

(12) 发明专利申请

(10) 申请公布号 CN 101960785 A

(43) 申请公布日 2011. 01. 26

(21) 申请号 200880127873. 1

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

(22) 申请日 2008. 12. 30

代理人 柯广华 王忠忠

(30) 优先权数据

61/009717 2007. 12. 31 US

12/215350 2008. 06. 26 US

(51) Int. Cl.

H04L 12/28(2006. 01)

H04L 12/56(2006. 01)

(85) PCT申请进入国家阶段日

2010. 08. 30

(86) PCT申请的申请数据

PCT/US2008/088569 2008. 12. 30

(87) PCT申请的公布数据

W02009/088880 EN 2009. 07. 16

(71) 申请人 北方电讯网络有限公司

地址 加拿大魁北克省

(72) 发明人 P·昂贝哈根 D·艾伦

B·格尔梅因 R·拉普

M·阿努马拉 N·布拉格

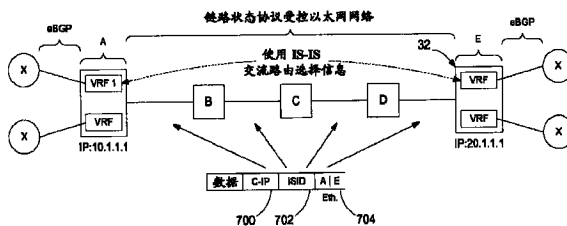
权利要求书 3 页 说明书 10 页 附图 5 页

(54) 发明名称

在链路状态协议受控以太网网络上实现 VPN

(57) 摘要

以太网网络上的节点实现链路状态路由选择协议。节点按 VRF 分配 IP 地址或 I-SID 值, 并使用 LSA 公告 IP 地址或 I-SID 值。当要在 VPN 上转发分组时, 入口节点识别分组的 VPN, 并在 VRF 的客户地址空间中执行 IP 查询以确定下一跳和出口节点上 VRF 的 IP 地址或 I-SID 值。入口节点预先考虑 I-SID 或 IP 报头以识别 VRF, 并创建 MAC 报头以允许将分组转发到以太网网络上的出口节点。当在出口节点接收到分组时, 从分组中剥离 MAC 报头, 并使用追加的 I-SID 或 IP 报头来识别出口 VRF。



1. 一种在链路状态协议受控以太网网络上实现第 3 层虚拟专用网络 (VPN) 的方法, 所述方法包括如下步骤:

给所述 VPN 分配 VPN I-SID;

在所述链路状态协议受控以太网网络中第一节点上为所述 VPN 建立第一 VRF;

给所述第一 VRF 分配所述 VPN I-SID;

由所述第一节点在第一链路状态协议链路状态公告中公告所述 VPN I-SID;

在所述链路状态协议受控以太网网络中第二节点上为所述 VPN 建立第二 VRF;

给所述第二 VRF 分配所述 VPN I-SID;

由所述第二节点在链路状态协议链路状态公告中公告所述 I-SID 的第二附件; 以及

在所述链路状态协议受控以太网网络上传送数据分组时使用所述 VPN I-SID 来允许接收所述分组的节点识别要用于所述分组的进一步处理的正确 VRF。

2. 如权利要求 1 所述的方法, 所述方法还包括如下步骤:

在所述链路状态协议受控以太网网络中其他节点上为所述 VPN 建立多个附加 VRF;

给所述其他 VRF 分配所述 VPN I-SID; 以及

由所述其他节点在链路状态协议链路状态公告中向所述其他 VRF 公告所述 I-SID 的附件。

3. 如权利要求 1 所述的方法, 其中使用 iBGP 来在所述 VRF 之间交流用识别 I-SID 标记的客户路由。

4. 如权利要求 3 所述的方法, 其中还通过 iBGP 交流识别所述 VRF 的 I-SID。

5. 如权利要求 1 所述的方法, 其中使用链路状态协议公告来在所述 VRF 之间交流用识别 I-SID 标记的客户路由。

6. 如权利要求 5 所述的方法, 其中还使用所述链路状态协议公告来交流所述链路状态协议受控以太网网络的路由选择信息。

7. 如权利要求 6 所述的方法, 其中所述链路状态协议是中间系统到中间系统 (IS-IS), 而所述链路状态协议公告具有类型长度值 (TLV), 所述类型长度值 (TLV) 向所述链路状态协议受控以太网网络上的节点指定所述 IS-IS 链路状态公告载送与 VPN 关联的客户路由。

8. 如权利要求 7 所述的方法, 其中所述 TLV 使 IPv4 及 IPv6 客户路由能够在所述 IS-IS 链路状态协议公告中载送。

9. 一种在链路状态协议受控以太网网络上实现因特网协议 (IP) 虚拟专用网络 (VPN) 的方法, 所述方法包括如下步骤:

在所述链路状态协议受控以太网网络中第一节点上为所述 IPVPN 建立第一 VRF;

在所述链路状态协议受控以太网网络内在链路状态公告中公告与所述第一节点的所述第一 VRF 和 MAC 地址关联的第一全局 IP 地址;

在所述链路状态协议受控以太网网络的中间节点中建立至所述 MAC 地址的转发状态; 以及

在所述链路状态协议受控以太网网络上第二节点具有要传送到所述链路状态协议受控以太网网络上所述第一节点的与所述 VPN 关联的分组时, 由所述第二节点对所述分组预先考虑第一报头, 所述第一报头包含与所述第一 VRF 关联的全局 IP 地址, 然后由所述第二节点对所述分组预先考虑第二报头, 所述第二报头包含所述第一节点的 MAC 地址, 并将具

有所述第一和第二报头的分组转发到所述链路状态协议受控以太网网络。

10. 如权利要求 9 所述的方法,其中所述链路状态协议受控以太网网络上的中间节点使用所述第二报头的 MAC 地址根据所述链路状态协议受控以太网网络的节点中用于该 MAC 地址的已安装转发状态转发所述 IP 分组。

11. 如权利要求 10 所述的方法,其中当所述第一节点接收到所述分组时,它读取所述 MAC 报头以获知所述 MAC 报头的目的地地址与其自身 MAC 地址匹配,然后读取所述第一报头的 IP 地址以识别所述第一 VRF。

12. 如权利要求 11 所述的方法,其中识别所述第一 VRF 的 IP 分组还包含第二客户端 IP 报头,所述第二客户端 IP 报头包含可结合所述第一 VRF 中的路由选择信息用于确定所述 IP 分组的客户面向接口的路由选择信息。

13. 如权利要求 9 所述的方法,其中使用 iBGP 在所述第一 VRF 与和所述 VPN 关联的已在所述链路状态协议受控以太网网络中其他节点上例示的其他 VRF 之间交流客户路由。

14. 如权利要求 9 所述的方法,其中使用链路状态协议公告在所述第一 VRF 与和所述 VPN 关联的已在所述链路状态协议受控以太网网络中其他节点上例示的其他 VRF 之间交流客户路由。

15. 一种在与 IP VPN 关联的 VRF 之间跨链路状态协议受控以太网网络交流路由选择信息的方法,所述方法包括如下步骤:

在以太网网络上实现链路状态路由选择协议,所述链路状态路由选择协议设计成使所述以太网网络上的节点能够交流包含与所述以太网网络的节点邻接性和链路状态有关的信息的链路状态公告,并能够构建包含所述链路状态协议受控以太网网络的同步拓扑视图的链路状态数据库;

在所述节点的第一节点上实现与 IP VPN 有关的第一 VRF;

在所述节点的第二节点上实现与所述 IP VPN 有关的第二 VRF;以及

使用所述链路状态公告来实现所述第一 VRF 与第二 VRF 之间 VPN 路由选择信息的交流。

16. 如权利要求 15 所述的方法,其中给所述第一 VRF 和所述第二 VRF 分配共同 I-SID,也使用所述链路状态公告来公告所述共同 I-SID。

17. 如权利要求 15 所述的方法,其中给所述第一 VRF 和第二 VRF 分配共同 I-SID,以及其中给与所述以太网网络上的 VPN 关联的业务预先考虑所述共同 I-SID,以使接收到所述业务的节点能够确定用于处理所述业务的正确 VRF。

18. 如权利要求 17 所述的方法,其中还给与所述 VPN 关联的业务预先考虑 MAC 报头,以允许根据所述链路状态协议受控以太网网络上预先安装的转发状态将所述业务从入口节点转发到出口节点,而不要求中间节点在所述网络的每一跳剥离所述 MAC 报头。

19. 如权利要求 16 所述的方法,还包括以下步骤:

由所述以太网节点的第一节点接收与所述 VPN 关联的客户路由选择信息;

用所述客户路由选择信息更新所述第一 VRF;以及

根据与所述第一 VRF 关联的导出策略,将在链路状态公告中的客户路由选择信息导出到所述第二 VRF。

20. 如权利要求 19 所述的方法,还包括如下步骤:

接收包含所导出客户路由选择信息的链路状态公告 ;以及
根据与所述第二 VRF 关联的导入策略,将所述客户路由选择信息导入到所述第二 VRF。

在链路状态协议受控以太网网络上实现 VPN

[0001] 相关申请的交叉引用

[0002] 本申请要求于 2007 年 12 月 31 日提交的美国临时专利申请号 61/009,717 的优先权,因此将其内容通过引用结合于此。

技术领域

[0003] 本发明涉及以太网网络,更具体来说涉及在链路状态协议受控以太网网络上实现 VPN。

背景技术

[0004] 数据通信网络可以包括多种计算机、服务器、节点、路由器、交换机、网桥、集线器、代理以及耦合且配置成彼此传递数据的其他网络装置。在本文中将这些装置称为“网络单元”。通过利用网络单元之间的一个或多个通信链路在网络单元之间传递如因特网协议 (IP) 分组、以太网帧、数据信元 (cell)、片断或数据位/字节的其他逻辑关联的协议数据单元来经由数据通信网络通信数据。当特定协议数据单元通过网络在源和目的地之间传播时,特定协议数据单元可由多个网络单元进行处理并跨多个通信链路。

[0005] 通信网络上的多种网络单元使用本文称为协议的预定义规则集彼此通信。使用不同协议来监管通信的多个不同方面,如信号应该如何形成以供网络单元之间的传输,协议数据单元应该像什么样、协议数据单元应该由网络单元如何处理或经由网络路由以及如路由选择信息的信息应该如何在网络单元之间交流 (exchange) 的多个方面。

[0006] 以太网是公知的联网协议,它由电气和电子工程师协会 (IEEE) 定义为标准 802.3 以及用于以太网桥接功能性的 802.1。在以太网网络体系结构中,连接到网络的装置在任何给定时间为使用共享通信路径能力而竞争。在使用多个网桥或节点来互连网络段的情况下,常常存在到同一目的地的多个可能路径。此体系结构的好处在于,它提供网桥之间的路径冗余性,并且允许以附加链路形式给网络增加容量。但是,为了防止形成回路 (loop),一般使用生成树来限制在网络上广播业务或洪泛 (flood) 业务的方式。生成树的特点在于,网络中任何一对目的地之间只有一个路径,因此“获知”与分组所经过的给定生成树关联的连通性是可能的。但是,生成树本身有局限性,并且常常导致生成树上链路的过度利用而不是生成树一部分的链路未被利用。

[0007] 为了克服以太网网络中固有的一些局限性,在 2006 年 10 月 2 日提交的标题为“Provider Link State Bridging”的申请号 11/537,775 中公开一种链路状态协议受控以太网网络,因此将其内容通过引用结合于此。正如该申请中更详细描述,并不在每个节点通过使用与透明桥接组合的生成树协议 (STP) 算法来利用获知的网络视图 (network view),而是在链路状态协议受控以太网网络中,形成网格网络的网桥交流链路状态公告 (advertisement) 以使每个节点能够具有同步的网络拓扑视图。这通过链路状态路由选择系统的公知机制实现。提供方链路状态桥接网络优选使用如 IEEE 802.1ah 中定义的 MAC 中 MAC (MAC-in-MAC) 封装,以便完全将客户 (C-MAC) 地址空间与提供方 (B-MAC) 地址空间

分离,使得链路状态路由选择系统仅暴露于在提供方控制下分配的 B-MAC 地址。通过使用 MAC 中 MAC 封装,网络中的网桥拥有同步的网络拓扑视图,拥有必要的单播和多播连通性知识,可以计算网络中任何一对网桥之间的最短路径连通性,并且可以单独根据计算的 network 视图填充 (populate) 其转发信息库 (FIB:forwarding information base),所有这些仅依据提供方管理的 B-MAC 地址。

[0008] 当所有节点已经计算它们在同步视图中的角色并填充它们的 FIB 时,网络将拥有从对等网桥集(出于无论哪种原因需要与任何给定网桥通信的那些网桥)到该任何给定网桥的无回路单播树;以及按在任何给定网桥托管 (host) 的服务实例将拥有从该任何给定网桥到同一对等网桥集或子集的全等且无回路的点到多点 (P2MP) 多播树。结果是给定网桥对之间的路径不限于通过 (transit) 生成树的根网桥,而且总结果可以更好利用网络连通性的宽度。本质上,每个网桥构成定义至该网桥的单播连通性和来自该网桥的多播连通性的一个或多个树的根。

[0009] 当客户业务进入提供方网络时,客户 MAC 地址 (C-MAC DA) 被解析为提供方 MAC 地址 (B-MAC DA),使得提供方可以使用提供方 MAC 地址空间在提供方网络上转发业务。此外,提供方网络上的网络单元配置成基于虚拟 LAN ID (VID) 转发业务,使得被寻址到同一目的地地址但具有不同 VID 的不同帧可以经由网络通过不同路径转发。操作中,链路状态协议受控以太网网络可以将一个 VID 范围与最短路径转发关联,使得单播和多播业务可以使用该范围中的 VID 进行转发,而业务工程路径可以跨网络在最短路径以外的路径上创建,并使用第二 VID 范围进行转发。

[0010] 链路状态路由选择协议的两个示例包括开放式最短路径优先 (OSPF:Open Shortest Path First) 和中间系统到中间系统 (IS-IS:Intermediate System to Intermediate System),但是还可以使用其他链路状态路由选择协议。IS-IS 在例如 ISO 10589 和 IETF RFC 1195 中进行了描述,由此将其各个内容通过引用结合于此。

[0011] 除了安装最短路径单播转发状态外,节点还可以为网络上的多播树安装转发状态。2007 年 2 月 5 日提交标题为“Multicast Implementation in a Link State Protocol Controlled Ethernet Network”的美国专利申请号 11/702,263 中更详细地描述了在链路状态协议受控以太网网络中实现多播的方法的示例,由此将其内容通过引用结合于此。正如该申请中描述的,可以使用链路状态公告来公告多播组成员关系以使多播组的转发状态被安装在网络上。具体来说,可以给给定多播组的每个源分配用于在网络上转发帧的目的地 MAC 地址 (DA)。如果网络上的节点确定它们在从多播源到通过链路状态协议公告多播组中“兴趣”的目的地节点中至少之一的最短路径上,则它们安装源 / 组树的转发状态。

[0012] 多播中的兴趣可以基于如 I-SID 的兴趣标识符群体 (community),使得网络上的节点将在它位于均已在与多播组关联的兴趣标识符群体中公告兴趣的源与目的地之间的最短路径上时为多播组安装转发状态。I-SID 通常与 802.1ah 关联,且意味着附加的 MAC 报头 (客户源和目的地 MAC 地址)。本文中容易理解并假设采用没有 C-MAC 报头的 I-SID。但是,转发状态基于与多播关联的多播 DA 和 VID。操作中,网络上的多个节点可以在特定 I-SID 中公告兴趣。网络上的节点跟踪哪些节点在哪个 I-SID 中公告了兴趣,并将在它们位于在特定 I-SID 中公告了兴趣的两个节点之间的最短路径上时为与 I-SID 关联的 DA/VID 对安装转发状态。这允许在不要求在网络上洪泛帧的情况下为兴趣群体安装转发状态。

[0013] 链路状态协议受控以太网网络工作在链路层（第 2 层）。即，入口（ingress）节点创建可用于跨以太网网络、例如从节点 A 到节点 E 交换（switch）帧的 MAC 报头。其他网络、如因特网协议（IP）网络工作在例如第 3 层（网络层）的更高层。IP 网络基于与 IP 分组关联的 IP 报头的 IP 地址转发分组。在常规 IP 网络中，在经由网络的每一跳（hop）执行 IP 查询。即，每个节点将剥离外部以太网报头，读取 IP 报头，并进行 IP 查询以路由分组通过网络。然后节点将添加新以太网报头以便在网络上将分组转发到下一节点。

[0014] 图 1 示出多协议标签交换（MPLS:multi-protocol label switching）网络中如何进行 IP 转发。MPLS 网络减少了要求在网络中在各个路由器执行的 IP 查询的数量。在 MPLS 网络中，将经由 MPLS 网络建立多个标签交换路径（LSP:Label Switched Path）。经由网络确定并创建 LSP 的特定方式是公知的。在图 1 所示的示例中，将假定标签交换路径包括节点 A、B、C、D、E。当分组从网络 X 到达边缘路由器 22A 时，边缘路由器 22A 将执行 IP 查询以确定应该使用哪个标签来将 IP 分组经由 MPLS 网络交换到 LSP 以到达网络 Y。边缘路由器 22A 还将确定分组在路径上的下一跳，并将 MAC 报头应用于分组以使分组被转发到在路径上的下一跳（路由器 22B）。

[0015] 路由器 22B 将剥离外部 MAC 报头并读取 MPLS 标签。当经由网络建立了 LSP 时，标签分配协议将已经创建路径上标签之间的关联，从而例如在特定 IP 转发等价类的标签 100 与标签 210 之间形成关联。标签关联允许路由器使用标签来转发分组而不是执行 IP 查询。因此，例如当路由器 22B 接收到具有 MPLS 标签 100 的分组时，它将用新 MPLS 标签 210 替换标签，并将分组转发到出站（outbound）接口。在转发分组之前，路由器 22B 将给分组添加新 MAC 报头，在本实例中新 MAC 报头是 SA = J、DA = K 以将源 MAC 地址识别为路由器 22B 上的 MAC J 以及将目的地 MAC 地址识别为路由器 22C 上的 MAC K。该过程将在沿标签交换路径上的每一跳重复，直到分组到达标签边缘路由器（节点 E）为止。

[0016] MPLS 因此允许在 MPLS 网络的边缘执行单次 IP 路由查询，并允许使用标签交换而不是 IP 查询来跨 MPLS 网络转发分组。执行初始 IP 查询并将标签分配给分组以将分组置于 LSP 的边缘路由器称为标签边缘路由器（LER）。MPLS 网络上执行标签交换的中间路由器一般称为标签交换路由器（LSR）。在每一跳通过在剥离 MAC 报头之后互换（swap）标签来实现 MPLS 网络中的转发。MPLS 转发仍要求在跨网络（当采用以太网链路时）的每一跳进行 MAC 报头剥离，要求每个 LSR 执行标签查询和标签互换，并要求每个 LSR 然后添加另一 MAC 报头以将分组沿 LSP 递送到下一 LSR。此过程要求每个节点上的处理和智能，因此这导致成本高的解决方案。此外，为使之有效，必须在开始时建立 LSP，这本身是计算成本高的过程。

[0017] 2007 年 12 月 31 日提交的标题为“IP Forwarding Across A Link State Protocol Controlled Ethernet Network”的美国专利申请号 12/006,258 中公开了一种使链路状态协议受控以太网网络能够实现网络层转发的方法，为此将其内容通过引用结合于此。此外，2007 年 12 月 31 日提交的标题为“MPLS Node Replacement Using A Link State Protocol Controlled Ethernet Network”的美国专利申请号 12/006,257 中还公开了一种使链路状态协议受控以太网网络能够替换 MPLS 网络中标签交换路由器的方法，由此同样将其内容通过引用结合于此。

[0018] 因为许多客户可以共享提供方的网络，所以已经开发了多种技术来允许将一个客户的数据与另一客户的数据隔离。按照惯例，这通过使用虚拟专用网络来实现。图 2 示出其

中已经跨网络建立一个或多个虚拟专用网络 (VPN) 的示范 MPLS 网络。可以使用对 IPv4VPN 的因特网工程任务组 (IETF) 请求评论 (RFC) 4364 或对 IPv6VPN 的 RFC 4659 中提出的过程在 MPLS 网络上建立 VPN。简言之,为了在 MPLS 网络上建立 VPN,参与到 VPN 的每个 LER 将实现虚拟路由选择和转发 (VRF) 表。VRF 用于保存 VPN 的路由。在 MPLS 网络内使用如内部边界网关协议 (iBGP) 的路由选择协议在 VRF 之间交流路由选择信息。当边缘节点从客户例如经由外部 BGP (e-BGP) 获知路由时, VRF 将根据 VRF 的路由导出 (export) 策略导出经 iBGP 到 VPN 上其他 VRF 的路由。当 VRF 接收到路由时,它将根据它的路由导入 (import) 策略导入路由。由此, VRF 将安装要用于在 VPN 上转发业务的路由选择信息。

[0019] 当在至 MPLS 网络的入口接收到分组时,入口 LER 将确定跨 MPLS 网络路由分组要使用的正确 VRF,并在 VRF 中执行 IP 查询确定分组的下一跳。入口 LER 将获取分组的标签,并经 MPLS 网络将分组转发到 LSP。入口 LER 还可以分配内部标签,内部标签将允许在出口 (egress) LER 识别正确 VRF。因为 IP VPN 在 MPLS 网络上盛行,并且能够集成链路状态协议受控以太网网络以与 IP 或 MPLS 网络一起工作或替代 IP 或 MPLS 网络,所以允许 IP VPN 在链路状态协议受控以太网网络上实现将是有利的。

发明内容

[0020] 链路状态协议受控以太网网络上节点实现如中间系统至中间系统 (IS-IS) 的链路状态路由选择协议。网络上的节点也称为链路状态分组的 IS-IS 链路状态公告 (LSA) 获知邻接性 (adjacency),并计算网络上所有节点对之间的最短路径。每个参与节点填充其转发信息库 (FIB) 以构造网络上各节点对之间的单播转发状态。还可以构造业务工程 (engineered) 路径,并将 TE 路径的转发状态安装到网络上的节点 FIB。

[0021] IS-IS 允许与特定网络单元中使用的交换或转发技术无关地交流拓扑信息以及第 2 层和第 3 层地址至特定网络位置和接口的绑定。链路状态协议受控以太网网络上的节点按 VRF 分配 IP 地址,然后在 IS-IS 链路状态公告 (LSA) 中公告与其 VRF 关联的 IP 地址。可以创建类型长度值 (TLV) 以向节点指示 LSA 包含 VRF 的 IP 地址。当接收到分组时,至链路状态协议受控以太网网络的入口节点识别分组的 VRF,并在 VRF 中执行 IP 查询以确定下一跳和网络上 VRF 的出口节点。入口节点创建具有源 IP 地址 = 入口 VRF IP 地址及目的地 IP 地址 = 出口 VRF IP 地址的 IP 报头。入口节点还创建 MAC 报头以将分组置于至链路状态协议受控以太网上出口节点的路径。链路状态协议受控以太网上的节点使用其 FIB 中的转发状态将分组转发给 DA/VID,而不在每一跳执行 MAC 剥离操作。当在出口节点接收到分组时,从分组中剥离 MAC 报头,并使用 IP 报头识别出口 VRF。然后,通过使用客户端 IP 报头中的信息来确定如何转发分组,在出口节点上在所识别 VRF 中执行 IP 查询。2007 年 11 月 6 日提交的标题为“Supporting BGPBased IP-VPN in a Routed Network”的美国专利申请号 11/935,563 中公开了在以太网网络上实现 IP-VPN 关联的附加细节,因此将其内容通过引用结合于此。

[0022] 在另一实施例中,给链路状态协议受控以太网网络上的节点所支持的每个 VRF 分配 I-SID。每个节点公告与其 VRF 关联的 I-SID 集。使用 iBGP 来在 VRF 之间交流路由,其中 iBGP 交流还对 I-SID 进行编码或具有预先配置的 I-SID 以路由目标绑定,使得 VRF 能够识别哪些路由受关注。在一个实施例中,使用 RFC 4364 以在 VRF 之间交流路由,并且根据

实施例,不仅使用 iBGP 来交流路由,而且还交流用于识别 VRF 的 I-SID 值。当入口节点接收到分组时,它识别入口 VRF 并执行 IP 查询以确定分组的下一跳。入口节点将确定与 VRF 关联的 I-SID,并将 I-SID 添加到分组以使出口节点能够确定要用于将分组转发出网络的正确 VRF。入口节点还将创建 MAC 报头以将分组置于至链路状态协议受控以太网上出口节点的路径。链路状态协议受控以太网上的节点使用其 FIB 中的转发状态将分组转发给 DA/VID,而不在每一跳执行 MAC 剥离操作。当在出口节点接收到分组时,从分组中剥离 MAC 报头,并使用 I-SID 来识别出口 VRF。然后,通过使用客户端 IP 报头中的信息,在出口节点上所识别 VRF 中执行 IP 查询以确定如何转发分组。

[0023] 在另一实施例中,与第二实施例中一样,给每个 VRF 分配 I-SID,但是不使用 iBGP 来在 VRF 之间交流路由。而是通过将路由信息包含在网络上交流的链路状态公告 (LSA) 中,在链路状态协议受控以太网上的节点之间本地交流路由信息。可以创建新类型长度值 (TLV) 以允许 LSA 基于按 I-SID(按 VRF) 载送路由信息。当跨链路状态协议受控以太网网络转发分组时,入口节点和出口节点以与前一实施例相同的方式使用路由选择信息。

附图说明

[0024] 在随附权利要求书中具体指出本发明的多个方面。附图中以示例方式图示本发明,附图中相似引用指示相似单元。附图仅出于说明目的公开本发明的多种实施例,而无意限制本发明的范围。为了清楚起见,并未在每个附图中标记每个组件。在附图中:

[0025] 图 1 是示出可如何在 MPLS 网络中实现 IP 转发的功能框图;

[0026] 图 2 是示出在 MPLS 网络中交流路由选择信息的功能框图;

[0027] 图 3 是配置成根据本发明实施例实现 IP VPN 的示范链路状态协议受控以太网网络的功能框图;

[0028] 图 4 是可根据本发明实施例能够在图 3 的链路状态协议受控以太网网络中实现 IP VPN 所实现的过程的流程图;

[0029] 图 5 是配置成根据本发明实施例实现 IP VPN 的示范链路状态协议受控以太网网络的功能框图;

[0030] 图 6 是可根据本发明实施例能够在图 5 的链路状态协议受控以太网网络中实现 IP VPN 所实现的过程的流程图;

[0031] 图 7 是配置成根据本发明实施例实现 IP VPN 的示范链路状态协议受控以太网网络的功能框图;

[0032] 图 8 是可根据本发明实施例能够在图 7 的链路状态协议受控以太网网络中实现 IP VPN 所实现的过程的流程图;

[0033] 图 9 是链路状态公告 (LSA) 可如何根据本发明实施例构造类型长度值 (TLV) 的一个示例的功能框图,类型长度值将允许 LSA 在与 IP VPN 关联的 VRF 之间载送路由选择信息;以及

[0034] 图 10 是配置成根据本发明实施例在链路状态协议受控以太网网络中使用的网络单元的可能实现的示意表示。

具体实施方式

[0035] 通过使至网络中交换域的入口节点将网络层地址映射到经由链路状态协议受控以太网网络的路径,可使用链路状态协议受控以太网网络实现网络层路由选择和转发。虽然本描述频繁地将 IP 地址称为示范网络层地址,但是本公开旨在同样应用于其他类型的网络层地址。因此,本发明不限于依赖于网络层寻址的 IPv4 或 IPv6 的实现,而是同样应用于其他形式的网络层寻址。

[0036] 将如 IP 地址的网络层地址映射到链路状态协议受控以太网网络中的路径,允许通过使至交换网络的入口节点将 IP 地址映射到链路状态协议受控以太网网络上的端点以太网 MAC 地址来进行在交换网络上的 IP 转发。当在链路状态协议受控以太网网络上的节点获知网络层地址时,节点可以公告网络层地址知识。公告网络层地址允许入口节点在接收到网络层分组时确定网络上哪个其他链路状态协议受控以太网网络节点能够到达该网络层地址。

[0037] 至链路状态协议受控以太网网络的入口节点创建用于将网络层分组转发到网络上端点 MAC 地址的 MAC 报头。如果入口节点是节点 A,并且预期出口节点是节点 E,则入口节点 A 将创建被寻址到与节点 E 关联的 MAC 地址的 MAC 报头。因为中间节点将已经安装最短路径转发状态以允许它们沿至该节点的最短路径转发被寻址到 MAC E 的分组,所以这些中间节点仅在其转发信息库 (FIB) 中执行 MAC 查询,并将分组转发到网络上的正确目的地。这些中间节点不要求在沿路径的每一跳剥离 MAC 报头也不要求创建新 MAC 报头。因此,链路状态协议受控以太网网络路径可用于实现经链路状态协议受控以太网网络的网络层转发。

[0038] 可以使用以太网交换来实现其中可在分组到达交换域中的转发实体时对分组执行单次网络层查询的网络层转发。可以将网络地址映射到 MAC 报头,MAC 报头可用于跨交换域传输分组而不要求在分组经过 (traverse) 交换域时进行进一步的网络层查询操作。不同于传统的路由网络,交换域允许网络层业务在不执行网络层查询的情况下经由多个以太网交换机进行转发。因此,能够获得不对所有节点要求网络层查询的 MPLS 功能目标。但是,使用链路状态协议受控以太网网络转发网络层业务具有优于 MPLS 的其他优点:它不要求在中间节点剥离 MAC 报头,没有互换模拟交换行为而添加的标签,也没有管理链路本地标签绑定所要求的关联信令开销。而是,可以在分组经过网络上的节点时使用同一 MAC 报头来经多个网络节点交换分组。

[0039] 图 3 是功能框图,示出根据本发明实施例示范选择形成经链路状态协议受控以太网网络的路径的节点以及交流路由选择信息以允许网络实现 IP VPN。在图 3 以及下文论述的图 5 和图 7 所示的实施例中,将假定节点 A-E 正参与到链路状态协议受控以太网网络,其中网络中的每个节点交流链路状态公告 (LSA) 以获知网络拓扑,并将最短路径转发状态 (多播和单播) 安装到其转发信息库中。链路状态协议受控以太网网络可以包括为避免使附图太复杂而未在这些示例中示出的众多其他节点。

[0040] 在图 3 所示的实施例中,通过使链路状态协议受控以太网网络上的入口和出口节点为每个所支持 IP VPN 实现 VRF,可以在链路状态协议受控以太网网络上实现 IP VPN。具体而言,要支持特定 IP VPN 的每个节点将为该 IP VPN 实现 VRF,使得它可以保密方式为 IP VPN 维持路由选择信息。在 IP VPN 领域中 VRF 是公知实体。客户端 (client) 路由可仅配置在 VRF 中,或使用如外部 BGP 的外部路由选择协议通过客户端公告来发现,然后由入口节

点将其存储在 IP VPN 的 VRF 中。链路状态协议受控以太网网络中不同节点上的 VRF 以常规方式交流路由,在本实施例中,使用如 iBGP 的内部路由选择协议。

[0041] 图 4 示出可用于允许在链路状态协议受控以太网网络上实现 IPVPN 的一个过程。如图 4 所示,根据本发明实施例,链路状态协议受控以太网网络上的每个节点将给它支持的每个 VRF 分配 IP 地址(从提供方的全局 IP 地址空间,与每个 IP VPN 内使用的客户 IP 地址空间分离)(400)。链路状态协议受控以太网网络上的节点一般公告它们知道的 IP 地址,并且根据本发明实施例还可以使用该同一机制公告分配给其 VRF 的全局 IP 地址(402)。可选地,可以使用唯一 TLV 来指定给 VRF 分配由 LSA 载送的 IP 地址而非普通 IP 地址。当其他节点接收到包含与 VPN 关联或给 VPN 分配的 IP 地址的 LSA 时,它们将在其链路状态数据库中把 IP 地址添加到其用于该节点的条目。因此,所有节点都知道与所有其他节点的 VRF 关联的提供方全局 IP 地址。具有相同 IP VPN 的 VRF 的节点以常规方式使用 iBGP 来交流路由以将 VPN 的路由选择信息填充到 VRF(406)。

[0042] 当至链路状态协议受控以太网网络的入口节点(如图 3 中的节点 A)在 UNI 接口上接收到 IP 分组(410)时,它将识别待用于执行 IP 查询的 VRF 以确定如何处理该 IP 分组(412),这常常从到达的物理或虚拟端口进行推断。在确定 VRF 之后,入口节点将在 VRF 中执行 IP 查询以确定分组的目的地(414)。假定出口节点是图 3 中的节点 E,则 VRF 中的 IP 查询将返回节点 E 作为分组下一跳。

[0043] 入口节点将通过把入口节点上 VRF 的 IP 地址用作源地址以及把出口节点上 VRF 的 IP 地址用作目的地地址来添加 IP 报头 302(416)。例如,在图 3 所示的网络中,IP 报头把节点 A 上 VRF 的 IP 地址用作源地址以及把节点 E 上 VRF 的 IP 地址用作目的地地址。IP 报头将允许出口节点当在出口节点 E 接收到分组时识别用于将分组转发出链路状态协议受控以太网网络的正确 VRF。

[0044] 入口节点还创建 MAC 报头以将分组置于经由链路状态协议受控以太网网络的交换路径上(418)。正如上文提到的,链路状态受控以太网网络上的节点将安装转发状态以沿至目的地 MAC 地址的最短路径交换分组。节点能够通过读取 MAC 报头的 DA 和 VID,然后转发分组而不要求在每一跳剥离和替换 MAC 报头来做到这一点。因此,可以跨链路状态协议受控以太网网络转发分组以到达其在图 3 中的目的地(节点 E)。

[0045] 当出口节点 E 接收到分组时,它将剥离外部 MAC 报头 304,并使用外部 IP 报头 302 来识别要用于转发分组的正确 VRF(420)。然后,出口节点将使用所识别 VRF 中的内部客户 IP 报头 300 执行 IP 查询以将分组转发到正确目的地(422)。

[0046] 图 5 和图 6 示出可以如何在链路状态协议受控以太网网络上实现 IP VPN 的另一示例。如图 5-6 所示,在本实施例中,链路状态协议受控以太网网络上的边缘节点向在本地例示为 VRF 的每个 VPN 分配 I-SID(600)。边缘节点将 I-SID 包含在结合链路状态协议受控以太网网络上使用的链路状态协议传送的链路状态公告(LSA)中。例如,如果正将 IS-IS 用作链路状态协议,则边缘节点将生成包含给在该边缘节点支持的 VPN 分配的 I-SID 的 LSA(602)。因此,如果 E 支持特定 VPN,则它将该 VPN 的 I-SID 分配给 VRF,并生成将传送到 A 的 LSA。A 将与节点 E 支持的 VRF 关联的 I-SID 添加到其对于节点 E 的链路状态数据库(LSDB)条目(604)。通过以标准方式使用 iBGP,交流多个网络节点上实现的相同 VPN 的 VRF 信息(606)。

[0047] 当 A 接收到分组 (610) 时,它将识别要用于执行 IP 查询的 VRF (612) 并在 VRF 中执行 IP 查询以确定下一跳。在此实例中,将假定查询为 E 中例示的此 VPN 返回经 VRF 可到达的 IP 地址 (614)。A 将 I-SID 502 添加到分组以识别用于转发 E 上分组的 VPN 以及最终 VRF (616),并创建 MAC 报头 504 以将分组置于经链路状态协议受控以太网网络至 E 的路径 (618)。当 E 接收到分组时,它将剥离外部 MAC 报头 504,并使用 I-SID 502 来为要用于转发分组的那个 VPN 识别 VRF (620)。E 将在所识别的 VRF 中使用客户 IP 报头 500 执行 IP 查询以将分组转发 (622) 到正确的面向客户接口。

[0048] 图 7-8 示出本发明的另一实施例,其中在用于链路状态协议受控以太网网络的 IS-IS 协议中交流 I-SID 标记的客户路由选择信息。由此,本实施例以与图 5-6 所示的实施例相似的方式工作,例外的是,在本实施例中使用链路状态路由选择协议在 VRF 之间交流路由选择信息而不要求使用如 iBGP 的内部网关路由选择协议来交流路由。

[0049] 具体而言,如图 7-8 所示,边缘节点将识别 VPN 的 I-SID 绑定到边缘节点支持的每个 VRF (800),并在链路状态协议受控以太网网络上在 LSA 中公告 I-SID (802)。当节点接收到包含 I-SID 的 LSA 时,它将 I-SID 添加到其对于发出 LSA 的节点的 LSDB 条目 (804)。当节点获知与 I-SID 关联的新路由时,它将路由填充到正确 VRF,且还生成包含路由的 LSA。可以使用新类型长度值 (TLV) 来在 VRF 之间载送路由选择信息,如下文结合图 9 所述。因此,端节点之间交流的 LSA 包括 VRF 的路由信息 (按 I-SID 的路由信息),使得边缘节点可以避免运行 iBGP (806)。

[0050] 当如节点 A 的边缘节点接收到分组时 (810),它将识别要用于对分组执行 IP 查询的 VRF (812)。边缘节点然后将在 VRF 中执行 IP 查询以确定分组的下一跳。在此示例中,将假定分组的下一跳是出口节点 E (814)。

[0051] 入口节点 (节点 A) 将 I-SID 702 添加到分组以在 VPN 的 E 上识别应该用于处理分组的 VRF (816)。A 还将创建 MAC 报头 704 以将分组置于经链路状态协议受控以太网网络至出口节点 (节点 E) 的交换路径上 (818)。

[0052] 当出口节点接收到分组时,它将剥离外部 MAC 报头 704,并使用 I-SID 702 来识别要用于转发分组的 VRF (820)。出口节点然后将在所识别 VRF 中使用客户 IP 报头 700 执行 IP 查询以将分组从链路状态协议受控以太网网络转发到正确的面向客户接口。

[0053] 正如上文结合图 7-8 提到的,可能有必要定义新类型长度值 (TLV) 以允许 IS-IS 链路状态公告 (LSA) 按 I-SID 载送路由信息。TLV 指定 LSA 的格式,以允许网络上的节点以正确方式解释 LSA 的字段。基本上,TLV 指定 LSA 的格式,并指定可以在 LSA 中运送的信息的类型。IS-IS TLV 的格式由 ISO 10589 定义。此标准规定 TLV 的第一个八位组 (octet) 对提供信息范围和信息格式的类型或“码点”进行编码。

[0054] 图 9 图示可以用于允许 IS-IS LSA 基于按 I-SID 在 VRF 之间运送路由选择信息的一个可能 ISIS LSA (TLV)。已假定 LSA TLV 的第一个八位组将指定值,该值将允许链路状态协议受控以太网网络上的系统理解 LSA TLV 中包含的字段的格式。

[0055] 在图 9 所示的示范 TLV 中,TLV 900 包含作为常规 TLV 中标准字段的若干字段。例如,TLV 可以包含系统 ID 字段 902、区域地址字段 904 以及与发出 LSA 的节点的相邻节点 (neighbor) 有关的信息 906。TLV 还包含供节点指定其 IPv4 接口 908 以及识别其知道的 IPv4 地址 908 和 IPv6 地址 910 的字段。TLV 还可以包含其知道的 I-SID 914。

[0056] 根据本发明的实施例, TLV 还可以包含将允许节点指定可用于 IPVPN、已分配给在该节点例示的 VRF 的 I-SID 的字段 (916)。如上文提到的, 允许节点给 VRF 分配 I-SID 将允许节点在跨链路状态协议受控以太网网络传送分组时将 I-SID 报头包含在分组上, 使得可以确定要由出口节点用于将分组转发出网络的正确 VRF。因此, TLV 允许节点通过公告分配给那些 VRF 的 IP VPN-I-SID 知识来公告特定 VRF 知识。在图 5-6 所示的实施例以及图 7-8 所示的实施例中, 将使用 IPVPN-I-SID 字段 916 来公告 VRF。

[0057] TLV 900 还包括允许 LSA 包含 IPv4 路由 918 和 IPv6 路由 920 的两个附加子字段。按 IP VPN-I-SID 包含这两个子字段, 使得可以基于按 IP VPN-I-SID 指定 IPv4 和 IPv6。这允许使用链路状态路由选择系统中实现的本地 (native) LSA 在 VRF 之间交流路由选择信息, 而不要求在 VRF 之间运行如 iBGP 的单独路由选择协议实例。VRF 可以在导出路由时应用 VRF 路由导出策略, 并且还可以在决定是否将路由选择信息导入到其 VRF 中时应用标准 VRF 导入策略。还可以使用其他 TLV 格式, 图示的实施例旨在提供可以如何将 LSA 格式化以载送与 IP VPN 关联的信息的示例。

[0058] 图 10 是如图 3、5 和 7 中网络单元 A-E 的网络单元的可能实现的示意表示, 网络单元配置成在链路状态协议受控以太网网络中使用。网络单元包括路由选择系统模块 80, 路由选择系统模块 80 配置成使用链路状态路由选择协议来与链路状态协议受控以太网网络中的对等方交流包含关于网络拓扑的路由选择和其他信息的控制消息。路由选择系统 80 接收的信息可以存储在链路状态数据库 88 中或以其它方式存储。如前文论述的, 信息交流允许网络上的节点生成同步的网络拓扑视图, 该视图允许路由选择系统模块 80 计算至网络上其他节点的最短路径。将路由选择系统 80 计算的最短路径规划到 FIB 82 中, FIB 82 填充有适当条目以用于基于计算的最短路径、多播树、业务工程路径条目以及基于其他条目来指引业务穿过网络。

[0059] 根据本发明的实施例, 路由选择系统 80 可以交流包含网络层可达性信息的路由更新。网络上的节点知道的网络层地址将被存储在网络单元上的链路状态数据库 88 中, 以允许出口节点在网络层分组到达时选择链路状态协议受控以太网网络上的正确出口节点。网络层地址知识还允许在网络上实现多播转发状态, 以便通过使节点安装相同 IP 多播组中感兴趣的节点对之间的转发状态而允许网络层多播由网络上的节点进行处理。

[0060] 根据本发明实施例的网络单元还包括 VRF 90, VRF 90 配置成包含用于要在链路状态协议受控以太网网络上实现的 VPN 的路由选择信息。VRF 可以如图所示地单独实现, 或者可以作为 FIB 82 的指定部分实现。

[0061] 当接收帧时, 如果分组的 DA 指示将分组寻址到节点, 则节点将从分组剥离外部 MAC 报头, 并查看内部 I-SID 或 IP 报头以确定用于为分组执行 IP 查询的 VRF。然后节点将访问正确的 VRF, 基于客户 IP 报头执行 IP 查询, 并将分组转发给客户 IP 报头的目的地地址。

[0062] 还应理解, 正如本领域技术人员将理解的, 所描述的模块仅出于说明目的, 并且可以通过在节点的模块之中组合或分配功能来实现所述模块。

[0063] 上述功能可以实现为存储在计算机可读存储器中并在计算机平台的一个或多个处理器上执行的程序指令集。但是, 对于本领域技术人员而言显而易见, 本文描述的所有逻辑可以通过使用分立组件、如专用集成电路 (ASIC) 的集成电路、结合如现场可编程门阵列

(FPGA) 或微处理器的可编程逻辑装置使用的可编程逻辑、状态机或包括它们的任何组合的任何其他装置来体现。可编程逻辑可以暂时或永久固定在如只读存储器芯片、计算机存储器、磁盘或其他存储媒体的有形媒体中。所有此类实施例应落入本发明的范围。

[0064] 应该理解,可以在本发明的精神和范围内对附图所示以及说明书中描述的实施例进行多种更改和修改。相应地,上文描述中包含以及附图中所示的所有内容应从说明而非限制的意义来解释。本发明仅按随附权利要求书及其等效中定义的进行限定。

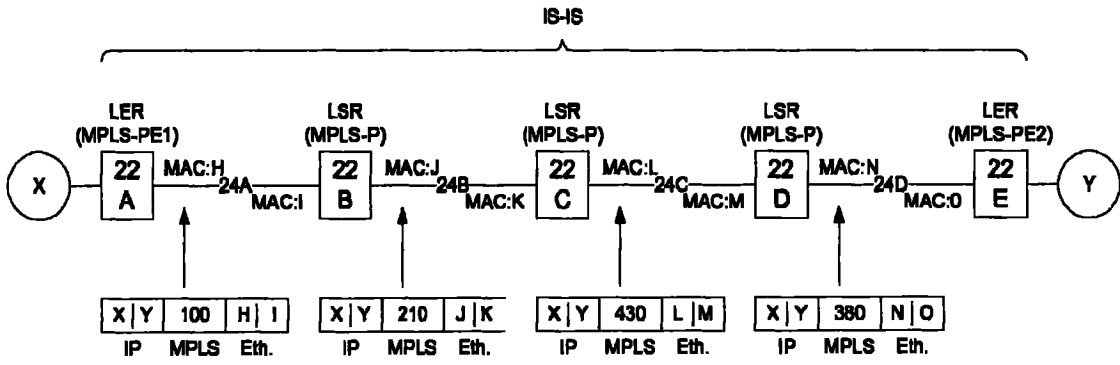


图 1

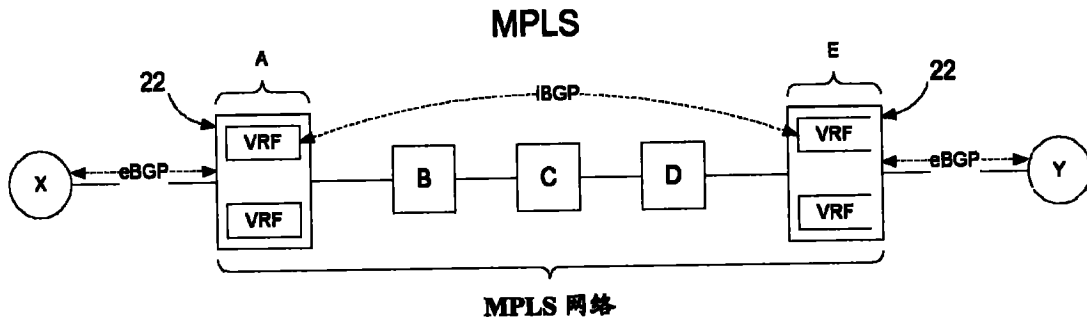


图 2

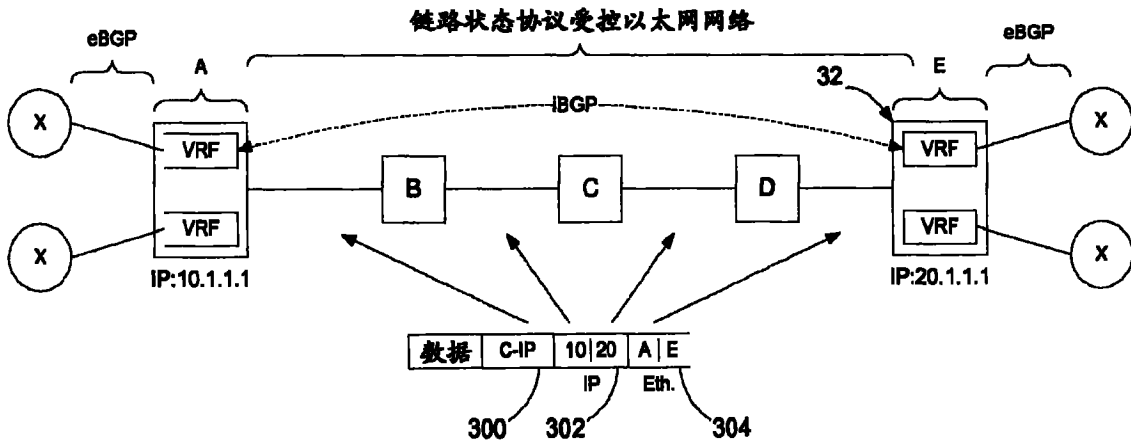


图 3

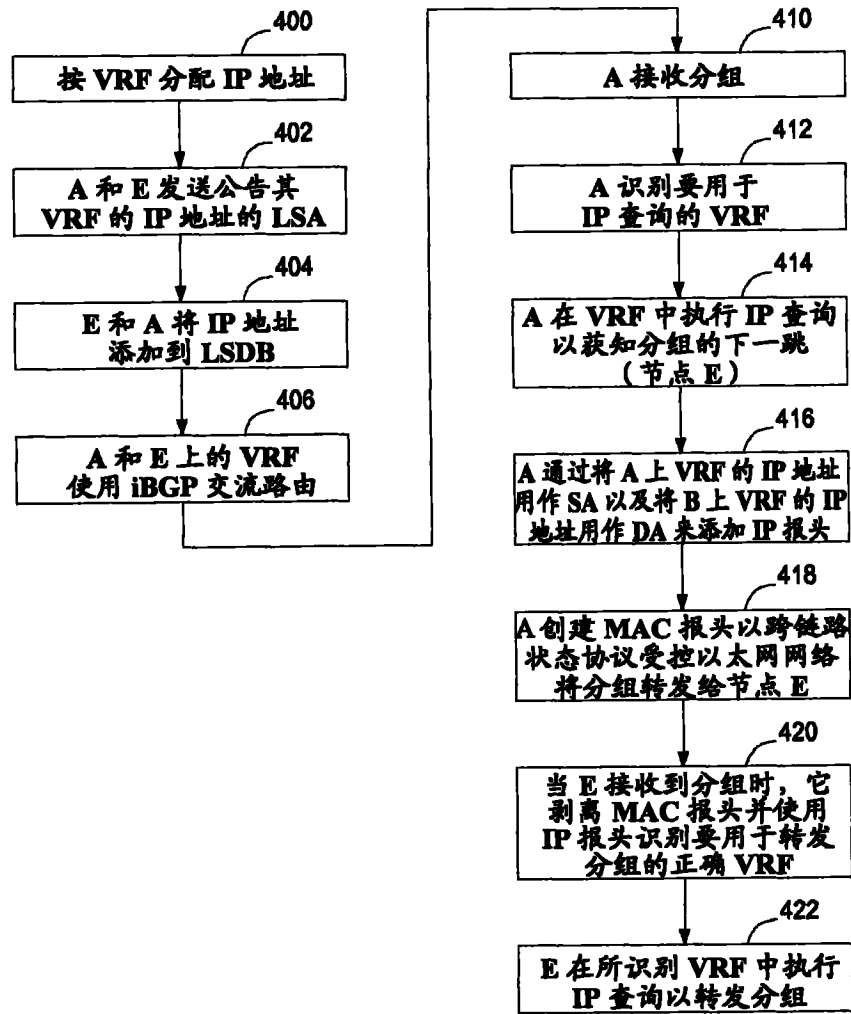


图 4

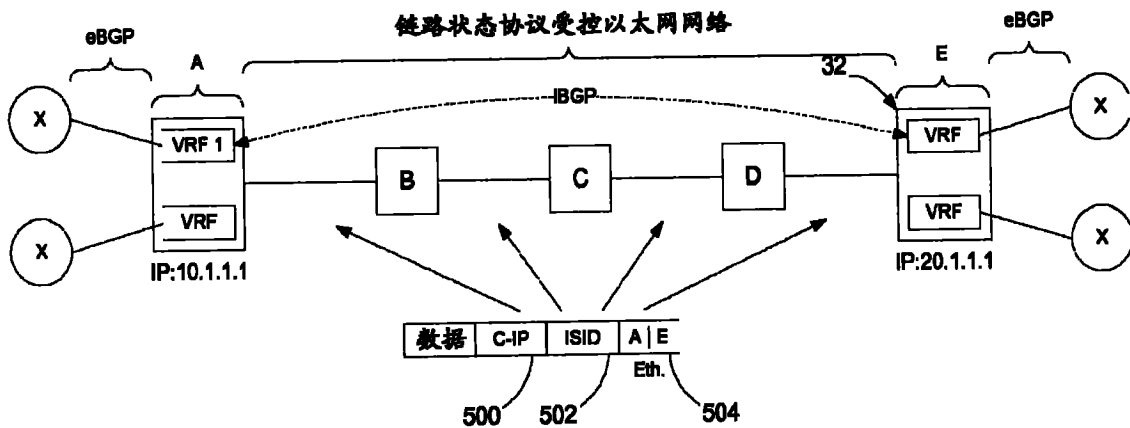


图 5

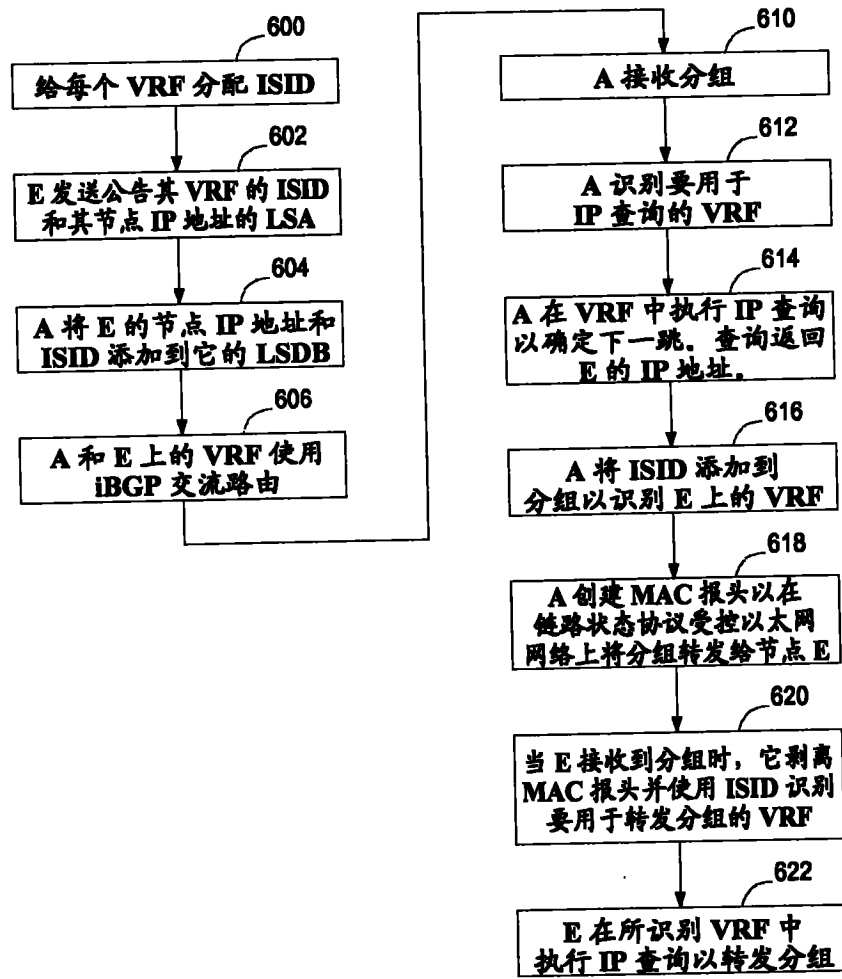


图 6

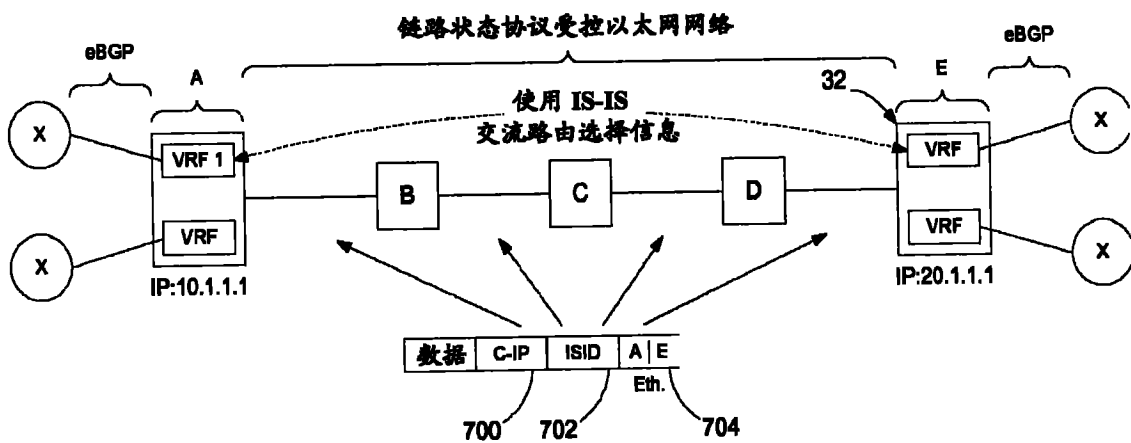


图 7

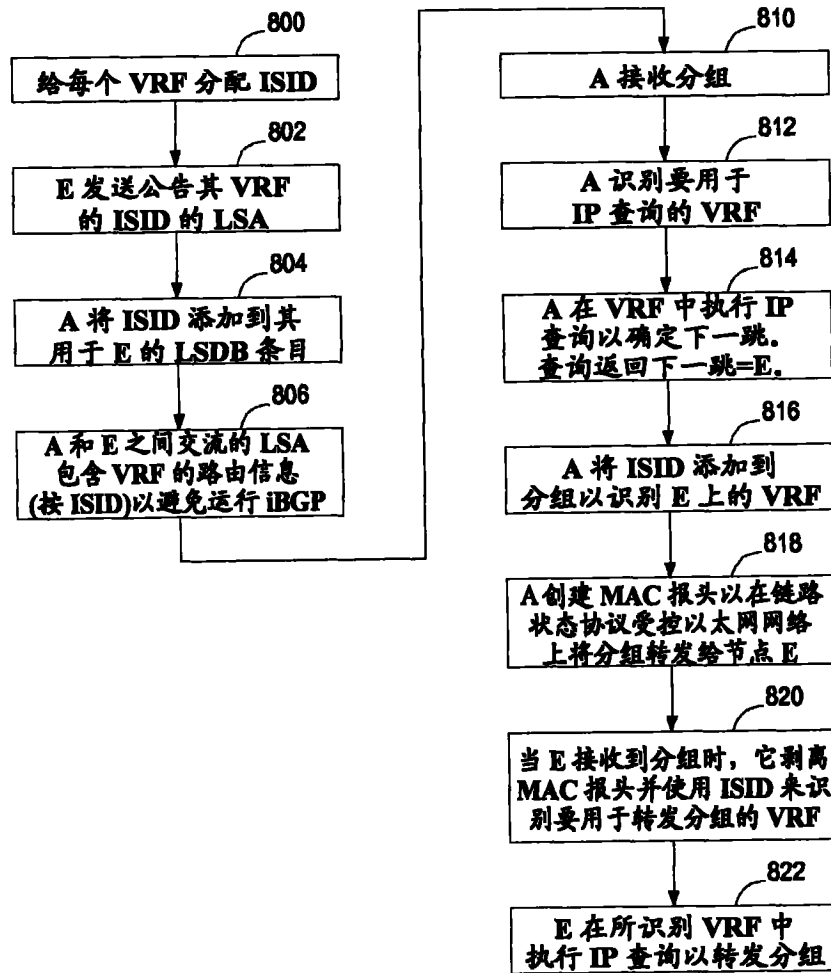


图 8

IS-IS LSA (TLV)

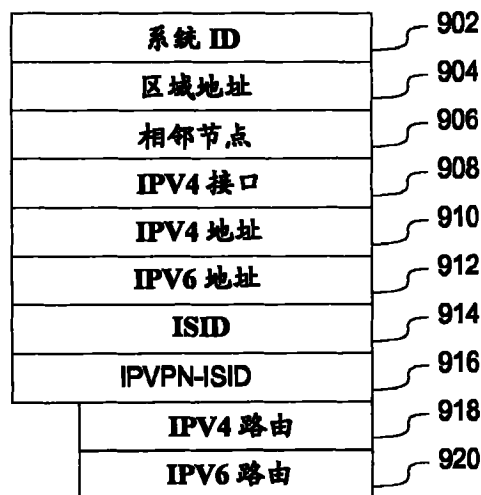


图 9

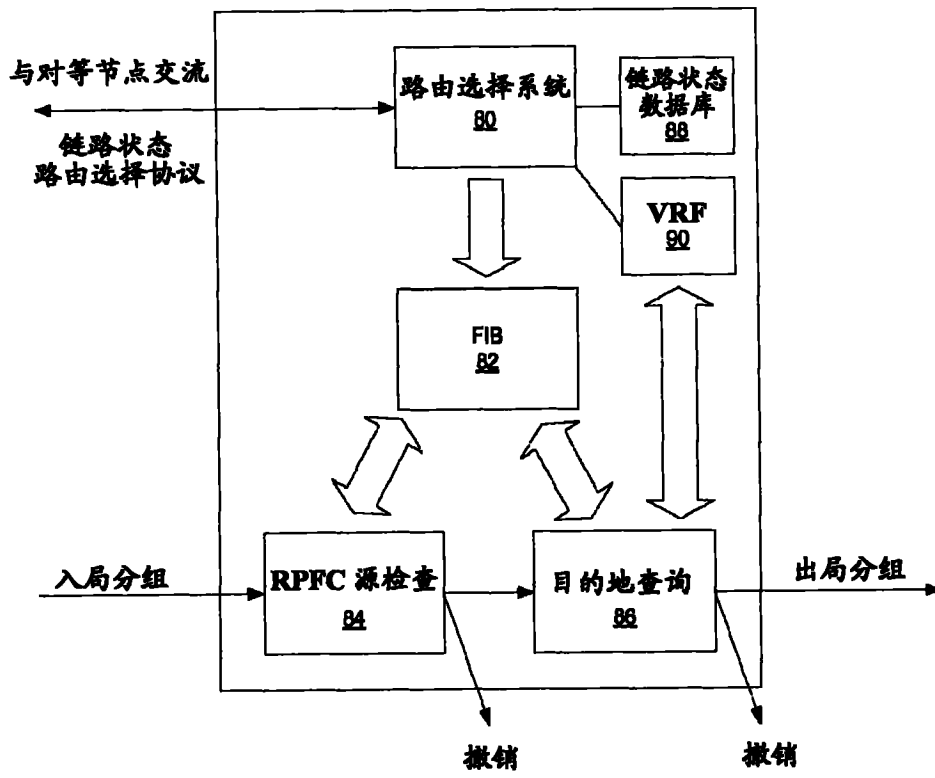


图 10