 TLV 内。在接收到具有区域 id 子 TLV 的通告后，头端节点利用其本地可达性信息以及来自目标节点的远程可达性信息计算路由，以创建

域间可达性信息。每个前缀的区域 id 被用于确定哪些前缀可以通过特定 TE-LSP 到达。这样，具有与尾端节点相同的区域 id 的前缀被通过 TE-LSP 路由到该尾端节点。头端节点给经由远程可达性信息获得的每个前缀指派一个等于（或约等于）由 RSVP TE 信令消息发现的 TE-LSP 的度量的度量。

根据本发明的另一方面，目标节点被沿 TE-LSP 定位，并且利用新的节点 id 子 TLV 通告在其域中处于尾端“下游”（从目标节点的视角看）的所有可达地址前缀，节点 id 子 TLV 将尾端节点标识为发起了前缀。节点 id 子 TLV 被包含在 IGP 通告的扩展可达性 TLV 内，IGP 通告是响应于例如目标节点的预配置或对 RSVP-TE 信令消息的扩展而被跨域发送到 IGP 节点的。关于后者，RSVP 扩展可以指定目标节点和访问控制列表两者，该访问控制列表限制传播的信息去往下游前缀。在接收到具有节点 id 子 TLV 的通告后，头端节点利用其本地可达性信息以及来自目标节点的远程可达性信息计算路由，以创建域间可达性信息。已被标识为尾端节点“下游”的前缀被通过 TE-LSP 路由到该尾端节点。头端节点给经由远程可达性信息获得的每个地址前缀指派一个等于由 RSVP TE 信令消息发现的 TE-LSP 的度量的度量。

根据本发明的另一方面，目标节点被沿 TE-LSP 定位，并且利用节点 id 子 TLV 以及新的前缀成本子 TLV 通告其域中从尾端节点可达的所有地址前缀。这里，前缀成本子 TLV 包含由不发起前缀的节点（即，目标节点）计算的来自尾端节点（由节点 id 子 TLV 标识）的每个前缀的度量（例如成本）。前缀成本子 TLV 被示例性地用在目标节点执行以尾端节点为根的特殊 SPF 计算以创建特殊 SPT 的情况下。前缀成本子 TLV 被包含在 IGP 通告的扩展可达性 TLV 中，IGP 通告是响应于例如目标节点的预配置或对 RSVP-TE 信令消息的扩展而被跨域发送到 IGP 节点的。在接收到该通告后，头端节点利用其本地可达性信息以及来自目标节点的远程可达性信息计算路由，以创建域间可达性信息。具体而言，对于经由具有节点 id 和前缀成本子 TLV 的 IGP 通告获得的每个地址前缀，头端节点给前缀指派一个等于在前缀成本子 TLV 中携带的度量的度量（除了 TE-LSP 度量

以外）。

有利地，新技术将所建立的跨越多个域的 TE-LSP 的尾端节点的可达性信息传播到 TE-LSP 的头端节点。通过向头端节点通知尾端节点的可达性信息，本发明的技术提供了取代次优路由技术（例如繁重的手工配置（例如静态路由或策略路由））的一种替换方案，其可以避免在这种次优路由技术中产生的某些风险和可能错误。

附图说明

本发明的以上和其他的优点可通过参考下面结合附图的描述得以更好地理解，在附图中相似的标号指示相同或功能类似的元件，附图中：

图 1 是可以根据本发明使用的示例性的区域计算机网络的示意性框图；

图 2 是可以结合本发明有利地使用的示例性路由器的示意性框图；

图 3 是可以被路由器流播的示例性 IGP 通告的示意性框图；

图 4 是可以结合本发明有利地使用的 RSVP Path 消息的一些部分的示意性框图；

图 5 是图示可以结合本发明有利地使用的 IGP 扩展 TLV 的格式的示意性框图；

图 6 是可以结合本发明有利地使用的示例性路由表的示意性框图；

图 7 是图示根据本发明的用于传播可达性信息的步骤序列的流程图；

图 8 是图示根据本发明的用于传播可达性信息的步骤序列的流程图；以及

图 9 是图示根据本发明的用于传播可达性信息的步骤序列的流程图。

具体实施方式

图 1 是示例性计算机网络 100 的示意性框图，计算机网络 100 包括分别具有示例性域内路由器 A 和 B 的区域 A1 和 A2 以及具有示例性域内路由器 C、D 和 E 的区域 A3。另外，A1 和 A2 共享区域边界路由器 ABR1 和 ABR2，而 A2 和 A3 共享 ABR3 和 ABR4。这里所用的区域（area）是

彼此共享完全网络拓扑信息但是不一定与区域外部的路由器共享该信息的路由器的集合。区域的集合可以被包含在单个自治系统（AS）内。这里所用的术语区域还包含术语“级别（level）”，该术语对于采用 IS-IS 作为其内部网关协议（IGP）的网络来说具有类似的含义，在这种情况下区域边界路由器 ABR1-4 被实现为第 1 级/第 2 级（L1L2）路由器。这些示例仅仅是代表性的。这里所用的术语区域和级别可以互换使用，并且 ABR、L1L2 路由器和（更一般地）IGP 节点的使用也可以互换。

数据分组可以利用预定义的网络通信协议在区域 A1-A3 之间交换，这些网络通信协议例如是传输控制协议/因特网协议（TCP/IP）、用户数据报协议（UDP）、异步传送模式（ATM）协议、帧中继协议、因特网分组交换（IPX）协议等。路由信息可以利用预定的 IGP（例如传统的距离向量协议或链路状态协议）通过使用链路状态通告或链路状态分组（“IGP 通告”）在区域的路由器之间分发。

图 2 是可以结合本发明有利地使用作为 IGP 节点（例如，域内路由器或边界路由器）的示例性路由器 200 的示意性框图。该路由器包括通过系统总线 250 互连的多个网络接口 210、处理器 220 和存储器 240。网络接口 210 包含用于通过耦合到网络 100 的物理链路传输数据的机械的、电的和信令电路。网络接口可以被配置为利用多种不同通信协议发送和/或接收数据，这些协议包括 TCP/IP、UDP、ATM、同步光网络（SONET）、无线协议、帧中继、以太网、光纤分布数据接口（FDDI）等等。

存储器 240 包括多个用于存储与本发明相关联的软件程序和数据结构的存储单元，这些存储单元可由处理器 220 和网络接口 210 寻址。处理器 220 可包括适合于执行软件程序并操作数据结构（例如路由表 600 和影子表（shadow table）650）的必要元件或逻辑。路由器操作系统 242（其某些部分一般驻留在存储器 240 中并由处理器执行）在功能上通过调用支持在路由器上执行的软件过程和/或服务的网络操作等等，来组织路由器。这些软件过程和/或服务包括路由信息库（RIB）245、流量工程（TE）模块 246、路由服务 247 和 RSVP 服务 249。本领域技术人员将清楚，其他处理器和存储器装置（包括各种计算机可读介质）可以用于存储和执行与这里

描述的创造性技术有关的程序指令。

路由服务 247 包含由处理器 220 执行的用于执行由一种或多种路由协议（例如 OSPF 和 IS-IS）提供的功能的计算机可执行指令。这些功能可以被配置为管理包含例如用于进行转发判决的数据的转发信息数据库（未示出）。RSVP 服务 249 包含根据本发明用于实现 RSVP 和处理 RSVP 消息的计算机可执行指令。RSVP 在可从 IETF 得到的 R. Braden 等人的 1997 年 9 月的注释请求（RFC）2205“*Resource ReSerVation Protocol (RSVP)*”以及题为“*RSVP-TE: Extensions to RSVP for LSP Tunnels*”的 RFC 3209 中有所描述，这两篇 RFC 在这里通过引用全文结合于此。

网络拓扑的改变可以利用链路状态协议（例如传统的 OSPF 和 IS-IS 协议）在路由器 200 之间传输。例如，假定 AS 内的通信链路发生故障，或者与网络节点相关联的成本值改变。一旦网络状态的改变被路由器之一检测到，该路由器就可以流播 IGP 通告，将改变传输到 AS 中的其他路由器。以这种方式，每个路由器最终“收敛”到网络拓扑的相同视图。

图 3 图示了可以由路由器 200 流播的示例性 IGP 通告 300。该分组包括通告节点字段 302、序列号字段 304、年龄字段 306、数据部分 320 和其他路由信息 312。通告节点字段 302 存储标识生成并最初广播 IGP 通告 300 的路由器的值。序列号字段 304 存储指示 IGP 通告的相对版本的序列号。一般来说，存储在字段 304 中的序列号对于 IGP 通告的每一新版本例如递增 1。因此，如果 IGP 通告 300 的序列号小于存储在 IGP 通告的先前接收的版本中的（即，由同一通告节点生成的）序列号，则该 IGP 通告 300 被认为是“陈旧的”（无效的）。因此，路由器 200 可以被配置为只存储并转发最新版本的 IGP 通告，例如，具有最大序列号的版本。

年龄字段 306 也存储可以用于确定 IGP 通告 300 是否有效的值。年龄值一般被初始化为非零整数值，经常以秒为单位。年龄值可以例如每一秒递减 1，直到年龄值达到零为止，从而指示 IGP 通告已变为无效。即，存储或流播 IGP 通告 300 的每一路由器 200 连续老化分组，直到年龄值等于零为止。本领域技术人员将意识到，也可以使用其他老化机制，例如从例如等于零的初始值开始递增 IGP 通告年龄值，直到年龄值达到已知的上限

为止。

数据部分 320 包括一对或多对邻居节点字段 308 和成本字段 310。每个邻居节点字段 308 存储指示可以从存储在通告节点字段 302 中的中间节点直接访问的网络节点的值（例如地址）。字段 310 存储已经例如由通告节点与在邻居节点字段 308 中标识的网络节点相关联的成本值。如图所示，每个邻居节点字段 308 只与一个相应成本字段 310 相关联。然而，注意，在其他实施例中，单个邻居节点可以与多个成本值相关联。其他路由信息 312 也可以被包括在 IGP 通告 300 中，例如校验和值、分组长度信息、标志值、服务类型度量等等和/或新 IGP 扩展对象 500（将在下面进一步描述）。一般来说，所接收的 IGP 通告被存储在路由器 200（未示出）的链路状态数据库（LSDB）中。

在一个实施例中，这里描述的路由器是实现多协议标签交换（MPLS）并充当标签交换路由器（LSR）的 IP 路由器。在一种简单的 MPLS 场景中，在网络的入口处，在将每个进入分组转发到下一跳路由器之前，基于其转发等价类给每个分组指派一个标签。在每个路由器处，利用在进入分组中找到的标签作为对包括该信息的标签转发表的索引（reference），来确定转发选择和新的替换标签。在网络出口（或在前的一跳）处，基于进入标签进行转发判决，但是可选地当分组被继续发送到下一跳时不包括标签。

分组以这种方式穿过网络所采取的路径被称为标签交换路径（LSP）。示例性的 TE-LSP 在图 1 中被示为头端节点（A）和尾端节点（C）之间的虚线。TE-LSP 的建立需要计算路径，沿路径信令，并修改沿路径的转发表。MPLS TE 在某些情况下建立了已经保证带宽的 LSP。示例性地，TE-LSP 可以通过使用 RSVP 协议（具体而言是 RSVP TE 信令消息）来通知。

根据 RSVP，为了在发送者（例如，头端节点 A）和接收者（例如，尾端节点 C）之间建立 TE-LSP（数据流），发送者可以将 RSVP 路径（Path）消息沿着某一路径（例如，单播路由）向下游逐跳地发送到接收者，以标识发送者并指示例如容纳数据流所需的带宽以及 TE-LSP 的其他

属性。Path 消息可以包含关于数据流的各种信息，例如包括数据流的流量特性。图 4 是可以结合本发明有利地使用的 RSVP Path 消息 400 的某些部分的示意性框图。消息 400 包含共同头部 410、发送者模板对象 420、流量规范（Tspec）对象 430 和 LSP-属性对象 440。应当注意，消息 400 可以包含其他对象，包括新的路由信息请求（RI-REQ）对象 800（将在下面进一步描述）。为了在接收者和发送者之间建立数据流，接收者可以将 RSVP 预留（Resv）消息（未示出）沿该路径向上游返回到发送者以确认 TE-LSP 的属性，并提供 TE-LSP 标签。应当注意，RSVP 的状态被按定时的间隔（例如每三十秒）刷新，其中 RSVP Path 和 Resv 消息被交换。该定时间隔可由系统管理员配置。

尽管这里描述的示例性实施例涉及 MPLS，但是还应当注意，本发明可以有利地应用于通用 MPLS（GMPLS），其不仅与基于分组和信元的网络有关，还与时分复用（TDM）和光网络有关。GMPLS 是公知的，并且在 2004 年 10 月的题为“*Generalized Multi-Protocol Label Switching (GMPLS) Architecture*”的 RFC 3945 和 2004 年 10 月的题为“*Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control*”的 RFC 3946 中有所描述，这两篇 RFC 的内容都通过引用整体结合于此。

为了计算跨多个域的路径，先前结合的美国专利申请 No. 10/767,574 描述了在分布式路径计算元件（PCE）体系结构中虚拟最短路径树（VSPT）的使用。注意，本领域技术人员将清楚，其他方法也可以用于计算 TE-LSP（例如，松散跳、显式路径等），并且这些方法在本发明的范围内。此外，路径计算请求（和响应）可以根据在 Vasseur 等人的“*RSVP Path Computation Request and Reply Messages*”的因特网草案（2004 年 7 月）中指定的协议来实现，该草案通过引用全文结合于此。

本发明涉及一种用于在计算机网络中将流量工程（TE）标签交换路径（LSP）的尾端节点的可达性信息传播到 TE-LSP 的头端节点的技术。TE-LSP 优选地跨越多个网络域，从而使尾端节点驻留在不同于（远离）头端节点的域的域中。域间信息传播技术采用内部网关协议（IGP）来将远程

可达性信息从驻留在与尾端节点同一域中的目标节点发送到头端节点。头端节点使用该远程信息来计算可以从尾端节点到达的路由（即，地址前缀和关联属性），以插入到其路由表中。

在这里描述的示例性实施例中，TE-LSP 是根据已知的显式路径（用户可配置）和/或 PCE 技术利用 RSVP TE 信令消息计算和建立的。具体而言，RSVP 服务 249 采用这种信令和技术来计算与所建立的 TE-LSP 相关联的一个或多个度量（例如成本）。然后，对 TE-LSP 的索引（标签）以及所计算的度量被存储在这里描述的影子表 650 中。其后，头端节点将 TE-LSP 索引插入到其路由表中，该 TE-LSP 索引被指派有 RSVP TE 信令消息所发现的度量。目标节点利用跨域（例如 IGP 区域/级别）发送的 IGP 消息传播（“通告”）与 TE-LSP 的尾端节点有关的远程可达性信息。目标节点被示例性地实现为区域边界路由器或第 1 级/第 2 级路由器，IGP 消息被实现为 OSPF 或 IS-IS 通告消息（“IGP 通告”）。注意，IGP 通告包括用于传送远程可达性信息的类型/长度/值（TLV）编码格式。

TLV 编码格式被用于标识被传输（传送）的信息类型（T）、要被传送的信息长度（L）以及传送的实际信息值（V）。包含在长度字段中的长度（L）参数一般是特定于实现方式的，并且可以表示从对象的类型字段的开头到结尾的长度。然而，长度通常表示值（V）字段的长度而不是类型（T）或长度（L）字段的长度。

图 5 是图示可以结合本发明有利地使用的 IGP 扩展对象 500 的格式的示意性框图。IGP 扩展对象 500 被示例性地实现为包含在 IGP 通告 300 中的 TLV，并且被扩展为携带可以从特定 IGP 节点到达的地址前缀。为此，“扩展可达性 TLV”500 被组织为包括包含预定类型值（例如，“类型 135”或“扩展 IP 可达性”TLV）的类型字段 505，如在以上引用的 RFC 3784 中定义的。长度字段 510 是可变长度值。值字段 515 示例性地包含可达地址前缀 520，以及一个或多个子 TLV 550 等等，这些子 TLV 各自具有类型字段 555、长度字段 560 和值字段 565。根据本发明，子 TLV 550 的字段以多种方式使用，包括这里所描述的方式。

根据本发明的一方面，IGP（目标）节点利用标识每个前缀所属的域

的新的区域标识符（区域 id）子 TLV 550a 来通告其域中所有可达的地址前缀。区域 id 子 TLV 550a 被包含在被跨域发送到 IGP 节点的 IGP 通告 300 的扩展可达性 TLV 500 内。具体而言，区域 id 子 TLV 被用于传送路由器的区域（级别）标识符（前缀）。每个区域/级别必须具有预定义的、唯一的区域 id，并且由区域/级别内的目标节点通告的每个前缀具有相同的区域 id 子 TLV。注意，前缀属性（例如区域 id 子 TLV 550a）在跨区域或级别传播前缀时由目标节点（例如边界路由器，比如 ABR 或 L1L2 路由器）保留。

在接收到具有区域 id 子 TLV 550a 的通告 300 后，头端节点利用其本地可达性信息以及来自目标节点的远程可达性信息来计算路由，以创建域间可达性信息。每个前缀的区域 id 被用于确定哪些前缀可以经由特定的 TE-LSP 到达。以这种方式，与尾端节点具有相同区域 id 的前缀被通过 TE-LSP 路由到该尾端节点。头端节点给经由远程可达性信息获得的每个前缀指派一个度量（例如成本），该度量等于或约等于由 RSVP TE 信令消息发现的 TE-LSP 的度量。更具体而言，头端节点在将所得到的信息插入到其路由表 600 中之前，将远程可达性信息加载到影子表 650 中。

图 6 是可以结合本发明有利地使用的示例性路由表 600 的示意性框图。路由表 600 被示例性地存储在存储器 240 中，并且包括各自包括多个字段的一个或多个条目 610，这多个字段用于存储可达目的地地址 612、到达该目的地的下一跳接口 614 和下一跳地址 616，以及到达目的地的关联度量（例如成本）。路由表 600 示例性地由 RIB 245 维护和管理。为此，RIB 245 维护由路由协议（例如 IGP）提供的路由（路径）的拷贝，以计算最佳路径/路由从而安装到路由表 600 中。

例如，假定目的地地址前缀 IP1 可以从节点 A 经由节点 C 到达。另外，将节点 A 连接到节点 C 的路径 A-C 的成本是“6”（例如经由图 1 的 ABR1 和 ABR3），并且到可达地址 IP1 的链路 C-N 的成本是“1”。条目 610N 的目的地地址字段 612 包含可达地址 IP1，并且下一跳字段 614、616 被分别填充以例如链路 A-ABR1 和节点 ABR1 的回送地址。注意，下一跳节点的回送地址出于很多原因被用作下一跳地址，这些原因包括作为一种

避免依赖于该节点的网络接口的可用性的方式。IP1 的成本是到可达地址的所有链路的成本，即“7”。

影子表 650 与条目 610N 的 IP1 相关联。如上所述，影子表 650 最初由 RSVP 服务 247 利用从 RSVP 信令消息获得的 TS-LSP 信息创建。然而，根据本发明，（IGP）路由服务 247（示例性地是 IS-IS）与 RSVP 服务 249 协作以利用从目标节点获得的远程可达性信息填充影子表 650 的某些字段。影子表 650 实质上包括与路由表 600 相同的格式，但是其目的地地址前缀可经由 TS-LSP 的目标节点到达。具体而言，影子表 650 的每个条目 660 可以包括多个字段，用于存储可以从目标节点到达的目的地前缀 662、对目标节点的 TS-LSP 的索引 664、目标节点的地址 666 以及从头端节点到可达前缀的成本度量 668。对于本发明的这一方面，影子表 650 存储等于 TE-LSP 的度量 668（TE-LSP 1）。然而，作为将远程可达性信息加载到影子表中的结果，每个前缀与具有匹配区域 id 的目标（尾端）节点相关联。示例性地，成本度量 668 是节点 A 和 C 之间的 TE-LSP 的成本，例如“4”。注意，TE-LSP 的成本度量可以大于、小于或等于 IP 路由链路的成本，并且值“5”和“7”应分别被当作示例。

根据本发明，当可达目的地地址匹配时，IGP/RSVP 服务与 RIB 245 协作以将影子表条目 660N 的内容与相应路由表条目 610N 合并。为了使路由表插入正确工作，要求头端节点在更新其路由表之前应用定时器，以使远程可达性信息（前缀）被加载到影子表 650 中。作为合并的结果，路由表条目 610N 的关联属性被更新为反映可由目标节点到达的属性。例如，路由表 600 的条目 610N 被更新以使得下一跳接口字段 614 包含来自条目 664 的 TE-LSP 索引，下一跳地址字段 616 包含来自字段 666 的节点 C，而度量字段 618 包含来自字段 668 的 TE-LSP 的成本（例如，值“4”）。

因而，更新后的路由表 600 包含可以从 TE-LSP 到达的前缀，从而流量可以沿 TE-LSP 被路由到这些前缀。注意，头端节点动态地计算这些路由，例如当接收到更新后的路由信息时，如上所述。另外，在本发明的一方面中，更新后的路由信息触发部分路由计算（PRC）（例如在 ISIS 的情况下），而不是全 SPF。

在 TE-LSP 变得不可用的情况下（例如，被手工去除或 TE-LSP 发生故障），来自影子表 650 的合并前缀和关联属性被从路由表 600 中去除。在本发明的一方面中，前缀是在 TE-LSP 还未被恢复之后、预定定时器期满之前被去除的。另外，在本发明的另一方面中，在将前缀重新关联到恢复后的 TE-LSP 之前，可以有利地使用等待恢复（WTR）定时器，以避免在资源紧张的情况下发生多个流量中断。

图 7 是图示根据本发明的用于传播可达性信息的步骤序列的流程图。序列 700 开始于步骤 705，并继续到步骤 710，在步骤 710 根据上述方法建立（设立）从头端节点到尾端节点的 TE-LSP。在步骤 715 中，头端节点将具有通过 RSVP TE 学习到的度量的 TE-LSP 插入到其路由表中。目标节点（例如，边界路由器，诸如 ABR 或 L1L2 路由器）在步骤 720 中发送具有新区域 id TLV 550a 的 IGP 通告。在步骤 725 中接收到 IGP 通告后，头端节点将区域 id 内容加载到影子表中，并检查这些内容以在步骤 730 中将通告的前缀的区域 id 与已知 TE-LSP 尾端节点的区域 id 相匹配。在步骤 735 中，头端将具有匹配区域 id 的前缀插入到其路由表 600 中。在这种情况下，前缀被给予作为 TE-LSP 的下一跳以及等于（或基本等于）TS-LSP 度量的度量。该序列在步骤 740 中结束。

根据本发明的另一方面，目标节点被沿 TE-LSP 定位，并且利用新的节点 id 子 TLV 550b 通告其域中尾端“下游”（从目标节点的视角看）的所有可达地址前缀，该节点 id 子 TLV 550b 将尾端节点标识为发起前缀。节点 id 子 TLV 550b 被包含在 IGP 通告 300 的扩展可达性 TLV 500 内，该 IGP 通告 300 是响应于例如目标节点的预配置或对 RSVP-TE 信令消息的扩展而被跨域发送到 IGP 节点的。关于后者，RSVP 扩展可以指定目标节点和访问控制列表两者，该访问控制列表限制传播的信息去往下游前缀。对 RSVP-TE 信令消息的扩展在先前结合的美国专利申请 No. 11/001,349 中有所描述。

具体而言，例如边界路由器（ABR 或 L1L2 路由器）之类的目标节点执行以目标节点为根的最短路径优先（SPF）计算以创建最短路径树（SPT）。通常，目标节点维护可以使用的当前 SPT。目标节点遍历 SPT

以定位开始于尾端节点的子分支，并利用与尾端节点的回送地址相对应的值给尾端节点下游（即，远离头端节点）的所有地址前缀加上标记（tag）。然后，目标节点生成一个或多个 IGP 通告 300，IGP 通告 300 包含这些加上标记的被编码到节点 id 子 TLV 550b 中的前缀。

示例性地，节点 id 子 TLV 550b 包含“发起”前缀的尾端节点的标识符（例如，如在上面通过引用结合的 RFC 3784 中指定的）。当前缀位于目标节点的 SPT 子分支中（即，超过了目标节点的标准 SPT 中的尾端节点）时，尾端节点被认为发起了前缀。由尾端节点发起的区域的所有前缀都与相同的节点 id 子 TLV 500b 相关联。注意，节点 id 子 TLV 是加性的，因为对于每个前缀它能够出现多次，从而允许同一前缀被标识为多个尾端节点的下游（从一个或多个边界路由器的角度看）。

在接收到具有节点 id 子 TLV 550b 的通告后，头端节点利用其本地可达性信息以及来自目标节点的远程可达性信息来计算路由，以创建域间可达性信息。已被标识为尾端节点“下游”的前缀被通过 TE-LSP 路由到该尾端节点。头端节点给经由远程可达性信息获得的每个地址前缀指派一个等于 TE-LSP 的度量的度量，TE-LSP 的度量是由 RSVP TE 信令消息发现的。而且，头端节点通常以与上面参考图 6 所述相同的方式计算其路由表 600，只不过路由服务 247 将前缀关联到目标（尾端）节点以插入到影子表 650 中，其中前缀的节点 id 子 TLV 550b 包含与尾端节点 id 匹配的节点 id。注意，在这种情况下，影子表不保存任何关联的度量。结果，下一跳接口和地址与 TE-LSP 的相同，并且每个前缀的关联成本将是 TE-LSP 的成本（例如“4”）。

图 8 是图示根据本发明的用于传播可达性信息的步骤序列的流程图。序列 800 开始于步骤 805，并继续到步骤 810，在步骤 810 根据上述方法建立（设立）从头端节点到尾端节点的 TE-LSP。在步骤 815 中，头端节点将具有通过 RSVP TE 学习到的度量的 TE-LSP 插入到其路由表中。目标节点（例如，边界路由器，诸如 ABR 或 L1L2 路由器）在步骤 820 中检测在其区域或级别内 TE-LSP 尾端节点的存在。示例性地，边界路由器可以通过截取头端节点和尾端节点之间的 RSVP TE 消息，或者通过接收请求

通告特定尾端节点的路由信息的显式请求（例如通过在先前结合的美国专利申请 No. 11/001,349 中描述的 RI-REQ TLV），来检测尾端节点。然后，边界路由器在步骤 825 中发送具有新的尾端节点 id TLV 550b 的 IGP 通告 300。在步骤 830 中接收到 IGP 通告后，头端节点将节点 id 内容加载到影子表 650 中，并检查这些内容以在步骤 835 中将通告的前缀的节点 id 与已知 TE-LSP 尾端节点的节点 id 相匹配。在步骤 840 中，头端将具有匹配节点 id 的前缀插入到其路由表 600 中。在这种情况下，前缀被给予等于 TE-LSP 的下一跳接口和等于（或基本等于）TE-LSP 度量的度量。该序列在步骤 845 中结束。

根据本发明的另一方面，目标节点被沿 TE-LSP 定位，并且利用节点 id 子 TLV 550b 以及新的前缀成本子 TLV 550c 通告其域中从尾端节点可达的所有地址前缀。这里，前缀成本子 TLV 包含由不发起前缀的节点（即，目标节点）计算的来自尾端节点（由节点 id 子 TLV 标识）的每个前缀的度量（例如成本）。前缀成本子 TLV 被示例性地用在目标节点执行以尾端节点为根的特殊 SPF 计算以创建特殊 SPT 的情况下。前缀成本子 TLV 550c 被包含在 IGP 通告 300 的扩展可达性 TLV 500 中，IGP 通告 300 是响应于例如目标节点的预配置或对 RSVP-TE 信令消息的扩展而被跨域发送到 IGP 节点的。

示例性地，前缀成本子 TLV 550c 包含由目标节点所计算的从尾端节点到达前缀的度量（例如成本），在这种情况下目标节点不是尾端节点。该度量被计算并存储在目标节点的特殊 SPT 中，并被放置到前缀成本子 TLV 550c 中以在 IGP 通告中发送。注意，该特殊度量只对 TE-LSP 的头端节点有意义。另外，节点 id 子 TLV 和前缀成本子 TLV 或者可以是单个组合的子 TLV，或者可以是两个单独的子 TLV 的连接。

在接收到这些通告后，头端节点利用其本地可达性信息以及来自目标节点的远程可达性信息计算路由，以创建域间可达性信息。具体而言，对于经由具有节点 id 和前缀成本子 TLV 的 IGP 通告获得的每个地址前缀，头端节点给该前缀指派一个与在前缀成本子 TLV 550c 中携带的度量相等的度量（除了 TE-LSP 度量以外）。

同样，头端节点通常以与上面参考图 6 所述相同的方式计算其路由表，只不过路由服务 247 将前缀关联到目标（尾端）节点以插入到影子表 650 中，其中前缀的节点 id 子 TLV 550b 包含与尾端节点 id 相匹配的节点 id。然而，另外，影子表包含从前缀成本子 TLV 550c 获得的与前缀相关联的度量（例如成本）。结果，下一跳接口和地址与 TE-LSP 的相同，并且每个前缀的关联成本是 TE-LSP 的成本（例如“4”）加上额外前缀成本（例如“1”），以获得值“5”。

图 9 是图示根据本发明的用于传播可达性信息的步骤序列的流程图。序列 900 开始于步骤 905，并继续到步骤 910，在步骤 910 根据上述方法建立（设立）从头端节点到尾端节点的 TE-LSP。在步骤 915 中，头端节点将具有通过 RSVP TE 学习到的度量的 TE-LSP 插入到其路由表中。目标节点（例如，边界路由器，诸如 ABR 或 L1L2 路由器）在步骤 920 中检测在其区域或级别内 TE-LSP 尾端节点的存在。示例性地，边界路由器可以通过截取头端节点和尾端节点之间的 RSVP TE 消息，或者通过接收请求通告特定尾端节点的路由信息的显式请求（例如通过上面引用的 RI-REQ TLV），来检测尾端节点。在步骤 925 中，边界路由器执行以尾端节点为根的 SPF 计算，并且在步骤 930 中，获得从尾端节点到达周围前缀的度量以置于新的前缀成本 TLV 550c 中。然后，边界路由器在步骤 935 中发送具有新的尾端节点 id TLV 550b 和新的前缀成本 TLV 550c 的 IGP 通告 300。在步骤 940 中接收到 IGP 通告后，头端节点将节点 id 和前缀成本内容加载到影子表 650 中，并检查这些内容以在步骤 945 中将通告的前缀的节点 id 与已知 TE-LSP 尾端节点的节点 id 相匹配。在步骤 950 中，头端将具有匹配节点 id 的前缀插入到其路由表 600 中。在这种情况下，前缀被给予等于 TE-LSP 的下一跳接口和等于（或基本等于）TE-LSP 度量加上在关联的前缀成本 TLV 550c 中包含的前缀度量的某一度量。该序列在步骤 955 中结束。

有利地，新技术将所建立的跨域多个域的 TE-LSP 的尾端节点的可达性信息传播到 TE-LSP 的头端节点。通过向头端节点通知尾端节点的可达性信息，本发明的技术提供了取代次优路由技术（例如手工配置（例如，

静态路由或策略路由））的一种替换方案，其可以避免在这种次优路由技术中产生的某些风险和可能错误。

前述描述涉及本发明的特定实施例。然而，应当清楚，可以对所描述的实施例进行其他变化和修改，且获得它的某些或全部优点。例如，明确地设想本发明的教导可以实现为软件（包括具有在计算机上执行的程序指令的计算机可读介质）、硬件、固件或其组合。因此，该描述只能被认为是示例性的，而不是以任何方式限制本发明的范围。因此，所附权利要求的目的是覆盖在本发明的真实精神和范围内的所有这些变化和修改。

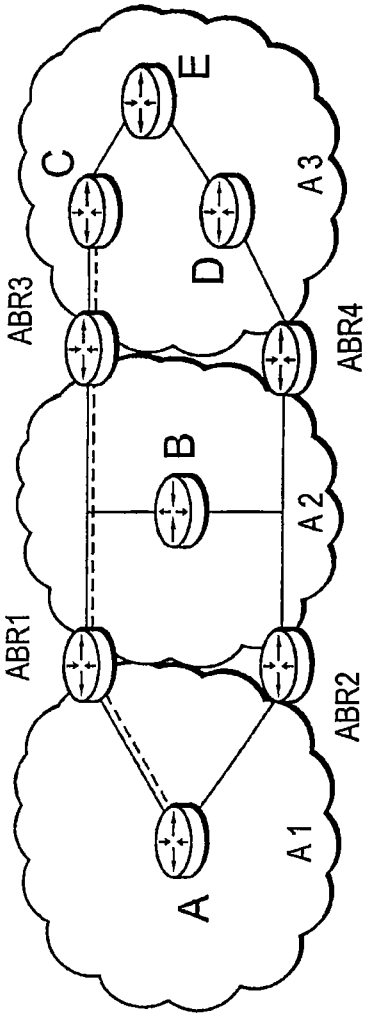


图1

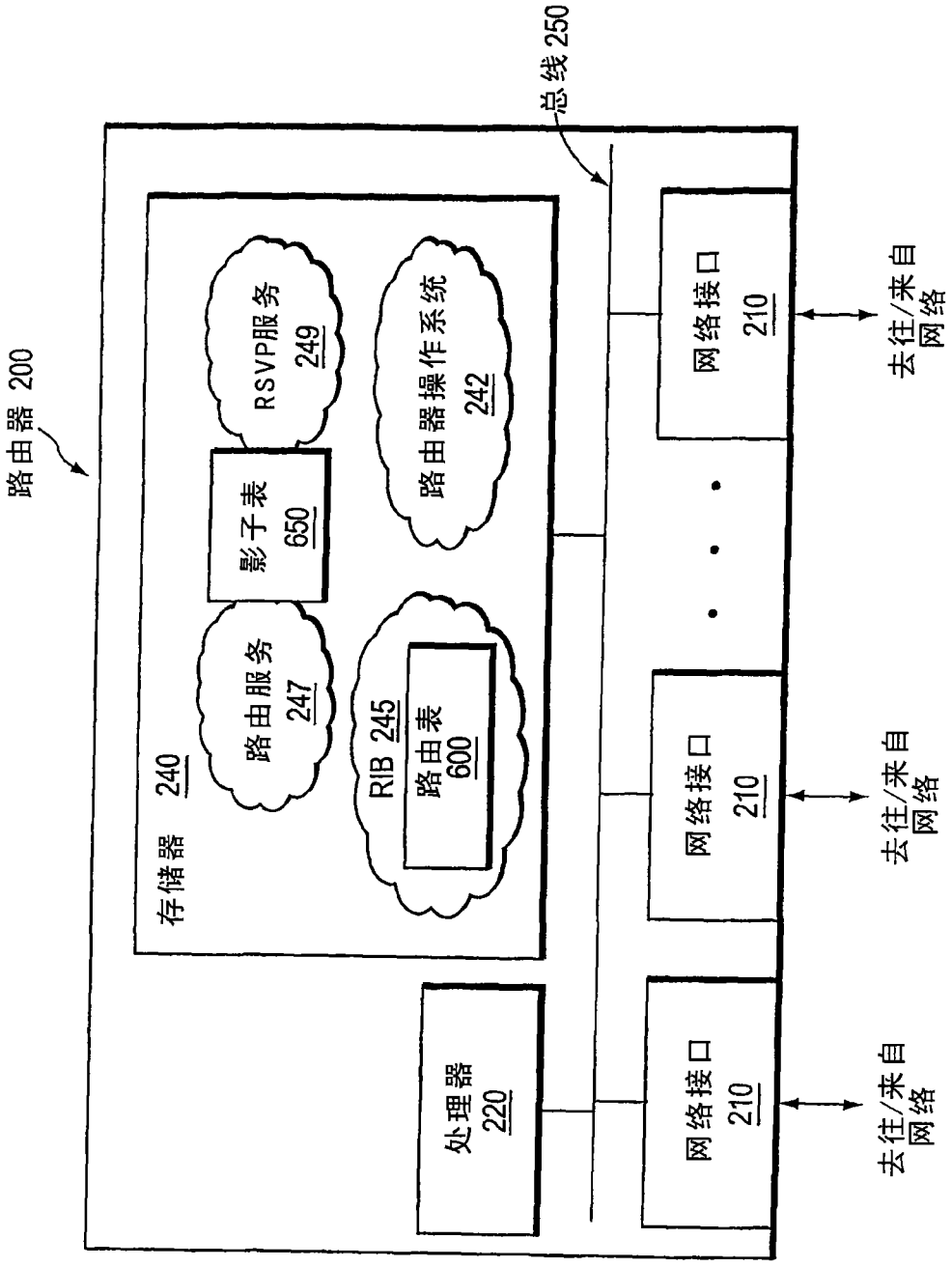

路由器 200
总线 250
存储器 240
RSVP服务 249
影子表 650
路由器操作系统 242
路由服务 247
RIB 245
路由表 600
处理器 220
网络接口 210
网络接口 210
网络接口 210
去往/来自网络
去往/来自网络
去往/来自网络

图2

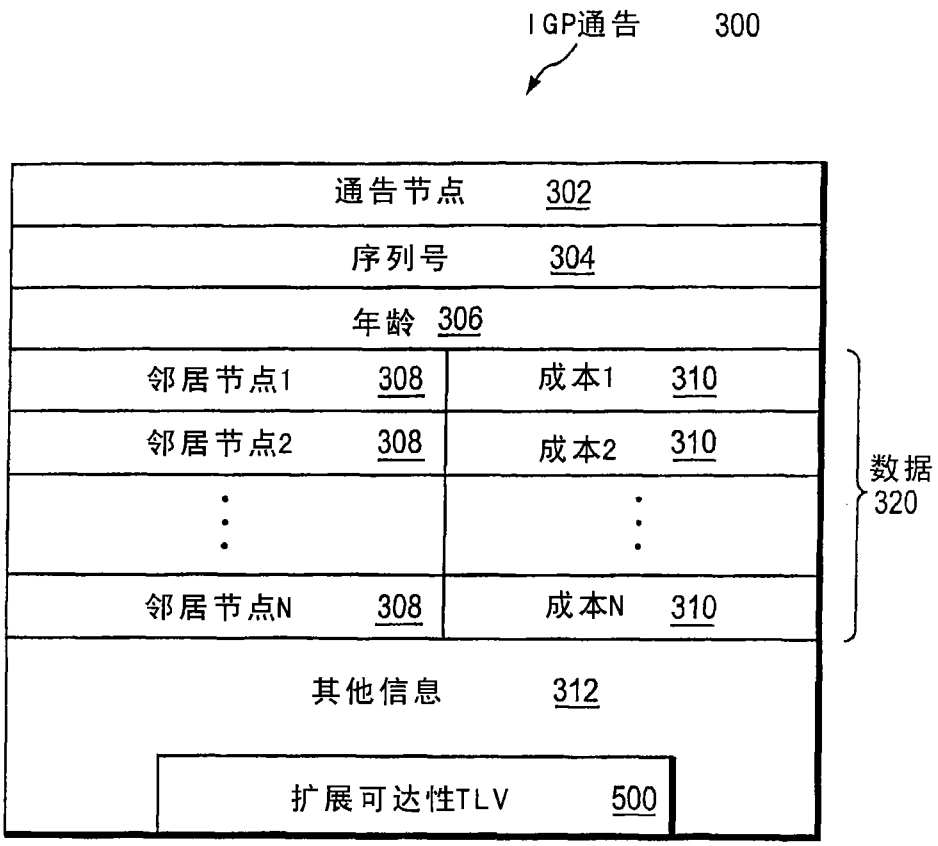

IGP通告 300
通告节点 302
序列号 304
年龄 306
邻居节点1 308
成本1 310
邻居节点2 308
成本2 310
邻居节点N 308
成本N 310
数据 320
其他信息 312
扩展可达性TLV 500

图3

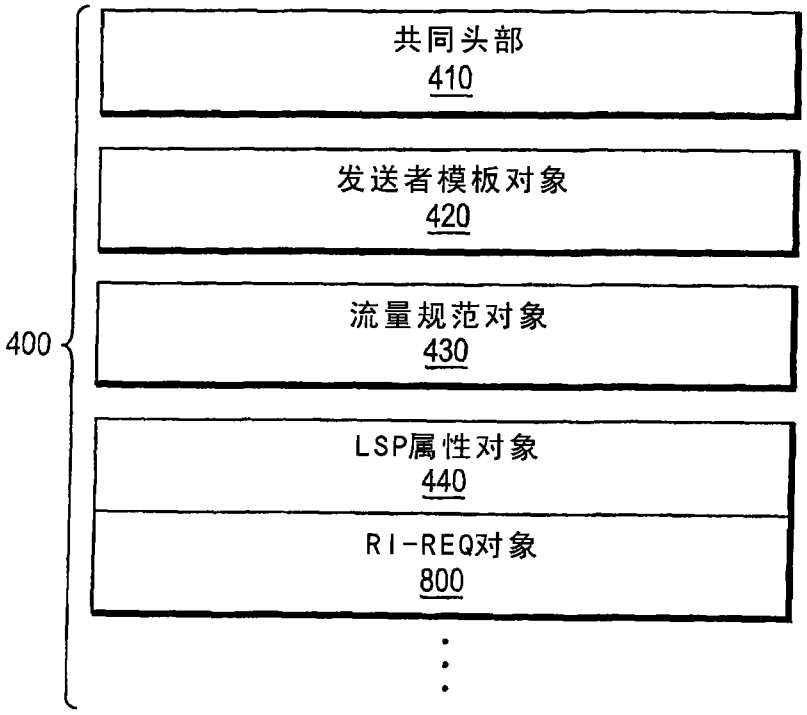


图4

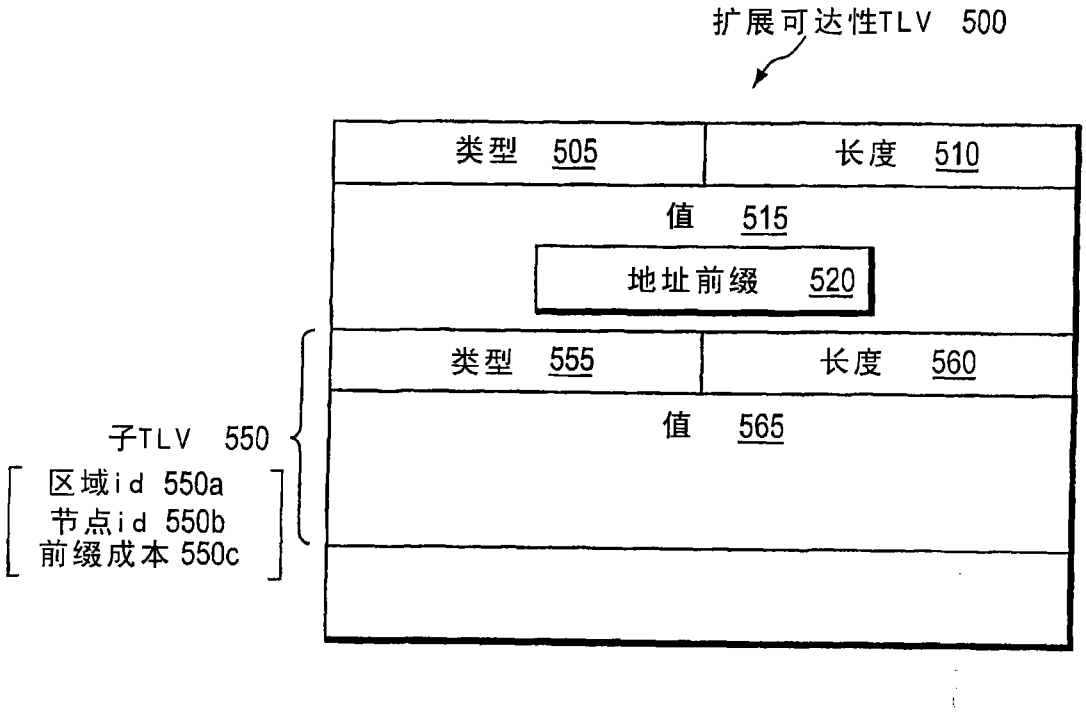

扩展可达性TLV 500
类型 505
长度 510
值 515
地址前缀 520
类型 555
长度 560
值 565
子TLV 550
区域id 550a
节点id 550b
前缀成本 550c

图5

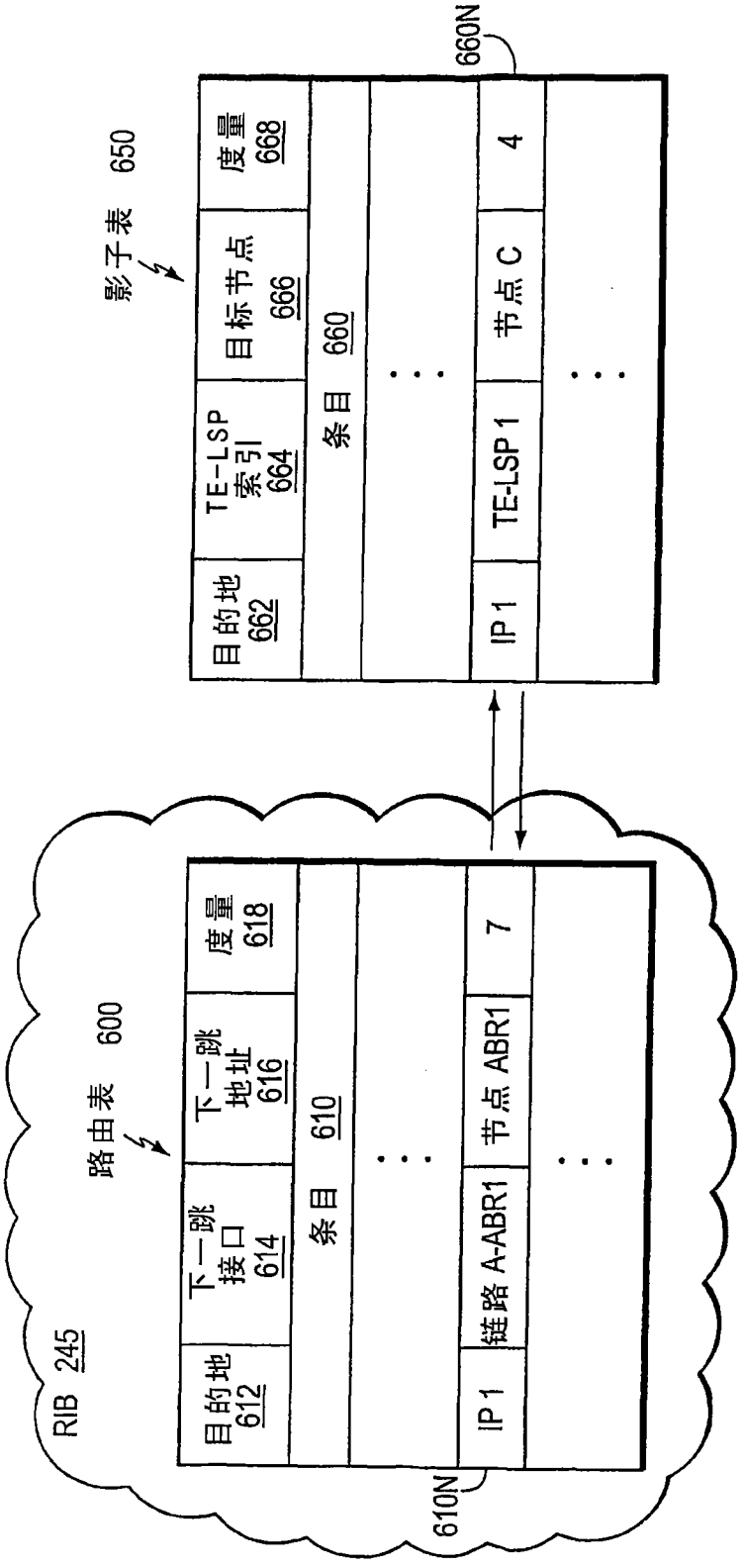

影子表 650
660N
目的地 662
TE-LSP 索引 664
目标节点 666
度量 668
条目 660
IP 1
TE-LSP 1
节点 C
4
RIB 245
路由表 600
610N
目的地 612
下一跳 接口 614
下一跳 地址 616
度量 618
条目 610
IP 1
链路 A-ABR1
节点 ABR1
7

图6

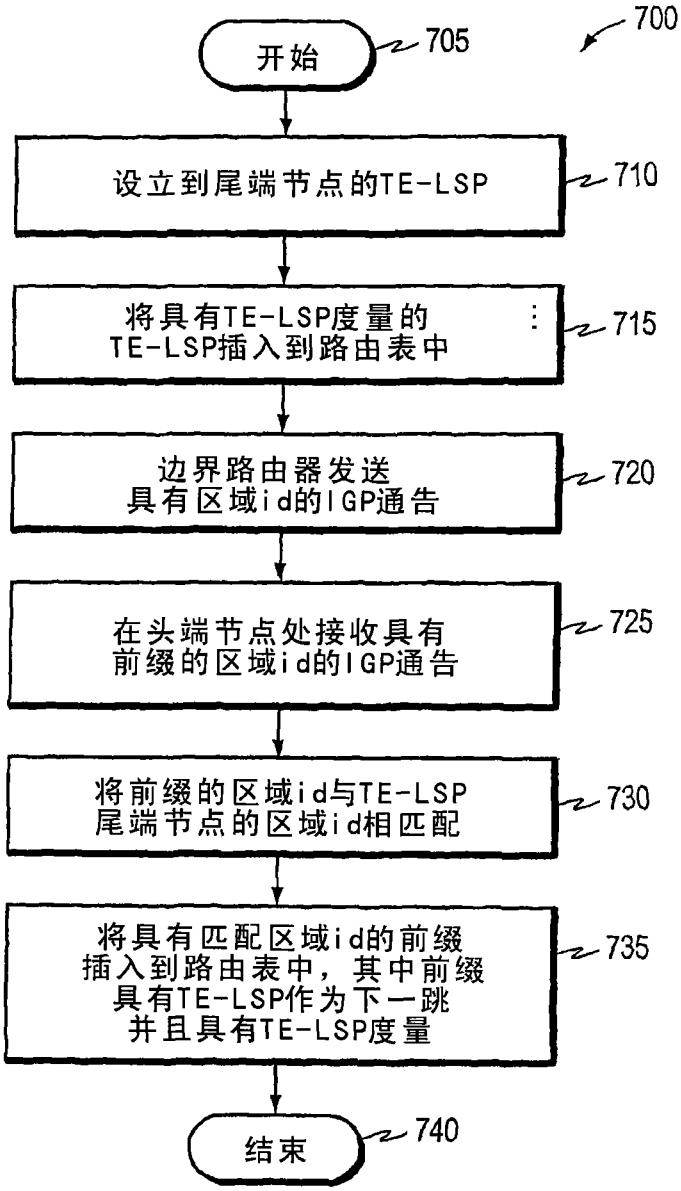

700
开始
705
设立到尾端节点的TE-LSP
710
将具有TE-LSP度量的
TE-LSP插入到路由表中
715
边界路由器发送
具有区域id的IGP通告
720
在头端节点处接收具有
前缀的区域id的IGP通告
725
将前缀的区域id与TE-LSP
尾端节点的区域id相匹配
730
将具有匹配区域id的前缀
插入到路由表中，其中前缀
具有TE-LSP作为下一跳
并且具有TE-LSP度量
735
结束
740

图7

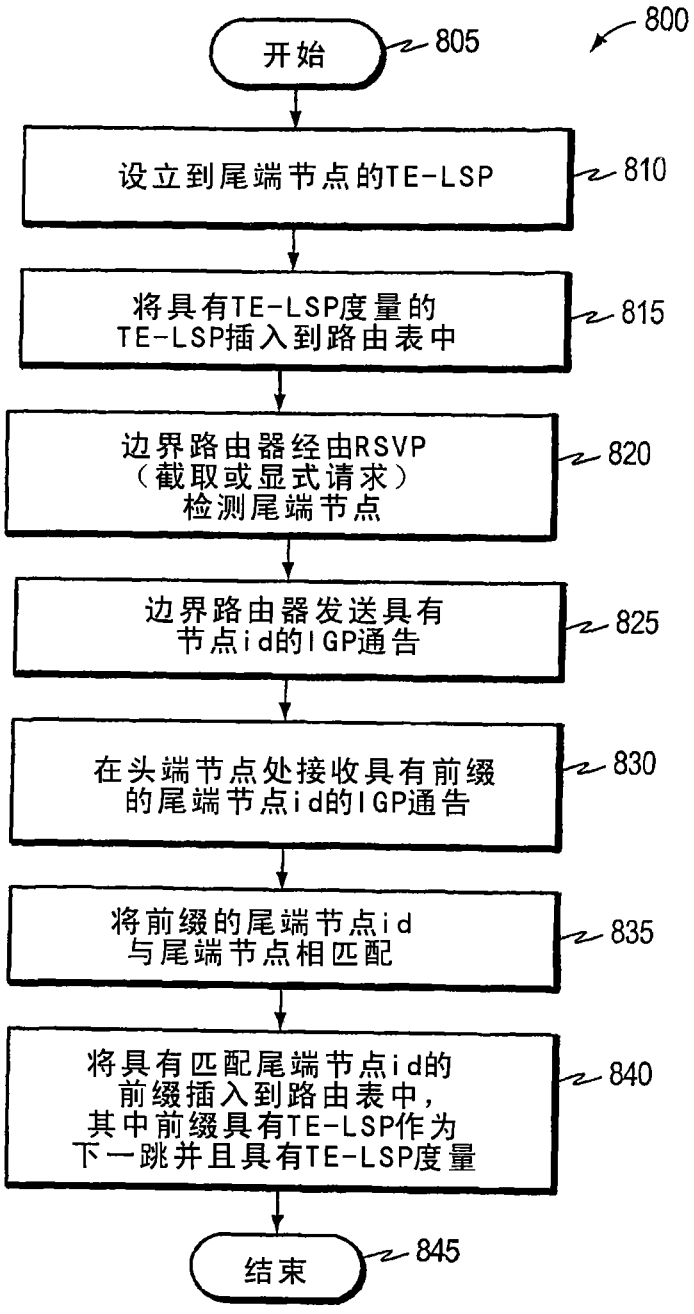

800
开始
805
设立到尾端节点的TE-LSP
810
将具有TE-LSP度量的
TE-LSP插入到路由表中
815
边界路由器经由RSVP
（截取或显式请求）
检测尾端节点
820
边界路由器发送具有
节点id的IGP通告
825
在头端节点处接收具有前缀
的尾端节点id的IGP通告
830
将前缀的尾端节点id
与尾端节点相匹配
835
将具有匹配尾端节点id的
前缀插入到路由表中，
其中前缀具有TE-LSP作为
下一跳并且具有TE-LSP度量
840
结束
845

图8

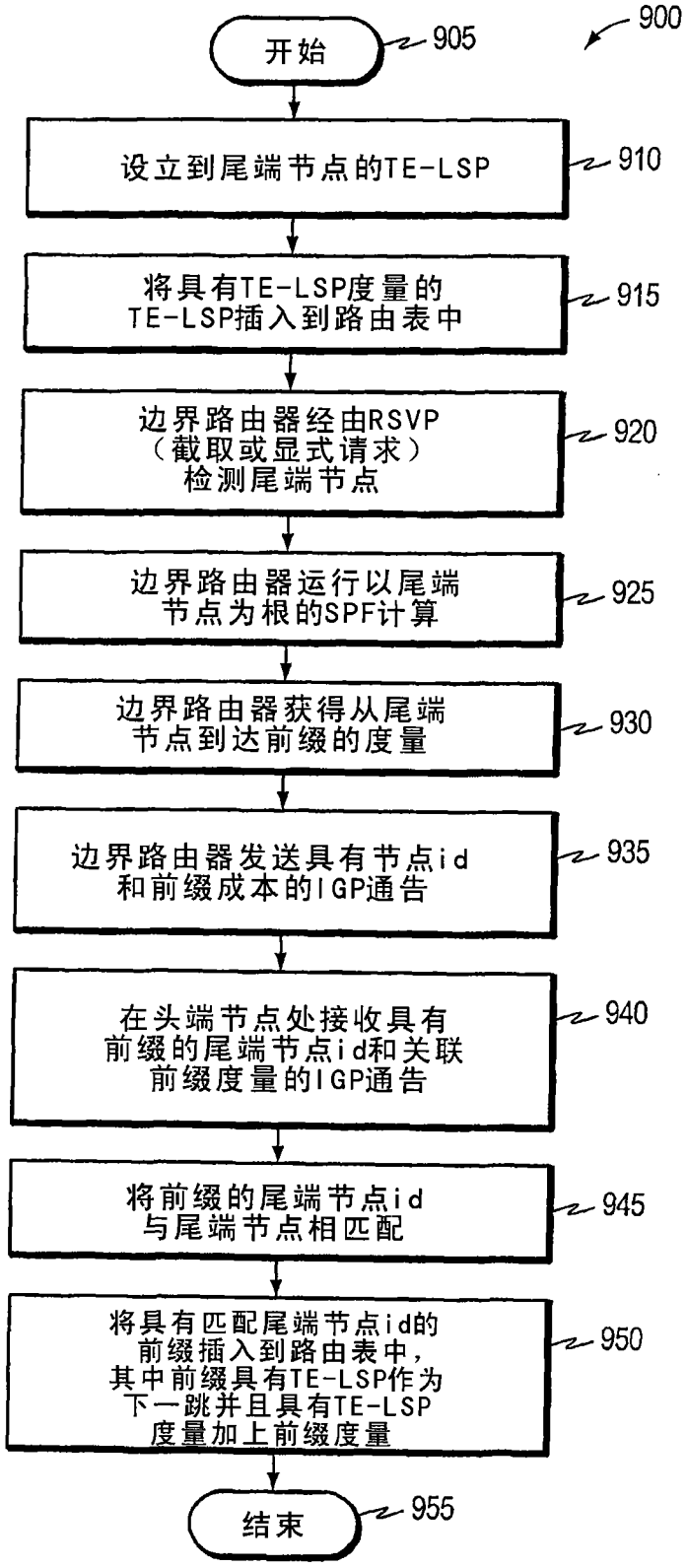

900
开始
905
设立到尾端节点的TE-LSP
910
将具有TE-LSP度量的
TE-LSP插入到路由表中
915
边界路由器经由RSVP
（截取或显式请求）
检测尾端节点
920
边界路由器运行以尾端
节点为根的SPF计算
925
边界路由器获得从尾端
节点到达前缀的度量
930
边界路由器发送具有节点id
和前缀成本的IGP通告
935
在头端节点处接收具有
前缀的尾端节点id和关联
前缀度量的IGP通告
940
将前缀的尾端节点id
与尾端节点相匹配
945
将具有匹配尾端节点id的
前缀插入到路由表中，
其中前缀具有TE-LSP作为
下一跳并且具有TE-LSP
度量加上前缀度量
950
结束
955

图9