(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 1890943 B
(45)授权公告日 2011.12.14

(21)申请号 200480036871.3

(22)申请日 2004.12.03

(30)优先权数据
10/742,562 2003.12.19 US

(85)PCT申请进入国家阶段日
2006.06.12

(86)PCT申请的申请数据
PCT/US2004/040386 2004.12.03

(87)PCT申请的公布数据
WO2005/062578 EN 2005.07.07

(73)专利权人 英特尔公司
地址 美国加利福尼亚州

(72)发明人 S·奥瓦德亚

(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 张政权

(51)Int.Cl.
*H04L 29/06*(2006.01)
*H04Q 11/00*(2006.01)

(56)对比文件
US 2003/0118053 A1,2003.06.26, 全文.
CN 1417968 A,2003.05.14, 全文.
US 2003/0198471 A1,2003.10.23, 说明书第0014至0039段.
Se-yoon Oh,etc..A Data Burst Assembly Algorithm in Optical Burst SwitchingNetworks.ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUE24 4.2002,24(4),312-315.
R.Rajaduray,etc..IMPACT OF BURST ASSEMBLY PARAMETERS ON EDGEROUTER LATENCY IN AN OPTICAL BURST SWITCHINGNETWORK.ANNUAL MEETING OF THE IEEE LASER & ELECTRO-OPTICS SOCIETY1.2003,155-56.

审查员 寇利敏

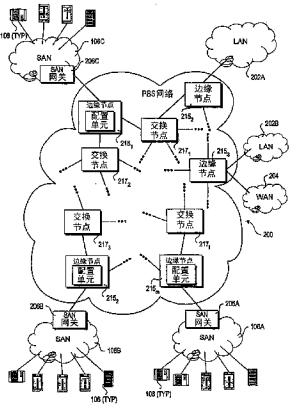
权利要求书 4 页 说明书 24 页 附图 20 页

(54)发明名称

用于服务器和存储区网络之间的光学连网的方法和架构

(57)摘要

一种用于经由光脉冲串交换(OBS)网络与SAN(存储区域网络和服务器区域网络)路由高速数据的方法和系统。包括边缘节点和交换节点的OBS网络组件耦合于SAN岛之间。在一个实施例中,OBS网络包括光子脉冲串交换(PBS)网络。在一种方案下,PBS边缘节点和SAN网关共同定位于与SAN的接口处,同时多个PBS交换节点部署于PBS边缘节点之间。在另一方案下,PBS交换/边缘节点共同定位于各自的SAN处。该方案采用用于经由选定的路径段路由数据的外部网关协议(EGP)。到达SAN和从SAN接收的数据被封装为光纤信道帧。经由PBS网络传送的数据被转换成具有封装的光纤信道帧的PBS帧。该方案还支持与诸如LAN和WAN的传统网络的接口。




CN 1890943 B

1. 一种用于在多个存储区域网络和 / 或服务器区域网络 SAN 之间传递数据的方法，包括：

经由光脉冲串交换 OBS 网络基础结构将第一 SAN 耦合到第二 SAN；

从所述第一 SAN 接收数据，所述数据根据第一 SAN 格式被配置；

将所述数据封装入一个或多个 OBS 数据脉冲串；

经过所述 OBS 网络将所述一个或多个 OBS 数据脉冲串从所述第一 SAN 发送到第二 SAN；以及

在所述第二 SAN 处提取所封装的数据，

其中，所述 OBS 网络包括光子脉冲串交换 PBS 网络，以及

其中，经由 OBS 网络基础结构将第一 SAN 耦合到第二 SAN 包括：

在用于第一和第二 SAN 中的每一个的各自 SAN 网关处共同定位各自 PBS 边缘节点模块；以及

将各自 PBS 边缘节点模块光耦合到至少一个 PBS 交换节点以形成一光通路，所述光通路包括在所述第一和第二 SAN 之间的至少两个光通路段。

2. 如权利要求 1 所述的方法，其特征在于，所述 OBS 网络包括波分复用 WDMPBS 网络。

3. 如权利要求 1 所述的方法，其特征在于，所述数据的第一 SAN 格式包括光纤信道 FC 帧，且将所述数据封装入一个或多个 OBS 数据脉冲串包括在所述一个或多个 OBS 数据脉冲串中的每一个中封装至少一个 FC 帧。

4. 如权利要求 1 所述的方法，其特征在于，由 SAN 网关和 PBS 边缘节点执行的各自操作由多模块单元中所包含的至少一个模块提供。

5. 如权利要求 4 所述的方法，其特征在于，所述至少一个模块包括经由模块化可再配置通信平台中的共用底板耦合的多个服务器刀片。

6. 如权利要求 1 所述的方法，其特征在于，还包括：

在用于第一和第二 SAN 的至少一个的各自 SAN 网关处共同定位至少一个外部网关协议路由器模块；以及

使用外部网关协议路由器确定一路径，其中所述一个或多个 OBS 数据脉冲串经由所述路径在第一和第二 SAN 之间传送。

7. 如权利要求 4 所述的方法，其特征在于，所述至少一个模块包括经由服务器刀片单元中的共用底板耦合的多个服务器刀片。

8. 如权利要求 1 所述的方法，其特征在于，还包括：

将第三和第四 SAN 光耦合到第一和第二 SAN，第三和第四 SAN 中的每一个在共同定位的 SAN 网关和 PBS 边缘节点处被光耦合；以及

确定一路径以光传送一个或多个光脉冲串，所述路径包括至少两个路径段的串联，每个路径段在各自 SAN 对之间耦合。

9. 如权利要求 8 所述的方法，其特征在于，还包括将共同定位的 PBS 边缘节点中的至少一个配置为外部网关协议路由器；以及

使用外部网关协议路由器确定所述路径。

10. 如权利要求 9 所述的方法，其特征在于，所述外部网关协议包括对包含 PBS 网络的规定的边界网关协议的扩展，且确定所述路径包括确定要用于每个路径段的光波长。

11. 一种使用光交换网络进行通信的系统，包括：

多个存储区域网络和服务器区域网络 SAN，它们每一个都包括各自的 SAN 网关；

多个光脉冲串交换 OBS 网络边缘节点，它们每一个与各自的 SAN 网关通信耦合；以及

多个 OBS 网络交换节点，它们按网状配置光耦合到 OBS 边缘节点的多个 OBS，

其中，所述 OBS 网络包括光子脉冲串交换 PBS 网络，OBS 网络交换节点包括 PBS 交换节点，且 OBS 边缘节点包括 PBS 边缘节点，

其中，PBS 边缘节点的至少一个包括与相应 SAN 网关共同定位的模块。

12. 如权利要求 11 所述的系统，其特征在于，还包括：

与多个 OBS 网络交换节点中的至少一个耦合的一个 OBS 网络边缘节点；

到包括局域网 LAN 和广域网 WAN 之一的传统网络的一个接口，它与 OBS 网络边缘节点通信耦合。

13. 如权利要求 11 所述的系统，其特征在于，每个 OBS 网络都包括波分复用 WDMPBS 网络。

14. 如权利要求 11 所述的系统，其特征在于，PBS 边缘节点的至少一个包括与相应的边缘网关协议路由器模块共同定位的模块。

15. 一种使用光交换网络进行通信的系统，包括：

多个存储区域网络和服务器区域网络 SAN，它们每一个都包括各自的 SAN 网关；以及

多个光脉冲串交换 OBS 网络交换和 / 或边缘节点，它们每一个都与各自的 SAN 网关通信耦合并光耦合到至少一个其它 OBS 网络交换和 / 或边缘节点，

其中，所述多个 OBS 交换和 / 或边缘节点的至少一个在各自的 SAN 网关处被共同定位。

16. 如权利要求 15 所述的系统，其特征在于，所述光脉冲串交换网络包括光子脉冲串交换 PBS 网络。

17. 如权利要求 16 所述的系统，其特征在于，所述光脉冲串交换网络包括波分复用 WDMPBS 网络。

18. 一种使用光交换网络进行通信的系统，包括：

包括具有布线通信信道的交换织构的底板，所述布线通信信道用于提供传播信息的媒介；

存储区域网络 SAN 网关模块，耦合在所述底板上用于提供与耦合到所述布线通信信道的 SAN 的接口以便在所述布线通信信道上传送数据；以及

耦合在所述底板上并与所述布线通信信道耦合的光输入 / 输出 I/O 模块，所述光 I/O 模块包括耦合到光脉冲串交换 OBS 网络的端口，所述光 I/O 模块从 SAN 网关接收 SAN 格式的数据，在一个或多个数据脉冲串内封装所述数据，并在所述 OBS 网络上光传送所述一个或多个数据脉冲串。

19. 如权利要求 18 所述的系统，其特征在于，还包括：

经由底板与所述布线通信信道耦合的传统接口模块，所述传统接口模块包括耦合到包括局域网 LAN 或广域网 WAN 之一的传统网络的端口。

20. 如权利要求 18 所述的系统，其特征在于，所述光脉冲串交换网络包括光子脉冲串交换 PBS 网络。

21. 如权利要求 20 所述的系统，其特征在于，所述 I/O 模块被配置成提供实现 PBS 边缘

节点的操作。

22. 如权利要求 20 所述的系统，其特征在于，所述 I/O 模块被配置成提供操作以实现 PBS 交换节点和 PBS 边缘节点的组合。

23. 如权利要求 22 所述的系统，其特征在于，所述 I/O 模块被配置成提供实现外部网关协议路由的操作。

24. 如权利要求 22 所述的系统，其特征在于，所述外部网关协议路由包括对边界网关协议的扩展，所述边界网关协议包括对波分复用 WDM 光脉冲串交换网络的规定。

25. 如权利要求 18 所述的系统，其特征在于，所述光 I/O 模块包括：

与所述布线通信信道耦合的总线桥路，以便从所述布线通信信道接收分组；

与所述总线桥路耦合的网络处理器单元，所述网络处理器单元将经由所述总线桥路接收的分组聚集成一脉冲串；

与所述网络处理器单元耦合的成帧器单元，其中所述成帧器用于将所述数据脉冲串封装入光脉冲串交换网络帧；以及

与所述成帧器单元和光学网络耦合的光输出接口，其中光输出接口用于在光脉冲串交换网络上发送包括所述光网络帧的光信号。

26. 如权利要求 25 所述的系统，其特征在于，还包括与所述总线桥路和网络处理单元耦合的通信整形器。

27. 如权利要求 25 所述的系统，其特征在于，还包括与所述成帧器单元和网络处理单元耦合的队列单元，其中所述队列单元用于存储数据脉冲串直到它们被调度为在光脉冲串交换网络上被发送。

28. 如权利要求 25 所述的系统，其特征在于，所述网络处理器单元形成控制脉冲串和数据脉冲串，其中所述控制脉冲串包括用于路由数据脉冲串通过光脉冲串交换网络的信息。

29. 如权利要求 25 所述的系统，其特征在于，还包括与所述成帧器单元和光脉冲串交换网络耦合的光输入接口，其中所述光输入接口用于将包括从光学网络接收的光脉冲串交换网络帧的光脉冲串信号转换成包括光脉冲串交换网络帧信息的电信号。

30. 如权利要求 25 所述的系统，其特征在于，所述成帧器单元也用于将来自光输入接口的光脉冲串交换网络帧信息去成帧。

31. 如权利要求 18 所述的系统，其特征在于，所述系统根据 PCI 工业计算机制造组 PICMG3.0 所定义的高级电信计算架构 ATCA 标准或之后的 ATCA 规范被配置。

32. 一种由光输入 / 输出 I/O 模块执行的方法，包括：

从存储区域网络和 / 或服务器区域网络 SAN 网关接收多个光纤信道帧；

将所述多个光纤信道帧封装入一个或多个光脉冲串交换 OBS 网络数据脉冲串；以及

将所述一个或多个 OBS 网络数据脉冲串发送到 OBS 边缘节点或交换节点之一，

其中，所述 OBS 边缘节点和所述 SAN 网关共同定位在单个单元中。

33. 如权利要求 32 所述的方法，其特征在于，所述光脉冲串交换网络包括光子脉冲串交换 PBS 网络。

34. 如权利要求 32 所述的方法，其特征在于，还包括：

经由 OBS 网络接收 OBS 数据脉冲串帧；

将所述 OBS 数据脉冲串帧去成帧以提取一个或多个封装的 FC 帧 ;以及

将所述 FC 帧提供给 SAN 网关。

35. 如权利要求 32 所述的方法,其特征在于,还包括 :

从一 FC 帧中提取路由数据,所述路由数据标识所述数据要路由到的目的地地址 ;

从一路由表中存储的路径中选择可用于达到所述目的地地址的一路径 ;以及

将其中封装所述 FC 帧的数据脉冲串转发到所选路径中的下一个中继段。

36. 如权利要求 35 所述的方法,其特征在于,光 I/O 模块包括接收数据的入口节点,且所述数据将被转发到 OBS 网络的出口节点,所述方法还包括 :

保留在入口节点和出口节点之间跨越的光通路 ;以及

在被保留的所述光通路上发送数据脉冲串。

# 用于服务器和存储区网络之间的光学连网的方法和架构

[0001] 相关申请的对照

[0002] 本申请涉及2002年4月17日提交的美国专利申请No.10/126091;2002年6月25日提交的美国专利申请No.10/183111;2002年12月24日提交的美国专利申请No.10/328571;2003年2月28日提交的美国专利申请No.10/377312;2003年2月28日提交的美国专利申请No.10/377580;2003年4月16日提交的美国专利申请No.10/417823;2003年4月17日提交的美国专利申请No.10/417487;2003年5月19日提交的美国专利申请No.(代理人案卷号42P16183),2003年6月18日提交的美国专利申请No.(代理人案卷号42P16552),2003年6月14日提交的美国专利申请No.(代理人案卷号42P16847),以及2003年8月6日提交的美国专利申请No.(代理人案卷号42P17373)。

## 技术领域

[0003] 本发明的领域一般涉及存储和/或服务器区域网络(SAN),且尤其涉及用于使用光交换网络在SAN之间传输数据的技术。

## 背景技术

[0004] 商业活动所产生和采集的数据量近些年呈指数增长,且这种增长被预计会在将来持续下去。数据是商业计算处理所依据的基础资源。为确保商业处理递交期望的结果,就必须访问数据。商业数据的管理和保护对于商业处理的可用性来说是至关重要的。管理覆盖诸如配置、执行和保护的多个方面,其范围从如果存储介质故障该怎么办到完成灾难恢复过程。

[0005] 在大型机的环境中,集中了存储管理。存储装置连接到大型机主机并直接由其中系统程序员(存储管理员)完全专注该任务的IT部门管理。按此方式管理存储是相对直接和容易的。

[0006] 客户机/服务器计算的出现产生了新的一组问题,诸如对台式机的管理成本迅速上升,以及新的存储管理问题。大型机环境中被集中的信息现在分散于一个或多个网络上并常被不佳地管理和控制。存储装置被分散并连接到各个机器;必须逐机器地安排容量增加;一个操作系统平台所获得的存储常不能用于其它平台上。

[0007] 数十年来,计算产业已认识到表达、处理和数据存储之间的割裂。客户机/服务器架构基于这三分层模型。顶层将台式机用于数据表达。台式机通常基于个人计算机(PC)。包括应用服务器的中间层进行处理。诸如电子邮件服务器或web服务器的应用服务器由台式机访问并使用底层上存储的数据,底层包括包含数据的存储装置。

[0008] 为解决以上问题,关于存储区域网络和服务器区域网络(这里都称作“SAN”)连网的技术和存储解决方案已且正在开发。SAN是允许网络连接基础结构所支持的距离内建立在存储装置和处理器(服务器)之间的直接连接的高速网络,所述网络连接基础结构最普通地包括光纤信道(FC)基础结构。在现今的SAN环境中,底层中的存储装置被集中和互连,它实际上表示后退到主机或大型机的中央存储模型。

[0009] SAN 可被认为是对存储总线概念的扩展，它使存储装置和服务器能使用与局域网 (LAN) 和广域网 (WAN) 中类似的元件 ：路由器、集线器、交换机、导向器和网关进行互连。SAN 可在服务器之间共享和 / 或专用于一个服务器。它可以支持同类（即，共同平台）和异类（混合平台）架构两者。

[0010] 图 1 中示出了一对异类 SAN 架构 100A 和 100B 的示例。每个架构都根据上述常规三层架构配置，包括客户机层、应用服务器层和存储层。客户机层包括各种类型的客户机计算机 102，诸如工作站、个人计算机、膝上计算机等。客户机层中的客户机计算机经由 LAN（局域网）或 WAN（广域网）106（对各自架构 100A 和 100B 中标记为 106A 和 106B) 连接到应用服务器层中的服务器 104。而服务器层中的服务器 104 经由各自的 SAN110A 和 110B 连接到存储层中的存储装置 108。

[0011] 异类架构支持各种服务器硬件和平台类型，并独立于平台分销商和操作系统类型。存储层 106 中的存储装置 108 用于存储可经由 SAN110A 和 110B 被访问的数据。一般，大多数类型的大容量存储装置可部署于 SAN 存储层中，只要该装置与 SAN 架构兼容。

[0012] 商业实体与更大企业的合并一般导致其中表示存储岛的各个 SAN 相互隔离的出现。为便于不同 SAN 之间的连续通信，必须采用有效的传输机制。在一种常规方案下，传输机制使用具有 IP（因特网协议）的以太网接口和交换机来做到，诸如图 1 所示的。为了在 SAN110A 和 SAN110B 之间接口，在 IP 网络 114 之间使用 SAN 网关 112A 和 112B。SAN 网关便于根据具体协议的数据再配置，以帮助跨网关的数据交换。

[0013] 虽然 SAN 一般被认为是高效网络，但 SAN 上发送的通信远不同于设计 IP 网络来处理的通信。IP 网络以路由为基础，并通常服务大量客户并且可以包括几百或甚至几千的路由器、交换机、网桥等。在 IP 协议下，通过将数据封装入包括首部的相对较小的分组来发送数据，所述首部在沿着数据源和数据目的地之间（诸如图 1 的 SAN110A 和 110B 之间）的路径的每个路由中继段 (hop) 处被检查。这包括大量开销。相反，SAN 通信通常包括很短的路径上发送的较大的有效负荷，通常是点对点。因此，SAN 被设计用于处理块通信，其中路由考虑是次要的。当使用 IP 网络在 SAN 之间发送数据时，这些较大的有效负荷必须在源 SAN 网关处被分成许多小得多的分组，分别地在 IP 网络上常沿着不同的路径被发送，并在目的地 SAN 网关处被再汇集。结果，使用常规传输机制（诸如 IP 网络）的经由 SAN 的数据传输是非常低效的且消耗有价值的带宽和网络资源。

## 附图说明

[0014] 本发明的前述各方面和许多优点将变得更易于理解，因为它们通过以下详细描述并结合附图将得到更好的理解，其中除非另外指明否则相同的标号贯穿各附图表示相同的部分 ：

[0015] 图 1 是说明典型存储区域网络 (SAN) 的组件和使用 IP 网络在 SAN 岛之间发送通信的常规技术的示意图。

[0016] 图 2 是示出根据本发明一个实施例的具有可变时隙的光子脉冲串交换 (PBS) 网络的简化框图，该网络连接到多个 SAN 和 LAN 网络。

[0017] 图 3 是示出根据本发明一个实施例的光子脉冲串交换 (PBS) 网络的操作的简化流程图。

[0018] 图 4 是示出根据本发明一个实施例的光子脉冲串交换 (PBS) 网络中使用的交换节点模块的框图。

[0019] 图 5 是示出根据本发明一个实施例的交换节点模块的操作的流程图。

[0020] 图 6 是示出根据本发明一个实施例的 PBS 网络中节点之间的 PBS 光脉冲串流的示图。

[0021] 图 7 是示出根据本发明一个实施例的 PBS 光脉冲串的一般 PBS 成帧格式的示图。

[0022] 图 8 是示出根据本发明一个实施例的图 7 的 PBS 成帧格式的进一步细节的示图。

[0023] 图 9a 是根据本发明一个实施例的网络架构的示意图，在该网络架构下用包括边缘节点处共同定位的 PBS 接口和 SAN 网关的 PBS 网络组件连网多个 SAN。

[0024] 图 9b 是根据本发明一个实施例的网络架构的示意图，在该网络架构下用包括用作边界网关协议 (BGP) 路由器的共同定位的 PBS 交换 / 边缘节点的 PBS 网络组件连网多个 SAN。

[0025] 图 9c 是从 BGP 路由器看的图 9b 的网络架构的示意图。

[0026] 图 10 是示出光纤信道如何被结构化为分层功能的分层集的示图。

[0027] 图 11 是示出光纤信道帧 (FC-2) 的格式的示图。

[0028] 图 12 是示出可以封装一个或多个光纤信道帧的 PBS 成帧格式的细节的示图。

[0029] 图 13 是示出根据本发明一个实施例的共同定位的 SAN 网关 /PBS 边缘节点单元的示图。

[0030] 图 14a 是示出根据本发明一个实施例的图 13 中描述的光学 PBS I/O 卡的框图。

[0031] 图 14b 是更详细地示出根据本发明一个实施例的图 17a 中描述的网络处理器单元和排列单元的框图。

[0032] 图 15 是示出根据本发明一个实施例的出口操作流程的流程图。

[0033] 图 16 是示出根据本发明一个实施例的出口操作流程的流程图。

[0034] 图 17 是示出 BGP 更新消息中的各种字段的示图。

[0035] 图 17a 是示出与常规 BGP 更新消息的通路属性相对应的各种字段的示图。

[0036] 图 17b 是示出根据本发明一个实施例的附加字段的示图，这些附加字段被添加到图 17a 的 BGP 更新消息的通路属性，它们使得外部路由能扩展到光脉冲串交换网络。

[0037] 图 18 是示出用于配置和初始化 PBS 网络以使能与该 PBS 网络耦合的多个 SAN 之间的基于 PBS 的数据传输的操作的流程图。

## 具体实施方式

[0038] 这里将描述使能使用光交换网络在 SAN 之间传输数据的技术的实施例。在以下描述中，阐述了许多特定细节，诸如对被实现用于光子脉冲串交换 (PBS) 网络的实施例的描述，以提供对本发明实施例的透彻理解。但相关领域的熟练技术人员将认识到，本发明可以在没有一个或多个这些特定细节或者用其它方法、组件、材料等的情况下实施。在其它实例中，未详细示出或描述公知的结构、材料或操作以使本发明的各方面更清晰。

[0039] 贯穿说明书对“一个实施例”或“一实施例”的引用表示联系该实施例描述的特殊特点、结构或特征包含于本发明的至少一个实施例中。因此，在整个说明书各处短语“在一个实施例中”或“在一实施例中”的出现不必都表示同一实施例。此外，在一个或多个实施

例中，特殊的特点、结构或特征可按任何合适的方式组合。

[0040] 根据这里描述的本发明的各方面，通过光交换网络便于在两个或更多不同 SAN 之间以及任选的其它传统网络类型（包括 LAN 和 WAN）的网络间通信。在以下的详细描述中，参考它们在光子脉冲串交换 (PBS) 网络中的使用来揭示本发明的实施例。PBS 网络是一种类型的光交换网络，通常包括高速中继段和跨度约束网络，诸如企业网。术语“光子脉冲串”这里用于表示具有类似路由要求的统计多路复用的分组（例如，因特网协议 (IP) 分组、以太网帧、光纤信道 (FC) 帧）。虽然概念上类似于基于主干的光学脉冲串交换 (OBS) 网络，但这些高速中继段和跨度约束网络的设计、操作约束和性能要求可以是不同的。但，可以理解，这里揭示的教示和原理也可应用于其它类型的光交换网络。

[0041] 常规光交换网络通常使用波长路由技术，该技术需要在光交换节点处进行光学信号的光 - 电 - 光 (O-E-O) 转换。光学网络中每个交换节点处的 O-E-O 转换不仅是很慢的操作（通常约 10 毫秒），而且是很昂贵且功耗大的操作，它潜在地形成了光交换网络的通信瓶颈。此外，当前的光交换技术不能有效地支持分组通信应用（例如因特网）中常发生的“猝发”通信。

[0042] 图 2 示出了示例性架构，在该架构下采用 PBS 网络 200 以便于 SAN106A、106B 和 106C，LAN202A 和 202B，以及 WAN204 之间的网络间通信。PBS 网络 200 包括多个节点，包括边缘节点 $215_1$-$215_M$ 以及交换节点 $217_1$-$217_L$。PBS 网络 200 可进一步包括与图 2 所示的交换节点互连的附加的边缘和交换节点（未示出）。在所示实施例中，边缘节点同时用作入口和出口节点。在任选的配置中，入口和出口节点可包括分开的节点。因此，以下分开描述入口和出口节点功能；可以理解，对入口或出口节点的参考也可应用于边缘节点。实际上，边缘节点提供“外部”网络（即 PBS 网络以外；图 2 实施例中的 SAN 106A-C，LAN202A 和 202B，以及 WAN204）和 PBS 网络的交换节点之间的接口。在该实施例中，用智能模块实现入口、出口和交换节点功能。

[0043] 在一些实施例中，入口节点执行接收光信号的光 - 电 (O-E) 转换，并包括电子存储器以缓冲接收信号直到它们被发送给合适的外部网络。此外，在一些实施例中，入口节点也在接收的电信号被发送给 PBS 网络 200 的交换节点 $217_1$-$217_L$ 前执行该电信号的电 - 光 (E-O) 转换。

[0044] 出口节点用光交换单元或模块实现，它们可配置成从 PBS 网络 200 的其它节点接收光信号并随后将它们路由到外部网络。出口节点也可从外部网络接收光信号并将它们发送到 PBS 网络 200 内的合适目的地节点，因此用作入口节点。在一个实施例中，出口节点执行所接收的光信号的 O-E-O 转换，并包括电子存储器来缓冲接收到的信号直到它们被发送给 PBS 网络 200 的合适节点。入口和出口节点也可从电域中实现的一个网络链接（例如，有线的以太网链接等）接收信号并将信号发送出去。

[0045] 交换节点 $217_1$-$217_L$ 用光交换单元或模块实现，它们各自被配置成从其它交换节点接收光信号并适当地将接收到的光信号路由到 PBS 网络 200 的其它交换和边缘节点。如下所述，交换节点执行光控制脉冲串和网络管理控制脉冲串信号的 O-E-O 转换。在一些实施例中，这些光控制脉冲串和网络管理控制脉冲串仅在预选的波长上传播。在这种实施例中，这些预选的波长不传播光“数据”脉冲串（与控制脉冲串和网络管理控制脉冲串相反）信号，尽管控制脉冲串和网络管理控制脉冲串可包括用于特殊一组光数据脉冲串信号的必

要信息。在一些实施例中，控制和数据信息在分开的波长上发送（这里也称作带外 (OOB) 信号发送）。在其它实施例中，控制和数据信息可在相同的波长上发送（这里也称作带内 (IB) 信号发送）。在另一实施例中，光控制脉冲串、网络管理控制脉冲串和光数据脉冲串信号使用不同的编码方案（诸如不同的调制格式等）在相同波长上传播。

[0046] 虽然交换节点 $217_1$-$217_L$ 可执行光控制信号的 O-E-O 转换，但在此实施例中交换节点不执行光数据脉冲串信号的 O-E-O 转换。相反，交换节点 $217_1$-$217_L$ 纯粹地执行光数据脉冲串信号的光交换。因此，交换节点可包括电子电路以存储和处理被转换为电子形式的进入的光控制脉冲串和网络管理控制脉冲串并使用该信息来配置光子脉冲串交换设定，并正确地路由与光控制脉冲串相对应的光数据脉冲串信号。基于新的路由信息替换先前的控制脉冲串的新控制脉冲串被转换成光控制信号，并被发送到下一个交换或出口节点。

[0047] 用于示例性 PBS 网络 200 的元件如下互连。SAN106A、106B 和 106C，LAN202A 和 202B，以及 WAN204 连接到 PBS 边缘节点 $215_1$-$215_M$ 中的相应一些。在所示实施例中，SAN 网关 206A、206B 和 206C 被分别用于便于 SAN106A、106B 和 106C 的通信接口。如以下详细描述的，在一个实施例中，SAN 网关和 PBS 边缘节点之间的“连接”实际上发生于同一“单元”内，因此共同定位 SAN 网关和 PBS 边缘节点的功能。在另一实施例中，基于光缆或电缆的链接可用于将 SAN 网关连接到 PBS 边缘节点。

[0048] PBS 网络 200 内，边缘节点 $215_1$-$215_M$ 经由光纤连接到交换节点 $217_1$-$217_L$ 中的一些。交换节点 $217_1$-$217_L$ 也经由光纤互相互连以形成包括边缘节点之间的多个光通路或光学链路的网状架构。理想地，有多个光通路以将交换节点 $217_1$-$217_L$ 连接到 PBS 网络 200 的每个端点（即，边缘节点是 PBS 网络 200 内的端点）。当一个或多个节点故障时，交换节点和边缘节点之间的多个光通路使能保护交换，或者可以使能诸如到目的地的主要和次要路径的特点。

[0049] 如以下结合图 3 所述的，PBS 网络 200 的边缘和交换节点被配置为发送和 / 或接收被波长多路复用的光控制脉冲串、光数据脉冲串和其它控制信号，以在预选波长上传播光控制脉冲串和控制标签并在不同的预选波长上传播光数据脉冲串或有效负荷。此外，PBS 网络 200 的边缘节点可发送光控制脉冲串信号同时从 PBS 网络 200 发送出数据（光或电的）。

[0050] 图 3 示出了根据本发明实施例的在 LAN 和 WAN 之间发送数据时 PBS 网络 200 的操作流程。该流程图反映了 PBS 网络所执行的一般传输操作。特别是，内部交换对于 SAN 和 LAN、WAN 或另一 SAN 之一之间的数据传输是一致的。以下描述用于 SAN 对接的附加措施。

[0051] 参考图 2 和 3，过程开始于框 300，其中 PBS 网络 200 从外部网络接收 IP 分组或以太网帧等。在一个实施例中，PBS 网络 200 在边缘节点 $215_1$-$215_M$ 之一处接收 IP 分组。接收到的分组可以是电子形式而非光学形式，或者按光学形式接收并随后转换成电子形式。在该实施例中，边缘节点以电形式存储接收到的分组。

[0052] 为清楚起见，PBS 网络 200 的操作流程的其余描述集中于从边缘节点 $215_2$（用作入口节点）到边缘节点 $215_3$（用作出口节点）的信息传输。其它边缘节点之间的信息传输基本类似。

[0053] 光脉冲串标签（即，光控制脉冲串）和光有效负荷（即，光数据脉冲串）从接收到的 IP 分组形成，如框 302 所描述的。在一个实施例中，边缘节点 $215_1$ 使用统计多路复用技术来从边缘节点 $215_2$ 中存储的接收到的 IP 分组形成光数据脉冲串。例如，由边缘节点 $215_2$

接收并在它们的路径上通过边缘节点 $215_3$ 传递到目的地的分组可被组装成一个光数据脉冲串有效负荷。

[0054] 接着，在框 304 中，特殊光学信道和 / 或光纤上的带宽被保留以通过 PBS 网络 200 传输光数据脉冲串。在一个实施例中，边缘节点 $215_2$ 保留通过 PBS 网络 200 的光数据信号路径中的时隙（即时分多路复用 (TDM) 系统的时隙）。该时隙可以是固定持续时间和 / 或可变持续时间，其中相邻时隙之间有均匀或不均匀的时间间隔。此外，在一个实施例中，为足以将光脉冲串从入口节点传输到出口节点的时间周期保留带宽。例如，在一些实施例中，边缘和交换节点维持所有使用的和可用的时隙的更新列表。时隙可在多个波长和光纤上分配和分布。这些保留的时隙这里也称作 TDM 信道。

[0055] 当边缘节点保留带宽时或当在传输光数据脉冲串后释放带宽时，网络控制器（未示出）更新该列表。在一个实施例中，网络控制器和边缘节点基于可用网络资源和通信模式使用各种脉冲串或分组调度算法执行该更新处理。周期性向所有边缘和交换节点广播的可用的可变持续时间 TDM 信道在与光控制脉冲串相同的波长上或在整个光学网络的不同共用预选波长上发送。网络控制器功能可驻留于边缘节点之一中，或者可以分布于两个或更多边缘节点上。

[0056] 光控制脉冲串、网络管理控制标签和光数据脉冲串随后通过光子脉冲串交换网络 200 在保留的时隙或 TDM 信道中传输，如框 306 所述。在一个实施例中，边缘节点 $215_2$ 沿着网络控制器确定的光学标签交换路径 (OLSP) 将控制脉冲串发送到下一个节点。在该实施例中，网络控制器在一个或多个波长上使用基于约束的路由协议（例如，多协议标签交换 (MPLS)）以确定对出口节点的最佳可用 OLSP。

[0057] 在一个实施例中，控制标签（这里也称作控制脉冲串）在光子数据脉冲串之前在不同的波长和 / 或不同的光纤上被异步发送。控制脉冲串和数据脉冲串之间的时间偏差允许每个交换节点处理该控制脉冲串并配置光子脉冲串交换机以在相应的数据脉冲串到达前正确地进行交换。术语光子脉冲串交换机这里用于表示不使用 O-E-O 转换的快速光交换机。

[0058] 在一个实施例中，边缘节点 $215_2$ 随后沿着路由（例如，交换节点 $217_1$）异步发送光数据脉冲串到交换节点，其中光数据脉冲串经历很少的时间延迟或不经历时间延迟且在每个交换节点内不作 O-E-O 转换。光控制脉冲串在发送相应的光数据脉冲串前被发送。

[0059] 在一些实施例中，交换节点可执行控制脉冲串的 O-E-O 转换，使得该节点能提取和处理标签中包含的路由信息。此外，在一些实施例中，TDM 信道以用于传播标签的相同的波长传播。或者，标签和有效负荷可利用不同的调制格式在同一光纤的相同波长上调制。例如，光学标签可利用不归零 (NRZ) 调制格式发送，同时光学有效负荷在相同波长上使用归零 (RZ) 调制格式被发送。光学脉冲串按类似方式从一个交换节点发送到另一交换节点，直到光控制和数据脉冲串终止于边缘节点 $215_3$。

[0060] 其余操作组属于出口节点操作（例如，边缘节点 $215_3$ 处执行的出口操作）。框 308 中，在接收到数据脉冲串后，出口节点分解它以提取被封装的数据（例如，IP 分组，以太网帧，光纤信道 (FC) 帧，等等）。在一个实施例中，出口节点将光数据脉冲串转换为出口节点可以处理以恢复每个分组的数据段的电子信号。这点处的操作流程取决于目标网络是光学 WAN 还是 LAN，如判断框 310 所述的。

[0061] 如果目标网络是光学WAN，则在框312中形成新的光控制和数据脉冲串信号。在该实施例中，边缘节点$215_3$准备新的光学标签和有效负荷信号。随后，在框314中，将该新的控制和数据脉冲串发送给目标网络（即，这种情况中的WAN）。在本实施例中，出口节点包括光学接口以将控制和数据脉冲串发送到光学WAN。

[0062] 但如果框310中目标网络被判定为是LAN，则逻辑进行到框316。因此，提取的数据分组或帧被处理，与相应的IP标签组合，并随后被路由到目标网络（即，这种情况中的LAN）。本实施例中，边缘节点$215_3$形成这些新IP分组。新IP分组随后被发送给目标LAN，如框318所示。

[0063] 图4示出了根据本发明实施例的用作PBS网络200中的交换节点的模块217。模块217包括一组光学波分多路分解器$400_1$-$400_A$，其中A表示用于传播有效负荷、标签和其它网络资源到该模块的输入光纤的数量。例如，在本实施例中，每个输入光纤都可以承载一组C个波长（即，WDM波长），尽管在其它实施例中输入光纤可承载不同数量的波长。模块217还包括一组N×N光子脉冲串交换机$402_1$-$402_B$，其中N是每个光子脉冲串交换机的输入/输出端口的数量。因此，在该实施例中，每个光子脉冲串交换机处波长的最大数量是A·C，其中$N \geq A \cdot C+1$。对于其中N大于A·C的实施例，额外的输入/输出端口可用于回送光信号用于缓冲。

[0064] 此外，尽管光子脉冲串交换机$402_1$-$402_B$被示作分开的单元，它们可以用任何合适的交换机架构实现为N×N光子脉冲串交换机。模块217还包括一组光学波分多路复用器$404_1$-$404_A$，一组光电信号转换器406（例如，光电检测器），控制单元407，和一组电光信号转换器408（例如，激光器）。控制单元407可具有一个或多个处理器以执行软件或固件程序。

[0065] 模块217的该实施例中的元件如下互连。光学多路分解器$400_1$-$400_A$连接到一组A个输入光纤，这些光纤传播来自光子脉冲串交换网络200的其它交换节点的输入光信号。光学多路分解器的输出引线连接到一组B个核心光交换机$402_1$-$402_B$以及连接到光信号转换器406。例如，光学多路分解器$400_1$具有连接到光子脉冲串交换机$402_1$-$402_B$的输入引线的B条输出引线（即，光学多路分解器$400_1$的一条输出引线连接到每个光子脉冲串交换机的一条输入引线）以及连接到光信号转换器406的至少一条输出引线。

[0066] 光子脉冲串交换机$402_1$-$402_B$的输出引线连接到光学多路复用器$404_1$-$404_A$。例如，光子脉冲串交换机4021具有连接到光学多路复用器$404_1$-$404_A$的输入引线的A条输出引线（即，光子脉冲串交换机$402_1$的一条输出引线连接到每个光学多路复用器的一条输入引线）。每个光学多路复用器还具有连接到电光信号转换器408的输出引线的输入引线。控制单元407具有与光电信号转换器406的输出引线或端口相连的输入引线或端口。控制单元407的输出引线连接到光子脉冲串交换机$402_1$-$402_B$的控制引线以及电光信号转换器408。如以下结合图5的流程图所述的，模块217用于接收和发送光控制脉冲串、光数据脉冲串和网络管理控制脉冲串。

[0067] 图7示出了根据本发明一个实施例的模块217的操作流程。参考图4和图5，模块217如下操作。

[0068] 模块217接收具有TDM控制和数据脉冲串信号的光信号。在本实施例中，模块217在一个或两个光学多路分解器处接收光学控制信号（例如，光控制脉冲串）和光数据信号（即，本实施例中的光数据脉冲串）。例如，可以在光学多路分解器$400_A$接收的光信号的第

一波长上调制光控制信号，而在光学多路分解器 $400_A$ 接收的光信号的第二波长上调制光数据信号。在一些实施例中，光控制信号由第一光学多路分解器接收而光数据信号由第二光学多路分解器接收。此外，在一些情况中，仅接收光控制信号（例如，网络管理控制脉冲串）。框 500 表示该操作。

[0069] 模块 217 将光控制信号转换成电信号。在该实施例中，光控制信号是光控制脉冲串信号，它由光学多路分解器从接收到的光数据信号中分离并被发送到光电信号转换器 406。在其它实施例中，光控制信号可以是网络管理控制脉冲串。光电信号转换器 406 将光控制信号转换成电信号。例如，在一个实施例中，TDM 控制信号的每个部分都被转换成电信号。由控制单元 407 接收的电控制信号被处理以形成新的控制信号。在该实施例中，控制单元 407 存储并处理控制信号中包含的信息。框 502 表示该操作。

[0070] 随后，模块 217 将处理后的电控制信号转换成新的光控制脉冲串。在该实施例中，控制单元 407 提供 TDM 信道校准，从而在期望的波长和 TDM 时隙模式中生成再转换的或新的光控制脉冲串。该新的控制脉冲串可在与框 500 中接收的控制脉冲串的波长和 / 或时隙不同的波长和 / 或时隙上被调制。框 504 表示该操作。

[0071] 随后，模块 217 发送光控制脉冲串到路径中的下一个交换节点。在该实施例中，电光信号发生器 408 发送新的光控制脉冲串到光学多路复用器 $404_1$-$404_A$ 的合适光学多路复用器以实现路由。框 506 表示该操作。

[0072] 随后，模块 217 根据控制信号中包含的路由信息将光数据信号（即，本实施例中的光数据脉冲串）路由到光学多路复用器 $404_1$-$404_A$ 之一。在该实施例中，控制单元 407 处理控制脉冲串以提取路由和计时信息，并将合适的 PBS 配置信号发送给一组 B 个光子脉冲串交换机 $402_1$-$402_B$ 以再配置每个光子脉冲串交换机来切换相应的光数据脉冲串。框 508 表示该操作。

[0073] 图 6 示出了根据本发明一个实施例的在示例性 PBS 架构 600 下节点之间的 PBS 光脉冲串流。架构 600 包括入口节点 610、交换节点 612、出口节点 614 和其它节点（出口、交换和入口，它们未被示出以使光脉冲串流的描述更清晰）。在本实施例中，示出的入口、交换和出口节点 610、612 和 614 的组件是使用机器可读指令实现的，这些指令使得机器（例如，处理器）执行允许节点将信息从 PBS 网络中的其它节点传入并传递到 PBS 网络中的其它节点的操作。本例中，光脉冲串流的光通路是从入口节点 610 到交换节点 612 再到出口节点 614。

[0074] 入口节点 610 包括入口 PBS MAC（媒体访问信道）层组件 620，它具有数据脉冲串汇编器 621、数据脉冲串调度器 622、偏差时间管理器 624、控制脉冲串构建器 626 和脉冲串成帧器 628。在一个实施例中，数据脉冲串汇编器 621 汇编数据脉冲串以便在 PBS 网络 200 上光学传送。在一个实施例中，根据许多不同的网络参数确定数据脉冲串的大小，所述参数诸如服务质量 (QoS)、可用光学信道的数量、入口节点处电子缓冲的大小、特殊的脉冲串汇编算法等等。

[0075] 数据脉冲串调度器 622 调度 PBS 网络 200 上的数据脉冲串传输。入口 PBSMAC 层组件 610 生成用于插入与正形成的数据脉冲串相关联的控制脉冲串的带宽请求。在一个实施例中，数据脉冲串调度器 622 还生成调度，以包括偏差时间（来自于以下所述的偏差管理器 624）从而允许 PBS 网络 200 中的各种节点在相关数据脉冲串到达前处理控制脉冲串。

[0076] 在一个实施例中，偏差时间管理器 624 根据各种网络参数确定偏差时间，这些参数诸如沿选定光通路的中继段的数量，每个交换节点处的处理延迟，用于特殊光通路的通信负荷，以及服务要求的等级。随后，控制脉冲串构建器 626 使用诸如所需带宽、脉冲串调度时间、带内或带外信号发送、脉冲串目的地地址、数据脉冲串长度、数据脉冲串信道波长、偏差时间、优先级等的信息构建控制脉冲串。

[0077] 脉冲串成帧器 628 使控制和数据脉冲串（在一些实施例中使用以下结合图 7、8 和 12 所述的成帧格式）成帧。脉冲串成帧器 628 随后在 PBS 网络 200 上经由物理光学接口（未示出）发送控制脉冲串，如箭头 650 所示。在该实施例中，控制脉冲串在带外 (OOB) 发送到交换节点 612，如图 6 中的光控制脉冲串 656 和 PBS TDM 信道 657 所指示的。随后，脉冲串成帧器 628 根据脉冲串调度器 622 所生成的调度在 PBS 网络上经由物理光学接口发送数据脉冲串到交换节点 612，如图 6 中的光学脉冲串 658 和 PBS TDM 信道 659 所指示的。光学脉冲串 656（控制脉冲串）和 658（数据脉冲串）之间的时间延迟在图 6 中被指示为 $OFFSET_1$。

[0078] 交换节点 612 包括 PBS 交换控制器 630，该控制器 630 具有控制脉冲串处理组件 632、脉冲串成帧器 / 去成帧器 634 和硬件 PBS 交换机（未示出）。光控制脉冲串 656 经由物理光学接口（未示出）和光交换机（未示出）被接收并被转换成电信号（即，O-E 转换）。控制脉冲串成帧器 / 去成帧器 634 使控制脉冲串信息去成帧并将控制信息提供给控制脉冲串处理组件 632。控制脉冲串处理组件 632 处理该信息，从而确定相应数据脉冲串的目的地、带宽保留、下一个控制中继段、控制标签交换等等。

[0079] PBS 交换控制器组件 630 使用该信息中的一些以控制和配置光交换机（未示出），从而在合适持续时间处将光数据脉冲串交换到合适信道处的下一个节点（即，本例中的出口节点 614）。在一些实施例中，如果保留带宽不可用，则 PBS 交换控制器组件 630 可采取合适的行动。例如，在一个实施例中，PBS 交换控制器 630 可以：(a) 确定不同的光通路以避免不可用的光学信道（例如，偏转路由）；(b) 使用 PBS 交换机结构内集成的缓冲元件（诸如光纤延迟线）延迟数据脉冲串；(c) 使用不同的光学信道（例如，通过使用可调波长转换器）；和 / 或 (d) 仅去掉同期数据脉冲串。PBS 交换控制器组件 630 的一些实施例也可发送否定的确认消息回到入口节点 610 以再发送去掉的脉冲串。

[0080] 但是，如果可以找到和为数据脉冲串保留带宽，则 PBS 交换控制器组件 630 提供硬件 PBS 交换机（未示出）的适当控制。此外，PBS 交换控制器组件 630 根据来自控制脉冲串处理组件 632 的更新保留带宽和可用 PBS 网络资源生成新的控制脉冲串。控制脉冲串成帧器 / 去成帧器 634 随后使再构建的控制脉冲串成帧，它随后经由物理光学接口（未示出）和光交换机（未示出）被光学发送到出口节点 614，如图 6 中 PBS TDM 信道 664 和光学控制脉冲串 666 所指示的。

[0081] 此后，当与接收 / 处理的控制脉冲串相对应的光数据脉冲串由交换节点 612 接收时，硬件 PBS 交换机已被配置为将光数据脉冲串交换到出口节点 614。在其它情况中，交换节点 612 可以将光数据脉冲串交换到不同的节点（例如，图 6 中未示出的另一交换节点）。来自入口节点 610 的光数据脉冲串随后被交换到出口节点 614，如 PBS TDM 信道 667 和光数据脉冲串 658A 所指示的。在该实施例中，光数据脉冲串 658A 仅仅是由硬件 PBS 交换机（未示出）再路由但可能在不同的 TDM 信道中发送的光数据脉冲串 658。光控制脉冲串 666

和光数据脉冲串 658A 之间的时间延迟由图 6 中的 $OFFSET_2$ 指示，它小于 $OFFSET_1$，例如是因为交换节点 612 中的处理延迟和其它计时错误。

[0082] 出口节点 614 包括 PBS MAC 组件 940，该组件具有数据多路分解器 642、数据脉冲串再汇编器 644、控制脉冲串处理组件 646 以及数据脉冲串去成帧器 648。出口节点 614 接收光控制脉冲串，如图 6 中的箭头 670 所指示的。脉冲串去成帧器 648 经由物理 O-E 接口（未示出）接收并去成帧控制脉冲。在该实施例中，控制脉冲串处理组件 646 处理被去成帧的控制脉冲串以提取相关的控制 / 地址信息。

[0083] 在接收到控制脉冲串后，出口节点 614 接收与接收到的控制脉冲串相对应的数据脉冲串，如图 6 中的箭头 672 所示。本例中，相对于控制脉冲串的末端，出口节点 614 在延迟 OFFSET2 后接收光数据脉冲串。按与以上针对接收到的控制脉冲串所述的相类似的方式，脉冲串去成帧器 648 接收并去成帧数据脉冲串。数据脉冲串再汇编器 644 随后处理被去成帧的数据脉冲串以提取数据（且如果数据脉冲串是分段的数据脉冲串，则再汇编该数据）。数据多路分解器 642 随后适当地多路分解提取的数据，用于发送到合适的目的地（它可以是 PBS 网络以外的网络）。

[0084] 图 7 示出了根据本发明一个实施例的用于 PBS 光脉冲串的一般 PBS 成帧格式 700。一般 PBS 帧 700 包括 PBS 一般脉冲串首部 702 和 PBS 脉冲串有效负荷 704（它可以是控制脉冲串或数据脉冲串）。图 7 还包括 PBS 一般脉冲串首部 702 和 PBS 脉冲串有效负荷 704 的展开图。

[0085] PBS 一般脉冲串首部 702 对于所有类型的 PBS 脉冲串是共同的，并包括版本号 (VN) 字段 710、有效负荷类型 (PT) 字段 712、控制优先级 (CP) 字段 714、带内信号发送 (IB) 字段 716、标签出现 (LP) 字段 718、首部纠错 (HEC) 出现 (HP) 字段 719、脉冲串长度字段 722 和脉冲串 ID 字段 724。在一些实施例中，PBS 一般脉冲串首部还包括保留字段 720 和 HEC 字段 726。以下针对具有 32 位字的成帧格式描述特殊字段大小和定义；但在其它实施例中，大小、顺序和定义可以是不同的。

[0086] 在该实施例中，PBS 一般脉冲串首部 702 是 4 字首部。第一个首部字包括 VN 字段 710、PT 字段 712、CP 字段 714、IB 字段 716 以及 LP 字段 718。该示例性实施例中的 VN 字段 710 是 4 位字段（例如，位 0-3），定义了用于使 PBS 脉冲串成帧的 PBS 成帧格式的版本号。该实施例中，VN 字段 710 被定义为第一个字的前 4 位，但在其它实施例中，它不需要是前 4 位、在第一个字中或限制于 4 位。

[0087] PT 字段 712 是定义有效负荷类型的四位字段（位 4-7）。以下示出了示例性有效负荷类型。

[0088] CP 字段 714 是定义脉冲串的优先级的 2 位字段（位 8-9）。例如，二进制“00”可表示正常优先级且二进制“01”表示高优先级。

[0089] IB 字段 716 是表示 PBS 控制脉冲串是正在带内还是 OOB 发送信号的 1 位字段（位 10）。例如，二进制“0”可表示 OOB 信号发送而二进制“1”表示带内信号发送。LP 字段 718 是用于表示是否已建立用于传送该首部的光通路的标签的 1 位字段（位 11）。

[0090] HP 字段 719 是用于表示首部纠错是否正用于该控制脉冲串中的 1 位（位 12）。不使用的位（位 13-31）形成当前不使用并保留用于将来使用的保留字段 720。

[0091] PBS 一般脉冲串首部 702 的第二个字包含 PBS 脉冲串长度字段 722，它用于存储与

PBS 脉冲串有效负荷 704 中字节数的长度相等的二进制值。在该实施例中，PBS 脉冲串长度字段是 32 位。

[0092] PBS 一般脉冲串首部 702 的第三个字包含 PBS 脉冲串 ID 字段 724，它用于存储用于该脉冲串的标识号。在该实施例中，PBS 脉冲串 ID 字段 724 是由入口节点（例如，图 6 中的入口节点 610）生成的 32 位。

[0093] PBS 一般脉冲串首部 702 的第四个字包括一般脉冲串首部 HEC 字段 726，它用于存储纠错字。在该实施例中，一般脉冲串首部 HEC 字段 726 是使用任何合适的已知纠错技术生成的 32 位。如图 7 所示，一般脉冲串首部 HEC 字段 726 是任选的，这在于如果不使用纠错则该字段可全部用零来填充。在其它实施例中，一般脉冲串首部 HEC 字段 726 不包含于 PBS 一般脉冲串首部 702 中。

[0094] PBS 脉冲串有效负荷 704 对于所有类型的 PBS 脉冲串都是共同的并包括 PBS 特定有效负荷首部字段 732、有效负荷字段 734 和有效负荷帧检验序列 (FCS) 字段 736。

[0095] 在该示例性实施例中，PBS 特定有效负荷首部 732 是 PBS 脉冲串有效负荷 704 的第一部分（即，一个或更多字）。通常，特殊有效负荷首部字段 732 包括用于与数据脉冲串相关的信息的一个或多个字段，它可以是该脉冲串本身或包含于与该脉冲串相关联的另一脉冲串中（即当该脉冲串是控制脉冲串时）。

[0096] 有效负荷数据字段 734 是 PBS 脉冲串有效负荷 704 的下一个部分。在一些实施例中，控制脉冲串没有有效负荷数据，所以该字段可被省去或全部包含零。对于数据脉冲串，有效负荷数据字段 734 可以相对较大（例如，包含多个数据分组或帧）。

[0097] 有效负荷 FCS 字段 736 是 PBS 脉冲串有效负荷的下一部分。在该实施例中，有效负荷 FCS 字段 736 是检错和 / 或纠错中使用的一个字的字段（即，32 位）。如图 7 所示，有效负荷 FCS 字段 736 是任选的，这在于如果不使用检错 / 纠错，则该字段可全部用零填充。在其它实施例中，有效负荷 FCS 字段 736 不包含于 PBS 脉冲串有效负荷 704 中。

[0098] 图 8 示出了根据本发明一个实施例的 PBS 光控制脉冲串成帧格式 800。为了更加清楚起见，图 8 包括 PBS 一般脉冲串首部 702 和 PBS 脉冲串有效负荷 704（先前结合图 7 描述的）的展开图，其中当为控制脉冲串的一部分时带有 PBS 有效负荷首部字段 732（以下描述）的进一步扩展。本例中，PT 字段被设定为“01”以表示该脉冲串是控制脉冲串。CP 字段被设定为“0”以表示该脉冲串具有正常优先级。IB 字段被设定为“0”以表示该脉冲串正使用 OOB 发送信号。LP 字段被设定为“0”以表示没有用于该控制脉冲串的标签。

[0099] 在 PBS 控制脉冲串的示例性实施例中，PBS 有效负荷首部字段 732 包括 :PBS 控制长度字段 802 ;扩展首部 (EH) 字段 806 ;地址类型 (AT) 字段 808 ;有效负荷 FCS 出现 (PH) 字段 810 ;控制信道波长字段 820 ;数据信道波长字段 822 ;PBS 标签字段 824 ;PBS 数据脉冲串长度字段 826 ;PBS 数据脉冲串开始时间字段 830 ;PBS 数据脉冲串使用期限 (time-to-live)(TTL) 字段 832 ;数据脉冲串优先级字段 834 ;PBS 数据脉冲串目的地地址字段 838 ;以及任选的扩展首部字段 840。

[0100] 该实施例中，PBS 有效负荷首部 732 的第一个字包括 PBS 控制长度字段 802，它用于存储控制首部按字节的长度。在该实施例中，PBS 控制长度字段 802 是通过控制脉冲串构建器 626（图 6）或控制脉冲串处理器 632（图 6）计算的 16 位字段（位 0-15）。在其它实施例中 PBS 控制长度字段 802 不需要是前 16 个位、在第一个字中或者限于 16 位。在该实

施例中，保留字段 804（位 16-27）包含于 PBS 有效负荷首部 732 内。在其它实施例中，这些位可用于其它字段。

[0101] PBS 有效负荷首部 732 的第一个字还包括 EH 字段 806，它在本实施例中用于指示扩展首部是否存在于脉冲串中。在该实施例中，EH 字段 806 是 1 位字段（位 28）。在其它实施例中，EH 字段 806 不需要是位 28 或者在第一个字中。

[0102] PBS 有效负荷首部 732 的第一个字还包括 AT 字段 808，它在本实施例中用于表示相关联的 PBS 数据脉冲串的目的地的地址类型。例如，地址类型可以是 IP 地址（例如，IPv4，IPv6）、网络服务存取点 (NSAP) 地址、以太网地址或其它类型的地址。在一个实施例中，AT 字段 808 是 2 位字段（位 29-30）。

[0103] PBS 有效负荷首部 732 的第一个字还包括 PH 字段 810，它用于表示有效负荷 FCS 是否存在于该脉冲串中。在该实施例中，PH 字段 810 是 1 位字段（位 31）。

[0104] PBS 有效负荷首部 732 的第二个字包括控制信道波长字段 820，它用于表示控制脉冲串被假定在其中被调制的 WDM 波长。在该实施例中，控制信道波长字段 820 是 16 位字段（位 0-15）。

[0105] PBS 有效负荷首部 732 的第二个字还包括数据信道波长字段 822，它被用于表示数据脉冲串在其中要被调制的 WDM 波长。在该实施例中，数据信道波长字段 822 是 16 位字段（位 16-31）。

[0106] PBS 有效负荷首部 732 的第三个字包括 PBS 标签字段 824，它用于存储用于正由脉冲串使用的光通路的标签（如果有）。在该实施例中，该标签是由标签管理组件生成的 32 位字。

[0107] PBS 有效负荷首部 732 的第四个字包括 PBS 数据脉冲串长度字段 826。本实施例中，PBS 数据脉冲串长度是 32 位字。

[0108] PBS 有效负荷首部 732 的第五个字包括 PBS 数据脉冲串开始时间字段 830。在该实施例中，PBS 数据脉冲串开始时间是由脉冲串调度器 622（图 6）生成的 32 位字。

[0109] PBS 有效负荷首部 732 的第六个字包括 PBS 数据 TTL 字段 832。在该实施例中，PBS 数据 TTL 字段 932 是由入口 PBS MAC 组件 620（图 6）生成的 16 位（位 0-15）字段。例如，在一个实施例中，入口 PBS MAC 组件 620 的脉冲串调度器 622（图 6）可以生成 TTL 值。

[0110] PBS 有效负荷首部 732 的第六个字还包括数据脉冲串优先级字段 832。在该实施例中，数据脉冲串优先级字段 832 是入口 PBS MAC 组件 620（图 6）生成的 8 位字段（位 16-23）。例如，在一个实施例中，入口 PBS MAC 组件 620 的脉冲串调度器 622（图 6）可生成数据脉冲串优先级值。此外，在该实施例中，PBS 有效负荷首部 732 的第六个字包括可在将来用于其它字段的保留字段 836（位 24-31）。

[0111] PBS 有效负荷首部 732 的第七个字还包括 PBS 数据脉冲串目的地地址字段 838。在该实施例中，PBS 数据脉冲串目的地地址字段 838 是可变长度字段，为清楚起见，示作单个 32 位字。地址的实际长度可根据 AT 字段 808 中指示的地址类型而变化。

[0112] PBS 有效负荷首部 732 的第八个字可以包括任选的扩展首部字段 840。该首部可用于保存将来使用的其它首部数据。在使用该首部时，EH 字段 806 被设定为 1。在该实施例中，以上已描述了有效负荷数据字段 734 和有效负荷 FCS 字段 736。

[0113] 图 9A 描述了示例性网络架构 900A，它支持经由光学脉冲串交换连网组件（在所

示的实施例中的 PBS 组件）的多个 SAN 岛之间的网络通信。网络架构 900 包括六个 SAN，分别标记为 $902_{1-6}$，它们经由多个 PBS 交换节点 $217_{1-3}$ 和光学链路 $904_{1-26}$ 而互连。在所示的实施例中，每个 SAN 都包括各自的 SAN 网关 $906_N$ 以及共同定位的 PBS 接口 $908_0$。SAN 网关和 PBS 接口共同提供了一 SAN 和 PBS 连网架构的内部 PBS 交换节点之间的接口。因此，这些共同定位的组件对于 PBS 交换节点表现为 PBS 边缘节点 $910_{1-6}$。

[0114] 为说明目的，光学链路 $904_{1-26}$ 成对示出，表示经由单根光纤在多个不同波长上或者经由多根光纤在单个波长上同时传送数据的能力。可以理解，单条光学链路在合适的 WDM 实现下可支持 1-N 个并存的波长。此外，一条以上光纤链路可用于连接一对节点，从而在链路故障的情况下或者为支持增加的通信提供了冗余。

[0115] 网络架构 900A 使得 $SAN902_{1-6}$ 能经由 PBS 结构相互通信。为支持该能力，有必要提供合适的通信接口以支持每个 SAN 和 PBS 网络基础结构的内部工作。如上所述，这是通过 SAN 网关和 PBS 接口的组合来实现的。为更好地理解该接口的 SAN 侧的基础操作，现在讨论基本 SAN 操作。有许多 SAN 资源对于连网技术领域的熟练技术人员方便可用，它们提供了以下讨论的 SAN 各方面的进一步细节。

[0116] SAN 操作被设计成支持各种不同的平台和连网技术。已开发了开放的标准以使各种供应商组件之间的网络能互操作，而不是使 SAN 成为限制性网络。用于 SAN 的基础数据传输是基于光纤信道 (FC) 标准。尽管名称意味着光纤链路的使用，但可以使用各种类型的光学和铜链路，包括同轴和双绞线链路。光纤信道是由美国国家标准协会 (ANSI) 开发的标准的集合组的一般名称 (X3T9.3 Task Group ofANSI :Fibre Channel Physical and Signaling Interface(FC-PH)) ；最新的 FC-PH 草案可在 http://www.t11.org/index.htm 获得。

[0117] 在光纤信道术语中，连接终端装置（即，服务器和存储装置）的网络基础结构称作织构 (Fabric)。光纤信道包括以相反方向发送的两个单向光纤并具有相关联的发送器和接收器，其中每个光纤都附着到一端处的一个端口的发送器以及另一端处的另一端口的接收器。当织构存在于配置中时，光纤可附着到节点端口 (N_Port) 以及织构的端口 (F_Port)。

[0118] 参考图 10，光纤信道被构成为分层功能的分层集。最底层 (FC-0) 定义系统中的物理链路，包括光纤、连接器、用于各种不同数据率的光学和电学参数。由于光纤链路中的光功率水平会超过由可应用的激光安全标准所定义的极限，还规定了安全系统 - 开放光纤控制系统 - 用于短波激光数据链路。本质上，破损光纤的检测造成激光器工作周期被自动减少以满足安全需要。

[0119] FC-1 层定义了包括串行编码和解码规则、特殊字符和错误控制的传输协议。光纤上传送的信息被每次 8 位地编码成 10 位传输字符。传输代码使用的主要原理是改善光纤上信息的传输特性。

[0120] 信号发送协议 (FC-2) 层用作光纤信道的传输机制。FC-2 定义了端口间要传递的数据的成帧规则、用于控制三个服务等级的不同机制以及用于管理数据传递序列的装置。为帮助链路上数据的传输，标准定义了以下构建块 ：有序集、帧、序列、交换和协议。这些是本领域熟练技术人员已知的。为了这里的实施例，FC 帧是 FC-2 的最重要方面，因此以下仅简要描述有序集、序列、交换和协议 ；它们每一个都是 SAN 领域中公知的。

[0121] 有序集是用于获得位和字同步的四字节传输字，它可以形成字边界校准。信号发

送协议定义了三种主要类型的有序集，包括帧定界符、原始信号和原始序列。

[0122] FC链接的基本构建块是帧。帧包含要发送的信息（即有效负荷）、源和目的地端口的地址以及链接控制信息。帧被广泛地分类为数据帧和链接控制（Link_control）帧。数据帧可用作链接数据（Link_Data）帧以及装置数据（Device_Data）帧，链接控制帧可分类为确认（ACK）和链接响应（Link_Response）（忙和拒绝）帧。织构的主要功能是从源端口接收帧并将它们路由到目的地端口。FC-2层的责任是将要发送的数据分解成帧大小，并再汇编这些帧。

[0123] FC帧1100的格式在图11中示出。每个帧都以帧定界符开始和结束。帧定界符（帧开始（SOF）定界符1101和帧结束（EOF）定界符1112）是紧邻帧内容之前或之后的有序集。帧首部1102紧接着SOF定界符1101。帧首部用于控制链接应用，控制装置协议传递，以及检测丢失或故障帧。最大2112字节长的数据字段1104包含要从源N_Port传递到目的地N_Port的信息。有效负荷可包括包含有附加链接控制信息的任选首部1106，并包括最大2048字节的数据有效负荷1108。4字节的循环冗余码校验（CRC）1110在EOF定界符1112之前。CRC用于检测传输错误。

[0124] 帧首部1102的进一步细节在图11的下半部处示出。帧首部包括控制CTL字段1114，继之以源和目的地地址字段1116和1118以及类型字段1120。包括序列计数（seq_cnt）字段1122和序列标识（seq_ID）字段1124的下两个字段包含序列信息。通过从一个N_Port单向发送到另一个的一个或多个有关帧的集合形成一序列。序列内的每个帧都用序列计数唯一地编号。通常在序列边界处进行较上协议层控制的错误恢复。

[0125] 交换_ID(exchange_ID)字段1126是最后一个帧首部字段。交换包括用于单个操作的一个或多个非并存序列。交换可以是两个N_Ports之间单向或双向的。在单个交换内，在任何一刻仅一个序列可以是活动的，但不同交换的序列可并发活动。

[0126] 协议关系到光纤信道提供的服务。协议可以是针对较高层服务的，尽管光纤信道提供其自身的一组协议以管理用于其数据传递的操作环境。协议由前述ANSI标准加以规定。

[0127] 流控制是用于调步多个N_Ports之间以及一N_Port和织构之间的帧流动的FC-2层控制过程以防止接收器处的超时运行。流控制取决于服务等级。等级1帧使用端对端流控制，等级3仅使用缓冲到缓冲，等级2帧使用这两种类型的流控制。

[0128] FC-3等级的FC标准旨在提供高级特点所需的共同服务。这些包括：分条(Striping)-用于并行使用多个N_Ports以在多条链路上发送单个信息单元来倍增带宽；查寻（hunt）组-一个以上端口响应于同一别名地址的能力。这通过减少到达忙N_Port的机会改善了效率；和多播-多播将单个发送提交到多个目的地端口。这包括发送给织构上的所有N_Ports（广播）或者仅发送给织构上N_Ports的一个子集。

[0129] FC-4是FC结构中的最高层，它定义了可以在FC上执行的应用程序接口。它规定了使用以下FC级的上层协议映射规则。FC同样擅长于传输网络和信道信息两者并允许在同一物理接口上并发地传输这两种协议类型。

[0130] 当前规定或提出了以下的网络和信道协议：小型计算机系统接口（SCSI）；智能外围接口（IPI）；高性能并行接口（HIPPI）成帧协议；因特网协议（IP）；用于计算机数据的ATM适应层（AAL5）；链路封装（FC-LE）；单字节命令代码集映射（SBCCS）；以及IEEE 802.2。

[0131] 为有效地适应 SAN 到 PBS 网络接口上的数据传送，提供了在 PBS 有效负荷内嵌入光纤信道帧的格式化机制。图 12 示出了根据一个实施例的包含多个 FC 帧的 PBS 数据脉冲串有效负荷 1200 的细节。PBS 一般脉冲串首部 702A 包括图 7 和 8 中示出的用于 PBS 一般脉冲串首部 702 的许多上述字段。更详细地，有效负荷类型字段 712A 可用于标识不同的有效负荷类型。在一个实施例中，使用以下的 4 位值：

[0132] 0000 无有效负荷

[0133] 0001 控制脉冲串

[0134] 0010 网络管理脉冲串

[0135] 0100 保留

[0136] 1XXX 数据有效负荷诸如

[0137] 1111IP 分组

[0138] 1001 以太网帧

[0139] 1101FC 帧

[0140] 1011 MPEG-1/2/4 视频帧

[0141] PBS 有效负荷首部 732A 包括 20 位保留字段 1202，以及段 ID(S-ID) 字段 1204，它用于存储再汇编分段数据脉冲串的标识符 (ID)。在该实施例中，段 ID 字段 704 是由控制脉冲串构建器 626( 图 6) 或控制脉冲串处理器 632 计算出的 8 位字段 ( 位 20-27)。

[0142] PBS 有效负荷首部 732A 还包括段脉冲串指示器 (SB) 字段 1208、串联有效负荷指示器 (CPI) 字段 1210 和有效负荷 PCS(PH) 字段 1212。这些字段分别用于表明：PBS 数据脉冲串是否被分段；脉冲串有效负荷是否被串联；以及有效负荷 FCS 是否存在。在所示实施例中，字段 1208、1210 和 1212 是 1 位字段 ( 分别为位 29、30 和 31)。在其它实施例中，这些字段可映射到不同的位，或者在与 PBS 有效负荷首部 732A 的第一个字不同的字中。与用于 PBS 控制脉冲串的 PBS 有效负荷首部不同，数据脉冲串的 PBS 有效负荷首部的该实施例仅具有一个 32 位字。但是，在其它实施例中用于 PBS 数据脉冲串的 PBS 有效负荷首部可以在长度上大于字。

[0143] 有效负荷数据 734A 被配置为一个或多个 FC 帧 1100，其中每个各自的帧都包括 PBS 脉冲串有效负荷长度 1214A。例如，所示实施例包括有效负荷中的三个 FC 帧 1100A、1100B 和 1100C，具有各自的 PBS 脉冲串有效负荷长度 1214A、1214B 和 1214C。每个 FC 帧都具有类似于以上参考图 11 所述的配置。PBS 脉冲串有效负荷长度 1214A、1214B 或 1214C 中的每一个都包含与各自的 FC 帧 1100A/B/C 的长度相对应的值。

[0144] 如上所述，在一个实施例中，由 SAN 网关和 PBS 接口提供的功能可共同定位于单个单元中。例如，图 13 示出了根据本发明一个实施例的模块可再配置 SAN 网关 /PBS 边缘节点单元 1300。在该实施例中，单元 1300 包括各自具有光学端口 $1304_1$ 和 $1304_2$ 的一对光学 PBS I/O 卡或模块 $1302_1$ 和 $1302_2$，具有传统网络端口 1308 的传统接口卡或模块 1306，多个可配置服务器模块 $1310_1$-$1310_N$( 仅示出其中两个 )，包括 FC 端口 1314 的一个或多个光纤信道接口卡 1312，底板 1316，连接器 $1318_1$-$1318_M$( 图 13 中仅连接器 $1316_1$-$1316_3$ 可见 )，以及机架 1320。在一些实施例中，单元 1300 可包括两个以上或以下的可配置服务器模块，以及两个以上或以下的 PBS I/O 卡。在其它实施例中，单元 1300 可不同于图 13 所示的实施例进行配置。以下结合图 14a 和 14b 描述光学 PBS I/O 模块 1302 的一个实施例。在一个

实施例中，各种模块和卡包括位于刀片服务器机架上的刀片服务器。在一个实施例中，单元1300根据高级电信计算架构（高级TCA或ATCA）标准（PICMG 3.0）(PCI工业计算机制造组）进行配置。

[0145] 在该实施例中，传统接口卡1306是用于利用GbE以太网协议与前缘路由器(LER)或其它LAN/WAN网络通信的千兆位以太网(GbE)卡。在其它实施例中，可以使用不同的传统协议。

[0146] 在该实施例中，服务器模块$1310_1$-$1310_N$是自包含高速服务器刀片，其中单个或多个服务器功能被作为单个集成刀片实现。

[0147] 在一些实施例中，底板1316包括电子交换织构，它具有缓冲器并具有与商业上可得到的刀片服务器系统中所使用的那些相类似的电子总线（参见图14a的交换织构1430）、电源和控制等。在一个实施例中，电子底板织构支持多个交换拓扑，诸如星形或双星形拓扑，以切换到合适的电接口，例如服务器模块中的外围组件互连(PCI)（例如，1999年1月25日的PCI规范v2.2）或快速PCI(PCI-Express)（例如，1999年9月27日的PCI-X规范v.1.0）、InfiniBand®（例如，2000年10月24日的InfiniBand® 1.0规范）接口。在其它实施例中，底板可包括其它类型的布线交换织构。这里使用的布线交换织构也可表示光交换织构或光学和电学交换织构的组合。

[0148] 单元1300的元件如下地互连。光学I/O模块$1302_1$和$1302_2$、传统接口模块1306、服务器模块$1310_1$-$1310_N$以及光纤信道接口卡1312经由连接器$1318_1$-$1318_M$连接到底板1316（以及前述电交换织构1430）。光学端口$1304_1$和$1304_2$连接到各自的PBS网络交换节点217（例如，图2中的PBS网络200的）。传统端口1308连接到传统网络(LAN或WAN)或LER（例如，参见图2）。机架1320容纳并物理支持这些模块、连接器和底板。机架1320还包括其它组件（例如，电源、一个或多个冷却风扇等），它们在图13中未示出以避免模糊本发明。

[0149] 操作中，单元1300可以用作SAN网关并使能通过给定SAN与各种存储装置主机的连接性。例如，在一个实施例中，经由本领域公知的常规SAN网关操作方便了SAN外的客户机与SAN内的数据主机之间的数据通信。支持这种类型的功能的SAN网关模块由若干供应商提供，包括，但不限于，IBM公司，White Plains，New York。例如，一个或多个服务器模块$1310_1$-$1302_N$可方便SAN网关操作。

[0150] 此外，单元1300可经由PBS网络以及光学I/O模块$1302_1$和$1302_2$向客户机提供服务。但是，与常规网络协议不同，光学I/O模块$1302_1$和$1302_2$从客户机接收光学PBS控制和数据脉冲串，它们随后如下所述地被O-E转换、去成帧、多路分解和路由。在一个实施例中，光学I/O模块$1302_1$和$1302_2$提供信息以按与服务器模块在底板1316上传递信息相同的方式经由底板1316将输入通信路由到正确的服务器模块。

[0151] 类似地，单元1300的服务器模块经由底板1316以及光学PBS I/O模块1302将信息传递给PBS网络。不同于常规网络协议装置，在一个实施例中，按与先前针对PBS网络200（图2）的入口节点描述的基本相同的方式，光学PBS I/O模块1302统计多路复用来自一个或多个服务器模块的输入通信流（例如，FC帧），以形成PBS控制和数据脉冲串。PBS脉冲串随后被成帧、调度、E-O转换和经由PBS网络发送到客户机，如先前针对PBS网络200所描述的。

[0152] 从传统网络进入到单元 1300 的用于通过 PBS 网络传送到目的地的通信在传统端口 1308 处由单元 1300 接收。如上所述，传统的网络可使用诸如 TCP/IP 或以太网协议的常规网络协议。在此实施例中，传统的网络是电 GbE 网络，虽然在其它实施例中可使用其它有线或无线网络。传统的接口模块 1306 以与任何服务器模块通过底板 1316 传送信息一样的方式将在传统端口 1308 接收的信息通过底板 1316 发送到光 I/O PBS 模块 1302。光 PBS I/O 模块 1302 以与上面对 PBS 网络 200 的入口节点描述基本上相同的方式将来自传统接口模块 1308 的信息构造成 PBS 脉冲串。然后，该 PBS 脉冲串如以前对 PBS 网络 200 描述地被调度、E-O 转换、并通过 PBS 网络发送到客户机。

[0153] 从 PBS 网络进入单元 1300 并用于传到 SAN 目的地的通信由 PBS 光学端口 1304 处的单元 1300 以光控制和数据 PBS 脉冲串的形式接收。光学 PBS I/O 模块 1302 O-E 转换 PBS 光学端口 1304 处接收的光控制和数据脉冲串，去成帧该 PBS 脉冲串，并将 PBS 数据脉冲串多路分解为例如构成 FC 帧 1100 的各个流。随后，将这些各个流经由底板 1316 传送到服务器模块中的合适一个。然后，用作 SAN 网关的该服务器模块将这些各个通信流经由光纤信道卡 1312 上的合适 FC 端口 1314 传送到 SAN。

[0154] 图 14a 示出了根据本发明一个实施例的光学 PBS I/O 模块 1302。在该实施例中，光学 PBS I/O 模块 1302 包括网络处理器单元 1402（该模块可具有多个网络处理器）、总线桥路 1404、队列单元 1406、成帧器单元 1408（具有框 14081 和 14082 所指示的成帧器和去成帧器功能）、E-O 接口 1410、O-E 接口 1416、网络处理器缓冲器 1420、通信整形器 (shaper) 1424 以及通信整形器缓冲器 1426。在一个实施例中，底板交换织构 1430 包括快速 PCI 总线，尽管在其它实施例中可使用任何其它的合适的总线。因此，可以使用商业上可得到的 PCI 桥路装置或芯片组实现总线 - 桥路 1404。

[0155] 在该实施例中，光学 PBS I/O 单元 1302 的前述元件如下互连。总线桥路 1404 连接到底板交换织构 1430 以支持经由互连 1438 的并行双向通信。总线桥路 1404 还经由电互连 1439 连接到通信整形器 1424。为了清楚起见，将图 14a 中的电互连 1438、1439 和其它信号互连描绘为单个互连线路（尽管连接可包括若干信号互连线路）。

[0156] 通信整形器 1424 分别经由互连 1440 和 1441 连接到网络处理器单元 1402 和缓冲器 1426。网络处理器单元 1402 分别经由互连 1442 和 1443 连接到队列单元 1406 和缓冲器 1420。接着，队列单元 1406 经由互连 1444 连接到 PBS 成帧器 / 去成帧器 1408。

[0157] 如图 14b 所示，在一些实施例中，网络处理器单元 1402 包括入口网络处理器 1460 和出口网络处理器 1462。因此，在光学 PBS I/O 模块 1302 的一些实施例中，互连 1440 和 1442 连接到入口网络处理器 1460。

[0158] 此外，如图 14b 所示，在一些实施例中，队列单元 1406 可包括数据队列 1470 和 1472、控制队列 1474 和 1475，以及与队列 1470、1472、1474 和 1475 的输出端口耦合的电交换机或多路分解器 1476。因此，在一些实施例中，队列 1470、1472、1474 和 1475 的输入端口经由交换机或多路复用器（未示出）连接到互连 1442。此外，在一些实施例中，交换机 1476 的输出端口可连接到互连 1444。

[0159] 在其它实施例中，网络处理器单元 1402 中可以使用不同数量的处理器（例如，单个处理器）。此外，在一些实施例中，可以在队列单元 1406 中使用不同数量的队列。例如，队列单元不需要包括一专用控制队列和 / 或两个数据队列。多个队列可用于提供用于构建

具有不同属性（诸如不同优先级）的多个脉冲串的存储。

[0160] 再参考图 14a，PBS 成帧器单元 1408 经由互连 1446 连接到 E-O 接口 1410。E-O 接口 1410 接着经由互连 1448 连接到 PBS 网络的其余部分。O-E 接口 1416 经由互连 1450 连接到 PBS 网络的其余部分。一般，O-E 接口 1416 可在一个互连的 SAN 上接收所有发送的波长 - 或者它具有可调光学脉冲串接收器或者具有多个固定波长的光学脉冲串接收器。O-E 接口 1416 还经由互连 1452 连接到成帧器单元 1408。成帧器单元 1408 还经由互连 1454 连接到网络处理器单元 1402。在一个实施例中，互连 1464 连接到网络处理器 1462（图 14b）。网络处理器单元 1402 经由互连 1456 连接到总线桥路 1404。以下结合图 15 和 16 描述从 PBS 网络和向 PBS 网络传递信息过程中光学 PBS I/O 模块 1302 的操作。

[0161] 参考图 14a-b 以及图 15 的流程图 1500，光学 PBS I/O 模块 1302 执行如下的 PBS 出口操作（即，将信息从 PBS 网络传递到常规网络和 / 或单元 1300 的服务器模块）。光学 PBS I/O 模块 1302 将经由互连 1450 从 PBS 网络接收的光学 PBS 脉冲串转换成电信号。在该实施例中，O-E 接口 1416 执行 O-E 转换。该操作流程由框 1502 表示。

[0162] 随后，将接收到的 O-E 转换的 PBS 脉冲串去成帧和多路分解。在该实例中，成帧器单元 1408 从 O-E 接口 1416 经由互连 1452 接收 O-E 转换的 PBS 脉冲串并去成帧该 PBS 脉冲串。例如，在一个实施例中，如以上参考图 7 和 8 所描述的，可使 PBS 脉冲串成帧。在其它实施例中，可使用不同的成帧格式。多路分解使得能将每个成帧的数据脉冲串分成相应的 IP 分组、以太网帧、FC 帧等。该操作流程由框 1504 表示。

[0163] 随后，处理 PBS 脉冲串中包含的信息。在该实施例中，网络处理器单元 1402 经由互连 1454 从成帧器单元 1408 接收去成帧和多路分解的 PBS 脉冲串并执行该处理。例如，在一些实施例中，网络处理器单元 1402 可提取地址和有效负荷信息，对首部和 / 或有效负荷信息进行纠错，连接有效负荷，再汇编分段的有效负荷等等。网络处理器单元 1402 可使用缓冲器 1420，以在以上处理操作期间临时存储信息。在一个实施例中，出口网络处理器 1462（图 14b）处理去成帧的脉冲串。该操作流程由框 1506 表示。

[0164] 随后，在底板交换织构 1430 上发送处理后的信息。在该实施例中，总线桥路 1404 从网络处理器单元 1402 经由互连 1456 接收处理后的信息并在底板交换织构 1430 上以合适的格式并用合适的总线控制信号（例如，根据 PCI 协议）将该信息发送到合适的目的地。该信息的目的地例如可以是与传统网络相连的装置（这种情况中信息被发送到传统接口模块 1306）或服务器模块（即，服务器模块 $1310_1$-$1310_N$ 之一）。该操作流程由框 1508 表示。

[0165] 流程图 1500 包括专用于转发要存储于 SAN 存储装置上的数据的框 1510-1514 中的附加操作。框 1508 中在底板上发送的数据由服务器模块 $1510_1$-$1510_N$ 之一接收。提供 SAN 网关功能的服务器模块随后标识数据要向其转发以便存储的 SAN 目的地。这些操作由框 1510 表示。根据框 1512 和 1514，数据被封装到 FC 帧，且该 FC 帧使用可应用的 SAN 数据传输技术被发送到目的地 SAN 存储装置。

[0166] 参考图 14a-b 以及图 16 的流程图 16，光学 PBS I/O 模块 1302 执行 PBS 入口操作；即如下地将信息从传统网络和 / 或单元 1300 的服务器模块传递到 PBS 网络。光学 PBS I/O 模块 1302 以电信号形式接收要在 PBS 网络上传送的信息。在该实施例中，总线桥路 1404 经由互连 1438 从底板交换织构接收信息。在该实施例中，该信息可经由传统接口 1306 来

自于传统网络或者来自于服务器模块 $1510_1$-$1510_N$ 之一。该操作由框 1602 表示。

[0167] 然后，将接收到的信息整形以帮助改善 PBS 网络（例如，图 3 的 PBS 网络 300）中的通信流。在该实施例中，通信整形器 1424 经由互连 1439 从总线桥路 1404 接收信息并整形该信息。例如，在一个实施例中，通信整形器 1424 对该信息执行操作以减少由于自相似效果引起的输入通信流的相关结构和长期依赖性。通信整形器 1424 可配置为执行本领域已知的任何合适的通信整形算法或技术。通信整形器 1424 可使用缓冲器 1426 以在执行通信整形操作的同时临时存储信息。该操作流程由框 1604 表示。

[0168] 随后，将整形后的信息多路复用入 PBS 控制和数据脉冲串。在该实施例中，网络处理器单元 1402 经由互连 1440 从通信整形器 1424 接收整形后的信息。随后，网络处理器单元 1402 处理该信息以形成并调度 PBS 控制和数据脉冲串，如以上针对 PBS 网络 300 中的入口节点所描述的。在其它实施例中，基于选定的脉冲串汇编算法将该信息汇编入合适的脉冲串大小以便在光学脉冲串网络（不必是 PBS 网络）上传送。在一个实施例中，入口网络处理器 1460（图 14b）处理通信整形信息。此外，在此实施例中，在控制和数据脉冲串正被形成时，网络处理器单元 1402 使用队列单元 1406 存储这些控制和数据脉冲串且直到它们被调度用于在 PBS 网络上传输。该操作流程由框 1606 表示。

[0169] 随后，将脉冲串封装入帧，用于在 PBS 网络上传输。在该实施例中，成帧器单元 1408 经由互连 1444 从队列单元 1406 接收脉冲串并执行成帧操作。在一个实施例中，如以上参考图 7 和 10 描述的那样使脉冲串成帧。在其它实施例中，可用使用不同的成帧格式。该操作流程由框 1608 表示。

[0170] 随后，将成帧的脉冲串转换成光信号并在调度的时间在 PBS 网络上传送。在该实施例中，E-O 接口 1410 经由互连 1446 从成帧器单元 1408 接收成帧的脉冲串（即 PBS 控制和数据脉冲串）。接着，E-O 接口 1410 执行 E-O 转换并在调度的时间并在 PBS 网络的保留的 PBS TDM 信道中发送光信号。该操作流程由框 1610 和 1612 表示。

[0171] 根据本揭示内容的其它方面，PBS 边缘、交换和路由设备可在 SAN 网关处共同定位。例如，图 9B 示出了一网络架构 900B，它包括了与图 9A 示出并在以上讨论的那些相类似的组件。但是，在该实施例中，PBS 交换模块 $217_{1-6}$ 在各自的 SAN 网关 $906_{1-6}$ 处共同定位。各种交换 PBS 交换模块 $217_{1-6}$ 经由光学链路 $904_{1-6}$ 被通信链接。

[0172] 尽管与图 9A 的实施例相比使用共同定位的 PBS 交换模块会需要附加的模块，但它消除了对独立的 PBS 交换节点的需要，从而形成了具有较低网络实现成本的更灵活的网络架构。通过与其共同定位的 SAN 网关的交互，PBS 交换模块动态地提供所请求的光通路，提前保留必要的带宽并根据通信优先级、其自身的分配资源和可用带宽调度要发送到其它 SAN 和 / 或其它 LAN/WAN 的 SAN 通信。结果，对 SAN 内基于 FC 的数据通信的影响最小。

[0173] 在一个实施例中，通过修改外部网关协议 (EGP) 来使能较大企业网络内 SAN 到 SAN 网络路由，当多个光通路可用时该外部网关协议用于确定到特定 SAN 网络的最佳可用路径。EGP 进行的路径选择是通过特殊 SAN 网络的相关属性来进行的。因此，不同 SAN 之间的每条光通路都被映射到给定的路径或交换连接。在一个实施例中，EGP 在专用控制光通路上运行但也可以在互连装置的分开的电（例如以太网）网络上运行。

[0174] 一方面，路由方案类似于用于因特网路由的方案，其中每个网络域作为一自主系统 (AS) 操作，且外部路由被用于通过使用仅意识到不同域之间的互连而不意识到关于每

个域内路由的任何信息的域间路由协议将数据路由到并通过各种 AS。特别是，用于因特网的路由域被称作边界网关协议 (BGP)，且本发明的实施例实现 BGP 协议的扩展版本，它包括便于基于 PBS 网络的路由的规定。

[0175] 在一个实施例中，PBS 网络的一个或多个共同定位的交换节点被指定为“外部网关协议”路由器，它们在它们与其它相邻 PBS 节点的接口连接上运行修改的 BGP 协议。因此，所有输出和输入数据通信到这些共同定位的交换节点之一通过PBS BGP路由器所指定的一个 SAN。在一个实施例中，每个外部网关协议路由器选择性地将其所有可能的路由通告给相邻 BGP 路由器中的一些或全部。在另一实施例中，每个 BGP 路由器都被允许根据相关属性以及其它标准（诸如带宽使用或端对端等待时间）将它发送的各种路由通告排列或定优先级。因此，在所有可用路径中的最佳路径选择中，SAN/PBS 网关可容易地影响 BGP 判断过程。跨 PBS 网络通告光通路路径的可用性是使用 BGP 更新 (UPDATE) 消息进行的。PBS 到 PBS 网络连接性不限于全光网络，也可以包括其它类型的光学物理链路，诸如 SONET/SDH 或 10Gb/s 以太网。

[0176] 图 9C 示出了从共同定位的 BGP 路由器方面观察的网络架构 900B，它包括用“$BGP_n$”标签示出的所有路由器。特别是，交换节点 $217_{1-6}$ 中的每一个都用作一 BGP 路由器，出于说明目的，它们通过各种路径段 $912_{1-8}$ 连接。在常规 BGP 路由下，每个路由器维护包括路径段的串联的一个路由表，其中每一个都共同地包括经过该路由器的一路径。但是，常规 BGP 路由不涉及基础传输机制，且不考虑路径段的调度使用。

[0177] 如上所述，在控制脉冲串从入口节点中继段到中继段地发送到出口节点用于具有可变时间供应的端对端单向带宽保留之后，数据脉冲串沿着与控制脉冲串相同的光通路被发送到出口节点（在某一时间差后）。但是，数据脉冲串透明地通过交换节点发送而不检查其内容。PBS 交换织构在动态保留的持续时间内提供输入和输出端口之间的连接，因此允许数据脉冲串被发送通过，其中保留的光通路构成耦合入口和出口节点的“虚拟光学回路”。从 PBS 边缘节点 BGP 路由器观察，该虚拟光学回路表现为 BGP 路由器端点之间的直接连接，如 BGP 路由器 $BGB_1$ 和 $BGP_4$ 之间的虚拟链路 $914_{1-3}$ 所描绘的。

[0178] 从路由的观点，在确认形成因特网的 AS 的数量远大于典型的企业网所使用的数量的情况下，BGP 路由网络架构 900B 粗略地类似于因特网上的 BGP 路由。但是，路由原理是类似的。这样，使用公知的设置和配置方法，许多路由实现将类似于常规 BGP 路由所遇到的。

[0179] BGP 是当前的实际标准域间路由协议。BGP 首先于 1989 年变成因特网标准并且最初在 RFC（请求注释）1105 中被定义。随后，它被采纳作为域间路由选择的 EGP。当前版本，BGP-4，于 1995 年被采纳并于 RFC1771 中定义。

[0180] BGP 是通过发送路由通告进行工作的通路矢量 (path-vector) 协议。路由信息存储于每个 BGP 路由器处作为目的地和到达该目的地的通路的属性的组合。路由通告指示网络可达性（即，表示连续 IP 地址的块的网络地址和网络掩码）。除了可达网络和用于达到该网络的路由器的 IP 地址之外（称作下一个中继段），路由通告还包含 AS 通路属性，它包含了可用于达到所宣告的网络的所有转接 AS 的列表。AS 通路的长度可被认为是路由度量。

[0181] BGP 更新 (UPDATE) 消息用于在网络内出现变化时提供路由更新。为了设置不同 PBS “岛”或网络之间的光通路，需要扩展标准 BGP 以传送必要的光通路路由信息给 BGP 路

由器。目的是充分利用现有的 BGP 属性，但将它们扩展以满足 PBS 网络的路由需要。

[0182] PBS LER（标签边缘路由器）被指定为主要的 PBS BGP 路由器以支持不同光学域之间的路由。如图 9C 所示，BGP 路由器 $BGP_{1-6}$ 中的每一个都是 PBS LER 候选，尽管任何数量的 BGP 路由器 $BGP_{1-6}$ 可实际作为 PBS LER 操作。PBS BGP 路由器将负责通过将光通路属性通告给其相邻 BGP 路由器来设置光通路，并构建和维护用于所有可能路径的路由信息库 (RIB，即路由表）。一般，PBS BGP 路由器和 PBS LER 可共同定位于同一网络节点处。

[0183] 图 17 示出了具有其相应字段的更新 (UPDATE) 消息的格式。更新消息包括不可能路由长度字段 1700、撤回路由字段 1702、通路属性长度字段 1704、通路属性字段 1706 和网络层可达性信息 (NLRI) 字段 1708。在更新消息中，在一对 BGP 广播器 (speaker)（即，经由单个中继段彼此连接的 BGP 路由器）之间通告路由：目的地是在 NLRI 字段 1708 中报告其 IP 地址的系统，且通路是同一更新消息的通路属性字段 1706 中报告的信息。

[0184] 不可能路由长度字段 1700 包括 2 个八位字节的无符号整数，它指示八位字节中撤回路由字段的总长度。其值必须允许如以下规定地确定网络层可达性信息字段 1708 的长度。0 值指示没有路由正从服务中撤回，并撤回路由字段不存在于该更新消息中。

[0185] 撤回路由字段 1702 是可变长度字段，它包含了用于正从服务中撤回的路由的 IP 地址前缀的列表。每个 IP 地址前缀被编码为 2 字节组，它包括单个八位字节长度字段后面跟着可变长度前缀字段。长度字段以位数指示 IP 地址前缀的长度。零长度指示匹配所有 IP 地址的前缀（其中前缀本身是零八位字节）。前缀字段包含 IP 地址前缀后面跟着足够的拖尾位，以使得该字段的末端落到一八位字节的边界。

[0186] 总通路属性长度字段 1704 包括 2 个八位字节的无符号整数，它以八位字节表示通路属性字段 1706 的总长度。0 的值表示该更新消息中没有网络层可达性信息字段。

[0187] 常规通路属性字段 1706 的细节在图 17a 的 1706A 处示出。通路属性的可变长度序列存在于所有更新 (UPDATE) 中。每个通路属性都是三倍可变长度。属性类型是两个八位字节的字段，它包括属性标记八位字节 1710A 后面跟着属性类型代码八位字节 1712。属性标记八位字节的高位位（位 0）是任选位 1714。它定义属性是任选的（如果是则设定为 1）还是公知的（如果是则设定为 0）。

[0188] 属性标记八位字节的第二高位位（位 1）是过渡位 (transitive bit)1716。它定义任选属性是过渡的（如果是则设定为 1）还是非过渡的（如果是则设定为 0）。对于公知属性，过渡属性必须设定为 1。

[0189] 属性标记八位字节的第三高位位（位 2）是部分位 1718。它定义任选过渡属性中包含的信息是部分的（如果是则设定为 1）还是完整的（如果是则设定为 0）。对于公知属性且对于任选非过渡属性，部分位必须设定为 0。

[0190] 属性标记八位字节的第四高位位（位 3）是扩展长度位 1720。它定义属性长度是一个八位字节（如果是则设定为 0）还是两个八位字节（如果是则设定为 1）。只有当属性值长度大于 255 个八位字节时，才使用扩展长度位 1720。

[0191] 属性标记八位字节的低位的四个位不使用，如保留字段 1722 所描绘的。它们必须是零（且在被接收到时必须被忽略）。

[0192] 属性类型代码八位字节 1712 包含属性类型代码。当前定义的属性类型代码在 RFC 1771 的部分 5 中讨论。

[0193] 如果属性标记八位字节1710的扩展长度位1720被设定为0，通路属性的第三个八位字节包含以八位字节计的属性数据的长度。如果属性标记八位字节的扩展长度位被设定为1，则通路属性的第三和第四八位字节包含以八位字节计的属性数据的长度。属性长度代码1724描述了这两种情况。通路属性的其余八位字节表示属性值1726并根据属性标记1710和属性类型代码1712进行解释。

[0194] 其中，更重要的属性类型代码中有ORIGIN(类型代码1)、AS_PATH(类型代码2)和NEXT_HOP(类型代码3)。ORIGIN是定义通路信息的来源的公知强制属性。AS_PATH是由AS通路段的序列构成的公知强制属性。每个AS通路段都由三元组表示。通路段类型是1个八位字节长的字段，而通路段长度是包含通路段值字段中AS数量的1个八位字节长的字段。通路段值字段包含一个或多个AS数，每个都被编码为2个八位字节长的字段。NEXT_HOP是公知强制属性(RFC1771)，它定义应用作到达更新消息的网络层可达字段中列出的目的地的BGP下一个中继段的路由器的IP地址。路由器进行递归查找以找出路由表中的BGP下一个中继段。

[0195] 根据将BGP路由扩展到光交换网络的各方面，按照一个实施例，图17b示出了包含附加信息（在粗体线的框中所示）的一组修改通路属性1706B的细节，所示附加信息用于指明光传输属性以便将BGP协议扩展到光交换网络。这些扩展包括PBS连接(PC)字段1726、可用波长属性字段1728和可用光纤属性字段1730。PC字段1726对应于属性标记八位字节1710B的位4。0值指示PBS连接不可用。1值指示PBS连接可用。

[0196] 可用波长属性字段1728中的值指示相邻PBS网络（光域）之间当前波长可用性的状态。如果该值为0，则无波长可用于所请求的光通路。任何包含的值都对应于可用于所请求的光通路的一个或多个波长。这意味着与PBS LER共同定位的BGP路由器可开始到特定目的地的光通路设置处理。

[0197] 可用光纤属性字段1730中的值指示相邻PBS网络之间当前光纤可用性的状态。0值指示该光纤不可用于所请求的光通路。这意味着该光纤由其它波长使用或者该光纤链路关闭。在任一情况中，必须选择备用路径。非零值指示光纤可由所请求的到达目的地地址的光通路使用。

[0198] 返回到图17，网络层可达性信息字段1708包括包含IP地址前缀列表的可变长度字段。网络层可达性信息的以八位字节计的长度不是明确地编码，但可以计算为：

[0199] 可达性信息被编码为长度(Length)(1八位字节)、前缀(Prefix)(可变长度)的形式的一个或多个2元组。长度字段指示IP地址前缀的位的长度。零长度表示匹配所有IP地址的前缀（本身为零八位字节的前缀）。前缀字段包含IP地址前缀，继之以足够的尾位以使得字段的末端落到一个八位字节的边界，其中尾位的值是不相关的。

[0200] BGP中的更新消息最相关于PBS BGP的设计和操作，因为它们将新的路由可用性信息从一个路由器传到另一路由器。例如，网络拓扑（从BPG路由器的立场）可通过经由相应的更新消息对相邻BPG路由器作出的通告表达。这些原理是网络路由领域内的熟练技术人员公知的。

[0201] 图18示出了概括前述设置和网络更新操作的流程图。设置过程开始于框1800，其中共同定位于各SAN网关的多个PBS交换/边缘节点模块被配置成使能相互之间的数据传输通路，因此在使能PBS连网基础结构上各SAN之间基于PBS的数据传输。一般，通信链路

可包括各光学 I/O 模块 1302 之间的一个或多个光纤链路。

[0202] 接着，在框 1802 中，从沿着跨多个 BGP 路由器的路径路由数据的立场，每个 SAN 都被建模为自主系统 (AS)。随后，选定的共同定位 PBS 交换 / 边缘模块被指定为用作 SAN 之间外部路由的 BGP 路由器，如框 1804 所述。

[0203] 框 1806 中，每个 BGP 路由器指定模块接收 PBS 网络内其它节点的路由可用性信息，它标识可用于在该节点和网络内的其它 BGP 路由器之间传输数据的路径。这么作是提供标识给定 PBS 网络内入口和出口 BGP 路由器之间的可用路径的路由信息。随后，在框 1808 中生成包含用于这些路径的通告的相应 BGP 更新消息，其中 BGP 更新消息具有图 17b 所示的通路属性格式。

[0204] 在这点上，在 BGP 路由器近邻之间交换包括光交换网络路由支持扩展的 BGP 更新消息，以更新每个 BGP 路由器中的外部路由表。这些操作在框 1810 和 1812 中执行。每个外部路由表都包含多个路由记录，每一个都指明到目的地网络的路径。特别地，每个路由记录都包括将顺序遇到以达到具有目的地地址的 SAN 处的入口节点 BGP 路由器的段中继段（即，BGP 路由器地址）的列表。该外部路由数据不包括 AS 内使用的内部路由的任何细节。

[0205] 一旦企业网被配置和初始化（即，建立 BGP 路由表），就可以通过将扩展 BGP 路由用于外部路由操作并将 IGP 路由机制用于给定 PBS 网络内的内部路由而在不同 PBS 网络之间以及不同 PBS 网络和非 PBS 网络之间传输数据。因此，路由类似于因特网所采用的路由，区别在于除常规的外部路由通告外在更新它们的路由表时现在路由器考虑光交换网络可用性信息。

[0206] 当用作沿着给定路径的中间节点时，PBS 交换 / 边缘节点模块将提供类似于以上讨论的 PBS 交换模块 217 的 PBS 交换机功能。同时，源 SAN 处的 PBS 交换 / 边缘节点模块将用作 BGP 路由器和 PBS 出口节点，其中目的地 SAN 处的 PBS 交换 / 边缘节点模块将用作 PBS 入口节点。

[0207] 返回到图 9a，在一个实施例中，前述 BGP 路由器功能可以在一个或多个 PBS 边缘节点 910 中实现，如通过 BGP 路由器模块 916 所描述的。在该实施例中，PBS 边缘节点 910 将提供 EGP 路由功能，并提供 PBS 边缘节点和共同定位的 SAN 网关操作。

[0208] 一般，BGP 路由器功能可由分开的服务器模块提供，或者可以集成于单元 1300 的现有组件上，诸如集成于光学 PBS I/O 模块 1302。如同前述 PBS 交换节点和边缘节点功能，路由器功能可通过硬件（例如编程逻辑）、软件或两者的组合而实现。更特别地，用于实现 PBS 交换节点、边缘节点、SAN 网关和 / 或 BGP 路由器功能的软件可体现为一组或多组指令或包含在某种形式的处理器核心（诸如网络处理器、服务器或 I/O 模块的处理器或其它类型的处理器）上执行的指令的模块。

[0209] 因此，本发明的实施例可用作或支持在某种形式的处理核心上执行或另外地在机器可读媒介之上或之内执行或实现的软件程序。机器可读媒介包括用于存储或传送可由机器（例如计算机）读取的形式的信息的任何机制。例如，机器可读媒介可包括诸如只读存储器 (ROM)；随机存取存储器 (RAM)；磁盘存储媒体；光学存储媒体；以及闪存装置等等。此外，机器可读媒介可包括传播信号，诸如电、光、声或其它形式的传播信号（例如，载波、红外线信号、数字信号等）。

[0210] 在前述说明书中，已描述了本发明的实施例。但显然，可对其进行各种修改和变化

而不背离所附权利要求书中所阐述的较宽精神和范围。因此，说明书和附图被认为是说明性而非限制性的。

[0211] 包含摘要中所描述的内容的所述本发明实施例的以上描述不被认为是穷尽性或者将本发明限制于所揭示的精确形式。虽然这里出于说明目的描述了本发明的特定实施例和示例，但各种等效修改也在本发明的范围之内，如相关领域的熟练技术人员能认识到的。

[0212] 可以根据以上的详细描述对本发明进行这些修改。以下权利要求书中所使用的术语不应认为将本发明限制于说明书和权利要求书中所揭示的特定实施例，相反，本发明的范围完全由以下的权利要求书确定，它是根据所建立的权利要求说明原则进行解释的。

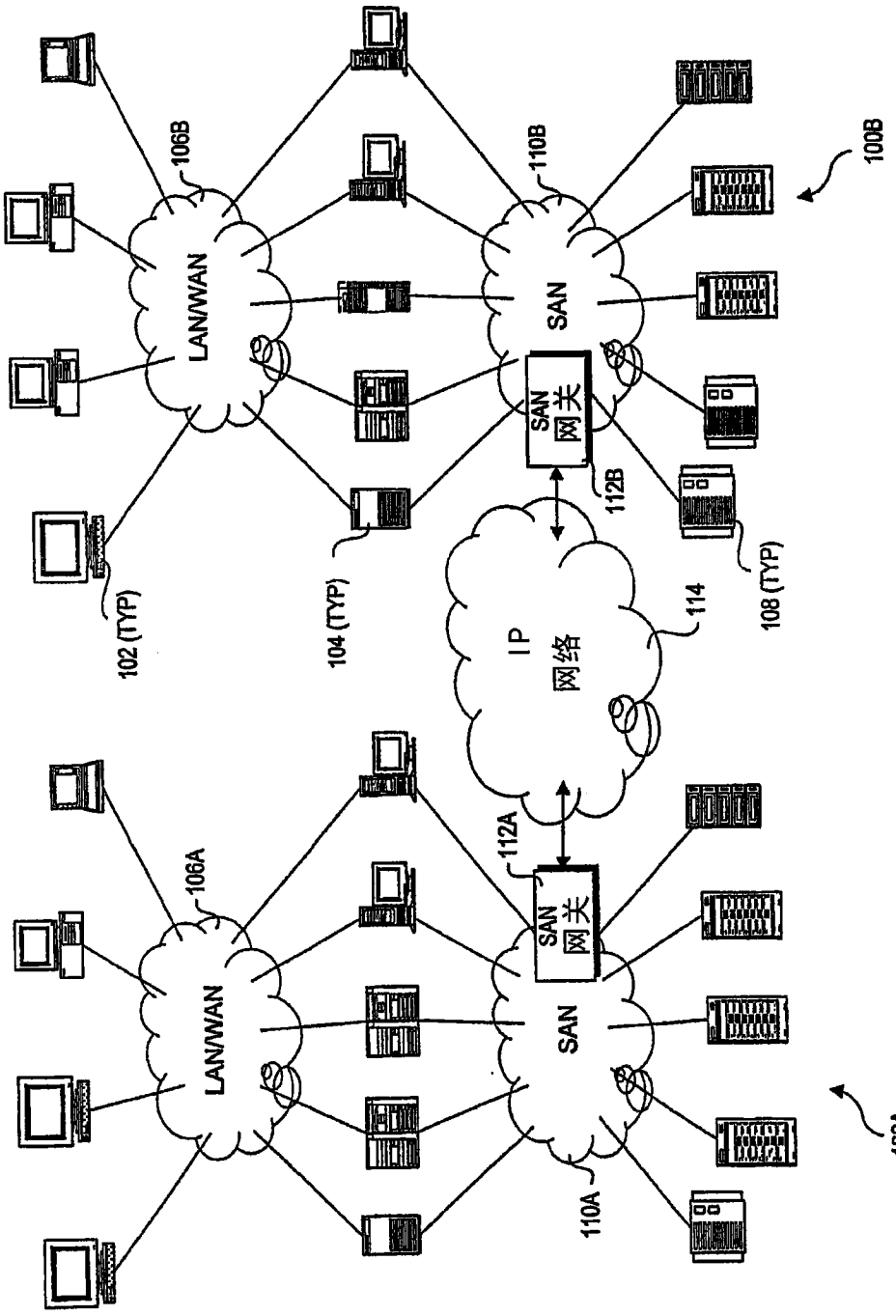

106B
LAN/WAN
110B
SAN
100B
SAN 网关
112B
102 (TYP)
104 (TYP)
108 (TYP)
IP 网络
114
106A
112A
SAN 网关
LAN/WAN
SAN
110A
100A

图 1
(现有技术)

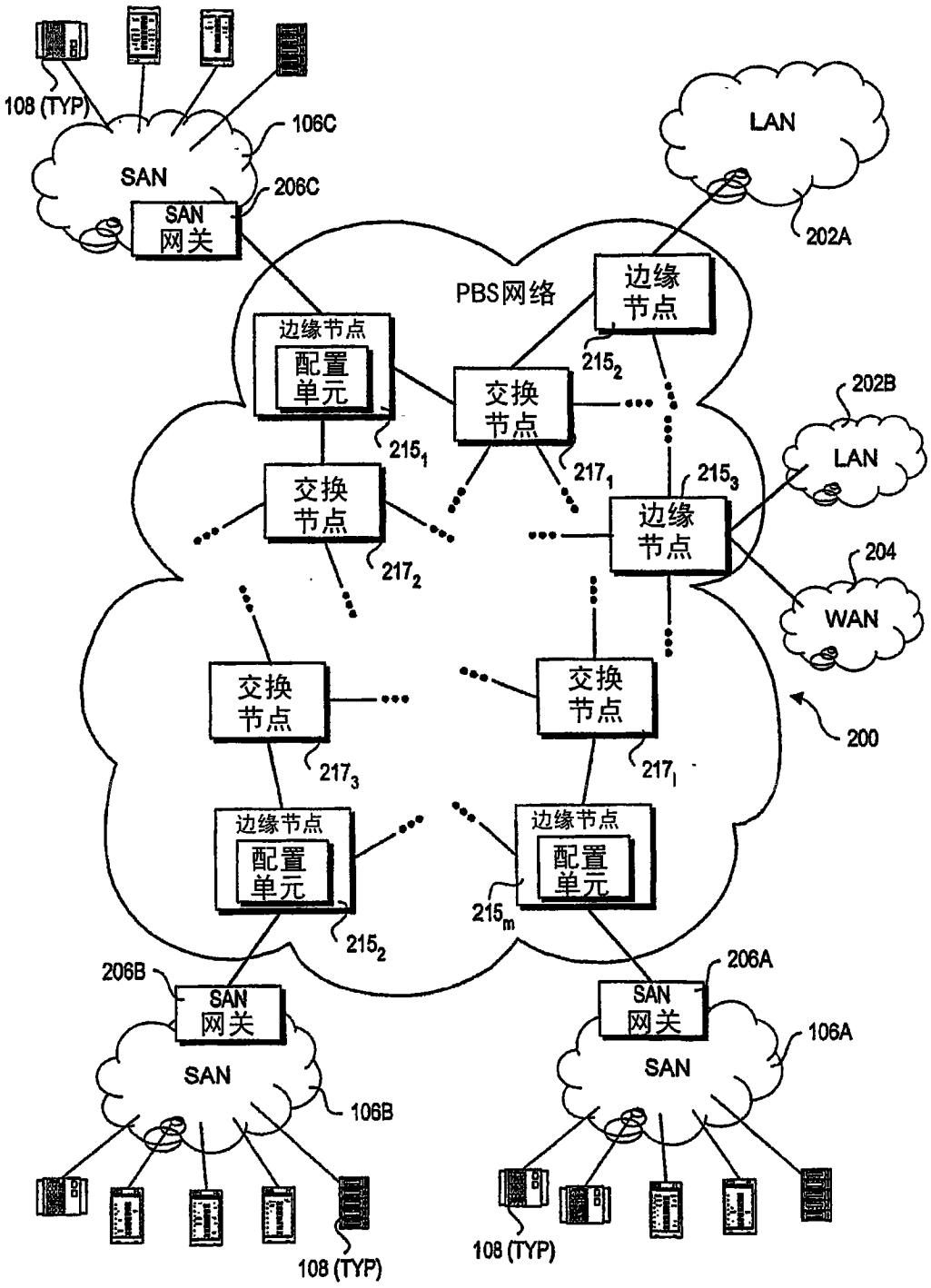

108 (TYP)
106C
SAN
206C
SAN
网关
LAN
202A
PBS网络
边缘
节点
$215_2$
边缘节点
配置
单元
交换
节点
$215_1$
$217_1$
$215_3$
202B
LAN
交换
节点
$217_2$
边缘
节点
204
WAN
交换
节点
交换
节点
200
$217_3$
$217_1$
边缘节点
配置
单元
边缘节点
配置
单元
$215_m$
$215_2$
206B
SAN
网关
206A
SAN
网关
106A
SAN
SAN
106B
108 (TYP)
108 (TYP)

图 2

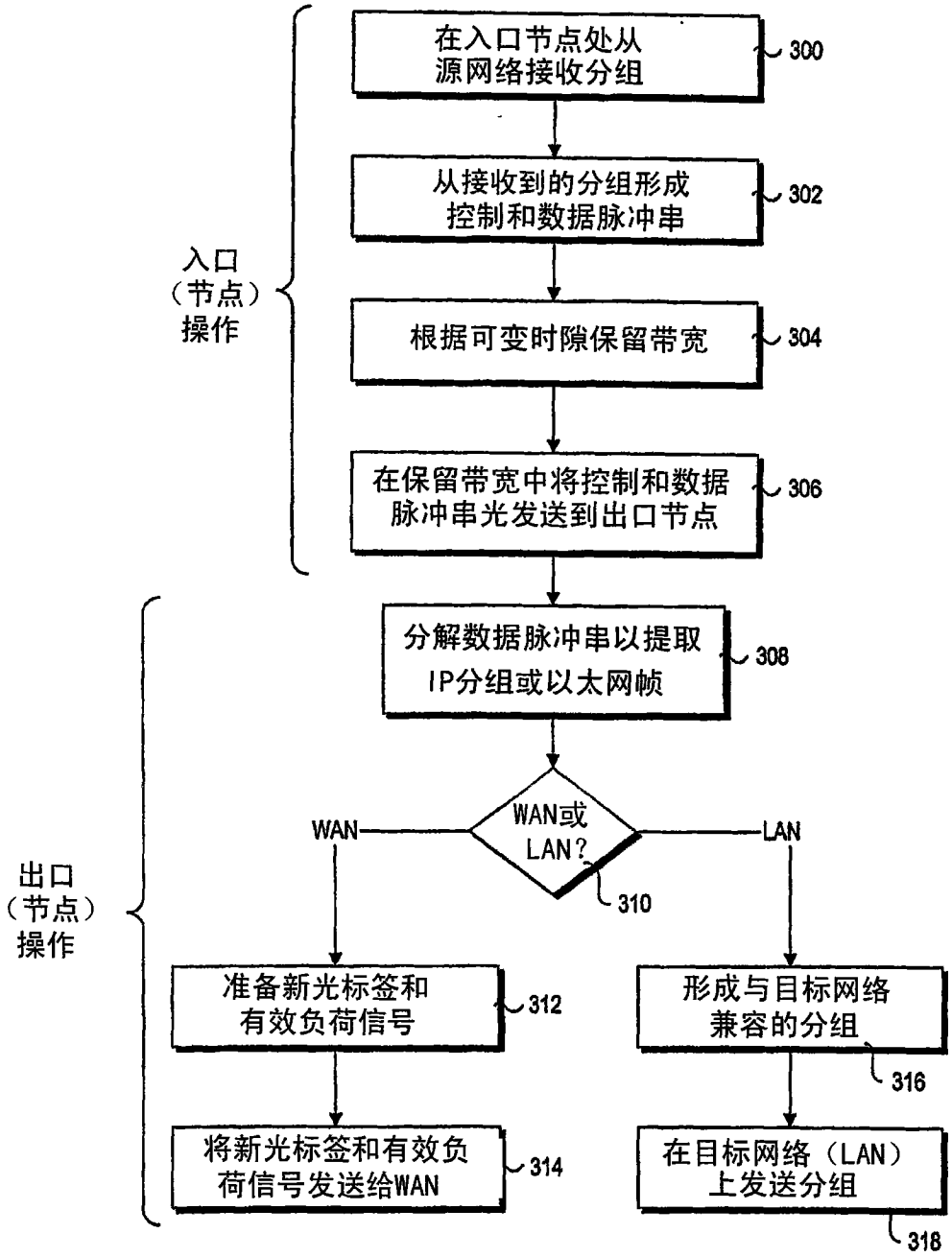

入口
（节点）
操作
在入口节点处从
源网络接收分组
300
从接收到的分组形成
控制和数据脉冲串
302
根据可变时隙保留带宽
304
在保留带宽中将控制和数据
脉冲串光发送到出口节点
306
出口
（节点）
操作
分解数据脉冲串以提取
IP分组或以太网帧
308
WAN或
LAN?
310
WAN
LAN
准备新光标签和
有效负荷信号
312
将新光标签和有效负
荷信号发送给WAN
314
形成与目标网络
兼容的分组
316
在目标网络（LAN）
上发送分组
318

图 3

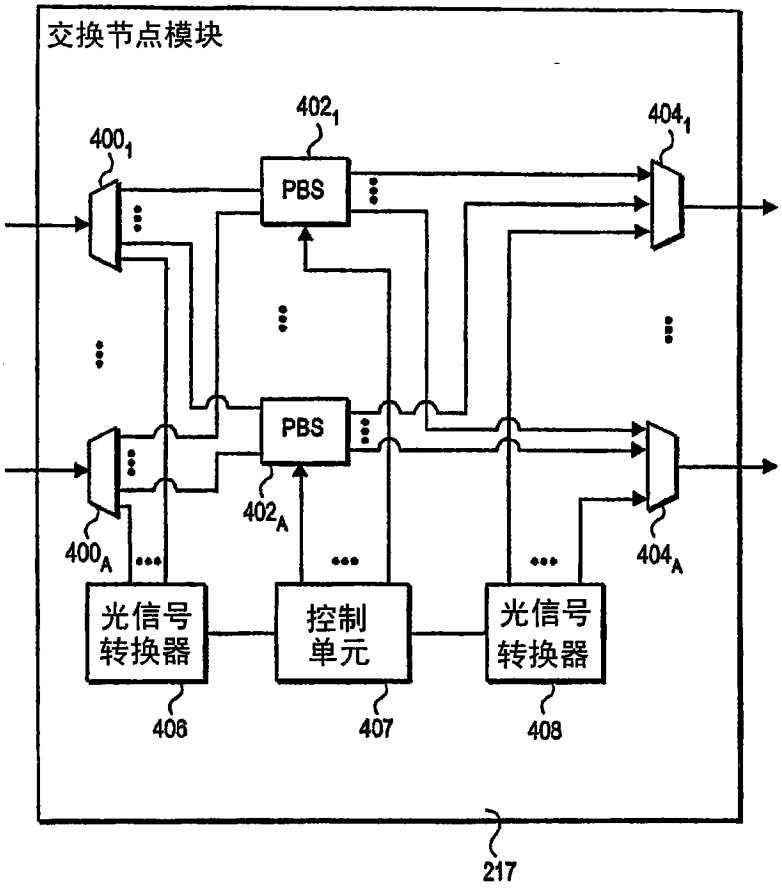

交换节点模块
$402_1$
$400_1$
$404_1$
PBS
PBS
$402_A$
$400_A$
$404_A$
光信号转换器
控制单元
光信号转换器
406
407
408
217

图 4

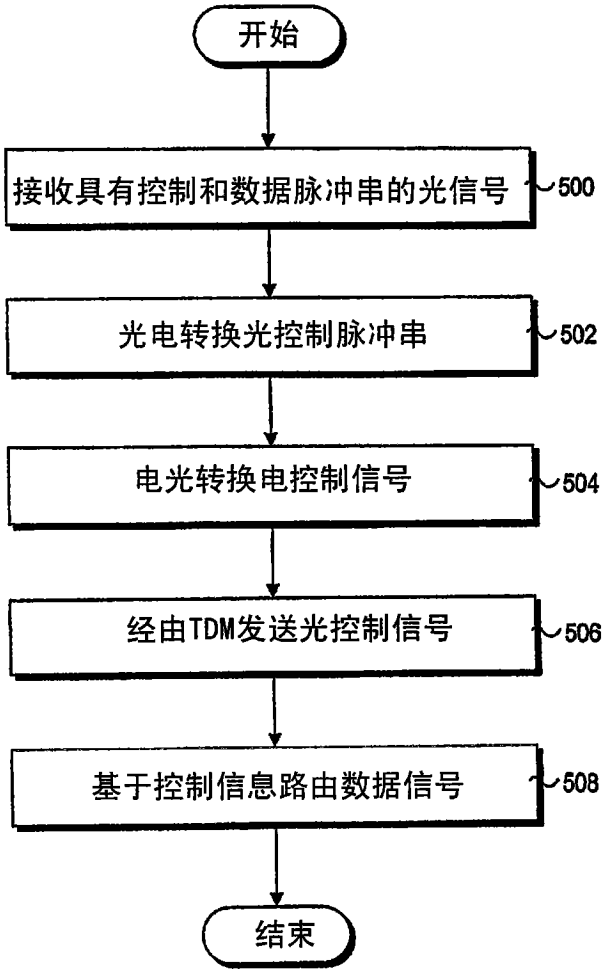

开始
接收具有控制和数据脉冲串的光信号
500
光电转换光控制脉冲串
502
电光转换电控制信号
504
经由TDM发送光控制信号
506
基于控制信息路由数据信号
508
结束

图 5

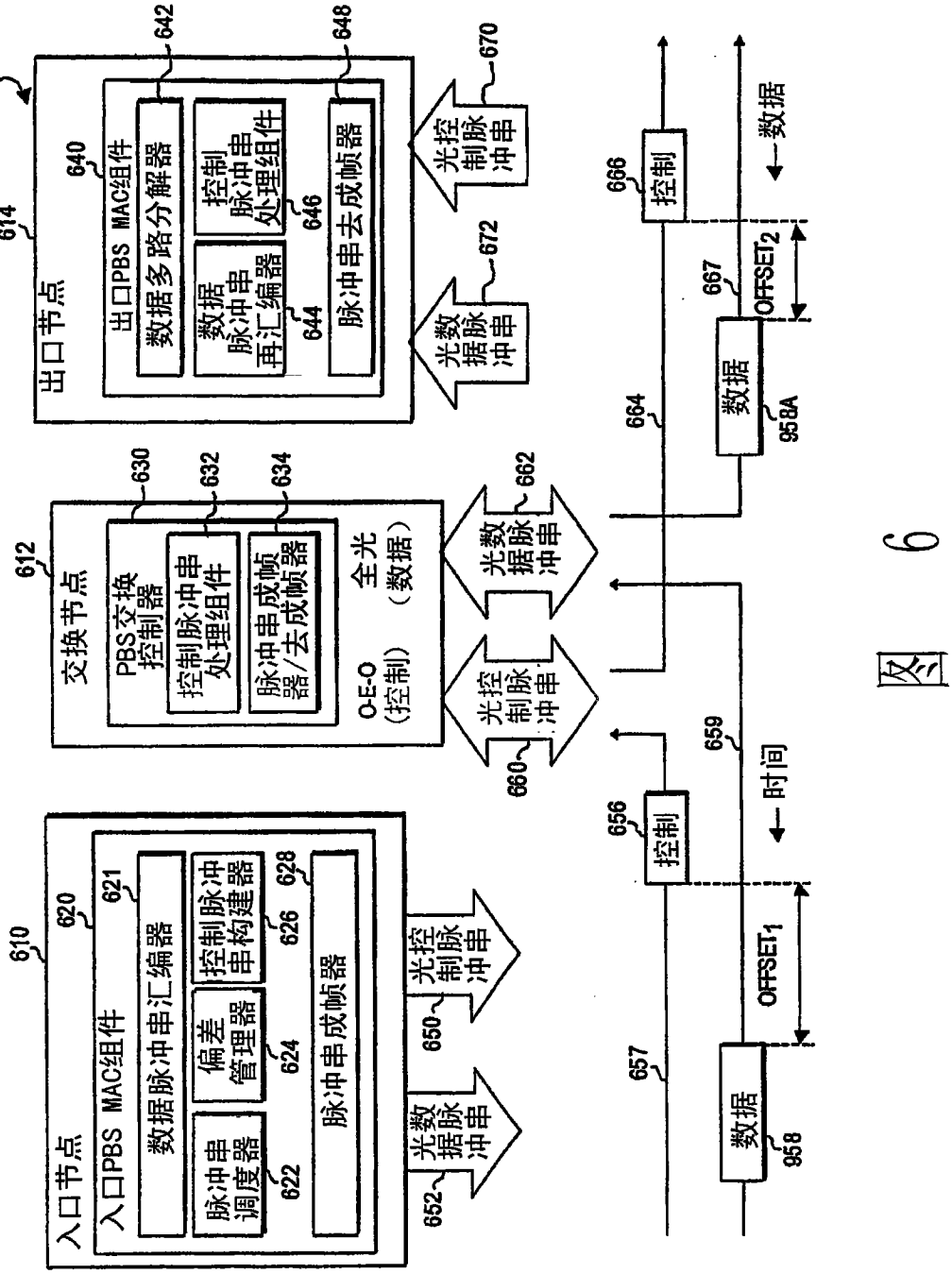

600
出口节点
614
出口PBS MAC组件
640
数据多路分解器
642
数据脉冲串再汇编器
644
脉冲串去成帧器
648
控制脉冲串处理组件
646
光控制脉冲串
670
光数据脉冲串
672
交换节点
612
PBS交换控制器
630
控制脉冲串处理组件
632
脉冲串成帧器/去成帧器
634
O-E-O (控制) 全光 (数据)
光数据脉冲串
662
光控制脉冲串
660
入口节点
610
入口PBS MAC组件
620
数据脉冲串汇编器
621
脉冲串调度器
622
偏差管理器
624
控制脉冲串构建器
626
脉冲串成帧器
628
光数据脉冲串
652
光控制脉冲串
650
664
666
控制
667
数据
OFFSET2
958A
656
控制
659
时间
OFFSET1
657
数据
958
数据

图 6

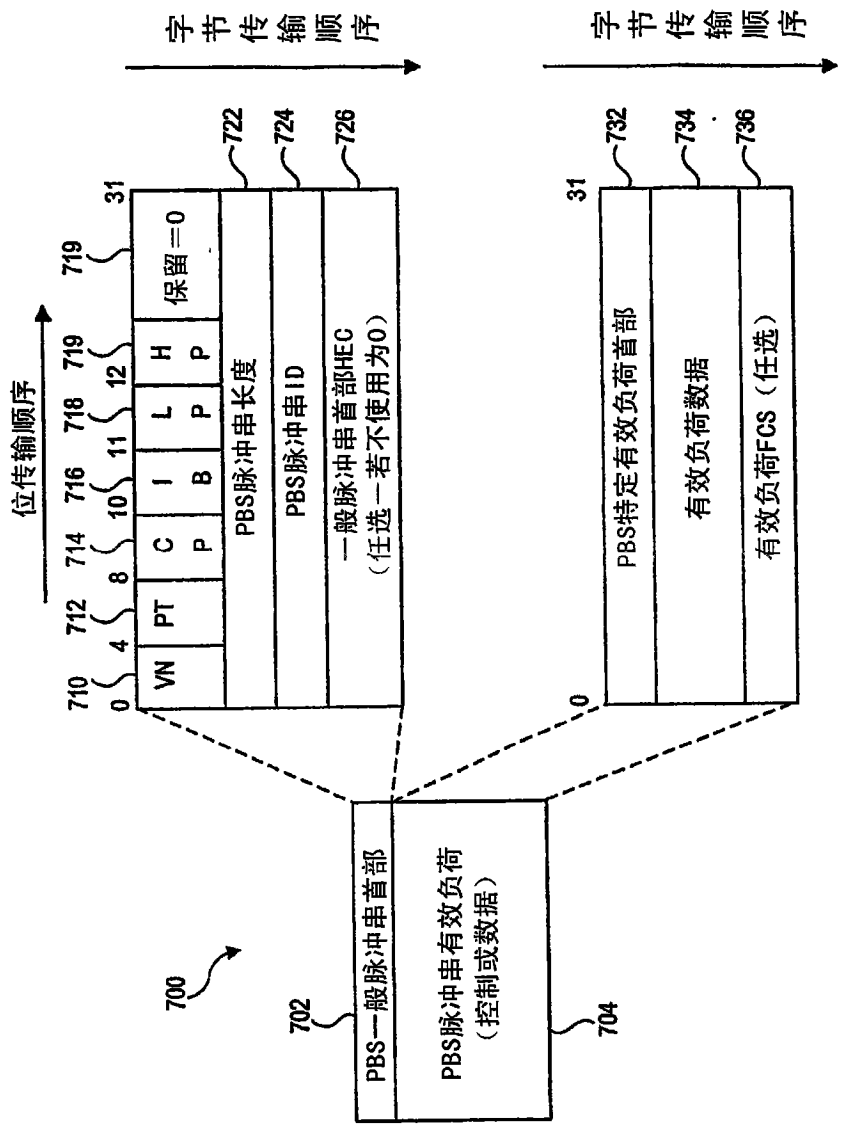

字节传输顺序
字节传输顺序
位传输顺序
31
0
4
8
10
11
12
710
712
714
716
718
719
719
719
722
724
726
732
734
736
VN
PT
C P
I B
L P
H P
保留=0
PBS脉冲串长度
PBS脉冲串ID
一般脉冲串首部HEC（任选－若不使用为0）
PBS特定有效负荷首部
有效负荷数据
有效负荷FCS（任选）
700
702
704
PBS一般脉冲串首部
PBS脉冲串有效负荷（控制或数据）

图 7

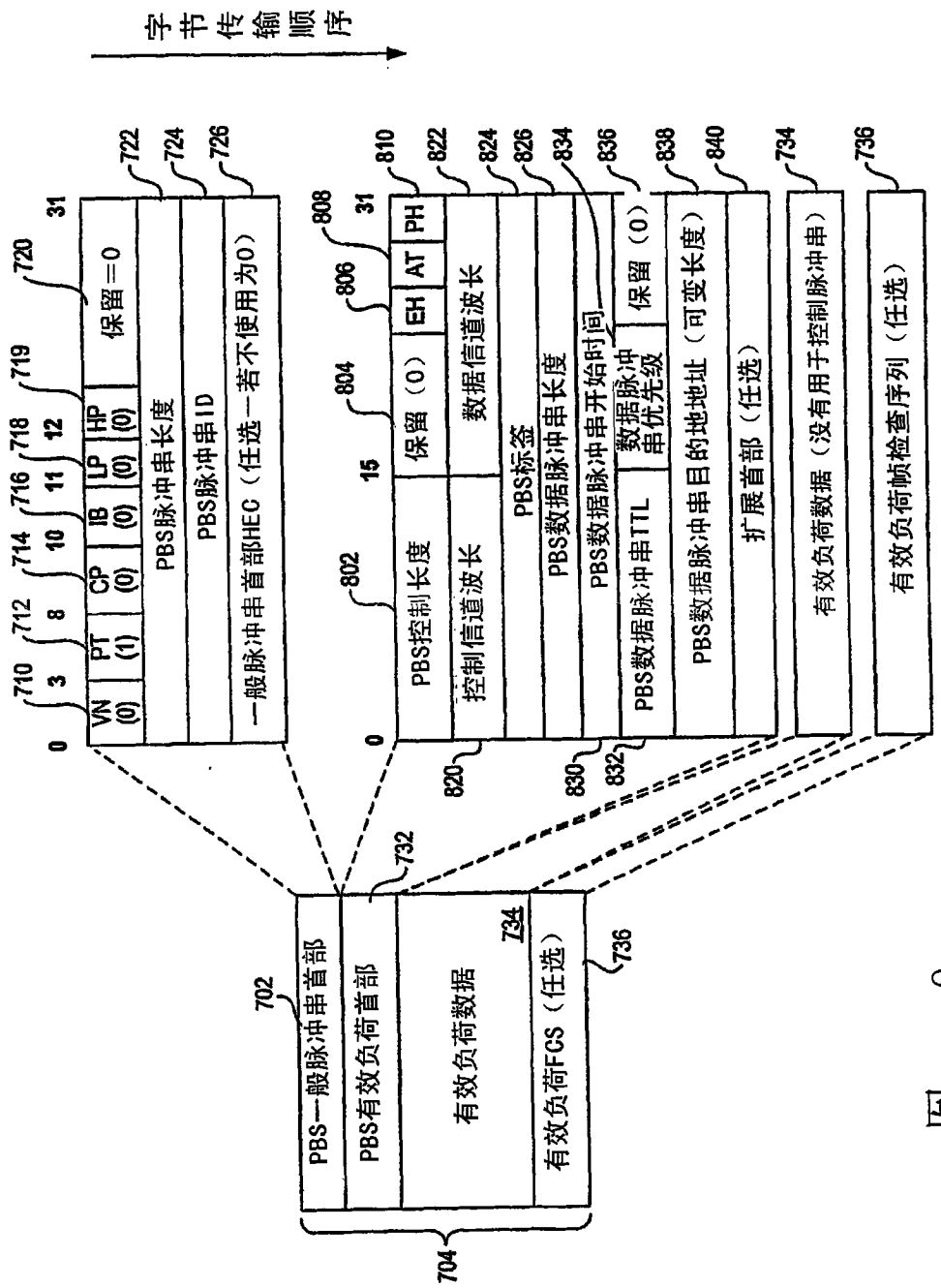

字节传输顺序
0
3
8
10
11
12
15
31
VN (0)
PT (1)
CP (0)
IB (0)
LP (0)
HP (0)
保留=0
PBS脉冲串长度
PBS脉冲串ID
一般脉冲串首部HEC（任选—若不使用为0）
PBS控制长度
保留（0）
EH
AT
PH
控制信道波长
数据信道波长
PBS标签
PBS数据脉冲串长度
PBS数据脉冲串开始时间
PBS数据脉冲串TTL
数据脉冲串优先级
保留（0）
PBS数据脉冲串目的地址（可变长度）
扩展首部（任选）
有效负荷数据（没有用于控制脉冲串）
有效负荷帧检查序列（任选）
PBS一般脉冲串首部
PBS有效负荷首部
有效负荷数据
有效负荷FCS（任选）
702
704
710
712
714
716
718
719
720
722
724
726
732
734
736
802
804
806
808
810
820
822
824
826
830
832
834
836
838
840

图 8

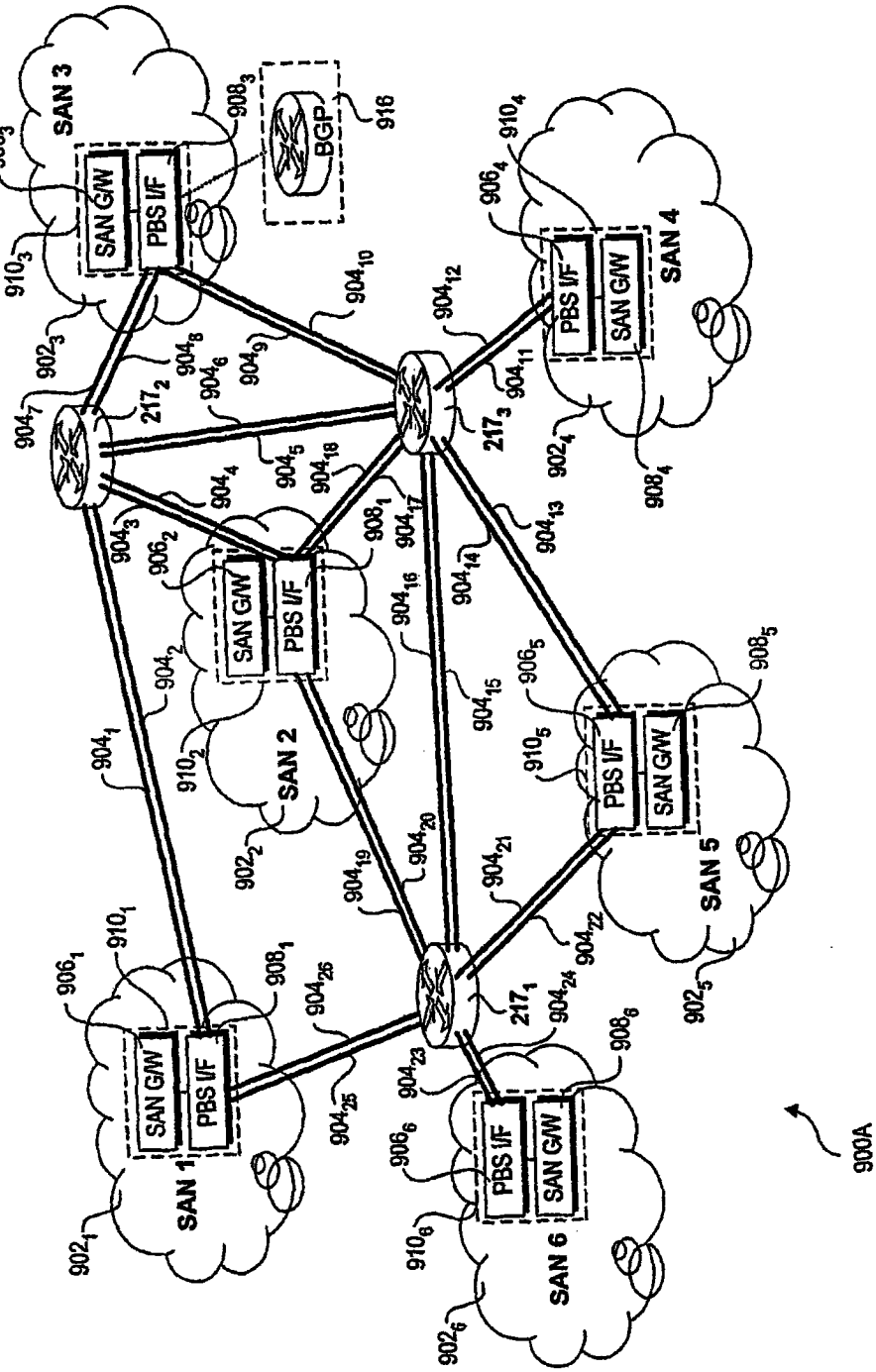



图 9a

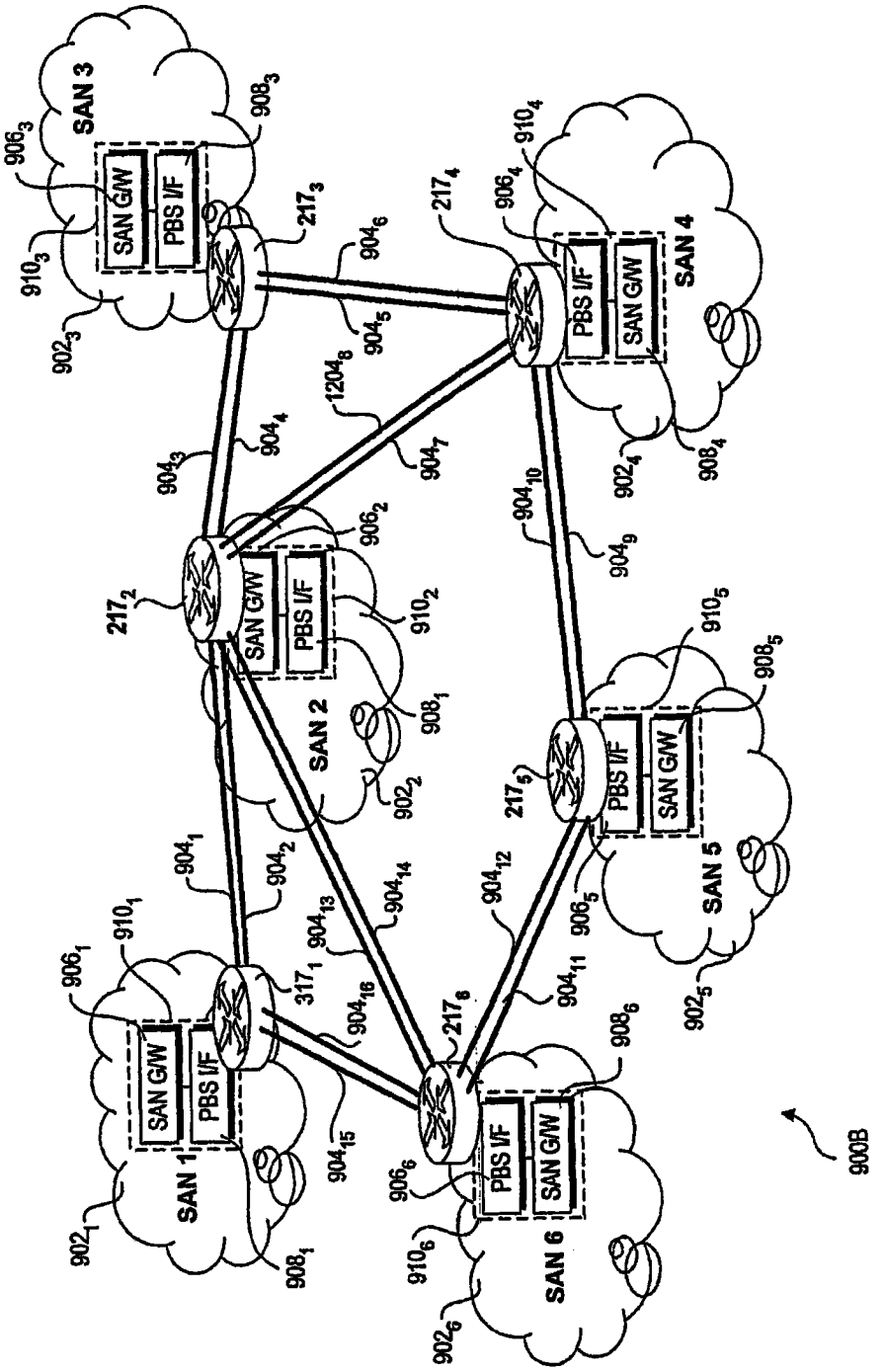



图 9b

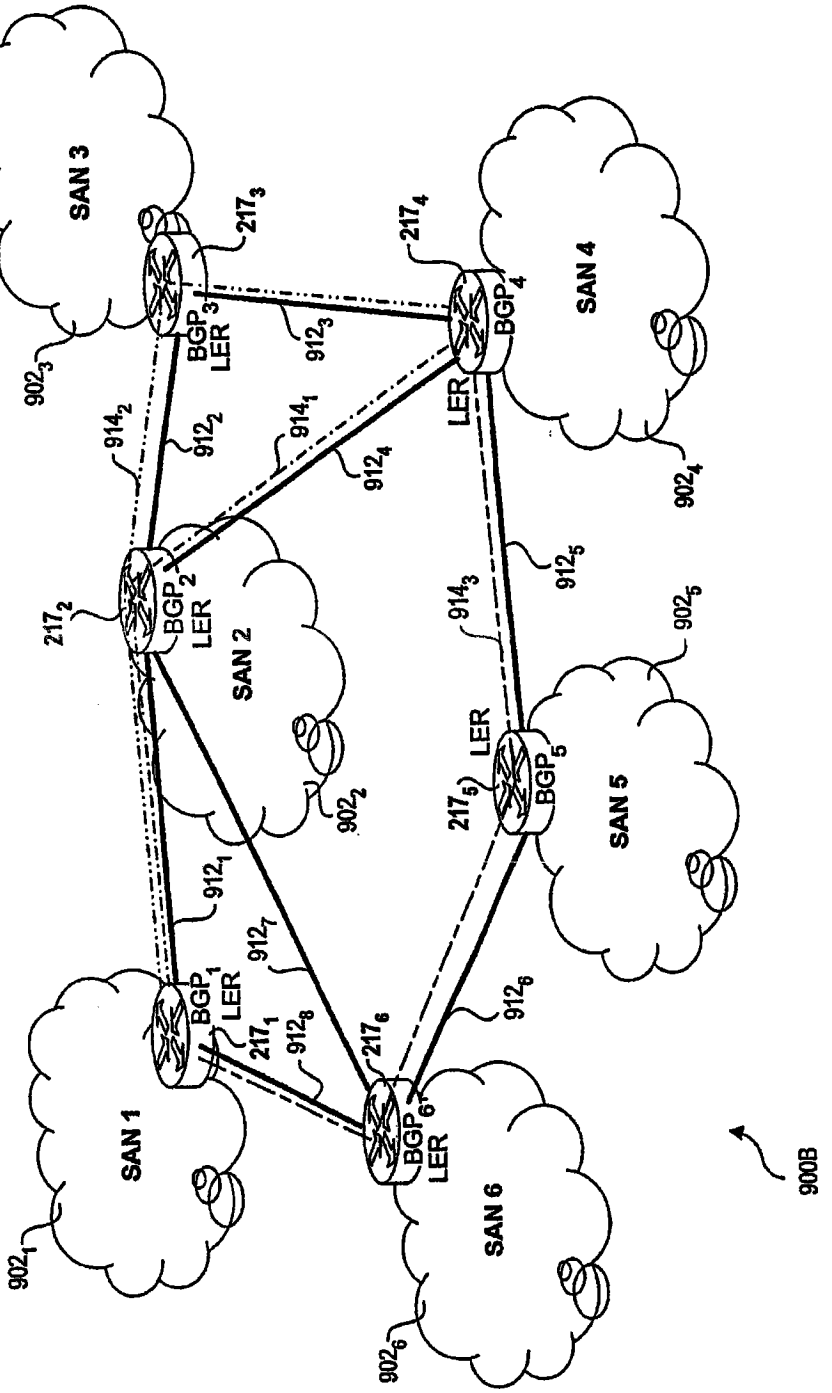

SAN 1
SAN 2
SAN 3
SAN 4
SAN 5
SAN 6
BGP LER 1
BGP LER 2
BGP LER 3
BGP 4
LER
BGP 5
LER
BGP LER 6
900B

图 9C

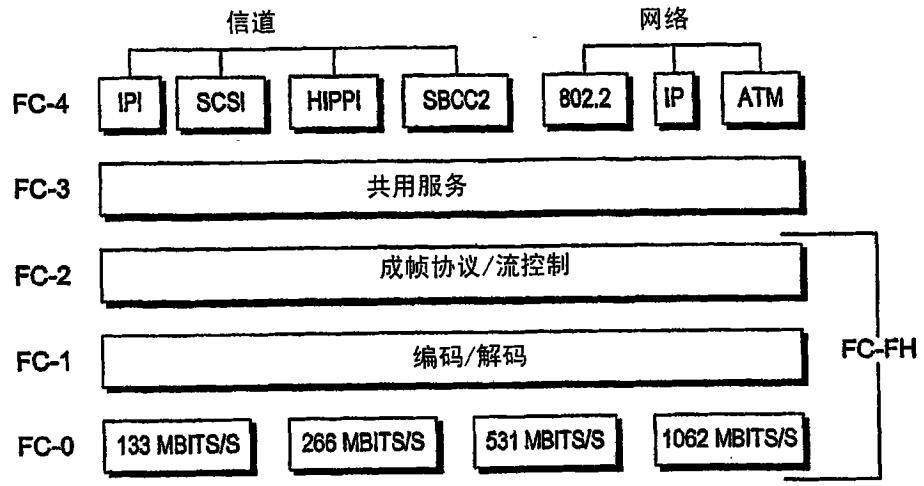

信道
网络
FC-4
IPI
SCSI
HIPPI
SBCC2
802.2
IP
ATM
FC-3
共用服务
FC-2
成帧协议/流控制
FC-1
编码/解码
FC-FH
FC-0
133 MBITS/S
266 MBITS/S
531 MBITS/S
1062 MBITS/S

图 10

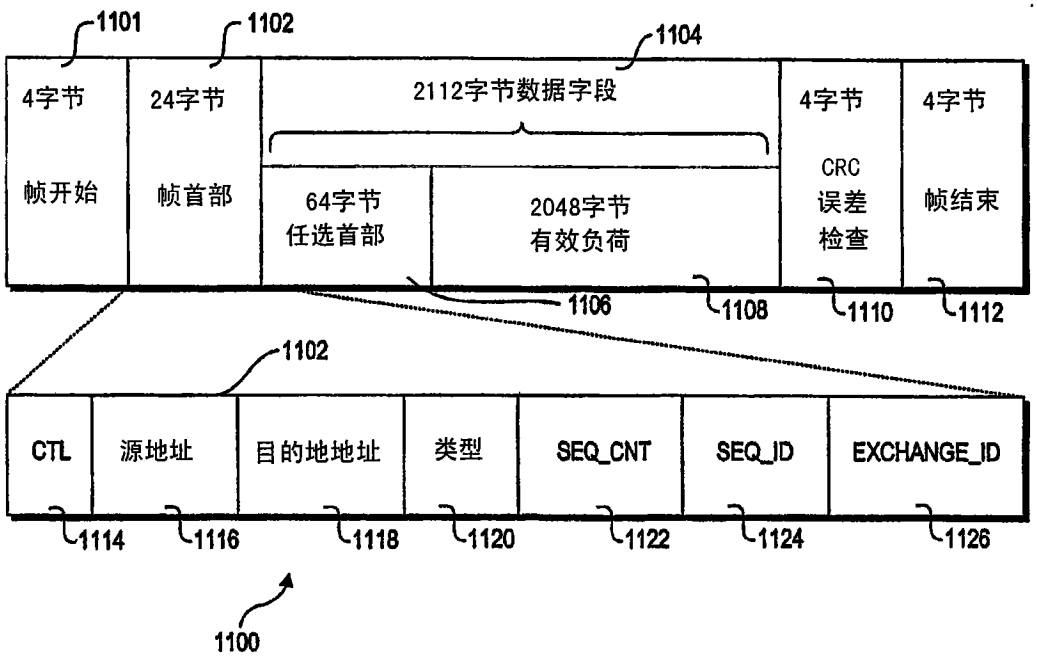

1101
1102
1104
4字节
帧开始
24字节
帧首部
2112字节数据字段
64字节
任选首部
2048字节
有效负荷
4字节
CRC
误差
检查
4字节
帧结束
1106
1108
1110
1112
1102
CTL
源地址
目的地地址
类型
SEQ_CNT
SEQ_ID
EXCHANGE_ID
1114
1116
1118
1120
1122
1124
1126
1100

图 11

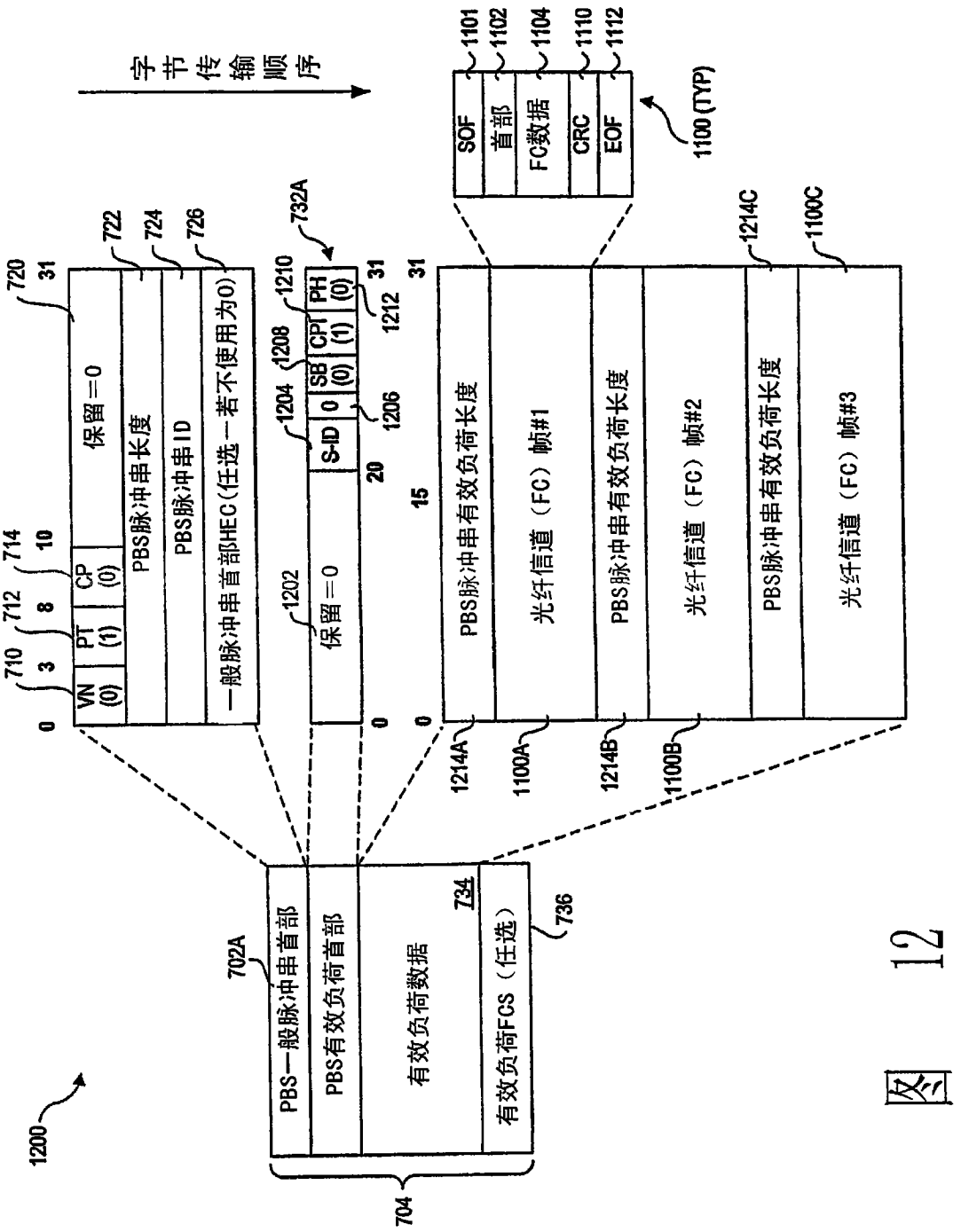

字节传输顺序
SOF
首部
FC数据
CRC
EOF
1100 (TYP)
PBS—般脉冲串首部
PBS有效负荷首部
有效负荷数据
有效负荷FCS（任选）
光纤信道（FC）帧#1
光纤信道（FC）帧#2
光纤信道（FC）帧#3
PBS脉冲串有效负荷长度
PBS脉冲串有效负荷长度
PBS脉冲串有效负荷长度
VN (0)
PT (1)
CP (0)
保留=0
PBS脉冲串长度
PBS脉冲串ID
一般脉冲串首部HEC(任选—若不使用为0)
保留=0
S-ID
SB (0)
CPT (1)
PH (0)
1200
704
702A
734
736
720
722
724
726
732A
1202
1204
1206
1208
1210
1212
1214A
1214B
1214C
1100A
1100B
1100C
1101
1102
1104
1110
1112

图 12

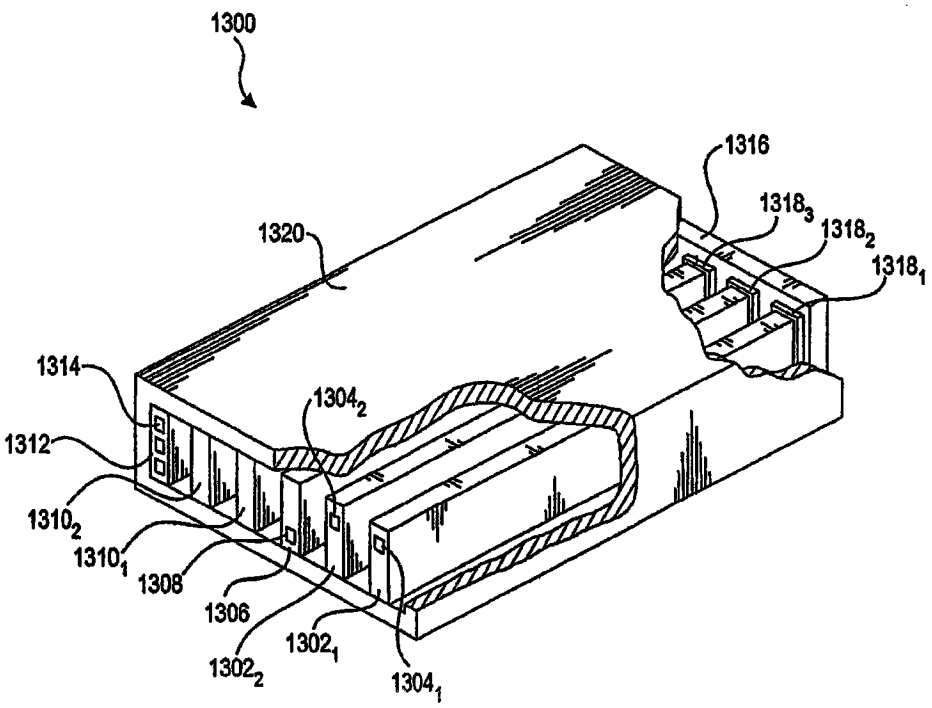

1300
1316
1318$_3$
1318$_2$
1318$_1$
1320
1314
1304$_2$
1312
1310$_2$
1310$_1$
1308
1306
1302$_1$
1302$_2$
1304$_1$

图 13

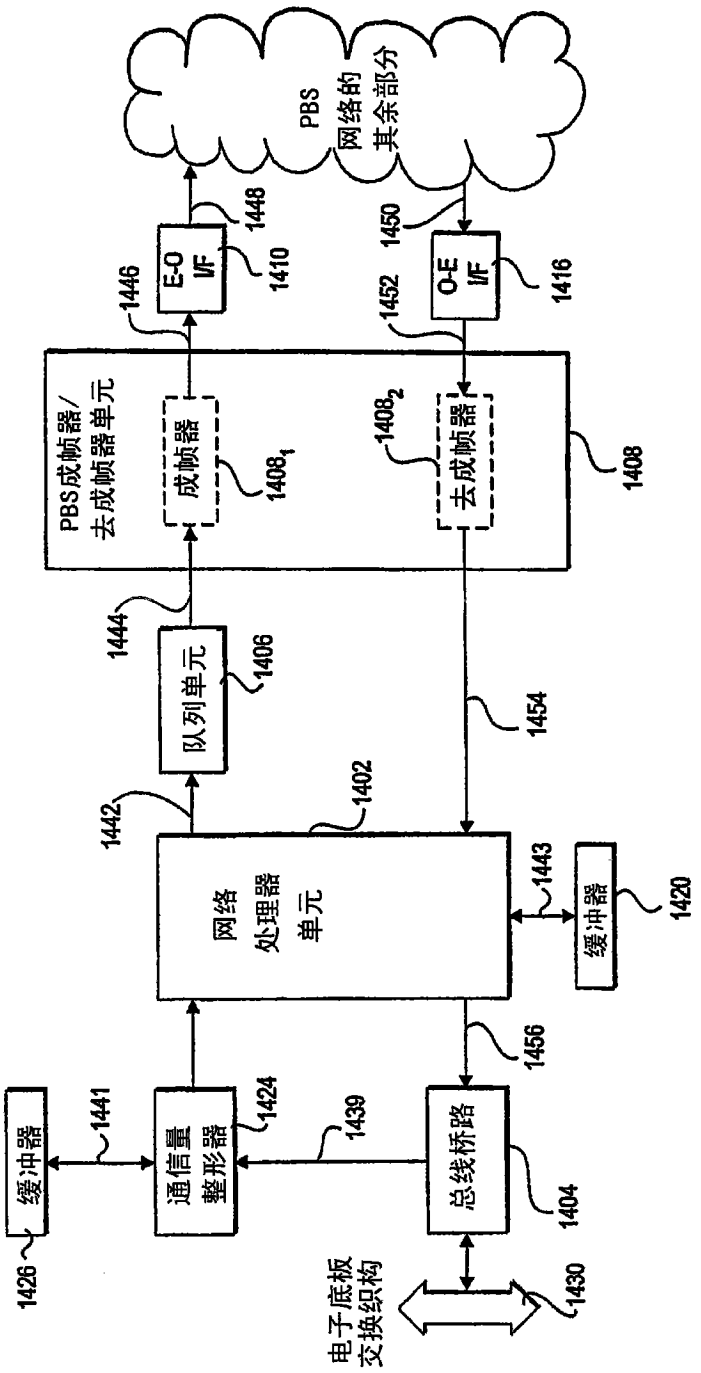

PBS
网络的
其余部分
1448
1450
E-O
I/F
1410
O-E
I/F
1416
1446
1452
PBS成帧器/
去成帧器单元
成帧器
1408₁
1408₂
去成帧器
1408
1444
队列单元
1406
1454
1442
1402
网络
处理器
单元
1443
缓冲器
1420
1456
缓冲器
1441
通信量
整形器
1424
1439
总线桥路
1404
1426
电子底板
交换织构
1430

图 14a

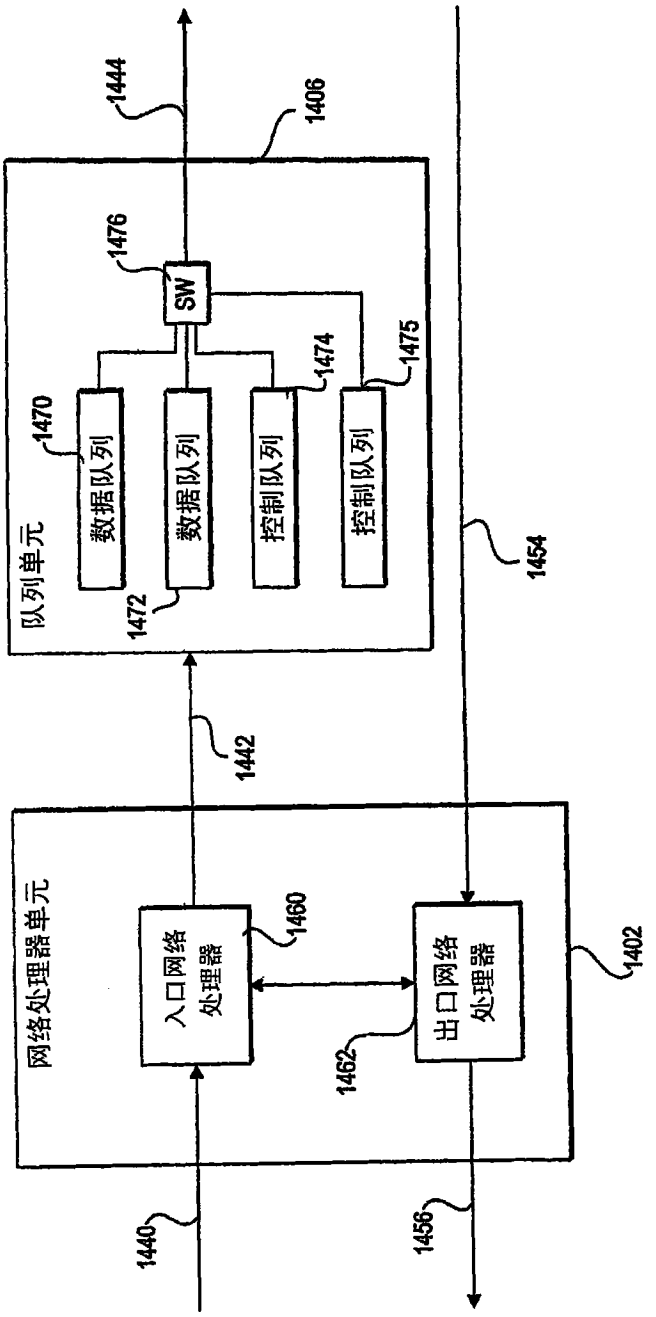

网络处理器单元
入口网络处理器
出口网络处理器
队列单元
数据队列
数据队列
控制队列
控制队列
SW
1402
1406
1440
1442
1444
1454
1456
1460
1462
1470
1472
1474
1475
1476

图 14b

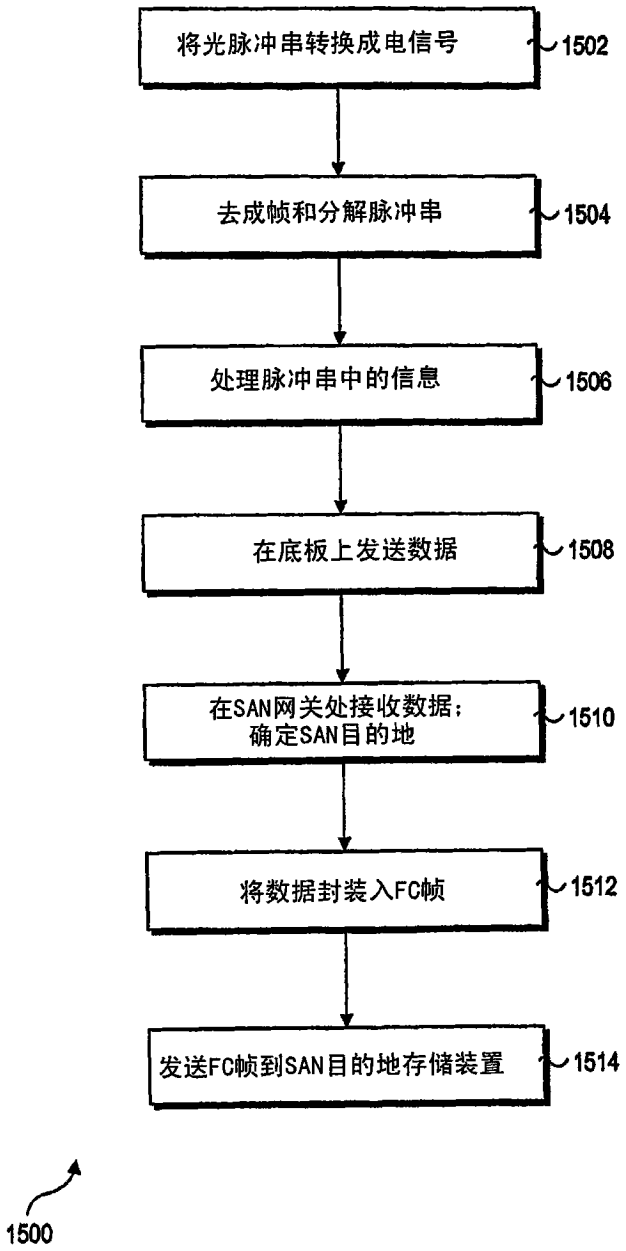

将光脉冲串转换成电信号
1502
去成帧和分解脉冲串
1504
处理脉冲串中的信息
1506
在底板上发送数据
1508
在SAN网关处接收数据；
确定SAN目的地
1510
将数据封装入FC帧
1512
发送FC帧到SAN目的地存储装置
1514
1500

图 15

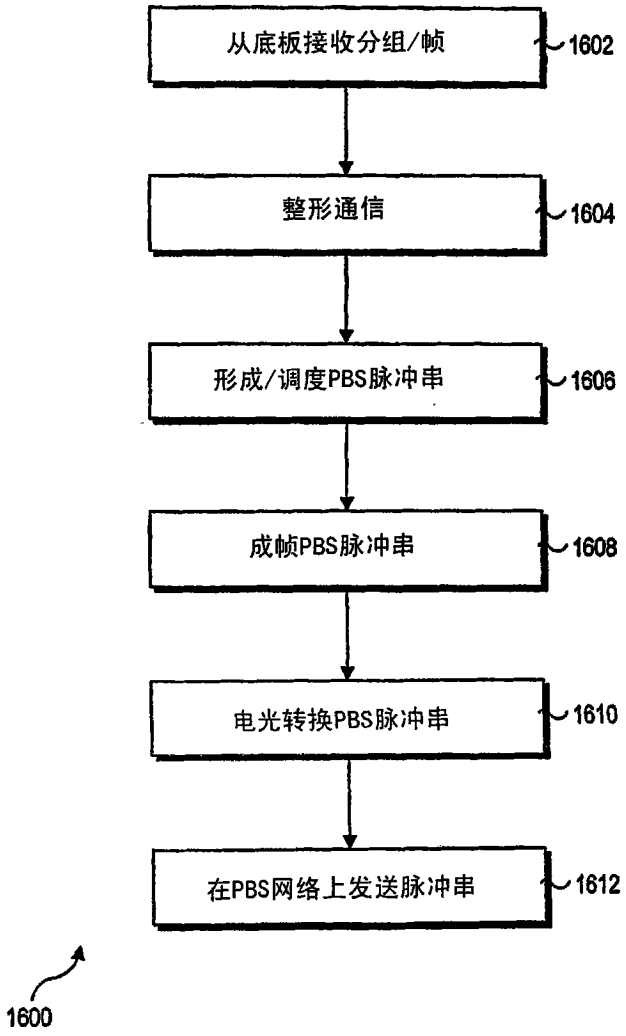

从底板接收分组/帧
1602
整形通信
1604
形成/调度PBS脉冲串
1606
成帧PBS脉冲串
1608
电光转换PBS脉冲串
1610
在PBS网络上发送脉冲串
1612
1600

图 16

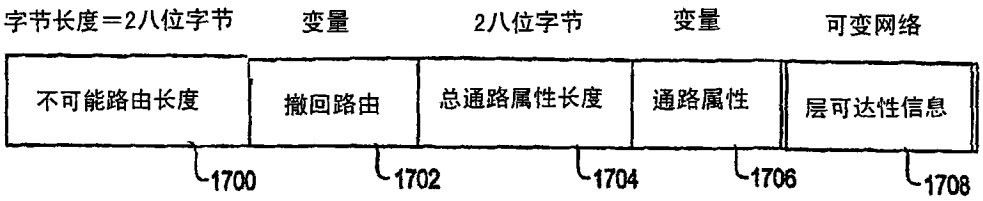

字节长度＝2八位字节
变量
2八位字节
变量
可变网络
不可能路由长度
撤回路由
总通路属性长度
通路属性
层可达性信息
1700
1702
1704
1706
1708

图 17

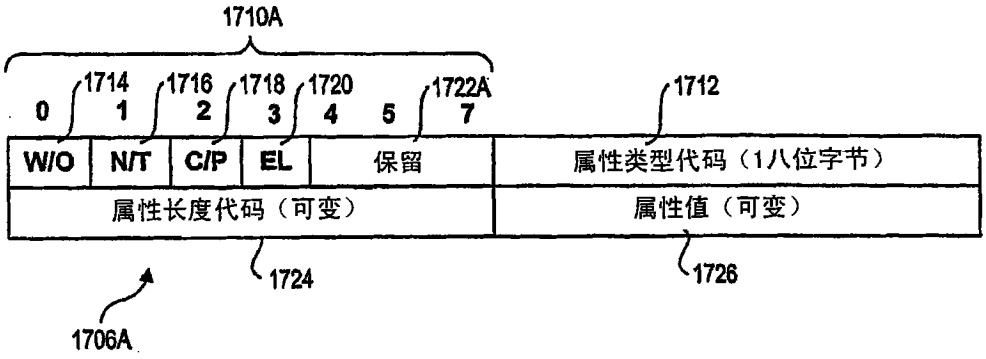



图 17a（现有技术）

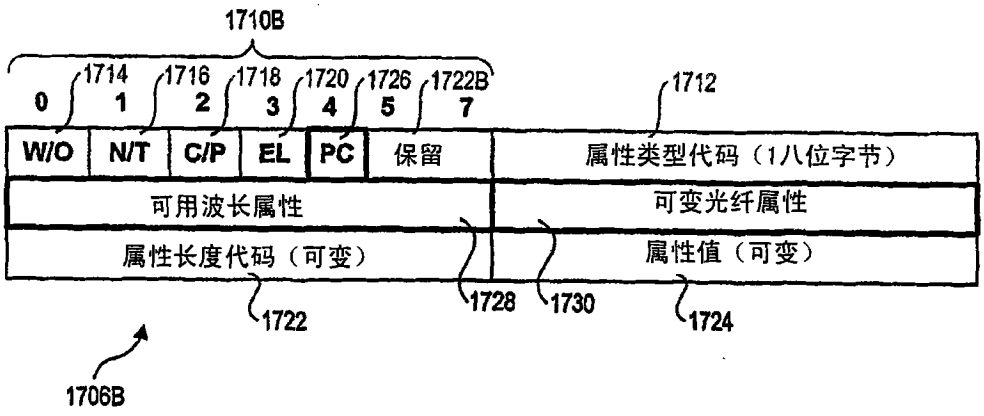



图 17b

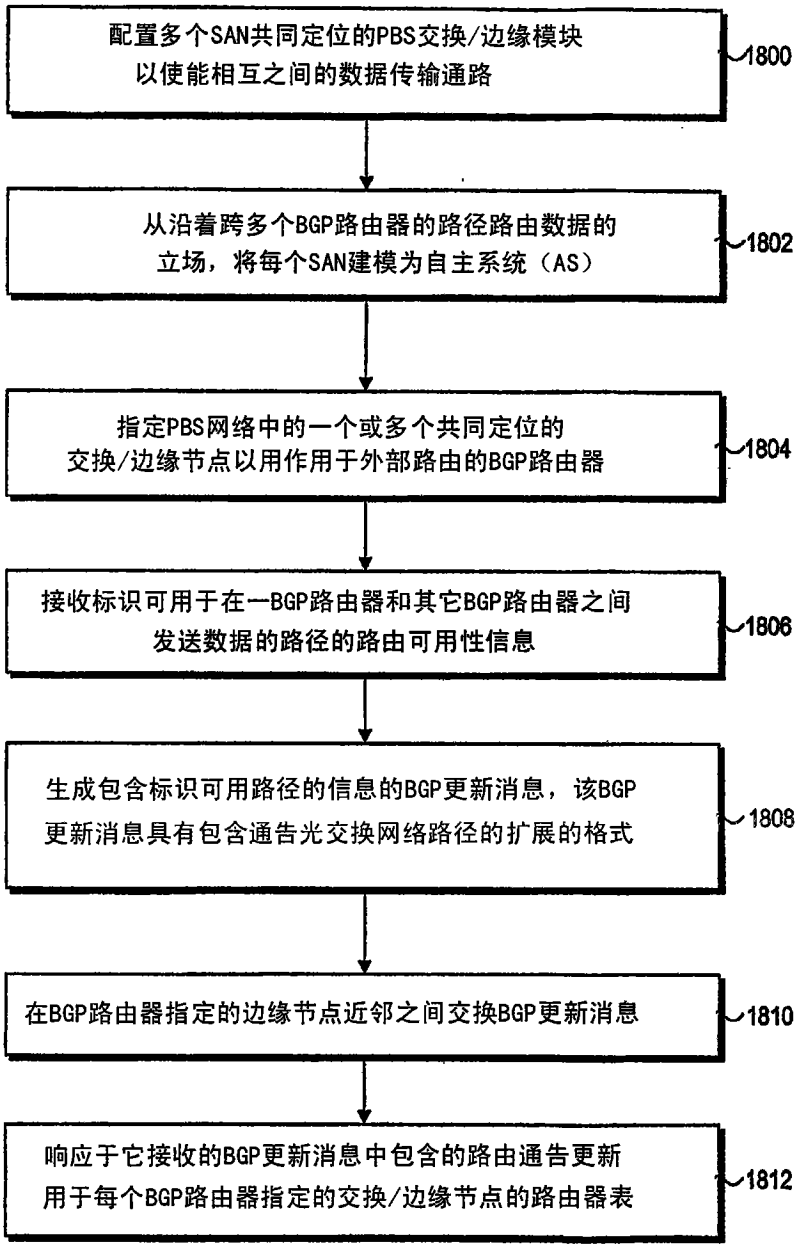

配置多个SAN共同定位的PBS交换/边缘模块
以使能相互之间的数据传输通路
1800
从沿着跨多个BGP路由器的路径路由数据的
立场，将每个SAN建模为自主系统（AS）
1802
指定PBS网络中的一个或多个共同定位的
交换/边缘节点以用作用于外部路由的BGP路由器
1804
接收标识可用于在一BGP路由器和其它BGP路由器之间
发送数据的路径的路由可用性信息
1806
生成包含标识可用路径的信息的BGP更新消息，该BGP
更新消息具有包含通告光交换网络路径的扩展的格式
1808
在BGP路由器指定的边缘节点近邻之间交换BGP更新消息
1810
响应于它接收的BGP更新消息中包含的路由通告更新
用于每个BGP路由器指定的交换/边缘节点的路由器表
1812

图 18