(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 1820463 B
(45)授权公告日 2010.12.01

(21)申请号 200380102077.X

(22)申请日 2003.10.21

(30)优先权数据
10/279,360 2002.10.24 US

(85)PCT申请进入国家阶段日
2005.04.25

(86)PCT申请的申请数据
PCT/US2003/033215 2003.10.21

(87)PCT申请的公布数据
WO2004/039015 EN 2004.05.06

(73)专利权人 思科技术公司
地址 美国加利福尼亚州

(72)发明人 诺曼·W·芬恩

(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 11258
代理人 王怡

(51)Int.Cl.
H04L 12/46(2006.01)
H04L 12/56(2006.01)

(56)对比文件
Interworking Task Group of IEEE802.1.P802.1s Draft Standard for local and metropolitan area networks-Amendments 3 to 802.1q Virtual Bridged Local Area Networks: Multiple Spanning Trees. Interworking Task Group of IEEE802.1,2002, 第6.1节, 7.5节, 8.6节, 8.10节, 11.2节, 13.3节, 13.4-13.5节,13.8节, 13.15节, 13.16节, 图13-3.

审查员 刘承恩

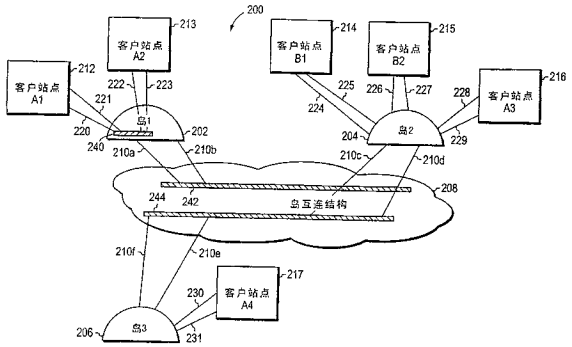
权利要求书 2 页 说明书 20 页 附图 10 页

(54)发明名称

大规模第2层城域网

(57)摘要

一种系统和方法,准许使用第2层(L2)交换技术创建超大城域网(MAN)。将不同组L2交换机逻辑地组织成岛。连接到每个岛的是多个客户站点,并且互连结构将岛耦合到一起。岛协作来向被耦合在一起的每组客户站点提供虚拟以太网连接(VEC)。客户通过用客户设备VLAN标识符(CE-VLAN ID)给它们的流量加标记或标签来标识它们的对应于VEC的流量。在每个岛中,将由客户的流量指定的CE-VLAN ID(并且从此指定相应的VEC)映射到唯一的MAN运营商设备VLAN ID(PE-VLAN ID)。为了防止形成环,岛运行代表多生成树协议(MSTP)的修改版本的MAN间控制协议。



CN 1820463 B

1. 一种在城域网 MAN 中防止第 2 层 L2 环的方法，所述城域网具有多个中间网络设备，并且提供多条虚拟以太网连接 VEC，每条代表虚拟的共享介质，所述方法包括步骤：

将所述多个中间网络设备组织成两个或更多管理组，每个包含一个或多个中间网络设备；

通过在每个管理组和互连结构之间提供冗余链路来使用所述互连结构以耦合所述两个或更多管理组；

在所述互连结构中定义一个或多个逻辑干线，每个逻辑干线代表连接两个或更多管理组的共享介质；

将每个虚拟以太网连接分配给不超过一个逻辑干线；并且

对于由管理组提供的每个虚拟以太网连接，将除了一个之外所有到所述互连结构所述冗余链路都阻塞。

2. 如权利要求 1 所述的方法，还包括步骤：

在每个管理组中，定义多个运营商设备虚拟局域网 PE-VLAN 标识符 ID；并且

在每个管理组中，将给定的 VEC 与刚好一个 PE-VLAN ID 相关联。

3. 如权利要求 2 所述的方法，其中，在两个管理组中定义并且与相同的 VEC 相关联的 PE-VLAN ID 不相同。

4. 如权利要求 1 所述的方法，其中，所述阻塞步骤包括步骤：

向由所述 MAN 的管理组的中间网络设备所生成的配置网桥协议数据单元 BPDU 消息提供分离的 VEC；并且

将多生成树 MST 配置标识符 ID 分配给所述 MAN 的每个中间网络设备，所述 MST 配置 ID 指定岛名。

5. 如权利要求 4 所述的方法，还包括步骤：

在管理组的给定的中间网络设备处，从逻辑干线接收一个或多个与所述 VEC 相关联的 BPDU 消息，每个接收到的消息指定 MST 配置 ID 和第一根；并且

倘若所述接收到的 BPDU 的 MST 配置 ID 与分配给所述给定的中间网络设备的 MST 配置 ID 相匹配，则在根据所述多生成树协议 MSTP 计算生成树实例中使用所述接收到的消息。

6. 如权利要求 5 所述的方法，其中

所述 MST 配置 ID 还指定配置名、修订级别和配置摘要，并且

如果所述指定的岛名、配置名、修订级别和配置摘要都匹配，则两个 MST 配置 ID 匹配。

7. 如权利要求 6 所述的方法，还包括步骤：

将根 ID 存储在所述给定的中间网络设备处；并且

如果所述接收到的 BPDU 的 MST 配置 ID 岛 ID 的岛名与被分配给所述给定的中间网络设备的 MST 配置 ID 的岛名不匹配，并且所述接收到的 BPDU 的第一根与由所述给定的中间网络设备存储的根 ID 不匹配，则丢弃所述接收到的 BPDU。

8. 如权利要求 4 所述的方法，其中

所述 MST 配置 ID 还指定配置名、修订级别和配置摘要；并且

所述阻塞步骤还包括步骤：如果所述接收到的 BPDU 的岛名与所述给定的中间网络设备的岛名匹配，但是所接收到的 BPDU 的配置名、修订级别和配置摘要中的一个或多个与被分配给所述给定的中间网络设备的配置名、修订级别和配置摘要中的各项之一不匹配，则

在根据快速生成树协议 RSTP 计算生成树实例中使用所述接收到的 BPDU。

9. 如权利要求 1 所述的方法，其中

将所述 MAN 的每个管理组标识为相应的岛，并且

将多个客户网络耦合到每个岛。

10. 如权利要求 9 所述的方法，其中，所述客户网络将配置网桥协议数据单元 BPDU 消息发送到它们各自的岛中，所述方法还包括将在所述客户网络中生成的 BPDU 消息不加修改地返回到所述客户网络的步骤。

11. 如权利要求 1 所述的方法，其中，所述阻塞步骤包括步骤：

配置逻辑干线，以使得每个中间网络设备通过所述逻辑干线只可与属于相同管理组的中间网络设备传输配置网桥协议数据单元 BPDU 消息；并且

在根据所述多生成树协议 MSTP 计算生成树实例中使用所述接收到的 BPDU 消息。

大规模第 2 层城域网

技术领域

[0001] 本发明涉及计算机网络，更具体地说，涉及大规模城域网。

背景技术

[0002] 包括企业、政府和教育机构在内的许多组织利用计算机网络，以便雇员和其他人可以共享并交换信息和 / 或资源。计算机网络一般包括依靠一种或多种通信介质互连的实体。实体可以由通过通信介质“发出”（即发送）或“吸收”（即接收）数据帧的任何设备组成，例如计算机。一种常见的计算机网络是局域网（“LAN”），它一般指单个建筑或者校园中的私有网络。LAN 一般应用诸如以太网、FDDI 或令牌环之类的数据通信协议 (LAN 标准)，这些数据通信协议定义由通信体系结构（即协议栈）的数据链路层和物理层执行的功能。

[0003] 常常使用一个或多个中间网络设备来将 LAN 耦合到一起，并且允许相应的实体交换信息。例如，可以使用网桥在两个或多个 LAN 或者终端站之间提供“交换”功能。一般来说，网桥是一个计算机，并且包括经由 LAN 被耦合到其他网桥、或者诸如路由器或主机之类的终端站的多个端口。一般将用来将网桥相互耦合的端口称为干线端口，而一般将用来将网桥耦合到终端站的端口称作接入端口。网桥桥接功能包括在源端口处从发送实体接收数据，并且将该数据传送到至少一个目的地端口，以转发到一个或多个接收实体。

[0004] 以太网

[0005] 以太网是现在最常使用的 LAN 标准。被称作 10Base-T 的原始以太网传输标准能够以每秒 10M(Mbs) 的速率发送数据。1995 年，国际电气电子工程师协会 (IEEE) 批准了被称作 100Base-T 的快速以太网传输协议，它能够以 100Mbs 的速率工作。然而，10Base-T 和 100Base-T 都受限于小 于 100 米的线缆长度。IEEE 的一个委员会（所谓的 802.3z 委员会）当前正在研究用于以 1000Mbs 的速率传输数据的 Gigabit 以太网，也被称作 1000Base-X（光纤信道）和 1000Base-T（长拉铜缆 (long haulcopper)）。除了极大增加的传输速率外，Gigabit 以太网也支持长至 3000 米的线缆长度。因此，Gigabit 以太网代表了潜在的以太网 LAN 的大小和范围的极大增长。

[0006] 生成树 (spanning tree) 算法

[0007] 大多计算机网络包括冗余通信路径，以使得任何给定链路发生故障不会隔离网络的任何部分。一般把这种网络称作网状网络或者部分网状网络。然而，存在冗余链路可能导致在网络中形成网络环路或者“环”。因为数据帧可能通过环无限循环，所以非常不希望存在环。

[0008] 此外，诸如网桥或交换机之类的某些设备复制目的地未知的帧，导致沿环增生数据帧。结果流量可能压垮网络。诸如路由器之类的工作在协议栈中较高层（例如传输控制协议 / 因特网协议（“TCP/IP”）参考模型的互联网层）的其他中间设备传递数据帧，并且获知与大多数网桥或交换机不同的网络上的实体的地址，以使得路由器一般不受持续的成环问题的影响。

[0009] 为了避免形成环，大多网桥和路由器执行生成树协议，该协议允许它们计算不

会形成环的活动网络拓扑（即树），并且该网络拓扑还连接网络中的每对 LAN（即该树是生成的）。IEEE 已经发布了定义由 802.1D 兼容设备执行的生成树协议的标准（IEEE Std.802.1D-1998™）。一般来说，通过执行 802.1D 生成树协议，网桥选举桥接网络中的单个网桥为“根网桥”。802.1D 标准利用每个网桥具有唯一的数字标识符（网桥 ID）这样的事实，指定根网桥为具有最小网桥 ID 的网桥。另外，对于耦合到任何网桥的每个 LAN，刚好选出一个网桥（“指定的网桥”）上的一个端口（“指定的端口”）。指定网桥一般是最靠近根网桥的网桥。根网桥上的所有端口都是指定端口，并且根网桥是所有下述 LAN 上的指定网桥，即根网桥具有端口到所述 LAN。

[0010] 每个非根网桥也从它的非指定端口中选择一个给出到根网桥的最低代 价路径的端口（它的“根端口”）。将选择的根端口和指定端口包括在活动拓扑中，并且置于转发状态中，以使得可以将数据帧转发向这些端口并且从这些端口转发，从而转发到将网络的网桥和终端站互连的 LAN′ 上。将未包括在活动拓扑中的端口置于阻塞状态中。当端口处于阻塞状态中时，将不把数据帧转发到该端口，也不从该端口接收数据帧。网络管理员也可以通过将端口置于禁用状态中，从而从生成树中排除该端口。

[0011] 为了获得运行生成树协议所必需的信息，网桥交换被称作配置网桥协议数据单元 (BPDU) 消息或者简写为 BPDU 的专用消息。BPDU 携带诸如假设的根和最低根路径代价之类的在计算活动拓扑中使用的信息。更具体地说，一旦启动，每个网桥最初就假定它自己是根网桥，并且因此发射 BPDU。一旦从相邻的设备接收到 BPDU，接收网桥就检查接收到的 BPDU 的内容，并且将其与存储在存储器中的相似信息（例如，假设的根和最低根路径代价）进行比较。如果来自接收到的 BPDU 的信息“优”于存储的信息，则网桥就采用较优的信息，并且在网桥从它的与在其上接收了“较优”信息的端口不同的端口发送的 BPDU 中使用该信息（将与接收端口相关联的代价添加到根路径代价）。尽管 BPDU 不是由网桥转发，但是根网桥的标识符最终还是如上所述被传播到所有网桥，并且被所有网桥所采用，允许它们选择它们的根端口和任何（多个）指定端口。

[0012] 为了使活动拓扑适应变化和故障，根网桥周期地（例如每 hello 时间地）发送 BPDU。响应于在它们的根端口上接收 BPDU，网桥从它们的指定端口发送它们自己的 BPDU（如果有的话）。从而，将 BPDU 周期地传播过整个桥接的网络，确认活动拓扑。当 BPDU 被更新和 / 或过时，并且重新计算活动拓扑时，端口可以从阻塞状态转变到转发状态，反之亦然。即，作为新 BPDU 信息的结果，先前被阻塞的端口可以获知它应当处于转发状态（例如，它现在是根端口或者指定端口）。

[0013] 快速生成树协议

[0014] 近来，IEEE 发布了定义快速生成树协议 (RSTP) 的新标准 (IEEEStd.802.1W-2001™)。RSTP 类似地选择桥接网络的一个网桥为根网桥，并且定义在 LAN 之间提供完整连通性的同时切断任何环的活动拓扑。根 据端口是否是活动拓扑的一部分，给每个网桥的每个独立端口分配端口角色。由 802.1w 规范标准定义的端口角色包括根端口、指定端口、替换端口和备份端口。将提供到根网桥的最佳（例如最小代价）路径的网桥端口分配为根端口角色。将提供到根网桥的替换（例如较高代价）路径的每个网桥端口分配为替换端口角色。对于每个 LAN，将提供从那个 LAN 到根网桥的最低代价路径的一个端口分配为指定端口角色，而将耦合到 LAN 的所有其他端口分配为根端口角色、备份端口角

色、或者在某些情形中为替换端口角色。在根网桥处，将所有端口分配为指定端口角色。

[0015] 将已经被分配为根端口角色和指定端口角色的那些端口置于转发状态中，而将已经被分配为替换和备份角色的端口置于一种状态中。如果所有分配为替换端口角色的端口都处于阻塞状态中，则可以将分配为根端口角色的端口快速转变到的转发状态。类似地，如果在当前分配为根端口角色的端口上发生了故障，倘若已将先前的根端口转变到丢弃或阻塞状态，则可以将分配为替换端口角色的端口重新分配为根端口角色，并且快速地转变到转发状态。倘若下游网桥的端口的角色与正被分配为指定端口角色的这个端口一致，则可以将下述端口快速转变到转发状态，所述端口是已被分配为指定端口角色或备份端口角色，而将被重分配为指定端口角色的端口。RSTP 提供了要被相邻网桥使用的明确的握手，以确认可以将新的指定端口快速转变到转发状态。

[0016] 与在 802.1D 规范标准中描述的 STP 相同，运行 RSTP 的网桥也交换 BPDU，以便确定将哪个角色分配给网桥的端口。在被应用到将指定端口快速转变到转发状态的握手中也利用 BPDU。

[0017] 虚拟局域网

[0018] 也可以将计算机网络分段到一系列逻辑网络中。例如，于 1995 年 2 月 28 日授权给 Ross 的美国专利 No.5,394,402（“402 专利”）公开了用于将交换机的任何端口与任何特定的网段相关联的布置。具体地说，根据“402 专利”，通过使用将端口与特定的 VLAN 标号 (designation) 虚拟地关联起来的虚拟局域网 (VLAN) 布置，可以将特定交换机的任何数目的物理端口与交换机中的任何数目的组相关联。更具体地说，交换机或集 线器将 VLAN 标号与它的端口相关联，并且还将那些 VLAN 标号与从已将 VLAN 标号分配给的任何端口发送来的消息相关联。

[0019] 用于每个端口的 VLAN 标号存储在交换机的存储器部分，以使得每次在给定的接入端口上接收到消息时，就将用于那个端口的 VLAN 标号与消息相关联。利用下述流处理元件实现关联，所述流处理元件基于在其上接收消息的具体接入端口，在存储器部分中查找 VLAN 标号。在许多情形中，最好将多个这些交换机互连，以便在网络中扩展端口的 VLAN 关联。那些具有相同 VLAN 标号的实体好像它们都是相同 LAN 的一部分一样工作。专门对 VLAN 配置的网桥进行配置来防止在具有不同 VLAN 标号的部分网络之间交换消息，以便保持每个 VLAN 的边界。虽然如此，诸如路由器之类的在第 2 层之上运行的中间网络设备可以在不同的 VLAN 段之间中继消息。

[0020] 除了“402 专利”之外，IEEE 发布了用于虚拟桥接局域网的 802.1Q 规范标准。为了保持通过 VLAN 敏感网络中的干线或者链路传输的消息的 VLAN 关联，Ross 和 IEEE Std.802.1Q-1998 规范标准都公开了将 VLAN 标识符 (VID) 字段附加到相应的帧。另外，Edsall 等的美国专利 No.5,742,604（它和本申请具有共同的所有人）公开了交换机间链路 (ISL) 封装机制，用于在维护帧的 VLAN 关联的同时在交换机之间有效地传输分组或包括 VLAN 修饰帧在内的帧。具体来说，可以利用快速以太网标准的 ISL 链路连接布置在每个交换机处的 ISL 接口电路。发送 ISL 电路在 ISL 头部和 ISL 检错信息中封装被传输的帧，而 ISL 接收电路剥离这种信息，并且恢复原始帧。

[0021] 多生成树协议

[0022] IEEE 也正在研究用于生成树协议的规范标准，该标准专门设计来与支持 VLAN 的

网络一起使用。在 IEEE 802.1s 草案规范标准中描述的多生成树协议 (MSTP) 将桥接网络组织为区域 (region)。在每个区域中，MSTP 建立这样的内部生成树 (IST)，该内部生成树提供到各个区域中的所有的网桥的连通性，以及到在其他区域中建立的 IST 的连通性。在每个 MSTP 区域中建立的 IST 也通过运行 STP 或 RSTP 的 IEEE Std.802.1Q- 1998 兼容网桥而提供在该 MSTP 区域之外建立的一个公共生成树 (CST) 的连通性。给定的 MST 区域的 IST 接收 BPDU，并将 BPDU 发送给 CST。因此，桥接网络的所有网桥由单个公共和内部生成树 (CIST) 连接。此外，从遗留网桥或者 IEEE 802.1Q 的观点来看，每个 MST 区域看起来就象 CST 上的单个虚拟网桥。

[0023] 在每个 MST 区域中，MSTP 兼容网桥建立多个活动拓扑，每个活动拓扑被称作多生成树实例 (MSTI)。MSTP 网桥也将每个 VLAN 分配或者映射到一个且只是一个 MSTI。因为可以将多个 VLAN 分配到不同的 MSTI，所以与不同的 VLAN 相关联的帧可以采用不同的路径通过 MSTP 区域。网桥可以为每个单个 VLAN 计算分离的拓扑，但是一般不计算，从而节省处理器和存储器资源。每个 MSTI 基本上是只在各个区域内存在的简单的 RSTP 实例，并且 MSTI 不在区域之外交互。

[0024] 和其他生成树协议一样，MSTP 使用 BPDU 来建 IST 和 MSTI，并且定义不同的 MSTP 区域的边界。网桥不对每个 MSTI 发送不同的 BPDU。相反，每个 MSTP BPDU 携带为在各个区域中定义的所有 MSTI 计算活动拓扑所需要的信息。此外，每个 MSTI 具有相应的标识符 (ID)，并且将 MSTI ID 编码到网桥 ID 中。即，如上所述，每个网桥具有唯一的 ID，并且这个 ID 由固定的部分和可设置的部分组成。利用 MSTP，将网桥 ID 的可设置部分进一步组织来包括系统 ID 扩展。系统 ID 扩展对应于 MSTI ID。从而，给定区域中的 MSTP 兼容网桥对于每个 MSTI 将具有不同的网桥 ID。对于给定的 MSTI，将对于那个实例具有最小网桥 ID 的网桥选举为根。从而，MSTP 兼容网桥可以是给定 MSTP 区域中的一个 MSTI 的根，而不是给定 MSTP 区域中其他 MSTI 的根。

[0025] 运行 MSTP 的每个网桥也具有由三个属性组成的单个 MST 配置标识符 (ID)：包括字母和数字的配置名，修订级别和 VLAN 映射表，该 VLAN 映射表将潜在的 4096 个 VLAN 中的每一个 VLAN 映射到相应的 MSTI。此外，每个网桥将它的 MST 配置 ID 加载到由该网桥发出的 BPDU。因为网桥只需要知道它们是否在相同的 MST 区域中，所以它们不在它们的 BPDU 中传播实际的 VLAN 到 MSTI 的表。相反，MST BPDU 只 携带 VLAN 到 MSTI 的表或映射的摘要。通过将公知的 MD-5 算法应用到 VLAN 到 MSTI 的表来生成摘要。当网桥接收到 MST BPDU 时，网桥就抽取包含在其中的包括有摘要的 MST 配置 ID，并且将它与网桥自己的 MST 配置 ID 相比较，以判断它是否与发送该 MST BPDU 的网桥在相同的 MST 区域中。如果这两个 MST 配置 ID 相同，则这两个网桥在相同的 MST 区域中。但是，如果这两个 MST 配置 ID 至少有一个不匹配的属性 (即，或者不同的配置名、不同的修订级别和 / 或不同的计算得的摘要)，则接收到 BPDU 的网桥推断出它在与发出 BPDU 的网桥不同的 MST 区域中。此外，如果指定网桥在不同的 MST 区域中，或者如果端口接收遗留的 BPDU，则认为 MST 网桥的端口在 MST 区域的边界处。

[0026] 图 1 是 MST BPDU 100 的高度简化的方框图。MST BPDU 100 包括与各个 LAN 标准 (例如以太网) 的媒体访问控制 (MAC) 层兼容的头部 102。除了其他以外，头部 102 包括目的地地址 (DA) 字段、源地址 (SA) 字段、目的地服务接入点 (DSAP) 字段和源服务接入点

(SSAP) 字段。DA 字段 104 携带有分配给生成树协议的唯一网桥多播目的地地址，并且 DSAP 和 SSAP 字段携带有分配给生成树协议的标准化的标识符。附加到头部 102 的是 BPDU 消息区域，该区域包括“外”部 104 和“内”部 106。外部 104 具有与 RSTP BPDU 消息相同的格式，并且被未实现 MSTP 的网桥识别为有效的 RSTP BPDU 消息。执行 MSTP 的网桥利用“内”部 106 来建立 IST 和 MSTI。内部 106 具有一组用于 IST 的生成树参数，以及一组用于由发出 MSTP BPDU 100 的网桥所支持的每个 MSTI 的参数。

[0027] 外部 104（也被称为 CIST 优先权向量）具有多个字段，包括协议标识符 (ID) 字段 108、协议版本 ID 字段 110、BPDU 类型字段 112、标签字段 114、CIST 根 ID 字段 116、外部路径代价字段 118、CIST 区域根 ID 字段 120、CIST 端 ID 字段 122、消息年龄字段 124、最大 (MAX) 年龄字段 126、hello 时间字段 128 和转发延迟字段 130。CIST 根标识符字段 116 包含被假设为公共和内部生成树的根的网桥的标识符，其可以在与发出 BPDU 消息 100 的网桥相同的 MSTP 区域中，也可以在另一个 MSTP 区 域中，或者在不运行 MSTP 的桥接网络的一部分中。外部路径代价字段 118 包含这样的值，该值代表从发出 BPDU 100 的网桥到在字段 116 中标识的 CIST 根，不通过在与发出 BPDU 消息 100 的网桥相同的区域中的任何其他网桥的最低代价。

[0028] 类似地，内部 106（也被称为 MSTI 优先权向量）也具有多个字段，包括版本 1 长度字段 132、空 (null) 字段 134、版本 3 长度 136、MST 配置 ID 字段 138、CIST 区域根 ID 字段 140、CIST 区域路径代价字段 142、CIST 网桥 ID 字段 144、CIST 端口 ID 字段 146、CIST 标签字段 148 和 CIST 跳字段 150。内部 106 还可以包括一个或多个可选的 MSTI 配置消息 152，每个 MSTI 配置消息组成另一个 MSTI 优先权向量或 M 记录。

[0029] 因为 RSTP 的版本 2 没有指定除版本 1 已经指定的那些字段外的任何额外的字段，所以 MST BPDU 没有版本 2 长度字段。

[0030] 如上所述，MST 配置 ID 字段 138 由三个子字段组成：配置名子字段 154、修订级别子字段 156 和 MD-5 校验和子字段 158。配置名子字段 154 携带在固定长度中编码的可变长度文本字符串，例如 32 八位组。修订级别子字段 156 携带在两个八位组的固定字段中编码的整数。MD-5 校验和子字段 158 携带通过将 MD-5 算法应用到网桥的 VLAN 到 MSTI 表而创建的 16 八位组的签名，其包含 4096 个连续的 2 个八位组的元素。

[0031] 每个 MSTI 配置消息 152 由下述多个字段组成，包括：CIST 区域根 ID 字段 160、CIST 区域路径代价字段 162、CIST 网桥 ID 字段 164、CIST 端口 ID 字段 166、CIST 标签字段 168 和 CIST 跳字段 170。MST 网桥利用包含在内部 106 的字段 140-150 中和每个 MSTI 配置消息 152 中的 STP 参数，以计算用于在各个区域中配置的每个 MSTI 的活动拓扑。

[0032] 城域网 (MAN)

[0033] 利用点到点链路、微波收发机、卫星连接等可以将多个 LAN 和 / 或终端站互连，以形成一般跨越数个城区、整个城市和 / 或例如 San FranciscoBay 区域的整个城市区域的城域网 (MAN)。MAN 一般互连位于物理上相互远离、但是仍在城市区域中的各个校园和 / 或建筑中的多个 LAN 和 / 或终端站。传统的 MAN 一般依赖于应用在现存公共交换电话网 (PSTN) 的同步光网络 (SONET) 之上运行的异步传输模式 (ATM) 的网络设备。和大多数 LAN 利用以太网标准一样，必须将在一个 LAN 中创建的网络消息或分组从以太网格式转换成 ATM 信元，以通过 SONET 链路传输。然后必须将 ATM 信元转换回以太网格式，以传递到目的地 LAN 或

终端站。这种将每个网络消息从以太网转换到ATM，并且再转换回去的需求要求MAN包括昂贵的网络设备。MAN供应者也不得不租借或者获得到SONET链路的接入。结果，建立和运行MAN可能是昂贵的。

[0034] 因此，需要一种系统和方法，用于更有效地建立并运行MAN。

发明内容

[0035] 提供了一种在城域网(MAN)中防止第2层(L2)环的方法，所述城域网具有多个中间网络设备，并且提供多条虚拟以太网连接(VEC)，每条代表虚拟的共享介质，所述方法包括步骤：将所述多个中间网络设备组织成两个或更多管理组，每个包含一个或多个中间网络设备；通过在每个管理组和互连结构之间提供冗余链路来使用所述互连结构以耦合所述两个或更多管理组；在所述互连结构内定义一个或多个逻辑干线，每个逻辑干线代表连接两个或更多管理组的共享介质；将每个虚拟以太网连接分配给不超过一个逻辑干线；并且对于由管理组提供的每个虚拟以太网连接，将除了一个之外所有到所述互连结构所述冗余链路都阻塞。

[0036] 简要地说，本发明集中于用于使用第2层(L2)交换技术建立超大城域网(MAN)的系统和方法。在示例性实施方式中，将不同组L2交换机逻辑地组织成岛。此外，将每个岛配置为分离的管理域。连接到每个岛的是多个客户站点，一般是局域网(LAN)。利用互连结构来将岛耦合到一起，以使得连接到第一岛的客户站点可以与连接到相同的岛或者连接到第二岛的客户站点通信。在示例性实施方式中，从配置来提供多标记交换协议之上的仿真VLAN(EVoMPLS)服务的多个第3层(L3)设备形成互连结构，其中EVoMPLS是MPLS之上的ATM仿真(ATM论坛标准af-lane-0021.000)的相仿体。或者，互连结构可以由使用802.1Q或者类似的标签的以太网LAN形成。岛协作来向正被耦合在一起的每组客户站点提 供虚拟以太网连接(VEC)。客户通过用客户设备VLAN标识符(CE-VLAN ID)给它们的流量加标记或标签来标识它们的对应于VEC的流量。在每个岛中，将由客户的流量指定的CE-VLAN ID(并且由此指定相应的VEC)映射到唯一的MAN运营商设备VLAN ID(PE-VLAN ID)。为在一个岛中给定的VEC选择的PE-VLAN ID可以与为在另一个岛中使用的为所述给定的VEC选择的PE-VLAN ID不同。对于每个横穿互连结构的VEC，建立岛间干线来在两个岛之间传输VEC流量。岛间干线是这样的逻辑结构，即，至少从岛的观点来看，它起到共享的介质的功能。具体地说，将利用岛间干线联合的岛配置来把相同的虚电路标识符(ID)(优选地作为MPLS标记)附加到置于岛间干线上的网络消息。在将在岛中从岛间接收到的网络消息发送到各个客户站点之前，剥离它们的虚电路ID标记和任何其他标记。

[0037] 可以预期由岛和互连结构组成的拼接而成的MAN对于任何标准生成树算法都太大了，而不能满意地防止形成环。为了防止在MAN中形成环，将岛配置来防止两个或多个VEC共享相同的岛间干线。岛还运行MAN间控制协议(IMCP)这种新协议，该协议代表多生成树协议(MSTP)的修改版本。具体地说，给布置在每个岛中的L2设备配置有新的多协议生成树(MST)配置ID，该配置ID除了配置名、修订级别和校验和之外，还包括岛名。此外，布置在相同岛中的L2设备都被给予相同的岛ID、配置名和修订级别。从而，每个岛将它自身标识为分离的MSTP区域。其次，对于跨越互连结构的每个VEC，每个岛中的L2设备也确保，除了一个之外所有其他将岛连接到互连结构的冗余链路都被阻塞。结果，切断了由在客

户站点和岛之间出现的冗余链路可能另外导致的环，而不管在客户站点中正运行的 STP 的版本。然而，对于不同的 VEC，阻塞的链路可以变化，由此在在岛和互连结构之间延伸的链路之间提供某一水平的负载共享。

[0038] IMCP 也施加数种新规则。具体地说，将在岛中接收到的这样的 BPDU 作为正常、匹配的 BPDU 处置，该 BPDU 的整个 MST 配置 ID 与接收 L2 设备的 MST 配置 ID 匹配。将接收到的这样的 BPDU 作为快速生成 树协议 (RSTP)BPDU 处置，该 BPDU 的岛名与接收 L2 设备的岛名匹配，但是它们的配置名、修订级别和 / 或配置摘要不匹配。这种规则允许相同岛中的网桥以与 802.1S 相同的方式运行，并且在重配置网桥期间维护连通性。如果从岛间干线接收到，则忽略下述接收到的 BPDU ：其岛名与接收 BPDU 的 L2 接收设备的岛名不匹配，并且其指定根 ID 与接收 L2 设备的根 ID 不匹配的 BPDU。这种规则有效地将岛的生成树相互去耦。如果在网桥的端口上而不是从岛间干线接收到，则接收到下述 BPDU 导致接收网桥将用于所有 VLAN 的各个端口阻塞，并且发出操作者警告，其中所述 BPDU 的岛名与接收 BPDU 的 L2 设备的岛名不匹配。这种规则防止除了岛间干线上的连接之外的岛之间的意外连接。在优选的实施方式中，L2 设备也通过阻塞用于所有 VLAN 的各个端口，并且发出操作者警告，来对接收这样的 BPDU 作出响应，这些 BPDU 的岛名与接收这些 BPDU 的 L2 设备的岛名不匹配，但是它们的指定根 ID 的确与接收 L2 设备的相匹配。这种规则允许岛检测以其他方式不会被检测到的岛之间的意外连接。

[0039] 附图说明

[0040] 下面的发明描述参考附图，其中 ：

[0041] 图 1（先前描述的）是根据多生成树协议的传统配置网桥协议数据单元的简化方框图 ；

[0042] 图 2 是超大城域网 (MAN) 的高度简化的示意图 ；

[0043] 图 3 是图 2 的 MAN 的岛的高度简化的示意图 ；

[0044] 图 4 是图 3 的岛的第 2 层 (L2) 设备的局部功能图 ；

[0045] 图 5 是配置标识符 (ID) 的高度简化的示意图 ；

[0046] 图 6 是 VLAN 映射表的高度简化的示意图 ；

[0047] 图 7 是岛间干线映射表的高度简化的示意图 ；

[0048] 图 8 是加标记的网络消息格式的高度简化的示意图 ；

[0049] 图 9-11 是图 2 的 MAN 的高度简化的局部示意图 ；和

[0050] 图 12 和 13 是根据本发明的另一个 MAN 的高度简化的示意图。

[0051] 具体实施方式

[0052] 图 2 是根据本发明的超大第 2 层 (L2) 城域网 200 的高度简化的示意图。如这里所使用的，术语“超大 MAN”指能够覆盖整个城市区域的 MAN，例如 San Francisco Bay 区域、Silicon Valley 等。MAN 200 包括多个岛，例如岛 202、204 和 206。如这里所述，每个岛包括一个或多个互连的第 2 层 (L2) 中间网络设备，例如网桥或交换机。一般来说，每个岛由相同的 MAN 运营商操作，并且代表分离的管理域。将 MAN 200 组织成不同的岛来增加可由 MAN 200 支持的超过 4096 的 VLAN 标号的数目。此外，可以将某些或者所有这些单独的岛分配给不同的管理员。

[0053] 利用岛互连结构 208 将岛耦合到一起。优选地，通过诸如岛间链路 210a-f 之类的

多个链路将每个岛耦合到岛互连结构208。此外，附着到每个岛的是诸如个人客户之类的一个或多个客户站点212-217。在示例性实施方式中，每个客户站点包括多个实体或主机，例如个人计算机、工作站、服务器等，它们都在相同的物理位置，并且都被互连来形成一个或多个局域网(LAN)，以使得实体可以相互发出或吸收数据帧。如这里所使用的，术语“相同的物理位置”指在单个校园或大约单个城区的区域中的单个建筑或多个建筑。可以通过一个或多个诸如网桥、交换机或路由器之类的客户操作的L2中间网络设备将在客户站点处的LAN互连。

[0054] 客户站点212-217一般属于不同的组织，例如组织A和组织B。具体地说，组织A包括客户站点212(A1)、213(A2)、216(A3)和217(A4)。组织B包括客户站点214(B1)和215(B2)。优选地通过多个站点链路220-231将每个客户站点212-217耦合到至少一个岛。如这里所述，客户从MAN 200获得各种服务，例如将它的地理上相互远离的站点互连。这样，位于一个客户站点的实体可以与另一个站点的实体通信。

[0055] 图2的MAN 200仅仅是为了说明目的，并不是要限制本发明。实际上，MAN 200一般包括更多的站点，例如数以千计。

[0056] 图3是岛的高度简化的示意图。岛202包括多个L2中间网络设备，例如交换机(S)302、304和306。每个交换机302、304和306都包括多个端口(P)402，至少某些端口被用来将交换机连接到客户站点。将其他 交换机端口402耦合到在交换机302、304和306之间延伸的岛内链路308-310。链路308-310可以是在交换机之间传输诸如帧之类的网络消息的点到点链路或共享介质链路。此外，每个交换机302-306优选地例如通过诸如端口0(P0)、端口1(P1)、端口2(P2)和端口3(P3)等的端口号识别它自己的端口402。从而交换机302-306能够将特定的端口与客户站点和/或耦合到它的其他交换机相关联。

[0057] 至少可以将每个岛的某些交换机划分到不同的类别。例如，每个岛具有一个或多个布置在该岛与一个或多个客户站点之间的边界处的运营商边界交换机。将运营商边界交换机直接耦合到客户站点。每个岛也包括一个或多个将岛连接到岛互连结构208的岛边界网桥。参考图3，交换机304是运营商边界网桥，交换机306是岛边界网桥，并且交换机302既是运营商边界网桥，也是岛边界网桥。

[0058] 用于与本发明一起使用的的合适的中间网络设备平台包括但不限于来自San Jose，California的Cisco Systems，Inc.的Catalyst 4000和6000系列商用交换机。

[0059] 图4是诸如交换机302之类的MAN运营商交换机的局部方框图。交换机302包括多个端口402a-402e，优选地通过数字(例如，P0-P4)识别每个端口。将一个或多个帧发送和接收对象(一般指404)与端口402a-d相关联，以使得可以捕获在给定的端口(例如P3)处接收到的包括帧在内的网络消息，并且可以将要由交换机302发送的帧传递到适当的端口(例如P1)来发送。帧接收和发送对象404优选地是诸如优先队列之类的消息存储结构。在示例性实施方式中，交换机302包括发送和接收电路，该电路包括一个或多个建立端口来交换网络消息的线卡和/或网络接口卡(NIC)，一个或多个下述管理卡(supervisor card)和一个或多个总线结构，所述管理卡具有中央处理单元(CPU)和/或微处理器，以及关联的存储器，用于执行运算并存储计算得到的结果。

[0060] 交换机302具有多个协议实体，包括至少一个多生成树协议(MSTP)实体408、至少一个转发引擎410和虚拟以太网信道(VEC)实体412。MSTP实体408优选地包括多个子组

件，包括端口角色选择状态 机 414、端口转变状态机 416、网桥协议数据单元 (BPDU) 消息发生器 418、岛边界确定引擎 420 和 MSTP 摘要生成器 422。岛边界确定引擎 420 优选地包括一个或多个比较器，例如比较器 423。优选地根据下述标准运行 MSTP 实体 408，所述标准是被这里描述的 MAN 间控制协议 (IMCP) 修改过的补充 802.1Q 规范标准的 IEEE 802.1s 多生成树协议 (MSTP) 草案，在这里通过引用而将它的当前草案 (IEEE 草案 P802.1s/D13™，2002 年 6 月 13 日）整体包含。MSTP 实体 408 包括（或与其通信）下述存储器设备或结构，例如 STP 存储器 424，它可以是易失性存储器或者非易失性随机访问存储器 (RAM) 或某些其他存储器。优选地将存储器 424 组织来包括多个记录或单元（未示出），用于存储生成树相关信息或参数，例如交换机配置 ID、数字网桥标识符 (ID)、为每个 MSTI 的每个端 402a-e 分配的路径代价、用于每个 MSTI 的每个端口 P0-P4 的当前或“最佳”生成树信息等。除了存储器 424 之外，STP 实体 408 还包括配置来存储 VLAN 到 MSTI 的映射的 VLAN ID(VID) 到多生成树实例 (MSTI) 翻译表 426。

[0061] VEC 实体 412 包括 VLAN 映射引擎 428、标签操纵引擎 430 和岛间干线引擎 432。VLAN 映射引擎 428 包括一个或多个 VLAN 映射表 600，VLAN 映射表 600 优选地以每个端口为基础将客户设备 VLAN ID(CE-VLAN) 映射到运营商设备 VLAN ID(PE-VLAN)。在替换实施方式中，可以存在用于每个端口（或若干个端口）的分离的 VLAN 映射引擎和分离的 VLAN 映射表，并且可以用 CE-VLAN 到 PE-VLAN 的不同映射配置每个 VLAN 映射表。

[0062] 岛间干线映射引擎 432 具有将 PE-VLAN ID 映射到 VEC 标识符 (ID) 的岛间干线映射表 700。优选地在将帧发送到岛互连结构 208 中之前，将 VEC ID 添加到该帧。为了在不同客户站点之间提供连通性，将 VEC 实体 412 配置来建立一个或多个用户网络接口 (UNI)，例如也由标号 436 指定的 UNI 01 和也由标号 438 指定的 UNI 02。如这里所述，每个 UNI 代表一个或多个 VEC 的终端点，并且从而可以被认为定义了一个或多个逻辑 VEC 端口。UNI 436 例如具有 3 个 VEC 端 440a-c。UNI 438 具 有 2 个 VEC 端口 442a-b。

[0063] 转发引擎 410 与帧发送和接收对象 404 通信，并且被耦合到至少一个存储与 MAN 200 的实体相对应的地址信息的过滤数据库 444（图 2）。具体地说，过滤数据库 444 具有多条记录（未示出），每条包含多个单元，包括目的地地址单元、目的地端口单元、过滤器数据库 ID(FID) 单元和相应的定时器单元。过滤器数据库 444 中的每条记录优选地对应于特定的网络实体。从消息的 PE-VLAN ID 导出的 FID 允许给定的目的地 MAC 地址与用于不同 PE-VLAN ID 的相同或不同的 MAC 地址对应。将转发引擎 410 配置来依赖于包含在转发数据库 428 中的信息，并且也依赖于由 MSTP 实体 408 所管理的各个端口 402 的生成树端口状态，来将诸如分组和 / 或帧之类的网络消息从源端口 402 交换或桥接到一个或多个目的地端口 402。转发引擎 410 也与 MSTP 实体 408 通信，并且中继在端口 402 处接收到的 MSTP 相关的消息。转发引擎 410 也与 VEC 实体 412 通信。

[0064] 本领域的技术人员应当理解，MSTP 实体 408、转发引擎 410 和 VEC 实体 412 每个都可以包括配置来形成时序逻辑电路的寄存器和组合逻辑。在示例性实施方式中，MSTP 实体 408、转发引擎 410 和 VEC 实体 412 优选地为包含关于这里描述的方法的程序指令的软件模块和库以及硬件元件的组合，所述软件模块和库可以由交换机 302 的一个或多个处理元件（未示出）执行。其他计算机可读介质也可以用来存储并执行这些程序指令。尽管如此，本领域的技术人员应当认识到，可以利用软件和硬件的各种组合，包括固件来实现本发明。

[0065] 将岛形成为分离的管理域

[0066] 最初，MAN 运营商将他们的设备（即 L2 交换机 302-306）组织为多个岛。在优选实施方式中，通过按照特定的方式配置交换机的 MST 配置 ID 来将它们组织成岛。更具体地说，MAN 运营商首先决定应当将它的哪些交换机组织到给定的岛中。然后 MAN 运营商将给定的岛中的每个交换机的 MST 配置 ID 配置为相同的值。图 5 高度简化地图示了根据本发明配置的 MST 配置 ID 500。MST 配置 ID 500 具有 32 字节的岛名字段 502、32 字节的配置名字段 503、2 字节的修订级别字段 504 和 2 字节的配置摘要字 段 506，优选地将 MD-5 校验和算法应用到 VID/MSTI 翻译表 426（图 4）来形成 MST 配置 ID。对于给定的岛中的每个交换机，MAN 运营商利用相同的值配置岛名字段 502、配置名字段 503 和修订级别字段 504。MAN 运营商也利用给定的岛建立 PE-VLAN 到 MSTI 的相同映射。从而，给定的岛中的每个交换机将生成相同的摘要值。交换机存储由 MAN 运营商在它们的 STP 存储器 424 处选择的 MST 配置 ID。

[0067] 在优选实施方式中，岛名字段 502 是 2 字节或更长，配置名字段 503 是 32 字节，修订级别字段 504 是 3 字节，并且配置摘要字段 506 为 2 字节。在替换实施方式中，岛名字段 502 和配置名字段 503 是组合的 32 字节，并且可以用诸如“#”符号之类的某些专门选出的字符将这两个值分离。

[0068] 可以通过在交换机处提供的命令行接口 (CLI) 本地执行交换机 302-306 的配置，或者通过公知的简单网络管理协议 (SNMP) 远程执行。

[0069] 例如，给布置在岛 202 中的交换机 302、304 和 306（图 3）都配置有相同的岛名（例如“ISLAND0001”）、相同的配置名（例如“MAN4452”）和相同的修订级别（例如“0001”）。另外，还将交换机 302、304 和 306 配置来使其具有相同的 PE-VLAN 到 MSTP 实例 ID 的映射。另一方面，将给布置在岛 204（图 2）中的每个交换机配置不同的岛名，例如“ISLAND0002”。可以给它们每个配置相同或不同的配置名和 / 或修订级别，并且一般将给它们每个配置不同的 PE-VLAN 到 MSTP 实例 ID 的映射。

[0070] 当 MAN 运营商初始化并运行它的多个交换机时，交换机将自动地（即不用人工干涉）将它们自身隔离到期望的岛中，作为它们的 MSTP 执行的一部分。更具体地说，因为已经给每个交换机 302、304 和 306 配置有相同地岛名、相同的配置名、相同的修订级别号和相同的 PE-VLAN 到 MSTP 实例 ID 的映射，所以它们将推断出它们都是相同 MSTP 区域的一部分，或者在这种情形中是相同岛的一部分。

[0071] 如这里所述，配置 UNI 来处置以双向中的一个方向从客户站点接收到的 BPDU 消息。更具体地说，UNI 或者丢弃从客户站点接收到的 BPDU 消 息，或者将接收到的 BPDU 消息处置为数据帧，并且将它们隧穿过岛，以使得可以由其他客户站点接收它们。类似地，UNI 不将由运营商边界网桥所生成的 BPDU 发送到客户站点中。因此，MAN 运营商的交换机 302、304 和 306 在计算任何（多个）活动拓扑中不与位于客户站点中的中间网络设备协作。

[0072] 此外，在每个岛中，MAN 运营商的交换机将建立在各个岛中定义的用于每个 MSTP 实例的活动拓扑。例如，假设 MAN 运营商在岛 202 中定义 10 个 MSTP 实例，并且至少将一个 PE-VLAN 分配给每个 MSTP 实例。作为它们的 MSTP 执行的一部分，交换机 302、304 和 306 将在岛 202 中建立内部生成树 (IST)，以及 10 个无环、活动拓扑。

[0073] 通过一个或多个岛链接多个客户站点

[0074] 假设操作站点 212、213、216 和 217(图 2)的客户希望互连这些站点。更具体地说,假设客户希望将站点 212 与 213、站点 212 与站点 216、以及站点 216 与 217 互连。客户优选地联系 MAN 运营商,并且要求这种服务。MAN 运营商依次配置它的设备(即,布置在岛 202、204 和 206 中的交换机)来建立期望的连接。

[0075] 根据本发明,MAN 运营商通过在每个标识出的客户站点之间建立虚拟以太网连接(VEC)来提供所请求的服务。VEC 模拟在两个或多个客户站点之间延伸的物理以太网链接或以太网桥接的 LAN。如这里所述,在每个岛中,存在 VEC 和 PE-VLAN ID 之间的一一对应。

[0076] 如图 2 图示的网络所示,将客户站点 212 和 213 都连接到相同的岛,即到岛 202。然而,将客户站点 212 和 213 每个都分别连接到不同的岛,即岛 202 和 204。类似地,将客户站点 216 和 217 每个都分别连接到不同的岛,即岛 204 和 206。为了提供所要求的连通性,除了其他以外,MAN 运营商优选地建立连接客户站点 212 和 213 的第一 VEC、连接客户站点 212 和 216 的第二 VEC、以及连接站点 216 和 217 的第三 VEC。第一 VEC 整个在岛 202 中。因此,MAN 运营商只需要在岛 202 中建立岛内链路 240 以由第一 VEC 使用。

[0077] 应当理解,岛内链路只是跨越单个岛的两个客户站点之间的互连的逻 辑表示,在优选实施方式中,其是根据 IEEE Std.802.1Q-1998 规范标准运行的 VLAN。岛内链路可以另外或者可替换地应用来自 Cisco Systems,Inc. 的 ISL 协议。

[0078] 另一方面,第二 VEC 必须跨越多个岛,即岛 202 和 204。因此,MAN 运营商必须建立连接岛 202 和 204 的岛间干线 242 来由第二 VEC 使用。类似地,第三 VEC 跨越多个岛,从而它也需要接入耦合岛 204 和 206 的岛间干线 244。

[0079] 岛内链路

[0080] 优选地如下进行创建耦合客户站点 212 和客户站点 213 的第一 VEC。在连接到客户站点 212 的运营商边界交换机 302 中,MAN 运营商建立用户网络接口(UNI),例如 UNI 438(图 4)。UNI 是客户站点与诸如岛之类的 MAN 运营商的网络之间的逻辑接口。由 MAN 运营商建立的每个 UNI 都具有一个或多个 VEC 端口,每个这种端口代表已被 MAN 运营商创建的相应的 VEC 的终端或端点。在 UNI 438 中,可以将 VEC 端口 442a 分配给第一 VEC。然后,MAN 运营商将 MAN 运营商设备 VLAN ID(PE-VLAN)分配给第一 VEC。如这里所述,PE-VLAN ID 是被附加到下述帧并且从而识别下述帧的 VLAN 标号,所述帧通过对应于各个 VEC(例如,连接客户站点 212 和 213 的第一 VEC)的各个岛(例如岛 202)传输。用于第一 VEC 的 PE-VLAN ID 可以是“4011”。

[0081] 当在由岛 202 耦合的客户站点 212 和 213 之间通信时,客户选择要由客户使用的客户设备 VLAN ID(CE-VLAN ID)。一般基于客户自己的站点和它的联网设备的需求选择可以是“0014”的 CE-VLAN ID。客户配置它自己的站点,以使得能够用选出的 CE-VLAN ID 来给在这些站点之一中创建的要被传递到其他站点(例如,站点 212)的所有网络消息(例如,以太网帧)加标签。MAN 运营商获知该选择,并且对在岛的边界处并连接到所述两个客户站点的交换机(即岛 202 的交换机 302 和 304)的 VEC 实体 412 进行配置,以将选择的 CE-VLAN ID(即“0014”)映射到各个 PE-VLAN(即“4011”)。具体地说,MAN 运营商配置 VEC 实体 412 的客户 VLAN 映射表 600。

[0082] 图 6 是图示了交换机 302 的 VLAN 映射表 600 的高度简化的示意图。至少逻辑地将表 600 组织为具有多个列和行的表或数组,其中在这些列和行的交叉处定义用于存储信

息的单元或记录。表 600 优选地具有 CE-VLAN ID 列 602、VEC 列 604、PE-VLAN ID 列 606、UNI 列 608 和 VEC 端口列 610。表 600 也具有多个行 614a-c。MAN 运营商优选地将自由行分配给第一 VEC，例如行 614a。在行 614a 处，MAN 运营商将选择的 CE-VLAN ID（即“0014”）加载到对应于列 602 的单元，将例如“001”的 VEC ID 加载到对应于列 604 的单元，将例如“4011”的已经分配给选择的 CE-VLAN ID 的 PE-VLAN ID 加载到对应于列 606 的单元，将分配给这个 VEC 的具体的 UNI（即 UNI 01）加载到对应于列 608 的单元，并且将特定的 VEC 端口（即 VEC 端口 0）加载到对应于列 610 的单元。MAN 运营商类似地配置在岛 202 和客户站点 213 的边界处的交换机 304 的 VLAN 映射表 600。

[0083] 通过使用选择的 CE-VLAN ID，在两个站点 212 和 213 中的终端站现在可以相互通信。例如，假设布置在站点 212 中的工作站希望与站点 213 中的工作站通信。站点 212 中的工作站将它的消息编码到一个或多个以太网帧中，并且在这些帧的 VLAN ID 字段中插入由客户选择的 CE-VLANID，例如“0014”。岛 202 中的交换机 302（在到客户站点 212 的边界处）接收这些加了 VLAN ID 标签的帧。最初将加了 VLAN ID 标签的帧提供给交换机的 VEC 实体 412，该 VEC 实体 412 访问它的 VLAN 映射表 600 来执行查找。具体地说，VEC 实体 412 搜索表 600 来确定接收到的帧属于的 VEC。VLAN 映射表 600 的行 614a 指示 CE-VLAN ID“0014”对应于 VEC “001”，并且这个 VEC 已经被映射到 PE-VLAN ID “4011”。

[0084] 在本发明的另一种实施方式中，VEC 实体的标签操纵引擎 430 用 PE-VLAN ID“4011”来加载这些帧的 VLAN ID 字段，替换 CE-VLAN ID“0014”。或者，标签操纵引擎 430 可以向消息添加一个新的 VLAN 标识符 (VID) 字段（未示出），并且用各个 PE-VLAN ID（即“4011”）来加载这个新的 VID 字段，不修改原始 VID 字段（携带 CE-VLAN ID）。

[0085] 然后，将现在用 PE-VLAN ID “4011”加标签的帧提供给 UNI，以经 由已经为这个 VEC 建立的 VEC 端口发送。这些帧在为该 VEC 建立的岛内链路 240 上传输，并且在交换机 304 处被接收。如这里所述，岛内链路 240 基本上对应于在已将 PE-VLAN ID“4011”映射到的岛 202 中定义的 MSTP 实例或者活动拓扑的一部分。就在交换机 302 和 304 之间布置的任何中间交换机或网桥对帧进行转发来说，这些中间交换机优选地不修改帧。即，将帧转发到岛 202 中的其他交换机的交换机不对帧的 VLAN 标签进行任何进一步的改变。

[0086] 在交换机 304 处，在代表创建来互连客户站点 212 和 213 的 VEC 的其他端的 VEC 端口上接收帧。当要从在交换机 304 处的 UNI 发送帧，即要将它们发送到到 202 之外时，要再次对它们进行转换。更具体地说，将这些帧提供给交换机 304 的对它的 VLAN 映射表 600 执行查找的 VEC 实体 412。在这里，VEC 实体 412 基于已经被用来对帧加标签的 PE-VLAN ID 搜索表 600。VEC 实体 412 确定对应于 PE-VLAN ID“4011”对应于 CE-VLAN ID“0014”。因此，标签操纵引擎 430 用 CE-VLAN ID“0014”加载帧的 VLAN ID 字段，替换 PE-VLAN ID“4011”。然后，将已经用它们的原始 VLAN ID 存储的帧从交换机 302 发送到客户站点 213 中。然后，基于由帧携带的目的地地址，将帧传递到作为目标的工作站。

[0087] 在其中一旦在岛 202 中接收到帧，就将新的 VID 字段添加到帧的实施方式中，在交换机 304 处的标签操纵引擎 430 在将帧发送到客户站点 213 中之前剥离新的 VID 字段。

[0088] 从客户站点 212 中的工作站到站点 213 的通信以相似的方式工作。具体地说，在交换机 304 处，通过将分配给 CE-VLAN ID 的 PE-VLAN ID 加载到用这一分配的 CE-VLAN ID 加标签的帧的 VLAN ID 字段中来修改这些帧。然后，这些帧沿着岛 202 中的岛内链路 240 传

输到交换机 302，交换机 302 在与客户站点 212 的边界处。然后，用帧的原始 CE-VLAN ID 恢复它们，并且将它们发送到客户站点 212 中以传递到客户站点 212 中作为目标的工作站。

[0089] 应当理解，在客户站点 212 和 213 中本来可以选择不同的 CE-VLANID 来与第一 VEC 一起使用。在这种情形中，优选地将 VLAN 映射表 600 配置来指定双方的 CE-VLAN ID。

[0090] 岛间干线

[0091] 优选地如下进行创建耦合客户站点 212 和客户站点 216 的第二 VEC。在连接到客户站点 212 的交换机 302 中，MAN 运营商或者建立新的 UNI，或者将现存的 UNI 分配给第二 VEC。因为已经将 UNI 438 分配给了客户站点 212 用于第一 VEC，所以 MAN 运营商可以重用这个存在的 UNI438，用于第二 VEC。虽然如此，但是必须为第二 VEC 提供在 UNI 438 处的新端口，例如 VEC 端口 442b，因为每个 VEC 必须具有它自己的 VEC 端口。然后，MAN 运营商选择并分配 PE-VLAN ID 给第二 VEC，以在岛 202 中使用。所选择的 PE-VLAN ID 将被用来标识通过对应于第二 VEC 的岛 202 传输的帧。假设 MAN 运营商选择 PE-VLAN ID “4027”用于岛 202 中的第二 VEC。

[0092] 当与布置在客户站点 216 中的网络实体通信时，客户选择用于由布置在客户站点 212 中的网络实体使用的 CE-VLAN ID。假设客户选择 CE-VLAN ID “0038”以在客户站点 212 中使用。客户配置布置在客户站点 212 中的它自己的联网设备，以使得用 CE-VLAN ID“0038”对诸如以太网帧之类的所有下述网络消息加标签，该网络消息是在那个站点中创建的，并且以站点 216 中的网络实体为目的地。客户也通知 MAN 运营商所选择 CE-VLAN ID。在响应中，MAN 运营商然后将交换机 302 的 VEC 实体 412 配置来将用选择的 CE-VLAN ID（即“0038”）加标签的帧映射到选择的 PE-VLAN ID（即正被映射到的“4027”），其中交换机 302 在岛 202 和客户站点 212 的边界处。具体地说，MAN 运营商配置在交换机 302 处的 VEC 实体 412 的 VLAN 映射表 600。

[0093] 更具体地说，MAN 运营商将自由行分配给第二 VEC，例如行 614b。在行 614b 处，MAN 运营商将选择的 CE-VLAN ID（即“0038”）加载到对应于列 602 的单元，将例如“002”的 VEC ID 加载到对应于 604 的单元，并且将例如“4027”的由 MAN 运营商所选择的 PE-VLAN 加载到对应于列 606 的单元。MAN 运营商也将分配给这个 VEC 的特定的 UNI（即 UNI 01）加载到对应于列 608 的单元，并且将选择的 VEC 端口（即 VEC 端口 1）加载到对应于列 610 的单元。

[0094] 在连接到客户站点 216 的岛 214 中，MAN 运营商建立让客户站点 216 中的网络实体访问第二 VEC 的 UNI。优选地在岛 204 和站点 216 之间的边界处的（多个）运营商边界交换机（即被直接连接到客户站点 216 的（多个）交换机）中提供 UNI（即经由链路 228 和 / 或 229）。MAN 运营商也在 UNI 中建立 VEC 端口来在岛 204 中终止第二 VEC。然后，MAN 运营商选择并分配 PE-VLAN ID 给第二 VEC，以在岛 204 中使用。所选择的 PE-VLAN ID 将被用来标识在对应于第二 VEC 的岛 204 中传输的帧。值得注意的是，选择来在岛 204 中使用的 PE-VLAN ID 可能与已选择来在岛 202 中使用的 PE-VLAN ID “4027”不同。实际上，假设 MAN 运营商选择 PE-VLAN ID “4017”用于岛 204 中的第二 VEC。

[0095] 和上面的一样，当与布置在客户站点 212 中的网络实体通信时，客户基于它自己的需求和要被布置在客户站点 216 中的网络实体要使用的它的联网设备的能力来选择 CE-VLAN ID。选择来在站点 216 中使用的 CE-VLAN ID 可以与选择来在客户站点 212 中使用的 CE-VLAN ID 相同，或者不同。假设客户选择 CE-VLAN ID “0018”来在客户站点 216 中

使用。客户配置布置在客户站点 216 中的它自己的互连设备，以使得用 CE-VLANID “0018” 对诸如以太网帧之类的所有下述网络消息加标签，该网络消息是在那个站点中创建的，并且以站点 212 中的网络实体为目的地。客户也通知 MAN 运营商所选择 CE-VLAN ID。然后，MAN 运营商将布置在岛 204 中的下述交换机的 VEC 实体 412 配置来将用 CE-VLAN ID“0018” 加标签的帧映射到选择来在岛 204 中使用的 PE-VLAN ID（即“4017”），其中所述交换机被直接连接到客户站点 216。具体地说，MAN 运营商配置相对于客户站点 216 的岛 204 的（多个）运营商边界交换机处的 VEC 实体 412 的 VLAN 映射表 600。

[0096] 行 614c（图 6）图示了如何配置岛 204 的各个（多个）运营商边界交换机处的 VLAN 映射表 600。更具体地说，MAN 运营商将选出的 CE-VLAN ID（即“0018”）加载到对应于列 602 的单元，将例如“002”的 VEC ID 加载到对应于 604 的单元，并且将例如“4017”的已被映射到选 出的 CE-VLAN ID 的相应的 PE-VLAN ID 加载到对应于列 606 的单元。MAN 运营商也将分配给这个 VEC 的特定的 UNI（例如，UNI 00）加载到对应于列 608 的单元，并且将 VEC 端口（例如，VEC 端口 0）加载到对应于列 610 的单元。

[0097] 如图所示，这个第二 VEC 穿过两个分离的到 202 和 204，以便提供两个选择的客户站点 212 和 216 之间的连通性。因此，第二 VEC 利用岛间干线 242，岛间干线 242 延伸通过岛互连结构 208，并且连接两个岛 202 和 204。在示例性实施方式中，岛间干线作为连通的共享介质以太网或桥接 LAN 运行，并且用多协议标记交换 (MPLS) 虚拟专用网 (VPN)、分组环、异步传输模式 (ATM) 仿真 LAN 或者其他这种技术建立。在岛互连结构 208 中形成 MPLS VPN。值得注意的是，穿过岛互连结构 208 的每个 VEC 都必须只使用单个岛间干线。但是，多个 VEC 可以使用相同的岛间干线。

[0098] 在示例性实施方式中，岛互连结构 208 可以是公知的因特网。

[0099] 如上所指示，每个 MAN 运营商的岛包括至少一个下述岛边界网桥，该岛边界网桥是提供从岛到岛互连结构 208 的直接连通、并且从而提供到 MAN 运营商的城域网的连通的交换机和网桥。例如，在岛 202 处，因为交换机 302 和 306 都分别经由岛间链路 210a 和 210b 提供到岛互连结构 208 的直接连通，所以它们都是岛边界网桥。对于诸如第二 VEC 之类的利用岛间干线的那些 VEC，MAN 运营商必须对位于正被互连的两个（或多个）岛的岛边界网桥中的 VEC 实体进行配置，以修改帧来发送过岛间干线。首先，MAN 运营商对岛间边界网桥的岛间干线映射表 700 进行配置。

[0100] 图 7 是岛间干线映射表 700 的高度简化的示意图。至少逻辑地将表 700 组织为具有多个列和行的表或数组，其中在这些列和行的交叉处定义用于存储信息的单元或记录。表 700 优选地具有 PE-VLAN ID 列 702、VEC 列 704 和岛间干线 ID 列 706。表 700 也具有多个行 710a-c。在岛 202 中的岛边界网桥处，MAN 运营商优选地将自由行分配给第二 VEC，例如行 710a。在行 710a 处，MAN 运营商将选择的用于岛 202 的 PE-VLAN ID （即“4027”）加载到对应于列 702 的单元，并且选择的用于第二 VEC 的 VEC ID（例如“002”）加载到对应于列 704 的单元。MAN 运营商用与要被附加到横穿岛互连结构 208 的网络消息的标签或标记相对应的岛间干线 ID 来加载对应于列 706 的单元。依赖于用来互连岛的（多个）具体的协议，选择可以包括多于一个标记或标签的岛间干线 ID。假设 MPLS 协议，更具体地说，MPLS 之上的仿真 VLAN(EVoMPLS) 是正被使用的协议，则为第二 VEC 的岛间干线选择唯一的 MPLS 标记，例如“6042”。

[0101] MAN 运营商也对在岛 204 的（多个）岛边界网桥处的岛间干线映射表 700 进行配置。行 710b（图 7）图示了将如何配置这个条目。具体地说，将选择来在岛 204 中使用的 PE-VLAN ID“4017”加载到对应于列 702 的单元，将公用的 VEC ID（即，“002”）加载到对应于列 704 的单元，并且将公用的岛间干线 ID 加载到对应于列 706 的单元。

[0102] 其次，MAN 运营商配置岛边界网桥来给帧加标签以通过岛间干线发送，并捕获和处理通过岛间干线接收到的帧。具体地说，当岛 202 的交换机接 302 收用对应于第二 VEC 的 CE-VLAN ID“0038”加标签的帧时，它用 CE-VLAN ID 来对它的 VLAN 映射表 600 执行查找，以得到相应的 PE-VLAN ID，即“4027”。然后，交换机 302 用相应的 PE-VLAN ID 来替换 CE-VLAN ID，并且将帧转发到岛 202 中（假设 VEC 正利用交换机 306 处的岛链路 210b）。在交换机 306 处接收到帧，交换机 306 封装接收到的帧以发送过岛互连结构 208。

[0103] 图 8 是用于发送过岛互连结构 208 的封装的帧 800 的高度简化的示意图。封装的帧 800 包括附加到原始以太网帧 804 的 MPLS 标记栈 802。如上所指示，如果在 UNI 处将 PE-VLAN ID 添加到 CE-VLAN ID，而不是替换它，则以太网帧 804 可以包括对应于 CE-VLAN ID 的 VLAN ID(VID) 字段 805。MPLS 标记栈 802 包括对应于由岛互连结构 208 所采用的介质的第 2 层 (L2) 头部 806、IP/MPLS 头部 808 和虚拟以太网电路 ID 字段 810。在 Reqeust for Comments(RFC)2684“ATM 自适应层 5 之上的多协议封装”(Multiprotocol Encapsulation over ATM Adaptation layer 5)(1999 年 9 月）中描述了与本发明一起使用的封装方案。岛边界网桥对它 的岛间干线映射表 700 执行查找来得到虚拟以太网电路 ID。具体地说，岛边界网桥定位于用来给接收到的帧加标签的 PE-VLAN ID 相对应的岛间干线 ID。在这里，PE-VLAN ID 是“4027”，匹配的岛间干线 ID 是“6042”。将这个获取到的值加载到虚拟以太网电路 ID 字段 810 中。

[0104] 然后将封装的帧发送到岛互连结构 208 上。岛间干线 242 将帧传递到岛 202-206 中的这样的所有端口，这些端口都是与由虚拟以太网电路 ID 指定的 VEC 相同的 VEC 的成员（除了在其上发送帧的端口）。从而，在岛 204 的（多个）岛边界网桥处接收封装的帧。岛 204 的岛边界网桥利用加载在封装的帧的虚拟以太网电路 ID 字段中的值来得到相应的 PE-VLANID，以在岛 204 中使用。在这里，虚拟以太网电路 ID 为“6042”，并且从而来自行 710b 的匹配 PE-VLAN ID 是“4017”。岛边界网桥也基于在其上接收到帧的端口的生成树状态，确定它是否可以接受接收到的帧。如果端口在用于 VLAN“4017”的阻塞生成树端口状态中，则丢弃帧。在这种情形中，在岛 204 处应当存在另一个岛边界端口，该端口处于用于 VLAN“4017”的转发生成树端口状态中，并且从而可以接收帧。

[0105] 在其处接受帧的岛 204 的岛边界网桥剥离 MPLS 标记栈，并且恢复原始的以太网帧 804。在帧的 VLAN ID(VID) 字段 805 中，岛边界网桥加载用于这个 VEC 的 PE-VLAN ID，即“4017”。然后，岛边界网桥在岛 204 内发送帧。在用于客户站点 216 的岛 204 的运营商边界网桥处接收帧。运营商边界网桥使用帧的 PE-VLAN ID 查找它的 VLAN 映射表 600，以得到相应的 CE-VLAN ID。在这里，PE-VLAN ID 是“4017”，并且从而匹配 CE-VLAN ID 是“0018”。因此，运营商边界网桥将 CE-VLAN ID 加载到以太网帧中替换 PE-VLAN ID。然后，岛 204 的运营商边界交换机将用 CE-VLAN ID 加了标签的帧发送到客户站点 216 中，以由目标网络实体接收。

[0106] 防止形成环

[0107] 如图 2 所示，优选地通过多个链路将每个客户站点耦合到它的各自的岛。另外，通过多个链路将每个岛连接到岛互连结构 208。为了利用 RSTP 的快速汇聚时间（相对于 802.1D 生成树协议），优选地每个岛具有最多 两个到任何一个岛间干线的连接或链路。使用多条链路防止任何客户站点和 / 或岛由于任何一致的链路发生故障而失去连通性。然而，相对于通信栈的某些较高层，由于客户站点和岛都在第 2 层 (L2) 运行，所以出现多个链路可能导致形成环。具体地说，因为在执行任何生成树协议时运营商边界网桥不与客户网络协作，所以每个 UNI 转变到用于在岛中定义的每个 PE-VLAN ID 的转发生成树状态。从而，UNI 不丢弃从客户网络接收到的任何帧，除非在 VLAN 映射表 600 中 CE-VLAN ID 未映射到任何 PE-VLAN ID。

[0108] 优选地通过让客户站点阻塞它的一个或多个端口来切断在客户站点与岛之间由于出现冗余的链路而形成的环。这可以通过将在客户站点中生成的 BPDU 与数据帧完全一样地处置来实现。更具体地说，可以将在运营商边界网桥处的 UNI 配置来检查从各个客户站点接收到的帧的目的地 MAC 地址，以确定该地址是否与由 BPDU 消息所用的目的地 MAC 地址相匹配。如果匹配，则运营商边界网桥同样地识别帧。在响应中，运营商边界网桥优选地将 BPDU 消息象数据帧一样传输过运营商网络。为了防止客户生成的 BPDU 被运营商的交换机弄错为由运营商生成的 BPDU，可以在到运营商网络的入口上修改客户 BPDU，并且在出口上恢复，例如通过修改它的目的地 MAC 地址。当在相同的或者另一个客户站点处回收到 BPDU 时，按照传统的方式根据客户站点中的具体的生成树协议处理它。结果，运营商网络对于客户站点看起来仅仅是共享的介质，并且通过运行生成树协议 (STP)，客户的互连设备将通过阻塞面对 UNI 之一的端口，或者阻塞客户站点内部的端口来切断环。

[0109] 或者，可以将运营商边界网桥配置来仅仅丢弃在 UNI 处接收到的 BPDU。在这种情形中，可能存在环，导致从客户站点进入岛的帧速率连续增加。如果岛监控正从客户站点接收的数据的速率，则如果这个速率超过某个阈值就可以触发警告。

[0110] 为了避免由在给定的岛和岛互连结构 208 之间出现多条连接而形成环，岛优选地运行新协议，即根据本发明的 MAN 间控制协议 (IMCP)。代表 MSTP 的修改版本的 IMCP 指定专门的规则和方法来有效地防止在 MAN 运营商的城域网的岛之间形成环。这种修改的版本阻止形成环，并且还避免不得不跨越整个 MAN（即，跨越所有的岛）运行生成树的单个实例。实际上，因为可能存在数百个岛（如果不是更多的话），并且因为在岛中定义的 VEC 的总数可能远超过 IEEE Std.802.1Q-1998 和 IEEE 草案 P802.1s/D13 规范标准所准许的 4096，所以跨越它们运行生成树实例即使不是不可能也是不实际的。

[0111] 如上面所指示，MAN 运营商配置每个岛中的交换机的 VID/MSTI 翻译表 426，以使得将每个 PE-VLAN ID 刚好与一个 MSTI 相关联。在给定的岛中，给定岛中的交换机将一般支持多个 MSTI 和一个 CIST。从而在由 MSTI 和 CIST 所定义的不同活动拓扑中间可以将对应于不同的 PE-VLANID 的流量进行共享负载。

[0112] 为了冗余（和共享负载）的目的，每个岛优选地最多具有两个到每个岛间干线的连接。例如，岛可以具有有两个到岛间干线的连接的单个岛间网桥，和 / 或两个每个都有一个到那个岛间干线的连接的岛间网桥。

[0113] 除了向和从耦合到岛间干线的端口（也被称作岛间端口）转发网络消息之外，岛边界网桥也生成并从它们的包括这些端口在内的端口发送 BPDU。具体地说，岛边界网桥的

MSTP 实体 408 的 BPDU 消息发生器 418 生成具有图 1 中所示的形式的 MST BPDU 消息 100。消息发生器 100 为了获得在生成 BPDU 时使用的信息而访问 MSTP 实体的 STP 存储器 424。具体地说，MSTP 实体的摘要发生器 422 使用 MD-5 算法，从它的 VID/MSTI 翻译表 426 产生摘要值。然后，BPDU 消息发生器 418 从 STP 存储器 424 获取岛名、配置 ID 和修订级别，并且创建 MST 配置 ID 500，优选地将它插入到 MST BPDU 100 的字段 138（图 1）中。类似地，BPDU 消息发生器 418 从 STP 存储器 424 获取 STP 参数值，以加载到 BPDU 100 的其他字段中。

[0114] 除了为每个横穿岛间干线的 VEC 建立的 VEC ID 之外，还定义仅由 IMCP 使用的额外的 VEC 和相应的 VEC ID。具体地说，与由岛边界网桥发送到岛间干线中的 BPDU 一起使用额外的 VEC ID。即，在 BPDU 被发送到岛互连结构 208 之前，在附加到 BPDU 的 MPLS 标记栈 802 中加载额 外的 VEC ID。在示例性实施方式中，岛中的每个网桥都运行 IMCP，并且具有到分配给给定的岛间干线的 VEC 的端口，还具有到额外的 VEC 的端口。将在诸如交换机 306 之类的岛边界网桥处接收到的携带有用于额外的 VEC 的 VEC ID 的封装消息识别为岛生成的消息。在响应中，虚拟以太网信道实体 412 剥离封装，恢复 BPDU，并且将 BPDU 传递给 MSTP 实体 408 来处理。

[0115] 行 710c 可以对应于用于下述额外的 VEC（即，VEC “301”）的条目，在岛间干线 242（即，“6042”）中用作为数据 VEC “002”。如图所示，没有给额外的 VEC 分配 PE-VLAN ID，因为不转发由岛边界网桥所接收的 BPDU。在封装的 BPDU 被发送到岛间干线 242 中之前，将分配的 VEC ID 加载到所述 BPDU 的虚拟以太网电路 ID 字段 810。

[0116] 如上所述，岛间干线在连通性方面象共享介质以太网或桥接的 LAN 那样工作。因此，“耦合”到岛间干线的所有其他交换机，以及也刚好被耦合到岛间干线的正发送 BPDU 的交换机的其他端口都接收发送到岛间干线的 BPDU。因此，从一个岛间端口发出并且用额外的 VEC ID 封装的 BPDU 被传递到所有耦合到岛间干线的岛间端口（除了在其上发送它们的之外）。此外，交换机利用接收到的 BPDU 中的信息来为在交换机处定义的每个 MSTI 计算活动拓扑。结果，对于每个 PE-VLAN ID，岛将阻塞除了到各个岛间干线的一个岛链路 210 外的所有岛链路。此外，因为将每个 VEC 与每个岛中的单个 PE-VLAN ID 相关联，所以将阻塞除了用于每个 VEC 的一个岛间链路 210 外的所以岛间链路。此外，转变到转发状态的特定的岛链路可以在 PE-VLAN ID 中变化。这提供了在岛间链路 210 之间负载共享的方法。

[0117] 当包括岛间网桥在内的布置在岛中的 MAN 运营商交换机接收 BPDU 时，它将 BPDU 传递到 MSTP 实体 408。如果在岛间端口上接收到 BPDU，则它将被封装在 MPLS 标记栈中。在这种情形中，象所有这种消息一样，将 BPDU 传递到 VEC 实体 412。VEC 实体 412 确定用额外的 VEC ID 封装了消息。在这种情形中，VEC 实体 412 剥离 MPLS 标记栈，并且将 BPDU 传递到 MSTP 实体 408。如果该 BPDU 是 MSTP BPDU，则 MSTP 实体 408 从字段 138（图 1）获取 MST 配置 ID，并且将它提供给比较器 423。比较器 423 将来自 BPDU 的 MST 配置 ID 与存储在 STP 存储器 424 中的交换机自己的 MST 配置 ID 相比较。如果所有四个值都匹配，即，它们具有相同的岛名、相同的配置名、相同的修订级别和相同的配置摘要，则交换机在它的活动拓扑计算中利用接收到的 BPDU。即，假定所述 BPDU 是由相同的岛中的另一个交换机或者由相同的交换机但是从不同的岛间端口发送的。

[0118] 在岛具有多个到岛间干线的连接的地方，这确保对于在给定的岛中定义的每个

PE-VLAN ID，只存在一个将岛连接到岛间干线的端口。换句话说，端口角色选择状态机 414 和端口转变状态机 416 只将一个这种端口转变到转发状态。对于这个 PE-VLAN ID，将所有其他端口都转变到阻塞状态。另外，将每个 VEC 映射到单个 PE-VLAN ID。从而，从耦合到各个岛间干线的单个端口只可以发送和接收与给定的 VEC ID 相关联的帧。另外，MAN 运营商配置岛，以使得岛之间仅有的连接是岛间干线，并且在不超过一个岛间干线上传输任何给定的 VEC。这些步骤的组合防止形成环。

[0119] 如果岛在刚好具有两个到给定的岛间干线的岛间端口，则优选地将它们配置为点到点链路，以利用 MSTP/RSTP 的快速生成树汇聚属性。

[0120] 如果 MST 配置 ID 500 的岛 ID 字段 502 与接收交换机存储的相匹配，但是 BPDU 的 MST 配置 ID 的任何其他部分都不匹配，即，配置名、修订级别和 / 或配置摘要值不同，则交换机将接收到的 BPDU 作为传统的 RSTP 处置。即，交换机利用 BPDU 的外部 104（图 1）中的信息来在计算单个 CIST 中与发出 BPDU 的网桥协作，但是忽略内部 106 中的信息。当 MAN 运营商在更新位于给定的岛中的交换机的 VID/MSTI 翻译表，并且从而更新修订级别的过程中时，可能发生这种情形。

[0121] 如果在接收到的 BPDU 中指定的岛 ID 502 与由交换机存储在它的 STP 存储器 424 中的岛 ID 不匹配，并且在 BPDU 的 CIST 根 ID 字段 116 中标识出的根 ID 与在 STP 存储器 424 中存储的相应的值不匹配，则 MSTP 实体 408 忽略并丢弃接收到的 BPDU。在这种情形中，假设布置在某些其他 岛中的交换机发送了接收到的 BPDU。它是将网桥连接到 BPDU VEC 来确保岛名的选择与由连接到相同的 BPDU VEC 的其他交换机所使用的名字一致的前提。MAN 运营商可以通过管理动作实现这点，例如，通过正确地设置各个交换机的 STP 存储器 424 的内容。

[0122] 如果在接收到的 BPDU 中指定的岛 ID 与交换机的岛 ID 不匹配，但是字段 116 中的根 ID 的确匹配，则 MSTP 实体 408 优选地将在其上接收到 BPDU 的端口转变到用于所有 VLAN 的阻塞状态，并且向 MAN 运营商发出警报。这种情况反映 MAN 运营商的岛的错误配置。具体地说，它表明用除了岛间干线之外（多个）链路互连了两个不同的岛。此外，如果在除了岛间端口之外的任何端口处，接收到其岛名与接收交换机的岛名不匹配的 BPDU，或者接收到不是 MST BPDU 的 BPDU，则对所有的 PE-VLANID 阻塞端口，并且发出操作者警报的信号。

[0123] 类似地，如果接收到不具有岛 ID 字段 502 的 BPDU，则丢弃它，并且在接收网桥的生成树计算中不依赖于它。

[0124] 图 9 是网络 200 的高度简化的局部方框图，图示了布置在岛互连结构 208 中并且配置来为第二 VEC 传输流量的岛间干线 242。如上所述，第二 VEC 在岛 202 和 204 之间延伸。此外，每个这些岛 202 和 204 都分别具有两个岛间链路 210a、210b 和 210c、210d。每个岛 202 和 204 通过将它的除了一个耦合到岛间干线 242 的端口之外的所有端口都置于阻塞状态中，防止如果不这样的话则由于存在岛间干线 242 而形成环。例如，可以将在岛 202 处耦合到岛链路 210b 的端口和在岛 204 处耦合到岛链路 210d 的端口每个都转变到阻塞状态，如点 902 和 904 所指示。另一方面，将对应于链路 210a 和 210c 的端口每个都转变到转发状态。

[0125] 图 10 是 MAN 200 的高度简化的局部方框图，图示了布置在岛互连结构 208 中并且配置来为第三 VEC 传输流量的岛间干线 244，其中第三 VEC 配置来经由岛 204 和 206 连接客

户站点 216 和 217(图 2)。经由岛间链路 210c、210d 和 210e、210f,将岛 204 和 206 耦合到岛间干线 244。每个岛 204 和 206 通过将它的除了一个耦合到岛间干线 244 的端口之外的所有端口都置于阻塞状态中,防止如果不这样的话则由于存在岛间干线 244 而形成环。例如,可以将在岛 204 处耦合到岛链路 210c 的端口和在岛 206 处耦合到岛链路 210e 的端口每个都转变到阻塞状态,如点 1002 和 1004 所指示。将对应于链路 210d 和 210f 的端口每个都转变到转发状态。

[0126] 假设还将岛 202、204 和 206 配置来提供第四 VEC,以互连客户站点 213、216 和 217。图 11 是 MAN 200 的高度简化的局部方框图,图示了布置在岛互连结构 208 中已经配置来为第四 VEC 传输流量的岛间干线 1100。在这里,所有四个岛间链路 210a-d 都连接到岛间干线 1100。为了防止形成环,每个岛 202、204 和 206 将它的除了一个耦合到岛间干线 1100 的端口之外的所有端口都置于阻塞状态中。例如,可以将在岛 202 处耦合到岛链路 210a 的端口、在岛 204 处耦合到岛链路 210d 的端口和在到 206 处耦合到岛链路 210f 的端口每个都转变到阻塞状态,如点 1102、1104 和 1006 所指示。将对应于链路 210b、210c 和 210e 的端口每个都转变到转发状态。

[0127] 如图 9-11 所示,尽管利用多个岛间链路将每个岛都耦合到岛互连结构,但是专门避免了形成环。另外,在岛间链路 210 之间负载共享流量。

[0128] 图 12 是根据本发明的另一个岛间干线 1200 的高度简化的示意图。为了清楚已经省略了岛间互连结构。岛间干线 1200 包括多个岛 1202-1210。此外,每个岛具有多个互连的网桥。如图所示,在岛 1202-1210 之间形成有三个 VEC 1212-1216,都在单个岛间干线上传输。岛 1202 只具有到 VEC 1212 的单个连接 1215。因此,如果丢失了连接 1215,则岛 1202 就失去与岛 1204 和 1210 的连通性。将岛 1208 的网桥组织成 1218a 和 1218b 两部分,每部分都由四个互连的网桥组成。然而,在形成岛 1208 内部的两部分 1218a 和 1218b 的网桥之间没有连接。相反,岛 1208 的两部分 1218a 和 1218b 利用 VEC 1214 和 1216 来互通信。类似地,在岛 1206 处,执行 IMCP 导致将两个网桥之间的链路 1219 阻塞。岛 1206 的两个网桥利用 VEC 1216 来互通信。

[0129] 图 13 是与图 12 相同的岛间干线 1200 的高度简化的示意图。然而,为了清楚已经省略了 VEC,并且图示了 BPDU VEC 或 BPDU 服务实例 1302。如上所述,被连接到岛间干线的岛 1202-1210 的每个岛间网桥具有 到 BPDU VEC 1302 的连接。如这里所述,岛间网桥利用 BPDU VEC 1302 来在它们之间交换 BPDU。岛间网桥在它们执行 IMCP 时使用这些接收到的 BPDU 来标识并阻塞到 VEC 的冗余链路。对于那些连接到给定的岛间干线上的多个 VEC 的岛间网桥,只需要单个到 BPDU VEC 1032 的连接。例如,在被连接到 VEC 1212 和 1214(图 2)的岛 1210 处的网桥 1304 只需要建立到 BPDU VEC 1302 的单个连接 1306。

[0130] 在每个岛内,岛边界网桥在它们的岛端口上运行 IMCP。对于岛内的其他端口,即非岛间端口,可以运行 IMCP 或者 MSTP。

[0131] 在替换实施方式中,可以使用未修改的 IEEE P802.1S/D13 多生成树协议,而不是 IMCP。在这种情形中,每个岛间干线必须具有用于每个岛的分离的 BPDU VEC。相反,必须将用于每个岛间干线的每个岛间网桥配置来附加到与在那个相同的岛中的其他岛间网桥相同的 BPDU VEC。此外,如果采用了多个岛间干线,则由每个岛间干线的 BPDU VEC 互连的岛的集合必须相同。从而,BPDU VEC 的连通性,而不是岛 ID 的比较定义了岛。由于在任何

BPDU VEC 上不同岛的岛间网桥未被互连，所以它们不能检测并报告未利用岛间干线的岛之间的错误连接。

[0132] 如上所述，存在不同类别的 VEC。上面所述的 VEC 对应于“类似于网桥的”VEC，在这种 VEC 中，接收到的帧的 CE-VLAN ID 在岛中被修改。另外，一旦在 UNI 处接收到对应于未用于客户 -MAN 交互的 L2 协议的网络消息，例如 IEEE Std.802.3-2000 暂停帧（也称作 802.3x 暂停帧），就将其丢弃。如上所述，布置在岛中的交换机在它们的 CIST 的计算中从不利用来自客户站点的 BPDU。利用“线形 (wire-like)”VEC，将用 CE-VLAN ID 加标签的帧透明地传输过 MAN，如同对应于未用于客户 -MAN 交互的 L2 协议的网络消息那样。

[0133] 还应当理解，岛可以由单个 L2 交换机组成。在这种情形中，将 PE-VLAN ID 局限于单个交换机。

[0134] 前面的描述已经集中于本发明的特定实施方式。然而，很清楚，可以对所描述的实施方式做出其他各种变化和修改，并且获得它们的某些或全部优点。因此，所附权利要求的目的是覆盖本发明的真实精神和范围中的 所有这种变化和修改。

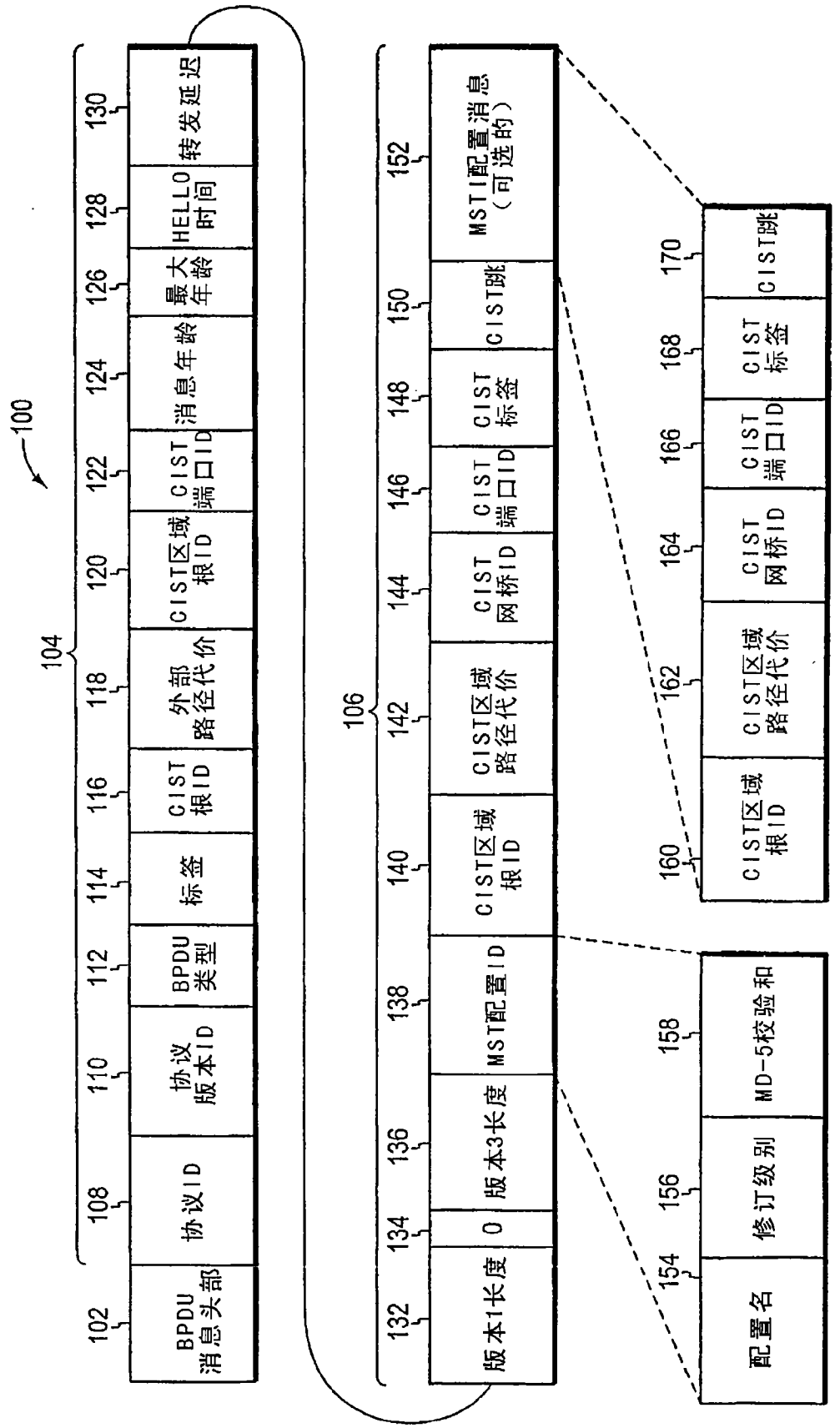

100
104
106
BPDU消息头部
协议ID
协议版本ID
BPDU类型
标签
CIST根ID
外部路径代价
CIST区域根ID
CIST端口ID
消息年龄
最大年龄
HELLO时间
转发延迟
102
108
110
112
114
116
118
120
122
124
126
128
130
版本1长度
0
版本3长度
MST配置ID
CIST区域根ID
CIST区域路径代价
CIST网桥ID
CIST端口ID
CIST标签
CIST跳
MSTI配置消息（可选的）
132
134
136
138
140
142
144
146
148
150
152
配置名
修订级别
MD-5校验和
CIST区域根ID
CIST区域路径代价
CIST网桥ID
CIST端口ID
CIST标签
CIST跳
154
156
158
160
162
164
166
168
170

图1
(现有技术)

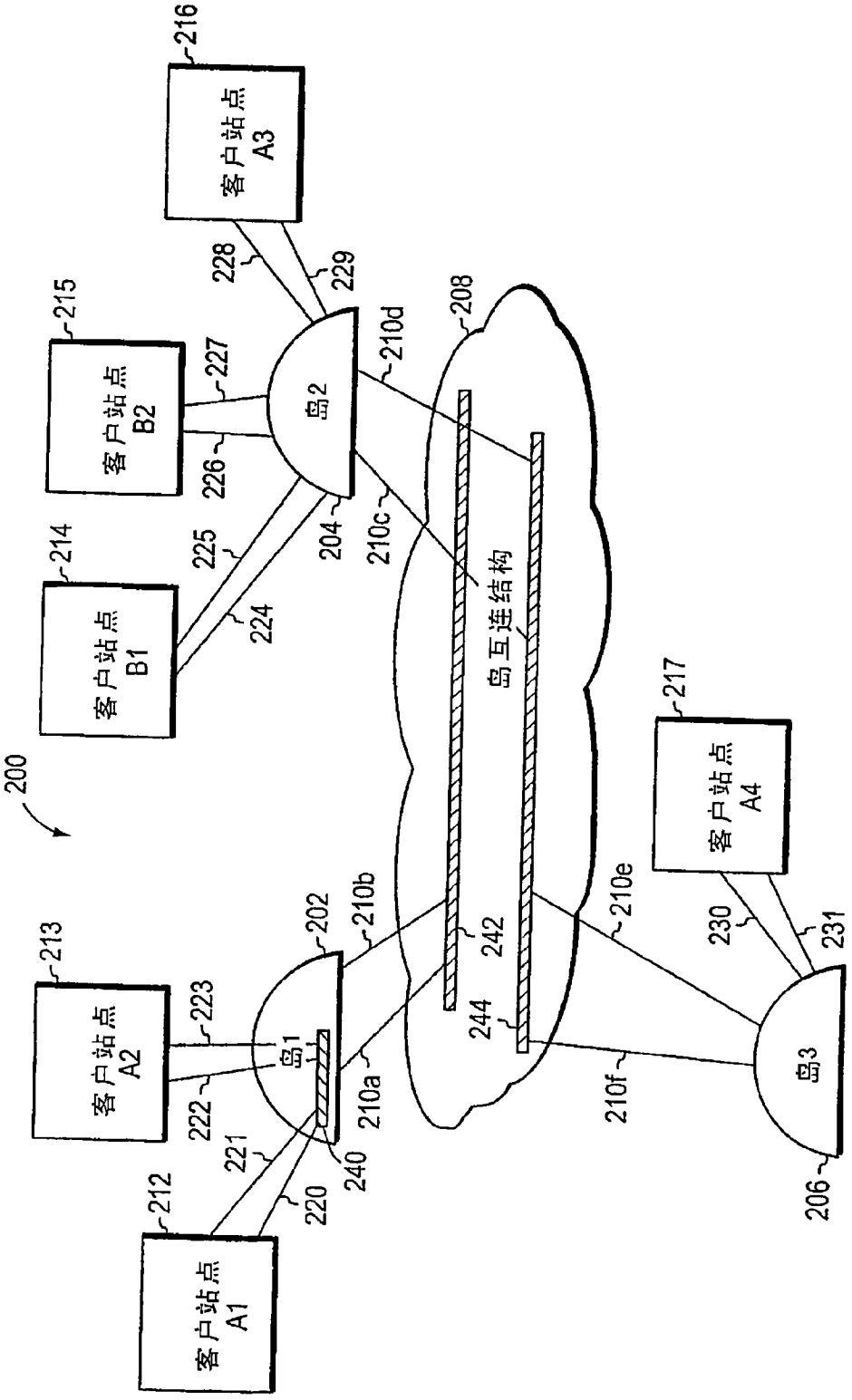

200
客户站点 A1
212
客户站点 A2
213
客户站点 B1
214
客户站点 B2
215
客户站点 A3
216
客户站点 A4
217
岛1
岛2
岛3
岛互连结构
202
204
206
208
210a
210b
210c
210d
210e
210f
220
221
222
223
224
225
226
227
228
229
230
231
240
242
244

图2

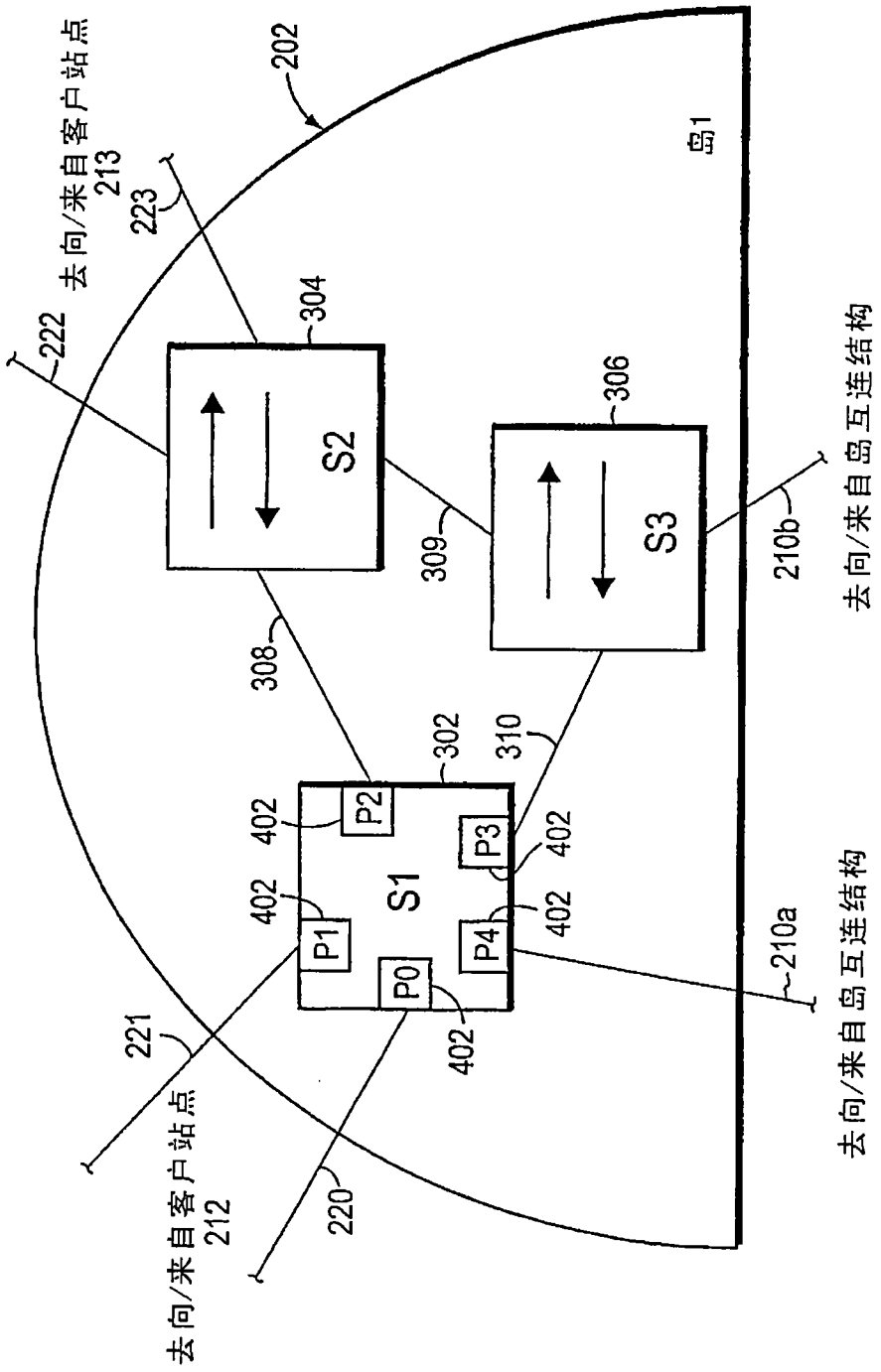

岛1
202
去向/来自客户站点
213
223
222
304
306
S2
S3
309
308
310
302
S1
P0
P1
P2
P3
P4
402
221
212
220
210a
210b
去向/来自岛互连结构
去向/来自岛互连结构
去向/来自客户站点

图3

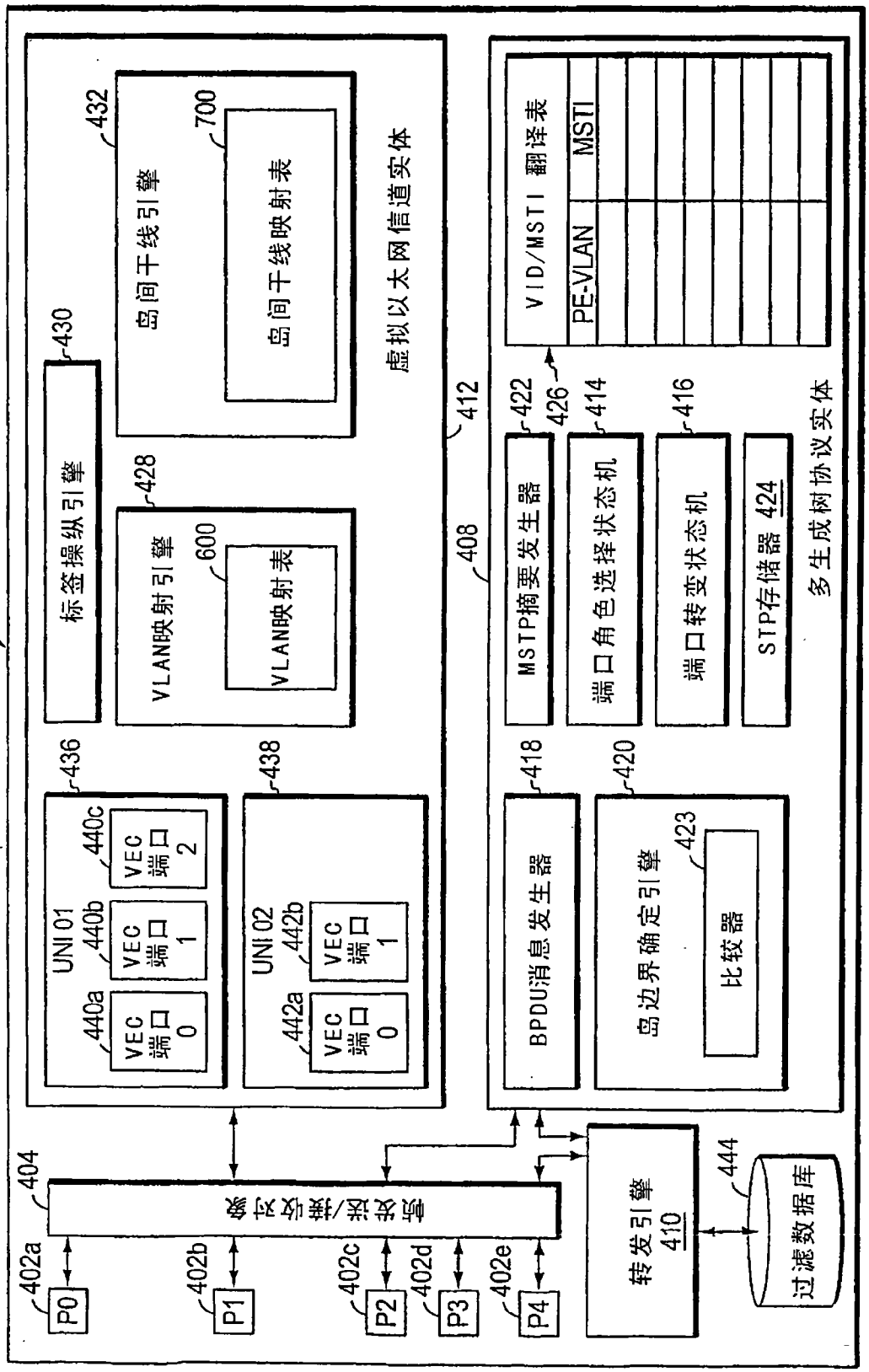

302
虚拟以太网信道实体
412
标签操纵引擎
430
岛间干线引擎
432
岛间干线映射表
700
VLAN映射引擎
428
VLAN映射表
600
408
UNI 01
436
VEC端口0
440a
VEC端口1
440b
VEC端口2
440c
UNI 02
438
VEC端口0
442a
VEC端口1
442b
BPDU消息发生器
418
岛边界确定引擎
420
比较器
423
MSTP摘要发生器
422
端口角色选择状态机
414
端口转变状态机
416
STP存储器 424
VID/MSTI 翻译表
426
PE-VLAN
MSTI
多生成树协议实体
转发引擎 410
444
过滤数据库
端口保护/禁止对象
404
402a
P0
402b
P1
402c
P2
402d
P3
402e
P4

图4

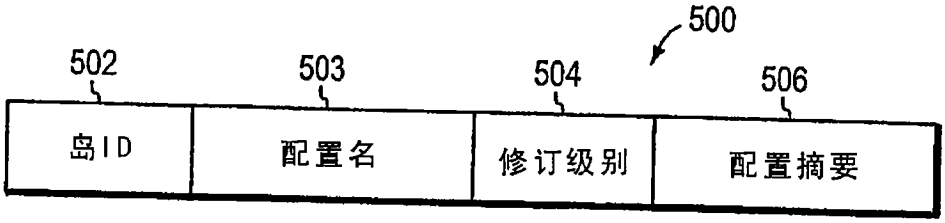


图 5

600

VLAN映射表

602 CE-VLAN ID	604 VEC	606 PE-VLAN ID	608 UNI	610 VEC端口	
0014	001	4011	01	00	614a
0038	002	4027	01	01	614b
0018	002	4017	00	00	614c

图 6

700

岛间干线映射表

702 PE-VLAN ID	704 VEC	706 岛间干线ID	
4027	002	6042	710a
4017	002	6042	710b
-	301	6042	710c

图 7

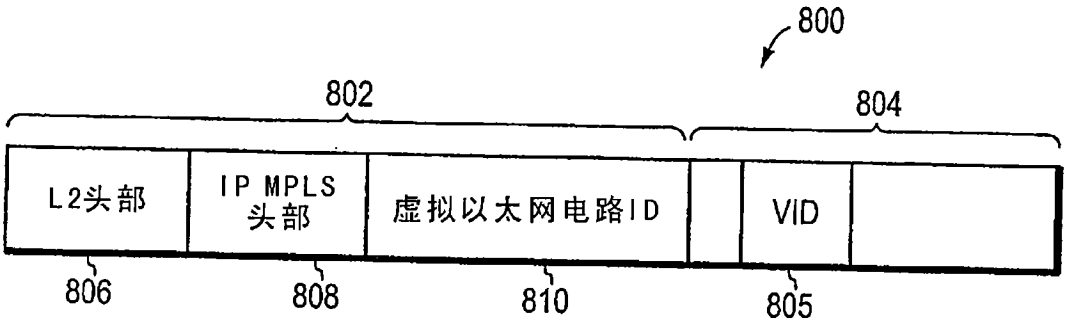

800
802
804
L2头部
IP MPLS
头部
虚拟以太网电路ID
VID
806
808
810
805

图 8

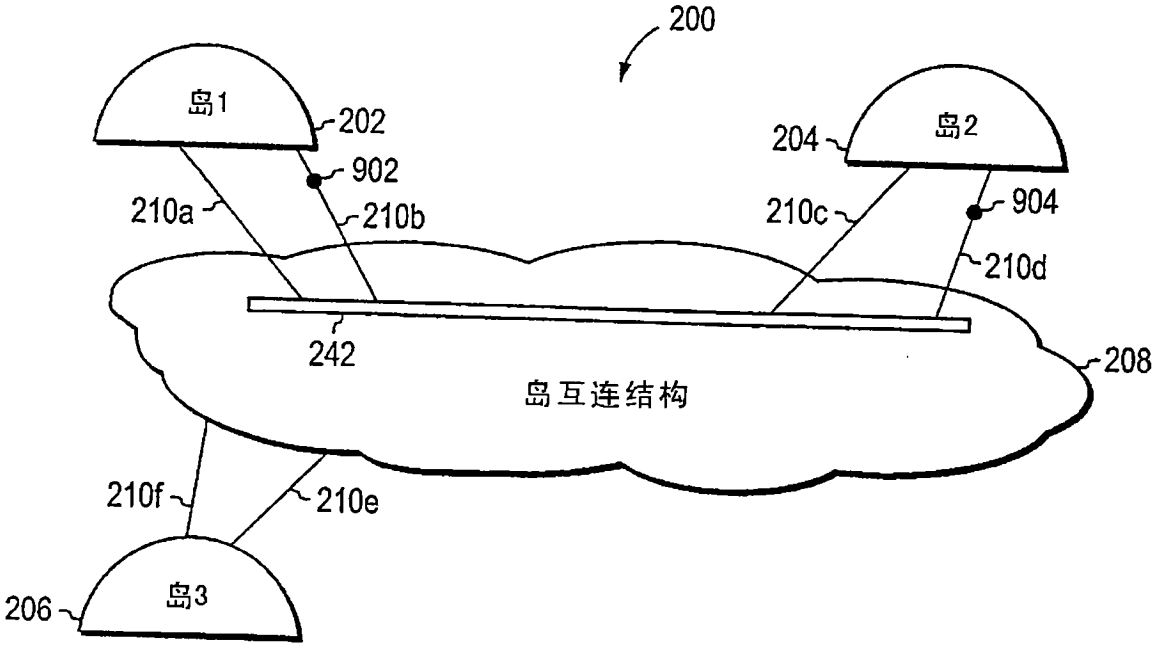

200
岛1
202
902
210a
210b
204
岛2
210c
904
210d
242
208
岛互连结构
210f
210e
206
岛3

图9

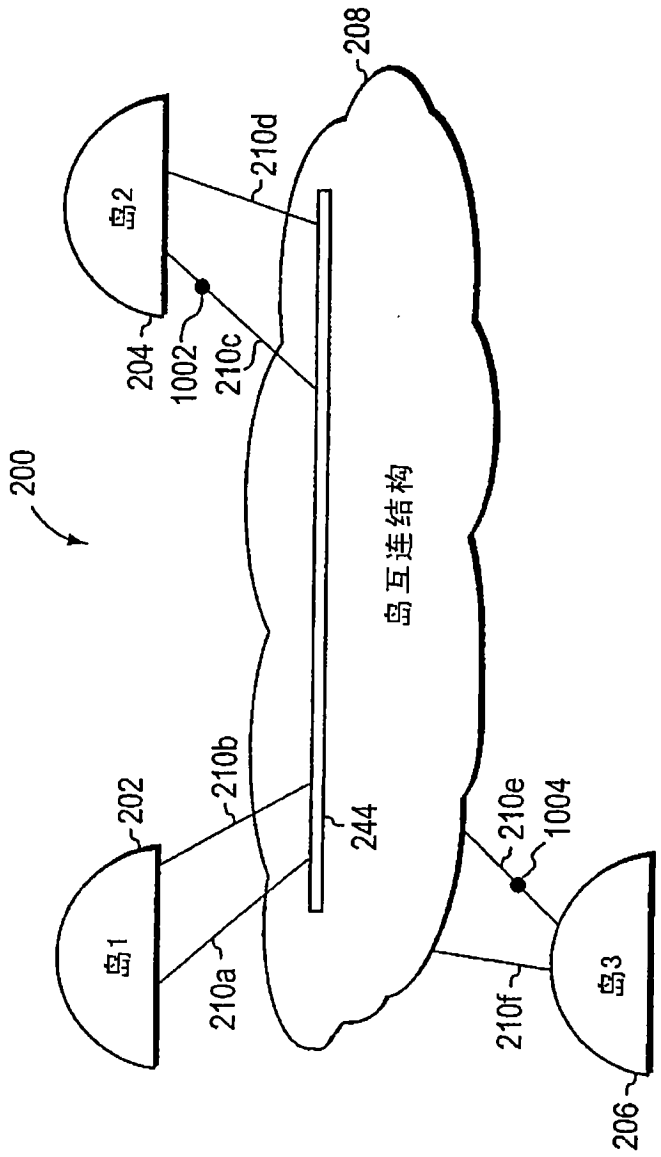

200
208
岛2
210d
204
1002
210c
岛互连结构
岛1
202
210b
244
210e
1004
210a
210f
岛3
206

图10

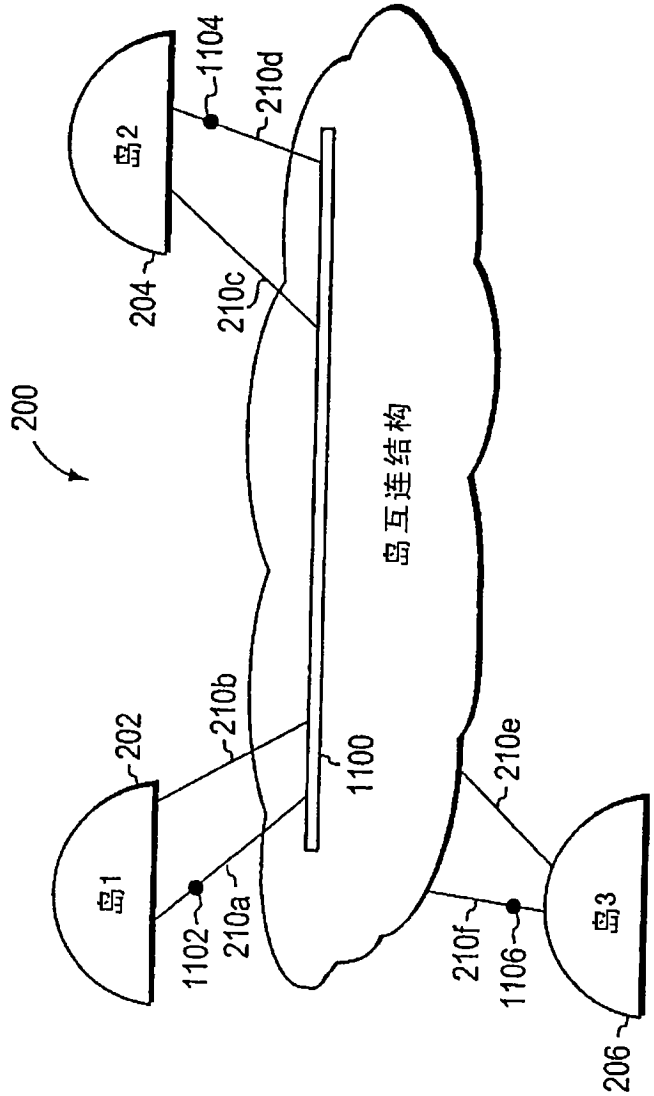

200
岛1
岛2
岛3
岛互连结构
202
204
206
210a
210b
210c
210d
210e
210f
1100
1102
1104
1106

图11

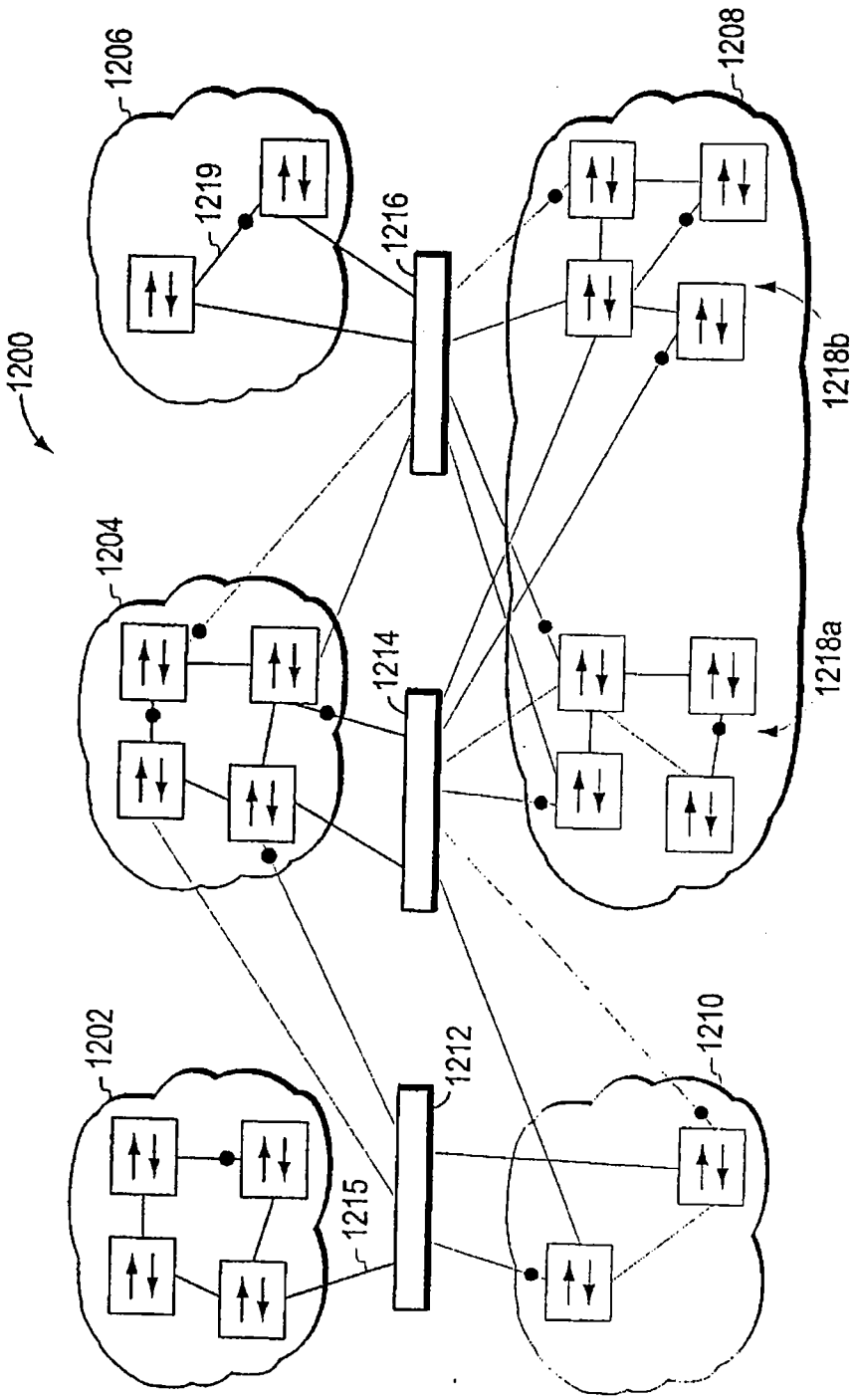

1200
1206
1219
1216
1208
1218b
1204
1214
1218a
1202
1212
1210
1215

图12

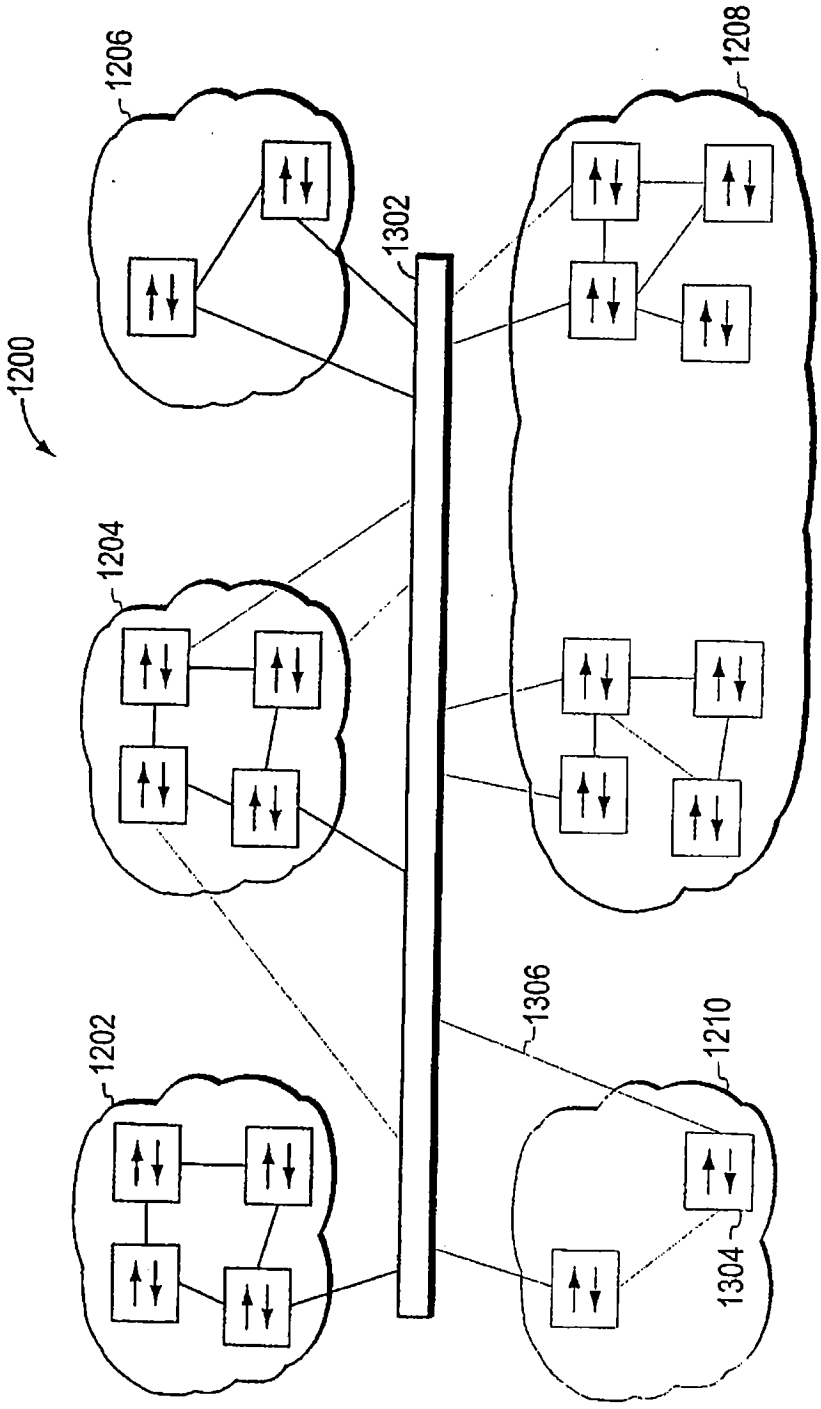


图13