(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 102197627 B
(45)授权公告日 2016.02.17

(21)申请号 200980143097.9

(22)申请日 2009.07.24

(30)优先权数据
12/201,799 2008.08.29 US
12/205,715 2008.09.05 US

(85)PCT国际申请进入国家阶段日
2011.04.28

(86)PCT国际申请的申请数据
PCT/US2009/051681 2009.07.24

(87)PCT国际申请的公布数据
WO2010/024993 EN 2010.03.04

(73)专利权人 极进网络有限公司
地址 美国加利福尼亚州

(72)发明人 苏尼尔·P·沙哈 阿尼尔·利姆
唐纳德·B·戈洛斯尔 吉姆·潘
科萨万·蒂鲁凡卡塔萨姆 朴基弘
曼普瑞特·S·桑杜
巴拉卡什·卡什耶普

(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 宋鹤

(51)Int.Cl.
H04L 12/70(2013.01)
H04L 12/18(2006.01)

(56)对比文件
WO 2008055427 A1,2008.05.15,
CN 101009689 A,2007.08.01,
CN 1863147 A,2006.11.15,

审查员 万红芳

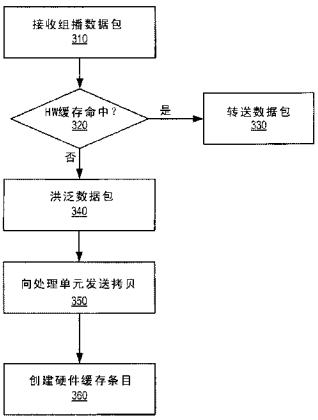
权利要求书2页 说明书8页 附图7页

(54)发明名称

组播流量收敛的改善

(57)摘要

在传输节点接收从源节点发送的组播数据包(310)。组播数据包包括源地址和组播组地址。对于组播数据包在传输节点检测硬件缓存缺失(320)。将组播数据包硬件洪泛到网络的端口(340)。洪泛包括基于虚拟局域网(VLAN)成员资格,将组播数据包的拷贝转送到传输节点的相邻节点。将组播数据包的缓存缺失拷贝发送到线外处理单元(350),在线外处理单元软件处理组播数据包的缓存缺失拷贝。处理包括经由硬件提取层建立用于组播数据包的硬件缓存条目(360)。不将缓存缺失拷贝转送到网络上。



CN 102197627 B

1. 一种用于组播收敛的方法包括：

在传输节点接收来自源节点的组播数据包，所述组播数据包具有源地址和组播组地址；

将所述组播数据包的源地址和组播组地址与硬件缓存的硬件条目进行比较；

响应于所述比较的结果不是所述组播数据包的源地址和组播组地址与所述硬件条目相匹配，对于所述组播数据包在所述传输节点检测硬件缓存缺失；

响应于检测到所述硬件缓存缺失来经由硬件将所述组播数据包洪泛到网络的端口，其中所述洪泛包括基于虚拟局域网 (VLAN) 成员资格，将所述组播数据包的拷贝转送到所述传输节点的相邻节点；

将所述组播数据包的拷贝发送到处理单元；

软件处理所述组播数据包的拷贝，而不将所述拷贝从所述处理单元转送到所述网络上，所述软件处理包括经由硬件提取层建立用于所述组播数据包的硬件缓存条目；以及

响应于未来组播数据包的源地址和组播组地址与用于所述组播数据包的硬件缓存条目相匹配，将所述未来组播数据包仅转送到所述传输节点的预定了组播流量的相邻节点。

2. 如权利要求 1 所述的方法，其中在所述传输节点检测到硬件缓存缺失是响应于所述比较未导致所述硬件条目与所述组播数据包的源地址和组播组地址的匹配的，并且其中经由硬件将所述组播数据包洪泛到网络的端口包括：

在所述组播数据包的拷贝被所述处理单元进行软件处理时进入初始硬件洪泛状态。

3. 如权利要求 1 所述的方法，其中所述处理单元是中央处理单元 (CPU) 或网络处理单元 (NPU) 中的一个。

4. 如权利要求 1 所述的方法，其中所述传输节点是以太网自动保护交换 (EAPS)。

5. 如权利要求 1 所述的方法，其中硬件洪泛通过对应的计时器开始。

6. 如权利要求 5 所述的方法，进一步包括当所述计时器到期时停止所述硬件洪泛。

7. 一种网络装置包括：

数据包处理器，该数据包处理器接收具有源地址和组播组地址的组播数据包；

硬件缓存，该硬件缓存存储用于组播流量的转送条目，所述转送条目基于源地址和组播组地址；

所述数据包处理器进一步：

将所述组播数据包的源地址和组播组地址与硬件缓存的转送条目进行比较；

响应于所述比较的结果不是所述组播数据包的源地址和组播组地址与所述转送条目相匹配，检测针对所述组播数据包的硬件缓存缺失，

响应于检测所述硬件缓存缺失来将所述组播数据包洪泛到网络的端口，其中洪泛包括至少部分地基于虚拟局域网 VLAN 成员资格，将所述组播数据包的拷贝转送到所述网络装置的相邻节点，以及

将所述组播数据包的拷贝发送到处理单元；

其中所述处理单元还用于软件处理所述组播数据包的拷贝，而不将所述拷贝转送到所述网络，所述软件处理包括经由硬件提取层建立用于所述组播数据包的硬件缓存条目，并且

其中所述数据包处理器还用于响应于未来组播数据包的源地址和组播组地址与用于

所述组播数据包的硬件缓存条目相匹配，将所述未来组播数据包仅转送到所述网络装置的预定了组播流量的相邻节点。

8. 如权利要求 7 所述的网络装置，其中所述数据包处理器还可操作用于将所述组播数据包的源地址和组播组地址与所述硬件缓存的转送条目进行比较。

9. 如权利要求 7 所述的网络装置，其中所述数据包处理器通过在所述组播数据包的拷贝被所述处理单元进行软件处理时进入初始硬件洪泛状态，来经由硬件将所述组播数据包洪泛到网络的端口。

10. 如权利要求 7 所述的网络装置，其中所述处理单元包括以太网自动保护交换 (EAPS) 部件，用于在 EAPS 网络上配置所述网络装置。

11. 如权利要求 7 所述的网络装置，其中所述处理单元包括命令行接口 (CLI)，用于在网络上配置所述网络装置。

12. 如权利要求 7 所述的网络装置，其中所述数据包处理器进一步包括计时器，其中当所述计时器到期时，通过所述数据包处理器的洪泛停止。

13. 一种用于组播收敛的设备包括 ：

用于在传输节点接收来自源节点的组播数据包的装置，所述组播数据包具有源地址和组播组地址 ；

用于将所述组播数据包的源地址和组播组地址与硬件缓存的硬件条目进行比较的装置 ；

用于响应于所述比较的结果不是所述组播数据包的源地址和组播组地址与所述硬件条目相匹配，对于所述组播数据包在所述传输节点检测硬件缓存缺失的装置 ；

用于响应于检测到所述硬件缓存缺失来经由硬件将所述组播数据包洪泛到网络的端口的装置，包括用于基于虚拟局域网 (VLAN) 成员资格，将所述组播数据包的拷贝转送到所述传输节点的相邻节点的装置 ；

用于将所述组播数据包的拷贝发送到处理单元的装置 ；

用于软件处理所述组播数据包的拷贝，而不将所述拷贝从所述处理单元转送到所述网络上的装置，包括用于经由硬件提取层建立用于所述组播数据包的硬件缓存条目的装置 ；以及

用于响应于未来组播数据包的源地址和组播组地址与用于所述组播数据包的硬件缓存条目相匹配，将所述未来组播数据包仅转送到所述传输节点的预定了组播流量的相邻节点的装置。

14. 如权利要求 13 所述的设备，其中在所述传输节点检测到硬件缓存缺失是响应于所述比较未导致所述硬件条目与所述组播数据包的源地址和组播组地址的匹配的，并且所述用于洪泛的装置包括 ：

用于在所述组播数据包的拷贝被所述处理单元进行软件处理时进入初始硬件洪泛状态的装置。

15. 如权利要求 13 所述的设备，其中所述传输节点是以太网自动保护交换 (EAPS)。

16. 如权利要求 13 所述的设备，其中硬件洪泛通过对应的计时器开始。

17. 如权利要求 16 所述的设备，进一步包括用于当所述计时器到期时停止所述硬件洪泛的装置。

组播流量收敛的改善

[0001] 本申请是2008年8月29日提交的名为“组播流量收敛的改善”(Improved Convergence of Multicast Traffic)的待审美国专利申请第12/201,799号的部份继续申请,并要求其优先权。

技术领域

[0002] 本申请公开的实施例涉及计算机联网,尤其涉及网络中拓扑结构变化之后组播流量的收敛。

背景技术

[0003] 组播是在互联网协议(IP)基础结构上用于点到多点通信的技术。组播利用网络基础结构,要求源仅向网络发出单个数据包,即使需要将其传送到多个目的地。通过使得网络节点复制数据包(仅当需要时)用于传送到多个接收器来完成所述利用。

[0004] 通过源和接收器使用组播组地址(例如IP组播)来发送和接收内容。源将组播组地址用作它们的数据包中的目的地IP地址。接收器使用组播组地址以向与组播组地址相关联的组播流量进行“预订”。换言之,这些接收器使用组播组地址以向网络传达接收寻址到该组播组的流量的愿望。

[0005] 互联网组管理协议(IGMP)是用于管理IP组播组成员资格的协议的实例。典型地,IP主机经由组播路由器发出广播到其他网络主机的IGMP询问。网络装置(例如第2层交换机)可以“监听”主机与路由器之间的对话,这是在现有技术中称为IGMP侦听的处理。当装置听到来自主机的组播组“加入”消息时,装置记录它听到消息的接口(例如端口)并将接口加入组中。类似地,当装置听到组播组“离开”消息或者响应计时器到期时,交换机将从组中移除该主机的交换机接口。通常将这些“加入”和/或“离开”消息称为“IGMP报告”。因此,IGMP侦听表基于IGMP报告将作为组播组成员的主机和/或接口的列表维护在硬件(例如缓存)中。

[0006] 当使用组播的网络中有拓扑结构变化时,将IGMP侦听表清空并发出IGMP询问,因此网络中的装置可以再获知IGMP侦听成员资格。如果在拓扑结构变化期间网络中有数据包,那么因为IGMP侦听表的清空,这些数据包将在硬件中经历缓存缺失。经历缓存缺失的所有数据包被发送到装置CPU用于软件中的转送,有时候将软件中的转送称为“慢通道”转送,因为软件中的转送比硬件中的转送数据包显著更慢。

[0007] 但是更麻烦的是如下的情况:在拓扑结构变化期间,在网络中有数千的数据包。在这种情况下,在CPU队列中有数千的数据包,消耗CPU资源,否则CPU资源可用于通过新的IGMP侦听表条目再规划硬件缓存。换言之,当网络中的组播数据包数量增加时,它增加了网络中组播流量的收敛时间(即,用于再规划硬件缓存以恢复组播流量中的缓存缺失的可接受比率的时间)。此外,上述情景假设了单个虚拟局域网(VLAN)。如果在网络中有大量(例如数千)的VLAN,则会将每个IGMP询问发送到每个VLAN,导致每个VLAN上的所有主机利用表示它们的IGMP成员资格的各自的IGMP报告作出响应。因此,CPU负担更多的慢通

道转送，进一步加重了组播收敛问题。

发明内容

[0008] 本申请公开的实施例有助于拓扑结构变化之后网络中组播流量的收敛。当网络节点接收组播数据包时，处理数据包。如果节点检测到针对数据包的缓存缺失，则将该数据包经由硬件洪泛到网络的端口上。此外，（导致了缓存缺失的）数据包的拷贝被发送到线外(out-of-line)处理单元。数据包的拷贝被处理，而不将拷贝从处理单元转送到网络上。对数据包的处理包括经由硬件提取层建立用于数据包的硬件缓存条目。

附图说明

[0009] 下面的描述包括附图的讨论，附图通过本发明实施例的实施方式的实例给出说明。附图应当理解作为实例，而不是作为限制。如同本申请中使用的，对一个或多个“实施例”的参考应当理解为描述包括在本发明至少一个实施方式中的特定特征、结构或特点。因此，本申请中出现的措辞例如“在一个实施例中”或“在替代实施例中”描述本发明的不同实施例和实施方式，不一定都表示相同的实施例。但是，它们也不一定相互排斥。

[0010] 图 1 是方框图，示出网络装置。

[0011] 图 2 是方框图，示出根据不同实施例使用组播的网络。

[0012] 图 3 是流程图，示出用于组播收敛的处理。

[0013] 图 4 是方框图，示出现有技术以太网自动保护交换 (EAPS) 系统。

[0014] 图 5 是方框图，示出采用以太网自动保护交换 (EAPS) 的网络装置。

[0015] 图 6 是流程图，示出用于 EAPS 环形网络中组播收敛的处理。

[0016] 图 7 是方框图，示出用于实践本申请所述各种实施例的适当计算环境。

具体实施方式

[0017] 在下面的描述中，提供多种细节。但是对本领域技术人员显而易见的是，不需要这些特殊细节也能实践本发明。在一些实例中，为了避免模糊本发明，将公知结构和装置示出为方框图形式而不是详细示出。

[0018] 本申请提供的方法、设备和系统有助于拓扑结构变化之后网络中组播流量收敛的改善。当网络节点接收组播数据包时，处理数据包。如果节点检测到关于数据包的缓存缺失，则数据包经由硬件洪泛到网络的端口上。此外，（导致缓存缺失的）数据包的拷贝被发送到线外处理单元。数据包的拷贝被处理，而不将拷贝从处理单元转送到网络上。对数据包的处理包括经由硬件提取层建立用于数据包的硬件缓存条目。

[0019] 在一些实施例中，当新的组播流开始在网络上发送流量时，缓存缺失出现并被检测到。不是简单地将缓存缺失转送到 CPU 以用于获知（在硬件中）并随后将其转送（即慢通道转送）到网络上（例如 IGMP 侦听使能），而是采用特殊的转送模式。在这种特殊的转送模式中，导致缓存缺失的数据包一开始经由硬件洪泛到网络的端口。此外，缓存缺失被发送到 CPU 以用于处理。处理包括获知每个数据包的（一个或多个）地址（例如源 IP 地址和组播组 IP 地址）以及在硬件中规划新的条目。通过规划硬件中的条目，具有相同源和组 (S,G) 地址的未来数据包会导致缓存命中并立即仅被转送出有实际接收器的那些端口。仅

在有实际接收器的端口上转送可称为“选择性转送”。被选择性转送的数据包会从该时刻起停止被洪泛。

[0020] 在各种实施例中，具体地，通过 CPU 的处理不包括将缓存缺失数据包拷贝从 CPU 转送到网络上（例如“慢通道转送”，“软件转送”等等）。这是因为慢通道转送会导致重复的数据包在网络上发出（假定导致缓存缺失的数据包已经洪泛到网络）。例如可经由命令行接口 (CLI)，由用户来控制特殊的转送模式。

[0021] 在采用用于具有环形拓扑结构（例如以太网自动保护交换 (EAPS) 环形拓扑结构）的网络的实施例中，响应于检测到网络中的拓扑结构变化，可自动触发类似的特殊转送模式。当 EAPS 环形网络中拓扑结构变化时，转送条目（称为转送数据库 (FDB) 条目）被清空并必须被再获知。因此，在这些实施例中，检测的拓扑结构变化会导致环形网络进入硬件洪泛模式，通过计时器确定硬件洪泛模式的上限。本申请使用的洪泛表示例如基于 VLAN 成员资格在网络装置的出站端口上发送数据包的拷贝。换言之，洪泛可导致数据包在网络装置的出站端口上转送，但是不一定在所有出站端口上。

[0022] 在洪泛周期里，导致缓存缺失的数据包也被转送到网络装置处理单元。处理单元创建针对每个缓存缺失数据包组播缓存条目，并且缓存条目经由硬件提取层在硬件中被规划。在硬件条目上产生命中的后续数据包将基于与条目相关联的所规划的出站端口在硬件中转送。通过这种方式，随着新的组播缓存条目被规划，缓存缺失将最终逐渐停止。

[0023] 当硬件洪泛模式计时器到期时，网络装置可以回复到它之前的操作模式。例如，如果在“初始硬件洪泛”模式之前，装置操作在 IGMP 侦听使能模式下，那么当计时器到期时，装置可以回复到 IGMP 侦听使能模式。在不使用 IGMP 侦听的实施例中，可能必须当时间到期时，将当前 / 最新的组播缓存列表推送到硬件提取层 (HAL)，以防止将来的缓存缺失。

[0024] 图 1 是方框图，示出根据各种实施例的网络装置。网络装置 110 包括一个或多个硬件表 120（例如缓存）、处理单元 130（例如中央处理单元 (CPU)、网络处理单元 (NPU) 等等）以及数据包处理器 140。数据包处理器 140 接收进入数据包。对于每个进入数据包（具有源地址和组播组地址），数据包处理器 140 进行硬件表 120 的查找，以确定进入数据包的源地址和组播组地址 (S，G) 是否与表中的其中一个条目匹配。如果匹配，则数据包处理器 140 将数据包转出到网络（例如基于组播组地址）。

[0025] 如果 (S,G) 地址组合与表 120 中的条目不匹配，则数据包处理器 140 将具有 (S,G) 地址组合的所有流量洪泛到网络并产生缓存缺失。这种缓存缺失事件或者简称缓存缺失被发送到处理单元 130，具体而言发送到内核 132。内核 132 将缓存缺失发送到组播 (MC) 管理器 134。在一些实施例中，洪泛会导致流量在网络装置 110 的所有出站端口上发出。在其他实施例中，例如基于 VLAN 成员资格，仅在出站端口上转送数据包。

[0026] 与数据包处理器 140 洪泛 (S，G) 流量结合，处理单元 130 工作以解决缓存缺失。MC 管理器 134 启动代理 IGMP 询问，代理 IGMP 询问在网络上被发出。基于响应于代理 IGMP 询问接收的任何 IGMP 报告，MC 管理器 134 通过用于 (S，G) 流量的转送条目规划硬件提取层 (HAL)136。HAL 136 随后配置硬件表 120 以包括转送条目。一旦对于 (S,G) 组已经“获知”（即规划）转送条目，则后续 (S,G) 流量将“命中”硬件表 120，并根据转送条目被转送（即没有洪泛）。重要的是注意，在处理之后缓存缺失没有转送回网络（就像标准 IGMP 侦听使能模式中的情况那样）。在本申请所述的实施例中，缓存缺失不是通过 CPU 以慢通道方

式转送到网络以避免重复数据包（就像所讨论的，因为数据包会已经由硬件洪泛转送到网络）。

[0027] 在各种实施例中，用户经由命令行接口 138 控制装置 110 使用的初始硬件洪泛模式。因此，用户可以使装置操作在正常 IGMP 侦听使能模式或侦听禁用模式下，或者用户可以操作装置以使用具有初始硬件洪泛模式的 IGMP 侦听使能。

[0028] 图 2 是方框图，示出采用组播和 IGMP 侦听的网络。在标准侦听使能情景中，网络 200 是这样的网络，其中组播发射器 (MC Tx)210 发出 IGMP 询问，以确定哪个节点（如果有的话）对接收针对特定组播组（例如组 ABC) 的组播流量感兴趣。组播接收器 (MC Rx)224 从 MC Tx 210 接收询问并报告它对接收针对组 ABC 的组播流量感兴趣。各种网络节点“监听”MC Tx 210 与 MC Rx 224 之间的通信，注意，通过针对组 ABC 更新它们各自的组播列表，MC Rx 224 对针对组 ABC 的组播流量感兴趣。例如，假定 MC Rx 经由端口 P2 可通信地连接到节点 222，节点 222 将端口 P2 加入它的针对组 ABC 的组播列表中。同样地，假定端口 P2 是流量将到达MC Rx 224 的端口，节点 220 将其端口 P2 加入它的针对组 ABC 的组播列表中。如图 2 所示，节点 218、216、214 和 212 也将它们各自的 P2 端口加入它们各自的针对组 ABC 的组播列表中。通过这种方式，到达节点 212 的组 ABC 流量将通过网络适当地转送到 MC Rx 224。此外，在标准侦听使能情景中，将缓存缺失发送到 CPU。这些数据包可通过 CPU 以慢通道方式转送，但是直到硬件提取层对硬件进行了规划，才会在硬件中转送它们。

[0029] 但是，在各种情景中，可在网络 200 上启动新的组播流而没有上述 IGMP 询问和 / 或报告。例如，如果针对组 XYZ 的流从节点 228 启动而没有任何前述 IGMP 询问和 / 或报告，则假定网络 200 中的节点都不具有用于组 XYZ 的任何组播列表条目，到达网络 200 的每个组 XYZ 数据包会导致缓存缺失。利用上述的标准 IGMP 侦听使能技术，这些缓存缺失将堵塞在网络节点的处理单元的慢通道转送队列中，消耗处理资源，延迟 XYZ 流的收敛。如果在网络 200 上，在相同或较近的时间启动多个新的不同的流，则收敛延迟增加。

[0030] 在各种实施例中，例如上述网络节点的缓存缺失会导致网络节点进入初始硬件洪泛的状态。本申请使用的“硬件洪泛”或“初始硬件洪泛”表示作为对缓存缺失的初始响应，硬件洪泛的动作。如上所述，缓存缺失被发送到网络装置处理单元（例如图 1 的处理单元 130) 用于处理。在通过处理单元处理缓存缺失的时间周期里，至少部分地发生初始硬件洪泛。一旦处理了缓存缺失（例如在硬件中已经针对 XYZ 组播组规划了新的条目），那么属于 XYZ 组的后续数据包就会“命中”缓存（即针对 XYZ 数据包的表查找会产生匹配）。缓存命中被接受（即它们将仅在对应的转送条目中定义的出站端口上被转送，与洪泛它们相反）。

[0031] 因此，当“命中”缓存的数据包的数量增加时，用于处理单元中缓存缺失的处理队列的尺寸缩小。初始硬件洪泛减小处理单元的处理负担，有助于（一个或多个）流收敛的加快。

[0032] 图 3 是流程图，示出用于组播收敛的处理。在 310，在网络装置接收组播数据包。在 320，网络装置至少部分地基于数据包的源 IP 地址和组的组播 IP 地址（即 (S,G) 地址组合）确定是否有针对数据包的缓存命中或缓存缺失。如果数据包导致缓存命中，则在 330，根据缓存条目中的转送信息转送数据包。如果数据包导致缓存缺失，则在 340，网络装置关于具有该 (S, G) 组合的流量进入初始硬件洪泛状态。硬件洪泛导致数据包根据 VLAN 成员资格转送到网络装置的出站端口。换言之，可以在网络装置的所有出站端口上转送数据包，

但是不一定是所有出站端口。

[0033] 除了洪泛数据包以外，在 350，还将导致缓存缺失的数据包的拷贝发送到网络装置上的处理单元。然后在 360，处理单元产生用于 (S，G) 组合的缓存条目。缓存缺失数据包的发送和处理可能在硬件洪泛开始之前、期间和 / 或之后发生。缓存条目允许具有相同的 (S,G) 组合的后续数据包在硬件中转送，不需要缓存缺失处理。如上所述，缓存缺失数据包的处理不包括慢通道转送，因为会在网络上产生重复的数据包。

[0034] 图 4 是方框图，示出现有技术以太网自动保护交换 (EAPS) 系统。EAPS 系统 400 包括一个或多个 EAPS 域 401。为了发送和接收 EAPS 系统控制消息 417，针对每个 EAPS 域 401 创建控制 VLAN 403。创建 EAPS 域 401 用于保护由一个或多个数据输送 VLAN 404 构成的群组。

[0035] EAPS 系统 400 在环形网络 402 上操作。环形网络 402 上的一个节点被指定为主节点 405。主节点 405 上的两个环形端口被指定为基本端口 406 和辅助端口 407。环形网络 402 上的所有其他节点都是传输节点 411，分别有自己的环形端口 412。每个主节点 405 和传输节点 411 分别具有转送数据库 (FDB)408 和 413，在数据库 408 和 413 中它们存储关于网络通信通道的信息。主节点 405 具有状态寄存器 409，用于存储环形网络 402 的状态。为了说明的目的，将环形网络 402 的状态或者描述为“失败”，表示在环形网络 402 中有故障或中断，或者描述为“完成”，表示环形网络没有终端，或者环形网络已经恢复，并且所有节点都正确地通信。传输节点 411 具有状态寄存器 414 和预转送计时器 415，传输节点在状态寄存器 414 中存储预转送状态。传输节点 411 还具有临时阻塞端口存储区 (TBP)416，传输节点在 TBP 中存储临时阻塞的端口的识别信息。

[0036] 主节点 405 和传输节点 411 利用控制消息 417 以经由控制 VLAN 403 通信。实施例中控制消息 417 的一些实例是健康检查消息、下行链接消息以及冲刷 FDB 消息。传输节点 411 将控制 VLAN 403 上发送的消息识别为控制消息 417，因为它具有与转送数据库 413 中的条目相对应的特殊 MAC(媒体访问控制) 地址。主节点和传输节点在将控制消息 417 拷贝到节点的中央处理单元 (CPU) 之前转送控制消息 417，在所述节点中，除了别的之外，它被记入日志以用于故障排除。在通过 CPU 处理之前转送控制消息 417 有助于故障之后以较之通过前述现有技术方法能实现的少得多的时间实现环形网络 402 的收敛。

[0037] 主节点 405 具有 hello 计时器 418，hello 计时器 418 是用于发送健康检查控制消息 417 的时钟。一旦 hello 计时器 418 启动，它就提示主节点 405 以定期的间隔 (例如每一秒钟) 在控制 VLAN 403 上发送健康检查消息 417。健康检查消息 417 环绕环形网络 402 转送，几乎立即返回主节点 405。当主节点 405 发送健康检查消息 417 时，它设定失败计时器 410。如果在健康检查消息返回主节点 405 之前失败计时器 410 到期，则主节点 405 确定在环形网络 402 中有故障。即使在故障期间也发送健康检查消息 417。当故障恢复时，主节点 405 立即知道，因为健康检查消息 417 的返回恢复。

[0038] 图 5 是方框图，示出根据各种实施例，采用以太网自动保护交换 (EAPS) 的网络装置。网络装置 510 包括一个或多个硬件表 520(例如缓存)、处理单元 530(例如中央处理单元 (CPU)、网络处理单元 (NPU) 等等) 以及数据包处理器 540。在一些实施例中，只要 EAPS 拓扑结构变化，EAPS 538 就向 MC 管理器 534 与其端口列表一起发送 VLAN 列表。MC 管理器 534 已具有已经在硬件提取层 (HAL)536 中规划的转送条目的列表和在它的基于 IGMP 报告

的软件表中的其他条目的列表。

[0039] 因此，当 MC 管理器 534 从 EAPS 538 接收 VLAN 列表和端口列表时，MC 管理器 534 将该信息与 IGMP 报告非常相似地来对待。换言之，对于已经在 HAL 536 中规划的转送条目，MC 管理器 534 会将 EAPS 出站端口（对应于特定组播流）添加到其缓存列表，并将更新的缓存列表发送到 HAL 536。区别在于通过 MC 管理器 534 接收 EAPS 出站端口列表，比 IGMP 报告快得多，因此显著减少（一个或多个）组播流的收敛时间。

[0040] 对于用于尚未针对已知组播流在 HAL 536 中规划的转送条目，MC 管理器 534 会将 EAPS 端口添加到它已有的组播组列表。对于不是在 HAL536 中的流，数据流仅当这些流到达网络装置时恢复，导致第 2 层 (L2) 缓存缺失并将 L2 缓存缺失发送到 MC 管理器 534。MC 管理器 534 会通过利用新的缓存条目更新 HAL 536 来处理缓存缺失。一旦新的缓存条目已经创建，用于（一个或多个）新流的数据流就会在硬件中交换。

[0041] 在另一实施例中，除了从 EAPS 538 接收端口列表（而不是等待 IGMP 报告）之外，MC 管理器 534 启动硬件洪泛状态并启动计时器。在一些实施例中计时器可以在 45 秒左右，但是根据网络的特定需要，它可以更短，也可以更长。作为启动洪泛状态的一部分，MC 管理器 534 规划 HAL536，以将相关 VLAN 设定为硬件洪泛模式。通过这种硬件洪泛状态 / 模式，如上所述，MC 管理器 534 接收 L2 缓存缺失，创建缓存条目并规划 HAL 536，并且随着硬件条目被填充，L2 缓存缺失将开始逐渐停止。但是，硬件（即数据包处理器 540 和（一个或多个）硬件表 520）会处于这样的状态 :它处于硬件洪泛模式（例如在特定 VLAN 上），但是同时，通过基于用于缓存命中的出站端口的转送，将接受硬件上的缓存命中。因此，对于那些在硬件中规划的条目，数据包将不会被洪泛或发送到处理单元 530。但是，对于所有其他有缓存缺失的组播数据包，这些数据包将在硬件中被洪泛，同时缓存缺失数据包的拷贝被发送到 CPU 530 以用于处理（即，获知和 / 或产生新的缓存条目，而不将数据包转回到网络）。

[0042] 一旦洪泛计时器到期，MC 管理器 534 就回复到标准 IGMP 侦听使能模式。

[0043] 应当注意，当 EAPS 538 向 MC 管理器 534 发送消息以启动硬件洪泛时，已有的缓存条目不应当被冲刷。否则，将产生另外的缓存缺失并发送到处理单元 530，导致进一步的处理压力。

[0044] 在涉及 EAPS 环的实施例中，有可能将没有接收器的出站端口添加到组播组列表。利用定期的 IGMP 老化，发送到没有接收器的端口的流量将最终移出，但是会比期望的更久。因此，可采用“快”移出。例如，如果正在使用 IGMP 询问，则在一定的时间周期里（例如 5-10 秒，或更短，或更长）应期望 IGMP 报告。用于接收报告的期望时间可用作“快”移出时间，以减少不必要流量的量。

[0045] 图 6 是流程图，示出 EAPS 环形网络中组播收敛的处理。在 610，检测 EAPS 网络中的拓扑结构变化。作为响应，在 620，至少一个节点清空其 IGMP 侦听表。在 630，发送代理 IGMP 询问以再获知组播转送条目。在 640，在某一时刻，在 640 检测 EAPS 网络上的缓存缺失。在 650，不是将缓存缺失数据包发送到处理单元以用于慢通道转送，而是在一个或多个出站端口经由硬件将数据包洪泛到网络。在 660，将缓存缺失数据包的拷贝发送到装置的处理单元，在 670，对于缓存缺失数据包在装置的处理单元创建硬件缓存条目。处理单元不会将缓存缺失数据包慢通道转送回网络。

[0046] 图 7 示出以计算机系统 700 的示例性形式，机器的示意性表示，在计算机系统 700

中可执行一组指令，用于使得机器进行本申请讨论的任意一个或多个方法。在替代实施例中，可将机器连接到（例如网络方式）局域网 (LAN) 中的其他机器、企业内部互联网、外部网络或者互联网。在客户机 - 服务器网络环境中，可在服务器或客户机的能力里操作机器，或者对等（或分布式）网络环境中将机器操作为个别机器。机器可以是个人计算机 (PC)、平板计算机、机顶盒、个人数字助理 (PDA)、蜂窝电话、或者能够执行一组指令（连续的或其他的）的任何机器，所述一组指令指定机器要采取的动作。此外，当仅示出单个机器时，也可以采用术语“机器”以包括机器（例如计算机）的任意集合，它们独立或联合地执行一组（或多组）指令，以进行本申请讨论的任意一个或多个方法。

[0047] 示例性计算机系统 700 包括处理器 702、内存 704（例如只读存储器 (ROM)、闪存、动态随机访问存储器 (DRAM) 例如同步 DRAM(SDRAM) 或 Rambus DRAM(RDRAM) 等等）、静态存储器 806（例如闪存、静态随机访问存储器 (SRAM) 等等）以及辅助存储器 818（例如数据存储装置），它们经由总线相互通信。

[0048] 处理器 702 表示一个或多个通用处理装置，例如微处理器、中央处理单元等等。具体而言，处理器 702 可以是复杂指令集计算 (CISC) 微处理器、精简指令集计算 (RISC) 微处理器、超长指令集架构 (VLIW) 微处理器、实现其他指令集的处理器或者实现指令集的组合的处理器。处理器 702 也可以是一个或多个专用处理装置，例如专用集成电路 (ASIC)、场可编程栅极阵列 (FPGA)、数字信号处理器 (DSP)、网络处理器等等。处理器 702 被配置为执行处理逻辑 126，处理逻辑 126 用于进行本申请讨论的操作和步骤。

[0049] 计算机系统 700 可进一步包括网络接口装置 716。计算机系统 700 还可包括视频显示单元 710（例如液晶显示器 (LCD) 或阴极射线管 (CRT)）、文字数字输入装置 712（例如键盘）以及指针控制装置（例如鼠标）。

[0050] 辅助存储器 718 可包括机器可读存储介质（或者具体而言计算机可读存储介质）724，上面存储一组或多组指令（例如软件 722），指令实现本申请所述的一种或多种方法或功能。在通过计算机系统 700 执行软件的过程中，软件 722 也可以完全或只是部分地置于内存 704 中和 / 或处理装置 702 中，内存 704 和处理装置 702 也构成机器可读存储介质。软件 722 可经由网络接口装置 716 进一步在网络 720 上传输或接收。

[0051] 虽然在示例性实施例中将机器可读存储介质 724 示出为单个介质，但是可以采用术语“机器可读存储介质”以包括存储一组或多组指令的单个介质或多个介质（例如收敛式数据库或分布式数据库，和 / 或相关联的缓存和服务器）。也可以采用术语“机器可读存储介质”以包括能够存储或编码一组指令的任意介质，所述一组指令用于通过机器执行，使得机器进行本发明的任意一个或多个方法。因此可以采用术语“机器可读存储介质”以包括，但不限于固态存储器、光学存储器或磁存储器。

[0052] 本申请描述了各种操作或功能，它们可实施或限定为软件代码或指令。这些内容可以是直接可执行的（“目标”或“可执行”形式）、源代码或差代码。可经由有代码或存储其上的指令的制造的物品提供本申请描述的实施例的软件实施方式，或经由操作通信接口的方法提供本申请描述的实施例的软件实施方式，以经由通信接口发送数据。机器可读介质或计算机可读介质可使得机器进行所述的功能或操作，并包括存储采用机器（例如计算装置、电子系统等等）可访问形式的信息的任意机构，例如可记录 / 不可记录介质（例如只读存储器 (ROM)、随机访问存储器 (RAM)、磁盘存储介质、光学存储介质、闪存装置等等）。通

信接口包括连接到任何硬连接的、无线的、光学的等等介质以与另一装置（例如存储器总线接口、处理器总线接口、互联网连接、磁盘控制器等等）通信的任意机构。可通过提供配置参数和/或发送信号来配置通信接口，以制备通信接口，提供描述软件内容的数据信号。经由发送到通信接口的一个或多个指令或信号可访问通信接口。

[0053] 本发明还涉及进行本申请操作的系统。该系统可特别构造用于要求的目的，或者可包括通用计算机，通过计算机中存储的计算机程序来选择性地启动或再配置通用计算机。可将这种计算机程序存储在计算机可读存储介质中，例如但不限于任意类型的盘，包括软盘、光盘、CDROM、磁光盘、只读存储器 (ROM)、随机访问存储器 (RAM)、EPROM、EEPROM、磁卡或光卡、或适合于存储电子指令的任意类型的介质，分别连接到计算机系统总线。

[0054] 本申请提供的方法和显示并非本来就涉及任何特定计算机或其他设备。通过根据本申请的教导的程序，可使用各种通用系统，或者可证明便于构造更专业的系统以进行所述方法的要求操作。这些系统的结构将出现，如同下面的描述所提出的。此外，本发明并非参照任何特定编程语言或操作系统来描述。应当理解，可使用各种编程语言来实现如同本申请所述的本发明的教导，并且可在多种操作系统里实现所述教导。

[0055] 本申请所述的各种部件可以是用于进行本申请所述功能的器件。本申请所述的每个部件包括软件、硬件或者它们的组合。本申请所述的操作和功能可实施为软件模块、硬件模块、专用硬件（例如专用硬件、专用集成电路 (ASIC)、数字信号处理器 (DSP) 等等）、内嵌控制器、硬连接电路等等。

[0056] 除了本申请所述之外，对本发明公开的实施例和实施方式可作出不脱离其范围的各种变型。因此，应当在示例性而不是限制性的意义上解释本申请的说明和实例。只能参照后附权利要求书来衡量本发明的范围。

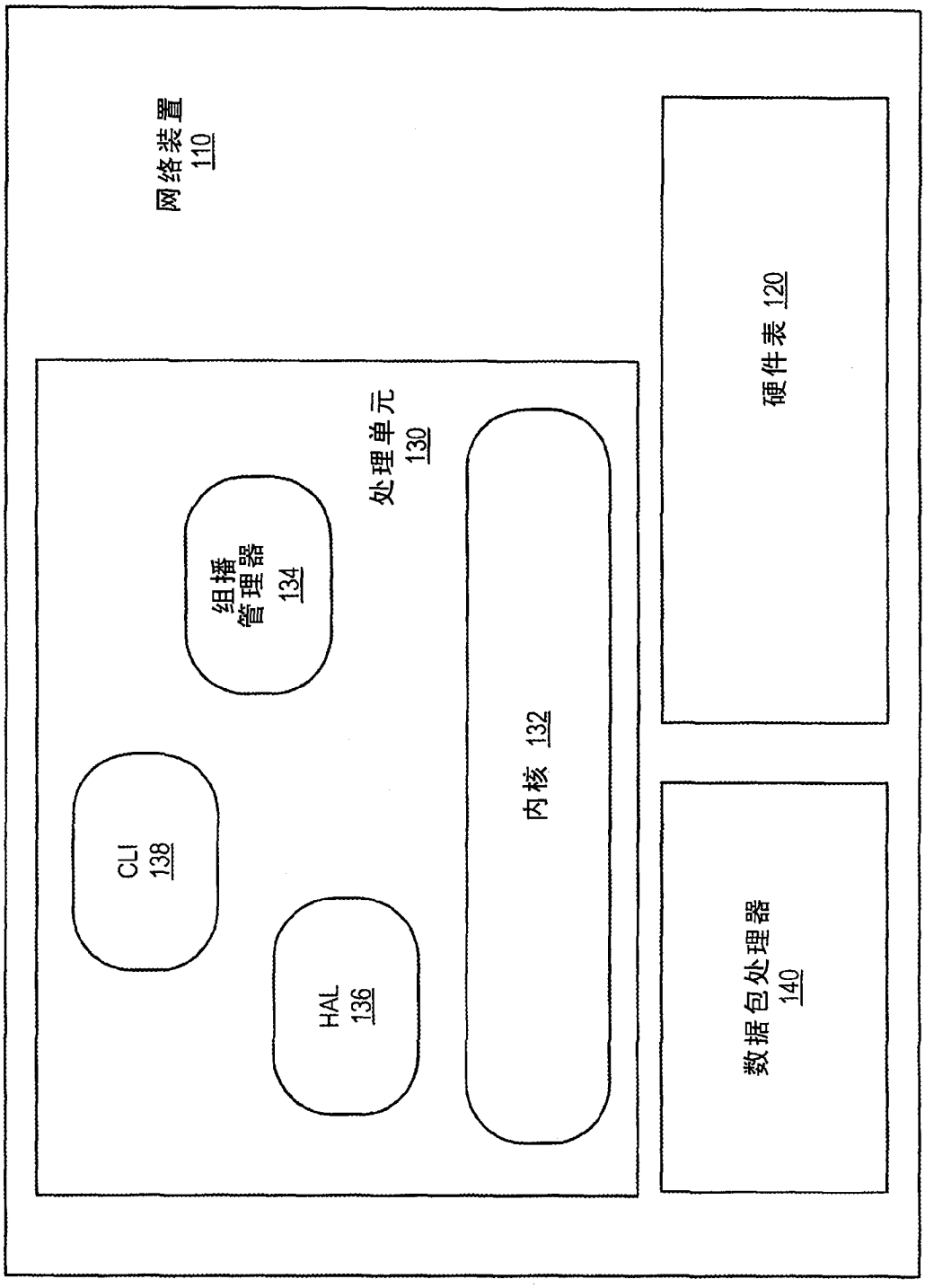

网络装置
110
处理单元
130
组播
管理器
134
CLI
138
HAL
136
内核 132
硬件表 120
数据包处理器
140

图 1

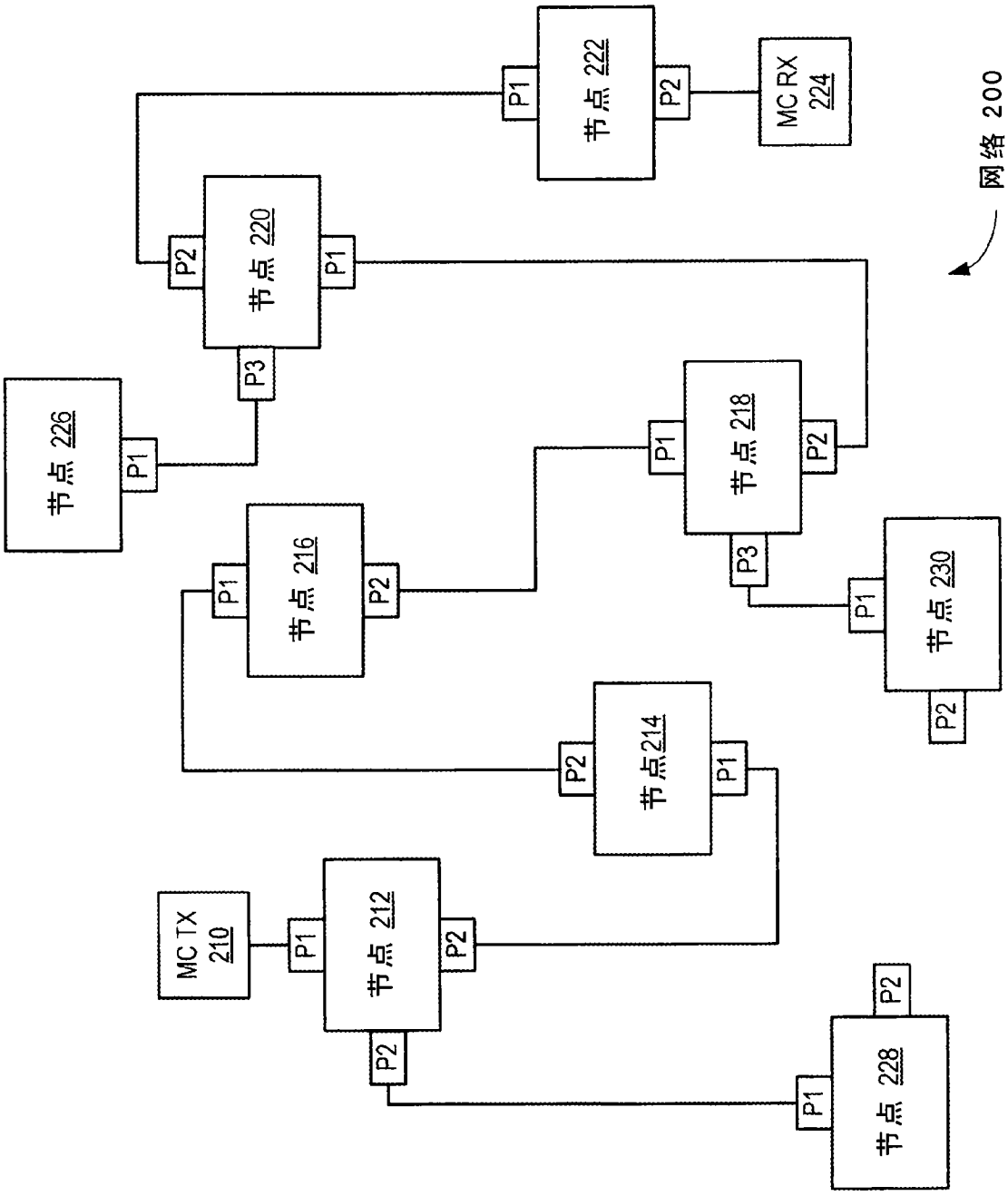

网络 200
节点 222
P1
P2
MC RX 224
节点 220
P2
P1
P3
节点 226
P1
节点 218
P1
P2
P3
节点 216
P1
P2
节点 230
P1
P2
节点214
P2
P1
MC TX 210
节点 212
P1
P2
P2
节点 228
P1
P2

图 2

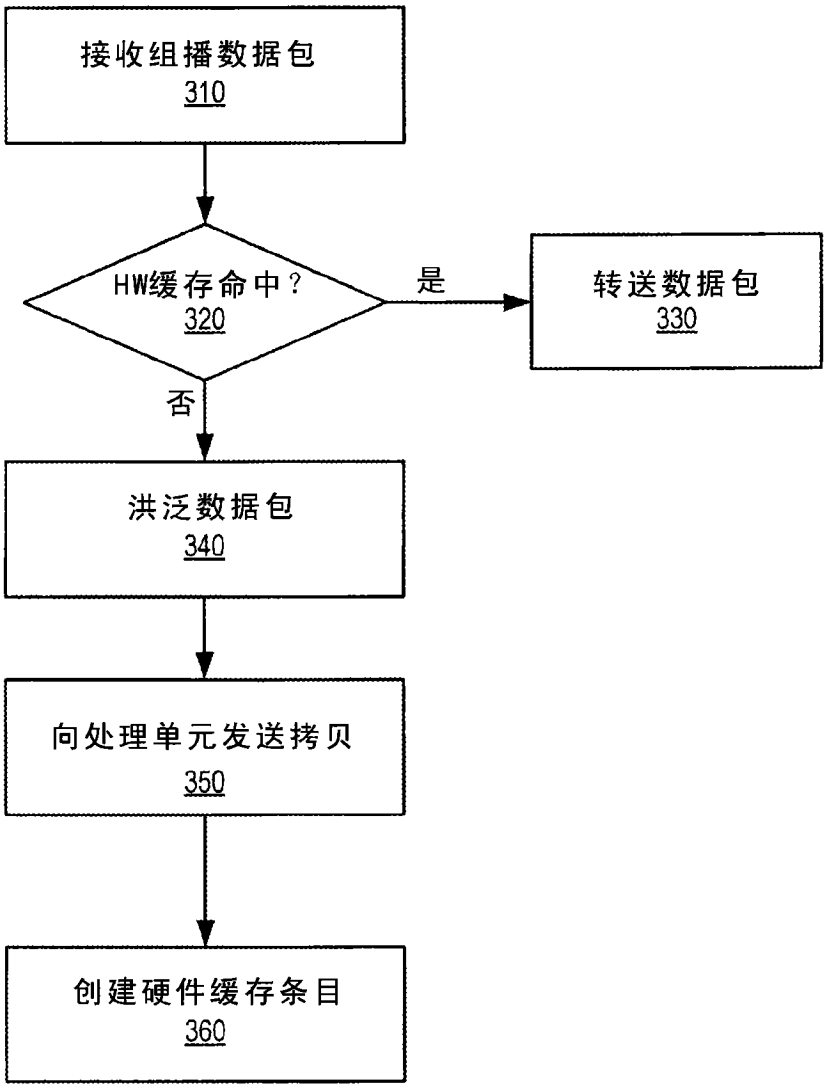

接收组播数据包
310
HW缓存命中？
320
是
转送数据包
330
否
洪泛数据包
340
向处理单元发送拷贝
350
创建硬件缓存条目
360

图 3

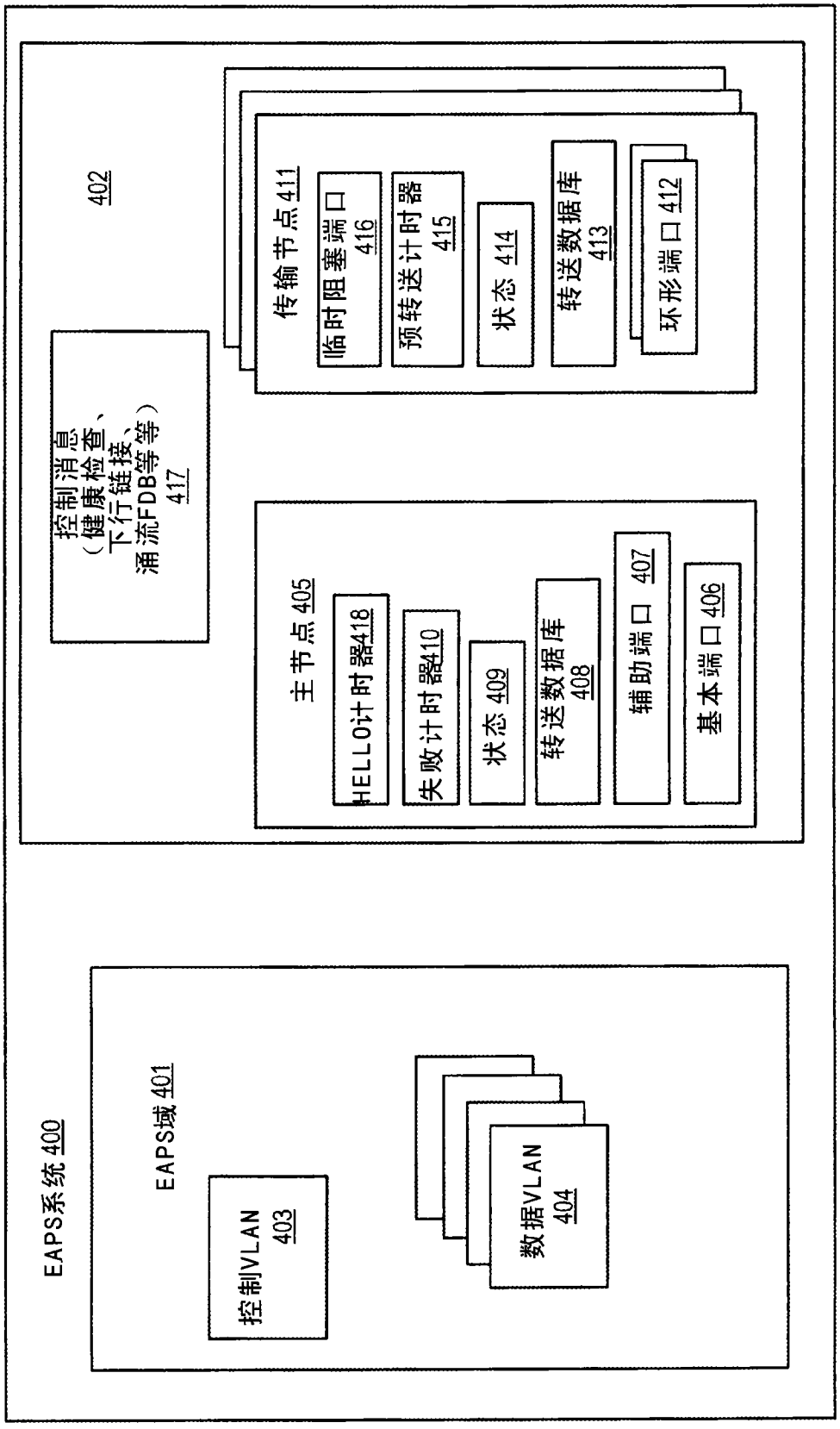

EAPS系统 400
EAPS域 401
控制VLAN 403
数据VLAN 404
402
控制消息（健康检查、下行链接、涌流FDB等等） 417
主节点 405
HELLO计时器418
失败计时器410
状态 409
转送数据库 408
辅助端口 407
基本端口 406
传输节点 411
临时阻塞端口 416
预转送计时器 415
状态 414
转送数据库 413
环形端口 412

图 4（现有技术）

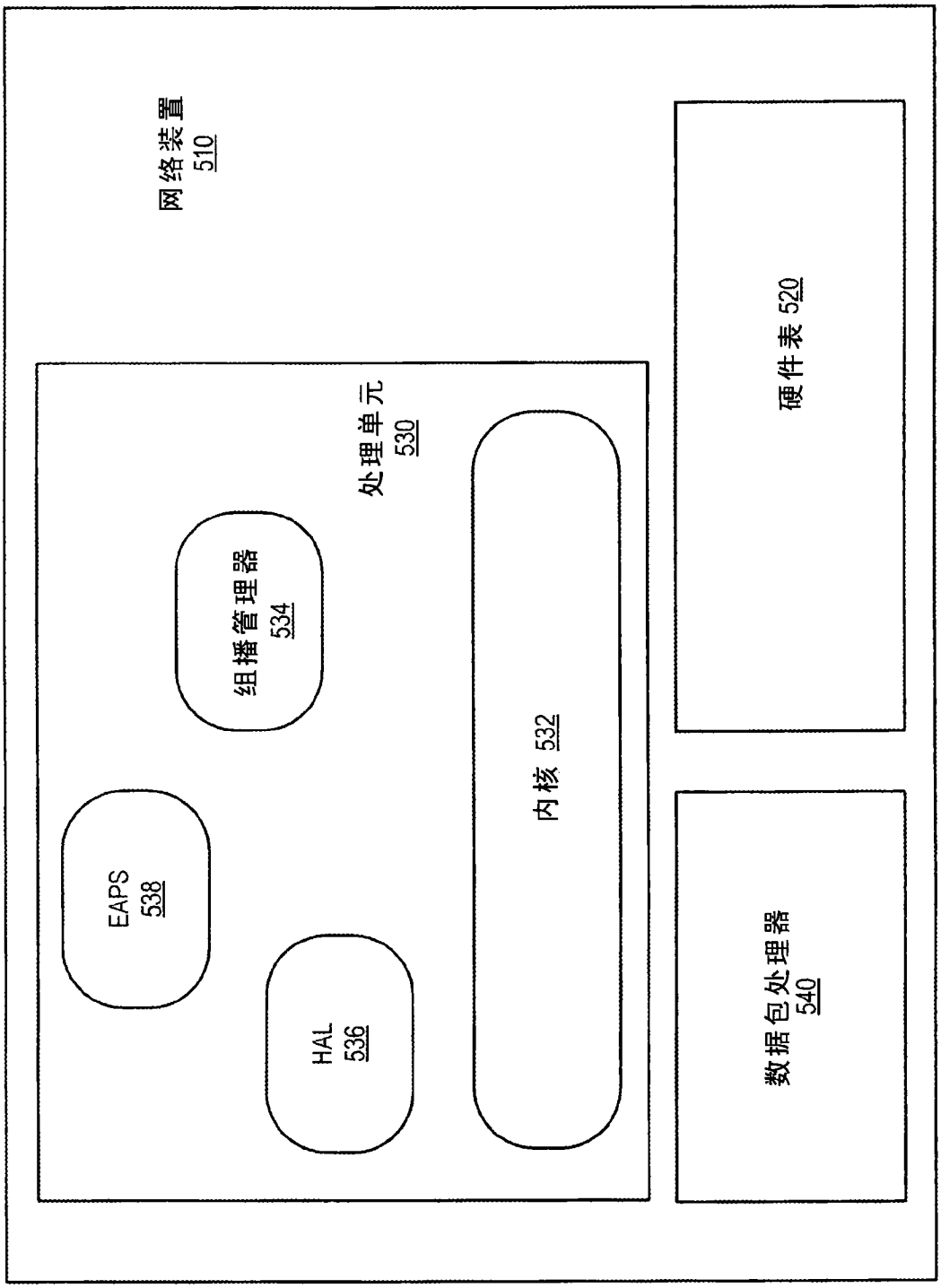

网络装置
510
处理单元
530
组播管理器
534
EAPS
538
HAL
536
内核 532
硬件表 520
数据包处理器
540

图 5

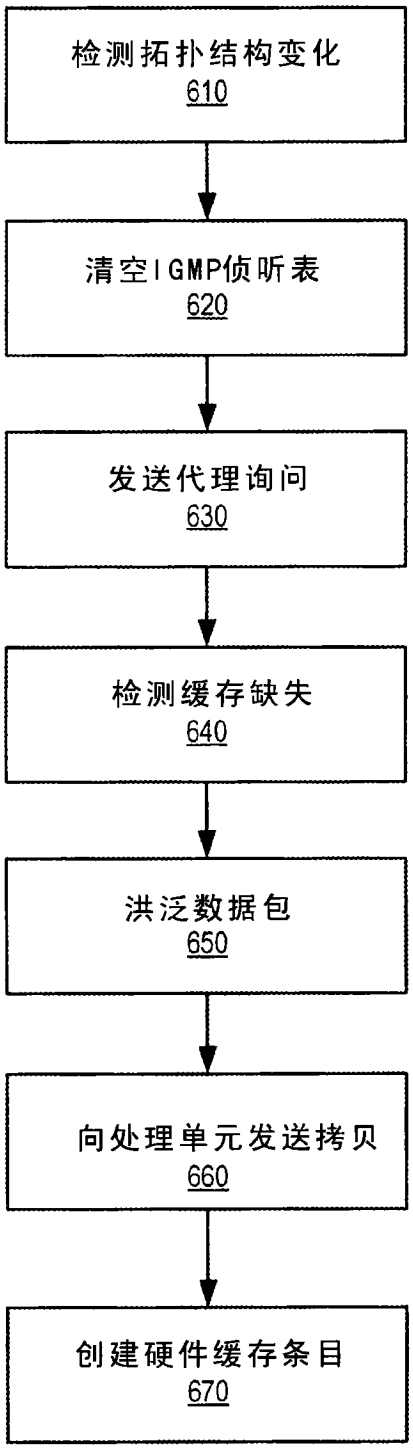

检测拓扑结构变化
610
清空IGMP侦听表
620
发送代理询问
630
检测缓存缺失
640
洪泛数据包
650
向处理单元发送拷贝
660
创建硬件缓存条目
670

图 6

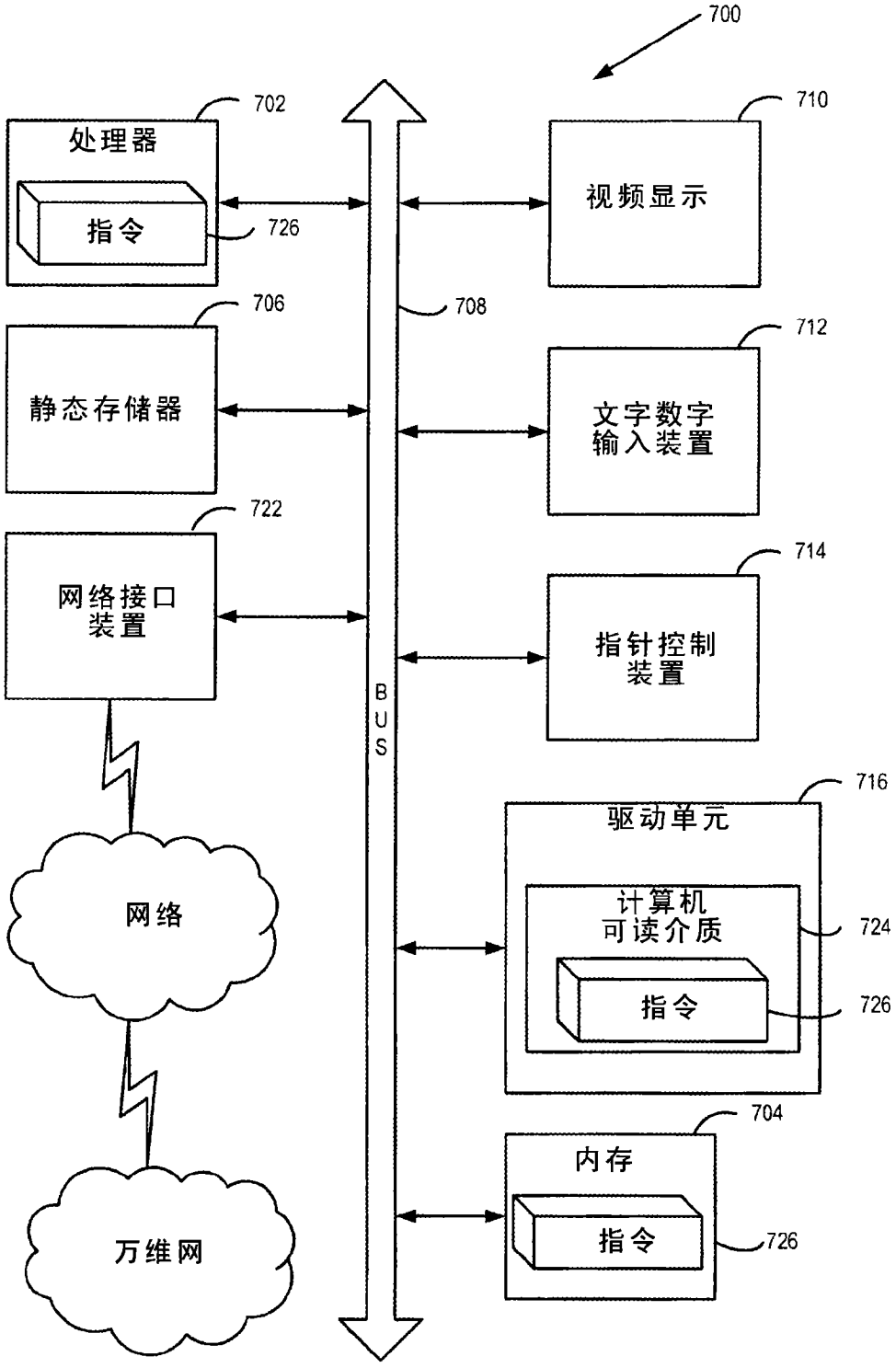

700
处理器
指令
726
702
710
视频显示
706
静态存储器
708
712
文字数字
输入装置
722
网络接口
装置
714
指针控制
装置
BUS
716
驱动单元
计算机
可读介质
724
指令
726
网络
704
内存
指令
726
万维网

图 7