



(12) 发明专利

(10) 授权公告号 CN 1806466 B

(45) 授权公告日 2011.05.18

(21) 申请号 200480016665.6

(22) 申请日 2004.03.12

(30) 优先权数据

10/417,823 2003.04.16 US

(85) PCT申请进入国家阶段日

2005.12.14

(86) PCT申请的申请数据

PCT/US2004/007633 2004.03.12

(87) PCT申请的公布数据

W02004/095874 EN 2004.11.04

(73) 专利权人 英特尔公司

地址 美国加利福尼亚州

(72) 发明人 史洛莫·欧瓦迪亚

克里斯琴·麦西奥科

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 钱慰民

(51) Int. Cl.

H04Q 11/00 (2006.01)

(56) 对比文件

EP 1089498 A2, 2001.04.04, 全文.

US 2002/0154320 A1, 2002.10.24, 全文.

EP 1073307 A2, 2001.01.31, 全文.

US 2002/0109878 A1, 2002.08.15, 全文.

WO 01/67694 A1, 2001.09.13, 全文.

CN 1409903 A, 2003.04.09, 全文.

WO 02/41663 A2, 2002.05.23, 全文.

审查员 李天星

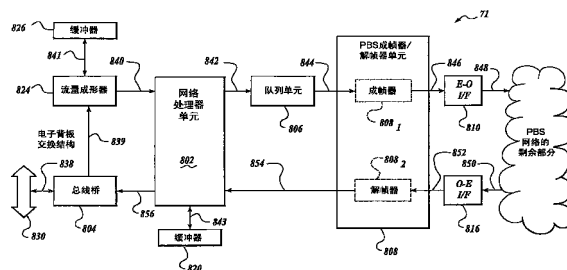
权利要求书 5 页 说明书 13 页 附图 11 页

(54) 发明名称

在基于 WDM 的光子突发交换网络中连网的多个高速服务器的体系结构、方法和系统

(57) 摘要

一种用在具有可变时隙配置的基于波分复用的光子突发交换 (PBS) 网络中的模块化、可重新配置的多服务器系统。该多服务器系统内的光高速 I/O 模块使得它能够在 PBS 网络中起到边缘节点的作用。光 I/O 模块统计复用从遗留网络接收的输入分组 (例如, IP 分组或以太网帧), 生成控制和数据突发, 然后调度这些突发在 PBS 网络上传输。光 I/O 模块然后根据流量优先级和可用的网络资源, 以光学方式发射并接收所调度的光突发。



1. 一种通信系统,包括:
包括具有有线通信信道的交换结构的背板,所述有线通信信道提供传播信息的媒介;
耦合到所述有线通信信道,以通过所述有线通信信道来传送分组的多个配置单元模块;
耦合到所述有线通信信道的遗留接口模块,所述遗留接口模块包括被耦合到遗留网络的端口;以及
耦合到所述有线通信信道的光输入/输出模块,所述光输入/输出模块包括被耦合到光网络的端口,所述光输入/输出模块将从所述有线通信信道接收的分组集总到突发中,以通过所述光网络以光学方式传送所述突发。
2. 如权利要求1所述的系统,其中所述分组选自自由因特网协议(IP)分组、以太网帧和SONET帧组成的组。
3. 如权利要求1所述的系统,其中所述光网络是光子突发交换网络。
4. 如权利要求1所述的系统,其中所述光输入/输出模块包括:
耦合到所述有线通信信道以从所述有线通信信道接收分组的总线桥;
耦合到所述总线桥的网络处理器单元,所述网络处理器单元将经由所述总线桥接收的分组集总到突发中;
耦合到所述网络处理器单元的成帧器单元,其中所述成帧器单元将所述突发封装到光网络帧中;以及
耦合到所述成帧器单元和所述光网络的光输出接口,其中所述光输出接口通过所述光网络传送包括所述光网络帧的光信号。
5. 如权利要求4所述的系统,其中所述网络处理器单元使用统计复用来集总分组。
6. 如权利要求4所述的系统,还包括耦合到所述总线桥和所述网络处理器单元的流量成形器。
7. 如权利要求4所述的系统,还包括耦合到所述成帧器单元和所述网络处理器单元的队列单元,其中所述队列单元在突发被调度通过所述光网络传送之前存储所述突发。
8. 如权利要求4所述的系统,其中所述网络处理器单元形成第一突发和第二突发,其中所述第一突发包括用于路由所述第二突发通过所述光网络的信息。
9. 如权利要求4所述的系统,其中所述光输入/输出模块还包括:
耦合到所述成帧器单元和所述光网络的光输入接口,其中所述光输入接口将从所述光网络接收的、包括光网络帧的光突发信号转换为包括所述光网络帧信息的电信号。
10. 如权利要求9所述的系统,其中所述成帧器单元还解帧来自所述光输入接口的光网络帧信息。
11. 如权利要求10所述的系统,其中所述网络处理器单元还处理来自所述成帧器单元的解帧后信息,以确定所述信息的目的地。
12. 如权利要求11所述的系统,其中所述总线桥还将包括处理后信息的分组通过所述有线通信信道从所述网络处理器单元传送到所述信息的目的地。
13. 如权利要求1所述的系统,其中所述多个配置单元模块包括存储模块。
14. 如权利要求1所述的系统,其中所述多个配置单元模块包括服务器模块,该服务器模块包括服务器。

15. 如权利要求 14 所述的系统,其中所述服务器模块包括耦合到所述服务器模块的服务器的存储单元。

16. 如权利要求 15 所述的系统,其中所述存储单元经由光纤信道 (FC) 接口耦合到所述服务器。

17. 如权利要求 14 所述的系统,其中所述服务器模块包括多个服务器。

18. 如权利要求 17 所述的系统,其中所述多个服务器中至少有一个服务器是一个服务器集群的一部分。

19. 如权利要求 1 所述的系统,其中所述交换结构包括光交换结构。

20. 如权利要求 1 所述的系统,其中所述交换结构包括电气交换结构。

21. 如权利要求 1 所述的系统,其中所述多个配置单元模块是可重新配置的。

22. 如权利要求 1 所述的系统,其中所述光网络是波分复用 (WDM) 网络。

23. 一种通信方法,包括:

将光突发信号转换为电突发信号,其中所述光突发信号是由光网络的边缘节点系统从所述光网络中接收的,所述边缘节点系统包括连接到背板的配置单元模块,所述背板包括具有有线通信信道的交换结构,其中所述交换结构包括光交换结构或电气交换结构中至少一个;

处理所述电突发信号,以获得数据及所述数据的目的地;以及

当所述目的地是所述配置单元模块时,将包括处理后信息的分组经由所述有线通信信道传送到所述配置单元模块。

24. 如权利要求 23 所述的方法,其中所述配置单元模块是耦合到所述有线通信信道,以接收通过所述有线通信信道传送的分组的多个配置单元模块的一部分。

25. 如权利要求 23 所述的方法,其中所述目的地与耦合到所述边缘节点系统的另一个网络相连。

26. 如权利要求 23 所述的方法,其中所述分组选自由因特网协议 (IP) 分组、以太网帧和 SONET 帧组成的组。

27. 如权利要求 23 所述的方法,其中所述光网络是光子突发交换 (PBS) 网络。

28. 如权利要求 24 所述的方法,其中所述多个配置单元模块包括存储模块。

29. 如权利要求 24 所述的方法,其中所述多个配置单元模块包括服务器模块,该服务器模块包括服务器。

30. 如权利要求 29 所述的方法,其中所述服务器模块包括多个服务器。

31. 如权利要求 30 所述的方法,其中所述多个服务器中至少有一个服务器是一个服务器集群的一部分。

32. 如权利要求 29 所述的方法,其中所述服务器模块包括耦合到所述服务器模块的服务器的存储单元。

33. 如权利要求 32 所述的方法,其中所述存储单元经由光纤信道 (FC) 接口耦合到所述服务器。

34. 如权利要求 24 所述的方法,其中所述多个配置单元模块是可重新配置的。

35. 如权利要求 23 所述的方法,其中所述光网络是波分复用 (WDM) 网络。

36. 一种通信系统,包括:

第一网络 ; 以及

耦合到所述第一网络的光网络, 其中所述光网络包括光子突发交换网络, 所述光网络还包括边缘节点, 该边缘节点包括 :

包括具有有线通信信道的交换结构的背板, 其中所述有线通信信道提供传播信息的媒介 ;

耦合到所述有线通信信道, 以通过所述有线通信信道来传送分组的配置单元模块 ;

耦合到所述有线通信信道的遗留接口模块, 所述遗留接口模块包括被耦合到遗留网络的端口 ; 以及

耦合到所述有线通信信道的光输入 / 输出模块, 所述光输入 / 输出模块包括被耦合到所述光网络的端口, 所述光输入 / 输出模块将从所述有线通信信道接收的分组集总到突发中, 以通过所述光网络以光学方式传送所述突发。

37. 如权利要求 36 所述的系统, 其中所述配置单元模块是耦合到所述有线通信信道, 以通过所述有线通信信道传送分组的多个配置单元模块之一。

38. 如权利要求 36 所述的系统, 其中所述分组选自自由因特网协议 (IP) 分组和以太网帧组成的组。

39. 如权利要求 36 所述的系统, 其中所述光输入 / 输出模块包括 :

耦合到所述有线通信信道以从所述有线通信信道接收分组的总线桥 ;

耦合到所述总线桥的网络处理器单元, 所述网络处理器单元将经由所述总线桥接收的分组集总到突发中 ;

耦合到所述网络处理器单元的成帧器单元, 其中所述成帧器单元将所述突发封装到光网络帧中 ; 以及

耦合到所述成帧器单元和所述光网络的光输出接口, 其中所述光输出接口通过所述光网络传送包括所述光网络帧的光信号。

40. 如权利要求 39 所述的系统, 其中所述网络处理器单元使用统计复用来集总分组。

41. 如权利要求 39 所述的系统, 还包括耦合到所述总线桥和所述网络处理器单元的流量成形器。

42. 如权利要求 39 所述的系统, 还包括耦合到所述成帧器单元和所述网络处理器单元的队列单元, 其中所述队列单元在突发被调度通过所述光网络传送之前存储所述突发。

43. 如权利要求 39 所述的系统, 其中所述网络处理器单元形成第一突发和第二突发, 其中所述第一突发包括用于路由所述第二突发通过所述光网络的信息。

44. 如权利要求 39 所述的系统, 其中所述光输入 / 输出模块还包括 :

耦合到所述成帧器单元和所述光网络的光输入接口, 其中所述光输入接口将从所述光网络接收的、包括光网络帧的光突发信号转换为包括所述光网络帧信息的电信号。

45. 如权利要求 44 所述的系统, 其中所述成帧器单元还解帧来自所述光输入接口的光网络帧信息。

46. 如权利要求 45 所述的系统, 其中所述网络处理器单元还处理来自所述成帧器单元的解帧后信息, 以确定所述信息的目的地。

47. 如权利要求 46 所述的系统, 其中所述总线桥还将包括处理后信息的分组通过所述有线通信信道从所述网络处理器单元传送到所述信息的目的地。

48. 如权利要求 37 所述的系统,其中所述多个配置单元模块包括存储模块。

49. 如权利要求 37 所述的系统,其中所述多个配置单元模块包括服务器模块,该服务器模块包括服务器。

50. 如权利要求 49 所述的系统,其中所述服务器模块包括多个服务器。

51. 如权利要求 50 所述的系统,其中所述多个服务器中至少有一个服务器是一个服务器集群的一部分。

52. 如权利要求 49 所述的系统,其中所述服务器模块包括耦合到所述服务器模块的服务器的存储单元。

53. 如权利要求 52 所述的系统,其中所述存储单元经由光纤信道 (FC) 接口耦合到所述服务器。

54. 如权利要求 36 所述的系统,其中所述交换结构包括光交换结构。

55. 如权利要求 36 所述的系统,其中所述交换结构包括电气交换结构。

56. 如权利要求 37 所述的系统,其中所述多个配置单元模块是可重新配置的。

57. 如权利要求 36 所述的系统,其中所述光网络是波分复用 (WDM) 网络。

58. 一种操作光网络的边缘节点装置以传输信息的方法,所述方法包括:

在所述边缘节点装置的光输入/输出模块处从所述边缘节点装置的配置单元模块接收分组,其中所述分组是经由将所述输入/输出模块通信地耦合到所述配置单元模块的有线通信总线接收的,所述边缘节点装置包括背板,所述背板包括具有所述有线通信总线的交换结构;

将经由所述有线通信总线从所述配置单元模块接收的所述分组选择性地集总成突发;以及

通过所述光网络来以光学方式传输所述突发。

59. 如权利要求 58 所述的方法,其中所述配置单元模块是耦合到所述有线通信总线以通过所述有线通信总线发送分组的多个配置单元模块之一。

60. 如权利要求 58 所述的方法,其中所述分组是从连接到所述边缘节点装置的另一个网络接收的。

61. 如权利要求 58 所述的方法,其中所述分组选自因特网协议 (IP) 分组、以太网帧和 SONET 帧组成的组。

62. 如权利要求 58 所述的方法,其中集总所接收的分组的操作包括形成第一突发和第二突发,所述第一突发包括用于路由所述第二突发通过所述光网络的信息。

63. 如权利要求 59 所述的方法,其中所述多个配置单元模块包括存储模块。

64. 如权利要求 59 所述的方法,其中所述多个配置单元模块包括服务器模块,该服务器模块包括服务器。

65. 如权利要求 64 所述的方法,其中所述服务器模块包括多个服务器。

66. 如权利要求 65 所述的方法,其中所述多个服务器中至少有一个服务器是一个服务器集群的一部分。

67. 如权利要求 64 所述的方法,其中所述服务器模块包括耦合到所述服务器模块的服务器的存储单元。

68. 如权利要求 67 所述的方法,其中所述存储单元经由光纤信道 (FC) 接口耦合到所述

服务器。

- 69. 如权利要求 58 所述的方法,其中所述交换结构包括光交换结构。
- 70. 如权利要求 58 所述的方法,其中所述交换结构包括电气交换结构。
- 71. 如权利要求 59 所述的方法,其中所述多个配置单元模块是可重新配置的。
- 72. 如权利要求 58 所述的方法,其中所述光网络是波分复用(WDM)网络。

在基于 WDM 的光子突发交换网络中连网的多个高速服务器的 体系结构、方法和系统

[0001] 相关申请的交叉引用：本发明与 2002 年 4 月 17 日递交的美国专利申请 No. 10/126, 091, 2002 年 6 月 25 日递交的美国专利申请 No. 10/183, 111, 2002 年 12 月 24 日递交的美国专利申请 No. 10/328, 571, 美国专利申请 No. (律师案卷号 No. 42P15724) 以及美国专利申请 No. (律师案卷号 No. 42P15725) 有关。

技术领域

[0002] 本发明的实施方案总地涉及光网络；更具体地说，涉及光子突发交换网络中多个服务器的高速连网。

背景技术

[0003] 电信网络（例如，因特网）中的传输带宽看上去一直在增长，并且正在寻求支持这种带宽需求的解决方案。这个问题的一种解决方案就是使用光纤网络，其中使用波分复用（WDM）技术来支持光网络中日益增长的对高数据率的需求。

[0004] 传统的光交换网络一般使用波长路由技术，它需要在光交换机中完成光信号的光-电-光（O-E-O）转换。在光网络中每个交换节点上的 O-E-O 转换不仅是非常慢的操作（一般约为 10 毫秒），而且还非常昂贵，潜在上产生了光交换网络的流量瓶颈。此外，目前的光交换技术还无法有效地支持在分组（packet）通信应用（例如，因特网）中经常遇到的“突发”流量。

[0005] 大的通信网络可以使用几个子网络来实现。例如，支持因特网流量的大网络可以被划分为许多个相对较小的、由因特网服务提供商（ISP）运营的接入网络，它们被耦合到多个城域网（光 MAN），而这些城域网又耦合到大的“骨干”广域网（WAN）。光 MAN 和 WAN 一般需要比局域网（LAN）更高的带宽，以便提供它们的终端用户所需要的足够的服务等级。然而，当 LAN 速度 / 带宽随着技术进步而提高时，就需要更高的 MAN/WAN 速度 / 带宽。

附图说明

[0006] 参考以下附图来描述本发明非限制性并且非穷尽性的实施方案，其中除非特别指出，否则在各个图中相近的标号表示相近的部分。

[0007] 图 1 是根据本发明的一个实施方案，图示了具有可变时隙配置的光子突发交换（PBS）网络的简化框图。

[0008] 图 2 是根据本发明的一个实施方案，图示了 PBS 网络的操作的简化流程图。

[0009] 图 3 是根据本发明的一个实施方案，图示了用在 PBS 网络中的交换节点模块的框图。

[0010] 图 4A-4B 是根据本发明的一个实施方案，图示了用在 PBS 网络中的光数据和控制突发的格式的图。

[0011] 图 5 是根据本发明的一个实施方案，图示了交换节点模块的操作的流程图。

- [0012] 图 6 是根据本发明的一个实施方案,图示了具有服务器边缘节点的 PBS 网络的图。
- [0013] 图 6A-6C 是根据本发明的一个实施方案,图示了在图 6 中描绘的配置单元的图。
- [0014] 图 7 是根据本发明的一个实施方案,图示了多服务器 /PBS 边缘节点单元的图。
- [0015] 图 8 是根据本发明的一个实施方案,图示了在图 7 中描绘的光 I/O 卡的框图。
- [0016] 图 8A 是根据本发明的一个实施方案,更详细地图示了在图 8 中描绘的网络处理器单元和队列单元的框图。
- [0017] 图 9 是根据本发明的一个实施方案,图示了出口操作流的流程图。
- [0018] 图 10 是根据本发明的一个实施方案,图示了入口操作流的流程图。

具体实施方式

[0019] 图 1 根据本发明的一个实施方案图示了光子突发交换 (PBS) 网络 10。术语“光子突发”在这里用来指具有类似路由需求的统计复用的分组 (例如,因特网协议 (IP) 分组或以太网帧)。光子突发一般包括光子标签和有效载荷,光子标签包括各个 IP 分组的头部和其他路由信息,有效载荷包括各个分组的数据段。

[0020] PBS 网络 10 的这个实施方案包括局域网 (LAN) 13₁-13_N 和骨干光 WAN (未示出)。另外, PBS 网络 10 的这个实施方案包括入口节点 15₁-15_M, 交换节点 17₁-17_L 和出口节点 18₁-18_K。PBS 网络 10 可以包括与图 1 中示出的交换节点互连的其他入口、出口和交换节点 (未示出)。入口和出口节点在这里也被称为边缘节点,这是因为它们在逻辑上位于 PBS 网络的边缘。边缘节点实际上提供了上述“外部”网络 (即,位于 PBS 网络之外) 和 PBS 网络的交换节点之间的接口。在该实施方案中,用智能模块来实现入口、出口和交换节点。该实施方案例如可以用作城域网,用于将城市区域内的大量 LAN 连接到大的光骨干网。

[0021] 在一些实施方案中,入口节点对接收的光信号执行光-电 (O-E) 转换,并且包括电子存储器,用于在接收信号被发送到适当的 LAN 之前缓冲这些信号。此外,在一些实施方案中,入口节点在接收的电信号被发送到 PBS 网络 10 的交换节点 17₁-17_M 之前,还执行这些电信号的电-光 (E-O) 转换。

[0022] 出口节点用光交换单元或模块来实现,它们被配置为从 PBS 网络 10 的其他节点接收光信号,并将它们路由到光 WAN 或其他外部网络。出口节点还可以从光 WAN 或其他外部网络接收光信号,并将它们发送到 PBS 网络 10 的适当节点。在一个实施方案中,出口节点 18₁ 对接收的光信号执行 O-E-O 转换,并且包括电子存储器,用于在接收信号被发送到 PBS 网络 10 的适当节点 (或者发送到光 WAN) 之前缓冲这些信号。

[0023] 交换节点 17₁-17_L 用光交换单元或模块来实现,这些单元或模块中的每一个都被配置为从其他交换节点接收光信号,并将接收的光信号适当地路由到 PBS 网络 10 的其他交换节点。如下所述,交换节点执行光控制突发和网络管理控制突发信号的 O-E-O 转换。在一些实施方案中,这些光控制突发和网络管理控制突发仅在预先选择的波长上传播。即使控制突发和网络管理控制突发可以包括一组具体的光数据突发信号的必要信息,预先选择的波长在这样的实施方案中也不传播光“数据”突发 (与控制突发和网络管理控制突发相对的) 信号。在一些实施方案中,控制和数据信息是在单独的波长上传输 (也被称为带外信令)。在其他实施方案中,控制和数据信息可以在相同的波长上发送 (也被称为带内信令)。在另一个实施方案中,光控制突发、网络管理控制突发和光数据突发信号可以使用不

同的编码方案（例如不同的调制格式等）在相同的波长上传播。在任何一种方法中，光控制突发和网络管理控制突发相对于其对应的光数据突发信号都是异步发送的。在另外一个实施方案中，光控制突发和其他控制信号作为光数据信号以不同的传输速率来传播。

[0024] 虽然交换节点 17_1-17_L 可以执行光控制信号的 O-E-O 转换，但是在该实施方案中，交换节点不执行光数据突发信号的 O-E-O 转换。相反，交换节点 17_1-17_L 对光数据突发信号只执行纯粹的光交换。因而，交换节点可以包括电子电路系统，该电路存储并处理已被转换成电形式的输入光控制突发和网络管理控制突发，并使用该信息来配置光子突发交换机设置，该电路还适当地路由对应于光控制突发的光数据突发信号。基于新的路由信息来取代先前控制突发的新控制突发被转换为光控制信号，并且它被传输到下一交换节点或出口节点。下面进一步描述交换节点的实施方案。

[0025] 示例性的 PBS 网络 10 的元件如下互连。LAN 13_1-13_N 被连接到对应的入口节点 15_1-15_M 。在 PBS 网络 10 内，入口节点 15_1-15_M 和出口节点 18_1-18_K 经由光纤与交换节点 17_1-17_L 中的一些相连。交换节点 17_1-17_L 在网状体系结构中经由光纤也彼此互连，以在入口节点之间，以及在入口节点 15_1-15_M 和出口节点 18_1-18_K 之间形成相当大量的光路或光链路。理想情况下，有一条以上的光路将交换节点 17_1-17_L 连接到 PBS 网络 10 的每一个端点（即，入口节点和出口节点是 PBS 网络 10 内的端点）。在交换节点、入口节点和出口节点之间的多条光路使得在一个或多个节点发生故障时可以实现保护交换，或者可以实现例如到目的地的主路由和副路由等特征。

[0026] 下面结合图 2 将描述，PBS 网络 10 的入口节点、出口节点和交换节点被配置为发送和 / 或接收光控制突发、光数据突发和其他控制信号，它们被波长复用，以便在预先选择的波长上传播光控制突发和控制标签，并且在不同的预先选择的波长上传播光数据突发或有效载荷。另外，PBS 网络 10 的边缘节点可以发送光控制突发信号，同时将数据发送到 PBS 网络 10 以外（光数据或电数据）。

[0027] 图 2 根据本发明的一个实施方案图示了 PBS 网络 10 的操作流。参考图 1 和 2，光子突发交换网络 10 如下运行。

[0028] PBS 网络 10 从 LAN 13_1-13_N 接收分组。在一个实施方案中，PBS 网络 10 在入口节点 15_1-15_M 处接收 IP 分组。所接收的分组可以是电形式的，而非光形式的，或者以光形式接收的，然后被转换成电形式的。在该实施方案中，入口节点电存储所接收的分组。框 20 代表了该操作。

[0029] 为清楚起见，对 PBS 网络 10 的操作流的剩余描述集中在信息从入口节点 15_1 到出口节点 18_1 的传输。从入口节点 15_2-15_M 到出口节点 18_1 （或其他出口节点）的信息传输基本与此类似。

[0030] 光突发标签 (label)（即，光控制突发）和光有效载荷（即，光数据突发）是由接收的分组形成的。在一个实施方案中，入口节点 15_1 使用统计复用技术，由存储在入口节点 15_1 中的接收 IP（因特网协议）分组形成光数据突发。例如，由入口节点 15_1 接收并且在其到达目的地的路径上必须穿过出口节点 18_1 的分组可以被组装成一个光数据突发有效载荷。统计复用通常是指由多个数据源基于这些源所使用的带宽的统计结果（例如，平均值）而非每个源需要的峰值带宽，来共享链路或信道的技术。例如，统计复用技术被以下文献公开：K. Kumaran 和 M. Mandjes 在 IEEE INFOCOM 2001 会议论文集中发表

的“Multiplexing Regulated Traffic Streams: Design and Performance (复用调节的流量流: 设计和性能)”; Su 和 G. de Veciana 在 IEEE INFOCOM 1998 会议论文集中发表的“On Statistical Multiplexing, Traffic Mixes, and VP Management (统计复用、流量混合、以及 VP 管理)”; B. Maglaris, D. Anastassiou, P. Sen, G. Karlson 和 J. D. Robbins 在 IEEE Transaction on Communication 36, 834-844, 1988 中发表的“Performance Models of Statistical Multiplexing in Packet Video Communications (在分组视频通讯重统计复用的性能模型)”; T. Brown 在 Kouvatsos, D. editor, Kluwer, 2000 的“Performance Evaluation and Application of ATM Networks (ATM 网络的性能评价和应用)”的第三章“Adaptive Statistical Multiplexing For Broadband Communication (用于宽带通讯的自适应统计复用)”。其他实施方案可以使用任何适合的统计复用技术。框 21 代表该操作。

[0031] 特定光信道和 / 或光纤上的带宽被保留来传送光数据突发通过 PBS 网络 10。在一个实施方案中, 入口节点 15₁ 在穿过 PBS 网络 10 的光数据信号路径中保留一个时隙 (即, TDM 系统的时隙)。该时隙可以是固定时长的和 / 或可变时长的, 并且在相邻时隙间具有均匀或非均匀的定时间隔。此外, 在一个实施方案中, 将带宽保留一段足够长的时间, 以将光突发从入口节点传输到出口节点。例如, 在一些实施方案中, 入口节点、出口节点和交换节点都保存着所有已用时隙和可用时隙的更新列表。时隙可以在多个波长和光纤上分配及分布。因此, 在不同的实施方案中可以为固定时长或可变时长的保留时隙 (这里也被称为 TDM 信道) 可以位于一根光纤的一个波长中, 和 / 或可以分散在多个波长和多根光纤上。框 22 代表该操作。

[0032] 当入口节点和 / 或出口节点保留带宽时, 或者在传输光数据突发后带宽被释放时, 网络控制器 (未示出) 对列表进行更新。在一个实施方案中, 网络控制器以及入口和出口节点基于可用的网络资源和流量模式, 使用各种突发或分组调度算法来完成这一更新过程。被周期性地广播到入口、交换和出口节点的可用的可变时长 TDM 信道在相同的波长上作为光控制突发来传输, 或者在整个光网络中在一条不同的公共预先选择波长上传输。网络控制器功能可以存在于入口或出口节点之一中, 或者可以分布在两个或更多的入口和 / 或出口节点上。在该实施方案中, 网络控制器是控制单元 37 (图 3) 的一部分, 控制单元 37 可以包括一个或多个处理器。

[0033] 然后, 光控制突发、网络管理控制标签和光数据突发在保留的时隙或 TDM 信道中被传输通过光子突发交换网络 10。在一个实施方案中, 入口节点 15₁ 沿着网络控制器所确定的光标签交换路径 (OLSP) 将控制突发传送到下一节点。在该实施方案中, 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, 通用多协议标签交换 (GMPLS) 草案因特网工程任务组 (IETF) 体系结构 -05 因特网 - 草案, 2003 年 3 月] 来确定到达出口节点的最佳可用 OLSP。

[0034] 在一个实施方案中, 控制标签 (这里也被称为控制突发) 在光子数据突发之前, 在一条不同的波长和 / 或不同的光纤上被异步传输。在控制突发和数据突发之间的时间偏移允许每个交换节点在相应的数据突发到来之前, 处理所述标签并且配置光子突发交换机进行适当的交换。术语“光子突发交换机”在这里被用来指不使用 O-E-O 转换的快速光交换机。

[0035] 在一个实施方案中, 入口节点 15₁ 然后将光数据突发异步发送到交换节点, 在每个

交换节点内光数据突发都几乎不经历任何时延,并且不经受 O-E-O 转换。光控制突发总是在传输对应的光数据突发之前发送。

[0036] 在一些实施方案中,交换节点可以执行控制突发的 O-E-O 转换,使得节点可以提取并处理包括在标签中的路由信息。此外,在一些实施方案中,TDM 信道在与用于传播标签的相同波长中传播。可替换地,可以使用不同的调制格式,在相同光纤中的相同波长上来调制标签和有效载荷。例如,可以使用非归零 (NRZ) 调制格式来传输光标签,而使用归零 (RZ) 调制格式来传输光有效载荷。光突发以类似的方式从一个交换节点传输到另一个交换节点,直到光控制突发和光数据突发终止在出口节点 18_i 上为止。框 23 代表该操作。

[0037] 此时的操作流取决于目标网络是光 WAN 还是 LAN。框 24 代表操作流中的这一分支。

[0038] 如果目标网络是光 WAN,则形成新的光标签和有效载荷信号。在该实施方案中,出口节点 18_i 准备新的光标签和有效载荷信号。框 25 代表该操作。

[0039] 然后,新的光标签和有效载荷被传送到目标网络(即,此时是 WAN)。在该实施方案中,出口节点 18_i 包括将光标签和有效载荷发送到光 WAN 的光接口。框 26 代表该操作。

[0040] 然后,如果在框 24 中目标网络是 LAN,则光数据突发被拆分,以抽取 IP 分组或以太网帧。在该实施方案中,出口节点 18_i 将光数据突发转换为电信号,出口节点 18_i 可以处理该电信号以恢复每个分组的数据段,框 27 代表了这一操作。

[0041] 抽取出的 IP 数据分组或以太网帧被处理,与对应的 IP 标签组合在一起,然后被路由到目标网络(即,此时是 LAN)。在该实施方案中,出口节点 18_i 形成这些新的 IP 分组。框 28 代表该操作。然后如框 26 所表示的,新的 IP 分组被传送到目标网络(即,LAN)。

[0042] PBS 网络 10 可以通过 TDM 信道所赋予的额外的灵活性,获得更高的带宽效率。虽然上面描述的示例性的实施方案包括利用入口、交换和出口节点将多个 LAN 耦合到光 WAN 骨干的光 MAN,但是在其他实施方案中,网络不一定是 LAN、光 MAN 或 WAN 骨干。也就是说,PBS 网络 10 可以包括多个相对较小的网络,它们被耦合到相对较大的网络,而这些较大的网络接着又耦合到骨干网。

[0043] 虽然这里描述了 WDM 的实施方案,但是在其他实施方案中,单个波长可被用于整个 PBS 网络。这些单波长的替换实施方案中的一些在每个节点之间具有多个光纤,以提供增高的带宽。

[0044] 图 3 根据本发明的一个实施方案图示了在光子突发交换网络 10(图 1)中用作交换节点的模块 17。在该实施方案中,模块 17 包括一组光波分解复用器(de-multiplexer) 30_1-30_A ,其中 A 代表用于向该模块传播有效载荷、标签和其他网络资源的输入光纤的数量。例如,在该实施方案中,每个输入光纤可以携带一组 C 个波长(即,WDM 波长),但是在其他实施方案中,输入光纤可以携带不同数量的波长。模块 17 还可以包括一组 $N \times N$ 的光子突发交换机 32_1-32_B ,其中 N 是每个光子突发交换机的输入/输出端口的数量。因此,在该实施方案中,在每个光子突发交换机上波长的最大数量是 $A \cdot C$,其中 $N \geq A \cdot C + 1$ 。对于 N 大于 $A \cdot C$ 的实施方案,额外的输入/输出端口可被用于回送光信号以进行缓冲。

[0045] 此外,虽然光子突发交换机 32_1-32_B 被显示为独立单元,但是可以使用任何适当的交换机体系结构将它们实现为 $N \times N$ 的光子突发交换机。模块 17 还包括一组光波分复用器

(multiplexer) 34₁-34_A、一组光-电信号转换器 36(例如,光检测器)、控制单元 37 和一组电-光信号转换器 38(例如,激光器)。控制单元 37 可以具有一个或多个执行软件或固件程序的处理器。

[0046] 模块 17 的这一实施方案的元件如下互连。光解复用器 30₁-30_A 与一组 A 条输入光纤相连,这些输入光纤传播来自光子突发交换网络 10(图 10) 中其他交换节点的输入光信号。光解复用器的输出导线被连接到一组 B 个核心光交换机 32₁-32_B,并被连接到光信号转换器 36。例如,光解复用器 30₁ 将 B 条输出导线连接到光子突发交换机 32₁-32_B 的输入导线(即,光解复用器 30₁ 的一条输出导线连接到每个光子突发交换机的一条输入导线),并且将至少一条输出导线连接到光信号转换器 36。

[0047] 光子突发交换机 32₁-32_B 的输出导线被连接到光复用器 34₁-34_A。例如,光子突发交换机 32₁ 将 A 条输出导线连接到光复用器 34₁-34_A 的输入导线(即,光子突发交换机 32₁ 的一条输出导线连接到每个光复用器的一条输入导线)。每个光复用器也将输入导线连接到电-光信号转换器 38 的输出导线。控制单元 37 使输入导线或端口连接到光-电信号转换器 36 的输出导线或端口。控制单元 37 的输出导线被连接到光子突发交换机 32₁-32_B 和电-光信号转换器 38 的控制导线。如下面结合图 5 的流程图所描述的那样,模块 17 被用来接收和发送光控制突发、光数据突发和网络管理控制突发。在一个实施方案中,光数据突发和光控制突发具有图 4A 和 4B 中所示的传输格式。

[0048] 图 4A 根据本发明的一个实施方案图示了用在 PBS 网络 10(图 1) 中的光数据突发的格式。在该实施方案中,每个光数据突发具有起始保护带 40、IP 有效载荷数据段 41、IP 头部段 42、有效载荷同步段 43(典型的是少数几比特)和结束保护带 44,如图 4A 所示。在一些实施方案中,IP 有效载荷数据段 41 包括用于形成突发的统计复用 IP 数据分组或以太网帧。虽然图 4A 显示有效载荷为相接的,但是模块 17 以 TDM 格式来发送有效载荷。此外,在一些实施方案中,数据突发可以在多个 TDM 信道上被分段。应当指出的是,在该实施方案中,光数据突发和光控制突发只在 PBS 网络 10 中具有局部意义,并可能在光 WAN 上失去它们的意义。

[0049] 图 4B 根据本发明的一个实施方案图示了用在光子突发交换网络 10(图 1) 中的光控制突发的格式。在该实施方案中,每个光控制突发具有起始保护带 46、IP 标签数据段 47、标签同步段 48(典型的是少数几比特)和结束保护带 49,如图 4B 所示。在该实施方案中,标签数据段 47 包括 IP 分组的所有必要的路由和定时信息,以便形成光突发。虽然图 4B 显示有效载荷为相接的,但是在该实施方案中,模块 17 以 TDM 格式来发送标签。

[0050] 在一些实施方案中,在 PBS 网络 10(图 1) 中还使用光网络管理控制标签(未示出)。在这样的实施方案中,每个光网络管理控制突发包括:类似于起始保护带 46 的起始保护带;类似于数据段 47 的网络管理数据段;类似于标签同步段 48 的网络管理同步段(典型的是少数几比特);和类似于结束保护带 44 的结束保护带。在该实施方案中,网络管理数据段包括为协调网络上的传输所必需的网络管理信息。在一些实施方案中,光网络管理控制突发是以 TDM 格式来发送的。

[0051] 图 5 根据本发明的一个实施方案图示了模块 17(图 3) 的操作流。参考图 3 和 5,模块 17 如下运行。

[0052] 模块 17 接收带有 TDM 标签和数据信号的光信号。在该实施方案中,模块 17 在光

解复用器中的一个或两个处接收光控制信号（例如，光控制突发）和光数据信号（即，在该实施方案中的光数据突发）。例如，光控制信号可以被调制在光解复用器 30_A 所接收的光信号的第一波长上，而光数据信号被调制在光解复用器 30_A 所接收的光信号的第二波长上。在一些实施方案中，光控制信号可以由第一光解复用器来接收，而光数据信号由第二光解复用器来接收。此外，在一些情况下，只接收光控制信号（例如，网络管理控制突发）。框 51 代表了该操作。

[0053] 模块 17 将光控制信号转换为电信号。在该实施方案中，光控制信号是光控制突发信号，它由光解复用器从接收的光数据信号中分离出来，并被发送到光 - 电信号转换器 36。在其他实施方案中，光控制信号可以是网络管理控制突发（前面参考图 4B 所描述的）。光 - 电信号转换器 36 将光控制信号转换为电信号。例如，在一个实施方案中，TDM 控制信号的每个部分都被转换为电信号。对控制单元 37 接收到的电控制信号进行处理，以形成新的控制信号。在该实施方案中，控制单元 37 存储并处理包括在控制信号中的信息。框 53 代表该操作。

[0054] 模块 17 然后基于包括在控制信号中的路由信息，将光数据信号（即，在该实施方案中的光数据突发）路由到光复用器 34₁-34_A 之一。在该实施方案中，控制单元 37 处理控制突发，以抽取路由和定时信息，并将适当的 PBS 配置信号发送到一组 B 个光子突发交换机 32₁-32_B，以重新配置每一个光子突发交换机来交换对应的光数据突发。框 55 代表该操作。

[0055] 模块 17 然后将处理后的电控制信号转换为新的光控制突发。在该实施方案中，控制单元 37 提供 TDM 信道校准，从而以期望的波长和 TDM 时隙模式来生成再次转换后的或新的光控制突发。可以在与在框 51 中接收的控制突发的波长和 / 或时隙不同的波长和 / 或时隙上调制新的控制突发。框 57 代表该操作。

[0056] 然后，模块 17 将光控制突发发送到路由中的下一交换节点。在该实施方案中，电 - 光信号转换器 38 将新的光控制突发发送到光复用器 34₁-34_A 中的适当光复用器，以获得路由。框 59 代表该操作。

[0057] 图 6 根据本发明的一个实施方案图示了具有服务器边缘节点的 PBS 网络 60。在该实施例中，PBS 网络 60 类似于 PBS 网络 10（图 1），除了用边缘节点 61₁-61_Y 来取代一个或多个入口和出口节点外。边缘节点 61₁-61_Y 分别包括配置单元 63₁-63_Y。术语“配置单元”使用在这里是指经由 PBS 网络 60 向客户提供服务的单元，例如服务器（例如，文件或应用）或存储。存储服务例如可以包括直接附接存储（DAS）、网络附接存储（NAS），或者如果 PBS 网络 60 正在采用基于 PBS 的存储域网络（SAN），则包括（一个或多个）存储设备。下面结合图 6A-6C 来描述配置单元的实施例。

[0058] 在该实施方案中，网络 13₁-13_N 经由标签边缘路由器（LER）62₁-62_X 与边缘节点 61₁-61_Y 相连。在一些实施方案中，网络 13₁-13_N 是 LAN、WAN、SAN 和 / 或其他 PBS 网络。在一些实施方案中，不用 LER，网络可以直接与边缘节点相连。PBS 网络 60 的节点（即，入口、出口、边缘和交换节点）如前面参考 PBS 网络 10（图 1）所描述的那样互连。

[0059] 为清楚起见，下面使用配置单元 62₁-62_X 是多服务器单元的实施例来描述 PBS 网络 60 的操作。在典型的应用中，多服务器单元可以支持高带宽和 / 或延迟敏感的应用，例如文件服务器、邮件服务器、媒体服务器、应用服务器、万维网服务器等应用。当配置单元被用于存储应用时，操作类似。

[0060] PBS 网络 60 的操作与具有传统服务器（例如服务器场或刀片服务器）的传统网络有很大不同。在传统网络中，服务器和连接到网络的其他设备没什么两样，要为带宽等网络资源而竞争。在一些应用中，服务器通过网络相互通信。然而，这些服务器几乎不提供任何网络配置 (provisioning) 和流量工程功能。因此，基于最大努力地传送网络上的大多数流量，代表了在服务器场之间的信息传输中的瓶颈。

[0061] 相反，PBS 网络 60 的这一实施方案通过修改传统的服务器场（或其他多服务器或存储单元）使其用作 PBS 网络中的边缘节点 61_1-61_Y ，从而避免了这一瓶颈。这些修改后的多服务器单元（即，配置单元 63_1-63_Y ）具有光 PBS 接口（例如，见图 6A）并将信息集总到光 PBS 控制和数据突发中以通过 PBS 网络发送。如前所述，PBS 网络提供高速连网，特别是在信息传输如同某些分组类型的通信协议（例如，IP 和以太网）一样具有“突发式”特点的应用中。下面参考图 7 来描述多服务器单元的一个实施方案。

[0062] 图 6A 图示了可用来实现配置单元 63_1-63_Y （图 6）中的一个或多个的配置单元 63A 的实施方案。多服务器单元包括 PBS 接口 65，它可以在 PBS 网络上光学形成 / 发送 PBS 控制和数据突发；并且可以接收 / 处理来自 PBS 网络的光 PBS 突发。下面结合图 7 更详细地描述一个实施方案。

[0063] 在一些实施方案中，配置单元 63A 的服务器（和 / 或 PBS 网络中的其他多服务器配置单元）可以被组织成一个或多个服务器集群。每个服务器集群被连接到 PBS 网络的其他客户视为单个服务器。这样的实施方案可以提供负载均衡（这可以改进性能）和故障恢复能力，以从例如服务器或其他硬件故障、服务器软件故障或连接故障中恢复过来。此外，集群技术可以用在这些实施方案中，以有选择地连接或从集群中断开（一个或多个）服务器，实际上是将其从 PBS 网络中断开。

[0064] 为了形成集群 (cluster)，集群中的每个服务器都包括集群软件，以检测故障，执行故障转移操作，标签执行负载均衡操作。集群软件还可以包括以下组件，其支持服务器与集群之间有选择的连接 / 断开，也就是实际上将服务器加入 PBS 网络 / 从 PBS 网络中除去。对于用在服务器中的几种操作系统，可以在商业上获得适合的集群软件。对于硬件，集群中的服务器经由网络（一般与客户所使用的网络相互分离）而互连，所述网络用于在集群的服务器之间通信 / 轮询。硬件也可以包括由集群中的服务器通过直接连接或者经由一个或多个集线器而共享的冗余存储设备。

[0065] 图 6B 根据本发明的一个实施方案图示了可以用来实现配置单元 63_1-63_Y （图 6）的配置单元 63B。在该实施方案中，配置单元 63B 包括服务器 66，其具有用作 DAS 单元的附接存储单元 67。服务器 66 包括用于通过 PBS 网络传输数据的 PBS 接口 65。存储单元 67 可以用一个或多个硬盘驱动器、磁带驱动器等来实现，并且一般经由各类接口与服务器 66 通信，这些接口例如包括小型计算机系统接口 (SCSI)（例如，公认标准委员会 INCITS 的技术委员会 2003 年 3 月 19 日公布的 SCSI 体系结构模型 3, T10 工作草案）、因特网 SCSI (iSCSI) 接口（例如，iSCSI 草案 20#，IP 存储工作组，因特网工程任务组 (IETF)，2003 年 1 月 19 日）、光纤信道 (FC) 接口（例如，用于 SCSI 的 FC 协议 (FCP-3) 版本 0，2002 年 12 月 12 日）等。在操作中，由存储单元 67 存储的信息对于经由服务器 66 连接到 PBS 网络的其他单元都是可用的。

[0066] 图 6C 图示了可被用来实现配置单元 63_1-63_Y （图 6）中的一个或多个的配置单元

63C 的实施方案。在该实施方案中,配置单元 63C 可以用作 PBS 网络 60 中的 NAS 单元。例如, PBS 网络 60 可以包括一个具有配置单元 63C 的边缘节点,而其余的边缘节点则具有类似配置单元 63A 的配置单元。这一配置允许所有的其他边缘节点直接经由 PBS 网络 60 中的光路来访问存储单元。然而,在该 NAS 应用中,在所有的其他 PBS 流量之上又增加了存储流量,这有可能降低 PBS 网络的性能。

[0067] 在另一种配置中,配置单元 63C 可被用来实现 PBS 网络 60 中的所有配置单元 63_1-63_N ,使得 PBS 网络 60 如同 SAN 一样运行。在该配置中,PBS 网络 60 提供专用于存储流量的网络。于是,其他网络(例如图 6 中的网络 13_1-13_N)可以利用边缘节点的出口能力来访问在 PBS 网络 60 中可用的存储。

[0068] 图 7 根据本发明的一个实施方案图示了模块化的、可重新配置的多服务器/PBS 边缘节点单元 70。在该实施方案中,单元 70 包括光 I/O 卡或模块 71,其具有 PBS 光端口 72、具有遗留端口 74 的遗留接口卡或模块 73、四个配置模块 75_1-75_4 (即,在该实施方案中的服务器卡或模块)、背板 77、连接器 78_1-78_6 (在图 7 中只能看到连接器 78_1-78_3)和机箱 79。在一些实施方案中,单元 70 可以包括多于四个或少于四个的服务器模块。在其他实施方案中,单元 70 可以与图 7 所示实施方案不同地进行配置。

[0069] 下面结合图 8 和 8A 来描述光 I/O 模块 71 的一个实施方案。

[0070] 在该实施方案中,遗留接口卡 73 是吉比特以太网(GbE)卡,用于使用 GbE 以太网协议与 LER 或其他 LAN/WAN 网络通信。在其他实施方案中,可以使用不同的遗留协议。

[0071] 在该实施方案中,服务器模块 75_1-75_4 是内含式高速服务器模块,其中单个或多个服务器被实现为单个集成模块。

[0072] 在一些实施方案中,背板 77 包括电子交换结构,其具有缓冲器,还具有电总线(见图 8)、电源、控制部分等,类似于在商业上可获得的刀片服务器系统中所使用的那些背板。在一个实施方案中,电总线支持星状或双星状网络拓扑,以交换到服务器模块中适当的电接口,例如外围元件互连(PCI)(例如,PCI 规范 v2.2,1999 年 1 月 25 日)或 PCI-Express(例如 PCI-X 规范 v.1.0,1999 年 9 月 27 日)接口、InfiniBand®(例如 InfiniBand® 1.0 规范,2000 年 10 月 24 日)接口。在其他实施方案中,背板可以包括其他类型的有线交换结构。这里所使用的有线交换结构也可以指光交换结构或者光和电交换结构的组合。

[0073] 单元 70 的这些实施方案如下互连。光 I/O 模块 71、遗留接口模块 73 和服务器模块 75_1-75_4 经由连接器 78_1-78_6 被连接到背板 77(和上述电交换结构 830)。光端口 72 与 PBS 网络(例如图 6 中的 PBS 网络 60)相连。遗留端口 74 与遗留网络或 LER(例如,见图 6)相连。机箱 79 收纳并在物理上支持所述模块、连接器和背板。机箱 79 还包括为避免模糊本发明而在图 7 中没有示出的其他单元(例如,电源、一个或多个冷却扇等)。

[0074] 在操作中,单元 70 可以用作向耦合到遗留网络的客户提供服务的传统多服务器单元。例如,在一个实施方案中,在客户和服务器模块(即,服务器模块 75_1-75_4 之一)之间的数据流量可以如同在传统的多服务器单元中一样,经由遗留接口模块 73、背板 77 和服务器模块 75_1-75_4 中适当的服务器来传送。

[0075] 此外,服务器模块(即,服务器模块 75_1-75_4 之一)可以经由 PBS 网络和光 I/O 模块 71 向客户提供服务。然而,与传统的多服务器单元不同,光 I/O 模块 71 从客户接收光 PBS 突发,然后就像前面对 PBS 网络 10(图 1)的入口节点所描述的那样,所述光 PBS 突发

被 O-E 转换,解帧并处理。光 I/O 模块 71 以和服务器模块将通过背板 77 传输信息同样的方式,经由背板 77 将路由输入流量的所有必要信息提供给适当的服务器模块(即,服务器模块 75₁-75₄之一)。

[0076] 类似地,单元 70 的服务器模块经由背板 77、光 I/O 模块 71 和 PBS 网络向客户提供所有必要的信息。与传统的多服务器单元不同,单元 70 的光 I/O 模块 71 以和前面针对 PBS 网络 10(图 1)的入口节点所描述的基本相同的方式,统计复用来自一个或多个服务器模块的输入流量流(例如,IP 分组、以太网帧),以形成 PBS 控制和数据突发。然后,如前面针对 PBS 网络 10 所描述的那样,PBS 突发被构造成帧、调度、E-O 转换并经由 PBS 网络传送到客户。

[0077] 从遗留网络(legacy network)进入单元 70,以经由 PBS 网络传输到目的地的流量在遗留端口 74 处由单元 70 接收。如前所述,遗留网络可以使用传统的连网协议,例如 TCP/IP 或以太网协议。在该实施方案中,遗留网络是电 GbE 网络,但是在其他实施方案中也可以使用其他有线或无线网络。遗留接口模块 73 以任何服务器模块通过背板 77 传输信息相同的方式,经由背板 77 将在遗留端口 74 处接收到的信息传送到光 I/O 模块 71。光 I/O 模块 71 以和前面针对 PBS 网络 10(图 1)的入口节点所描述的基本相同的方式,将来自遗留接口模块 73 的信息形成为 PBS 突发。然后,就像前面针对 PBS 网络 10 所描述的那样,所述 PBS 突发被调度、E-O 转换并经由 PBS 网络传送到客户。

[0078] 相反,从 PBS 网络进入单元 70,以经由遗留网络传输到目的地的流量在 PBS 光端口 72 处以光控制和数据 PBS 突发的形式由单元 70 接收。光 I/O 模块 71 O-E 转换在 PBS 光端口 72 处接收的光控制和数据突发,解帧该 PBS 突发,并且将 PBS 控制和数据突发解复用为例如由 IP 分组和 / 或以太网帧组成的单独的流。接着,各个单独的流经由背板 77 被传输到遗留接口模块 73。然后,遗留接口模块 73 将各个单独的流量流传输到遗留网络。

[0079] 图 8 根据本发明的一个实施方案图示了光 I/O 模块 71(图 7)。在该实施方案中,光 I/O 模块 71 包括网络处理器单元 802(该单元可能具有多个网络处理器)、总线桥 804、队列单元 806、成帧器单元 808(具有成帧器和解帧器的功能,如方块 808₁ 和 808₂ 所示)、E-O 接口 810、O-E 接口单元 816、网络处理器缓冲器 820、流量成形器 824 和流量成形器缓冲器 826。在一个实施方案中,背板交换结构 830 包括 PCI Express 总线,但是在其他实施方案中也可以使用其他适合的总线。因此,总线桥 804 可以使用在商业上可获得的 PCI 桥设备或芯片组来实现。

[0080] 在该实施方案中,光 I/O 模块 71 的这些元件如下互连。总线桥 804 经由互连 838 与背板交换结构 830 相连,以支持并行双向流量。总线桥 804 也经由电互连 839 与流量成形器 824 相连。为清楚起见,电互连 838、839 和图 8 中的其他信号互连被描绘为单个互连线(即使连接可以包括几条信号互连线)。

[0081] 流量成形器 824 分别经由互连 840 和 841 与网络处理器单元 802 和缓冲器 826 相连。网络处理器单元 802 分别经由互连 842 和 843 与队列单元 806 和缓冲器 820 相连。队列单元 806 又经由互连 844 与 PBS 成帧器 / 解帧器 808 相连。

[0082] 如图 8A 所示,在一些实施方案中,网络处理器单元 802 包括入口网络处理器 860 和出口网络处理器 862。因而,在光 I/O 模块 71 的一些实施方案中,互连 840 和 842 都连接到入口网络处理器 860。

[0083] 此外,如图 8A 所示,在一些实施方案中,队列单元 806 可以包括数据队列 870 和 872、控制队列 874 以及耦合到队列 870、872 和 874 的输出端口的电气开关或解复用器 876。因此,在一些实施方案中,队列 870、872 和 874 的输入端口经由开关或复用器(未示出)被连接到互连 842。此外,在一些实施方案中,开关 876 的输出端口可以与互连 844 相连。

[0084] 在其他实施方案中,在网络处理器单元 802 中可以使用不同数量的处理器(例如,单个处理器)。此外,在一些实施方案中,在队列单元 806 中可以使用不同数量的队列。例如,队列单元不需要包括专用控制队列和/或两个数据队列。多个队列可被用来为使用不同的属性(例如不同的优先级)建立多个突发而提供存储。

[0085] 再次参考图 8,PBS 成帧器单元 808 经由互连 846 与 E-O 接口 810 相连。E-O 接口 810 又经由互连 848 与 PBS 网络 60(图 6)的剩余部分相连。O-E 接口 816 经由互连 850 与 PBS 网络 60 的剩余部分相连。O-E 接口 816 还经由互连 852 与成帧器单元 808 相连。成帧器单元 808 还经由互连 854 与网络处理器单元 802 相连。在一个实施方案中,互连 864 与网络处理器 862(图 8A)相连。网络处理器单元 802 经由互连 856 与总线桥 804 相连。下面结合图 9 和 10 来描述光 I/O 模块 71 在向 PBS 网络及从 PBS 网络传输信息时的操作。

[0086] 参考图 8 和 9,光 I/O 模块 71 如下执行 PBS 出口操作,即,将信息从 PBS 网络传输到遗留网络和/或单元 70(图 7)的服务器模块。光 I/O 模块 71 将经由互连 850 从 PBS 网络接收的光 PBS 突发转换成电信号。在该实施方案中,O-E 接口 816 执行 O-E 转换。这个操作流由框 902 来表示。

[0087] 然后,所接收的 O-E 转换后的 PBS 突发被解帧。在该实施方案中,成帧器单元 808 经由互连 852 从 O-E 接口 816 接收 O-E 转换后的 PBS 突发,并解帧该 PBS 突发。例如,在一个实施方案中,如在上述美国专利申请 No. (律师案卷号 No. 42P15725) 中描述的那样,PBS 突发可以被构造成帧。在其他实施方案中,可以使用不同的成帧格式。这个操作流由框 904 来表示。

[0088] 然后对包括在 PBS 突发中的信息进行处理。在该实施方案中,网络处理器单元 802 经由互连 854 从成帧器单元 808 接收解帧后的 PBS 突发,并且执行处理。例如,在一些实施方案中,网络处理器单元 802 可以抽取地址和有效载荷信息,对头部和/或有效载荷信息执行纠错,级联有效载荷,重新组装分段后的有效载荷,等等。网络处理器单元 802 可以使用缓冲器 820 在以上处理操作期间临时存储信息。在一个实施方案中,出口网络处理器 862(图 8A)处理解帧后的突发。这个操作流由框 906 来表示。

[0089] 然后,通过背板交换结构 830 来传送处理后的信息。在该实施方案中,总线桥 804 经由互连 856 从网络处理器单元 802 接收处理后的信息,并通过背板交换结构 830 将该信息以正确的格式,并且利用适当的总线控制信号(例如,根据 PCI 协议)传送到正确的目的地。该信息的目的地例如可以是与遗留网络相连的设备(此时信息被传送到遗留接口模块 73)或服务器模块(即,服务器模块 75₁-75₄之一)。这个操作流由框 908 来表示。

[0090] 参考图 8 和 10,光 I/O 模块 71 如下执行 PBS 入口操作,即,从遗留网络和/或单元 70(图 7)的服务器模块传输信息到 PBS 网络。光 I/O 模块 71 接收将以电信号的形式通过 PBS 网络 60(图 6)来传送的信息。在该实施方案中,总线桥 804 经由互连 838 从背板交换结构接收信息。在该实施方案中,该信息可以是经由遗留接口 73 来自遗留网络,或者是来自服务器模块 75₁-75₄(图 7)之一。这个操作流由框 1002 来表示。

[0091] 然后,接收的信息被成形,以帮助改善 PBS 网络 60(图 6)中的流量流。在该实施方案中,流量成形器 824 经由互连 839 从总线桥 804 接收信息,并且使该信息成形。例如,在一个实施方案中,流量成形器 824 对信息执行操作,以减少由自相似效应引起的输入流量流的长期依存关系以及相关结构。流量成形器 824 可以被配置来执行本领域中公知的任何适当的流量成形算法或技术。流量成形器 824 可以使用缓冲器 826 在执行流量成形操作时临时存储信息。这个操作流由框 904 来表示。

[0092] 然后,成形后的信息被形成为 PBS 突发。在该实施方案中,网络处理器单元 802 经由互连 840 从流量成形器 824 接收成形后的信息。网络处理器单元 802 然后处理该信息,以如同上面针对 PBS 网络 10(图 1)中的入口节点所描述的那样,形成并调度 PBS 控制和数据突发。在其他实施方案中,信息被组装成任何适当的集总格式,以通过光突发网络(不一定是 PBS 网络)来传送。在一个实施方案中,入口网络处理器 860(图 8A)处理流量成形后的信息。此外,在该实施方案中,网络处理器单元 802 使用队列单元 806 在控制和数据突发正在被形成时,在它们被调度来通过 PBS 网络传送之前存储它们。这个操作流由框 1006 来表示。

[0093] 然后,突发被封装到帧中,以通过 PBS 网络来传送。在该实施方案中,成帧器单元 808 经由互连 844 从队列单元 806 接收突发,并执行成帧操作。在一个实施方案中,如上述美国专利申请 No. (律师案卷号 No. 42P15725)所描述的那样,突发被构造成帧。在其他实施方案中,可以使用其他不同的封装格式。这个操作流由框 1008 来表示。

[0094] 成帧后的突发然后被转换为光信号,并在排定的时间上通过 PBS 网络来传送。在该实施方案中,E-0 接口 810 经由互连 846 从成帧器单元 808 接收成帧后的突发(即,PBS 控制和数据突发)。E-0 接口 810 然后执行 E-0 转换,并在排定的时间上,在 PBS 网络的保留 PBS TDM 信道中传送光信号。这个操作流由框 1010 和 1012 来表示。

[0095] 这里描述了用于实现光子突发交换网络的方法和装置的实施方案。在以上描述中,阐述了很多具体的细节,以提供对本发明的实施方案的完整理解。然而,本领域的技术人员将会认识到,没有这些具体的细节,或者采用其他方法、组件、材料等也可以实现本发明的实施方案。此外,没有详细示出或描述公知的结构、材料或操作,以免模糊了该描述。

[0096] 在整个说明书中提及“一个实施方案”或“实施方案”是指结合该实施方案所描述的具体特征、结构或特性至少包括在本发明的一个实施方案中。因此,短语“在一个实施方案中”或“在实施方案中”在整个说明书中不同地方的出现不一定是指同一实施方案。此外,在一个或多个实施方案中,可以以任何适当的光的方式将具体的特征、结构或特性组合起来。

[0097] 因此,本发明的实施方案可以被用作或者支持在某种形式的处理核(例如计算机的 CPU 或模块的处理器)上执行的,或者以其他方式在机器可读介质上或内实施或实现的软件程序。机器可读介质包括用于以机器(例如,计算机)可读的形式来存储或传输信息的机构。例如,机器可读介质例如可以包括只读存储器(ROM);随机访问存储器(RAM);磁盘存储介质;光存储介质;和闪存器件等。此外,机器可读介质可以包括传播信号,例如电、光、声或其他形式的传播信号(例如载波、红外信号、数字信号等)。

[0098] 在以上说明书中已描述了本发明的实施方案。然而清楚的是,在不偏离在所附权利要求书中阐述的更宽精神和范围的情况下,可以对其做出各种修改和改变。因此,说明书

和附图应被视为示意性的而非限制性的。

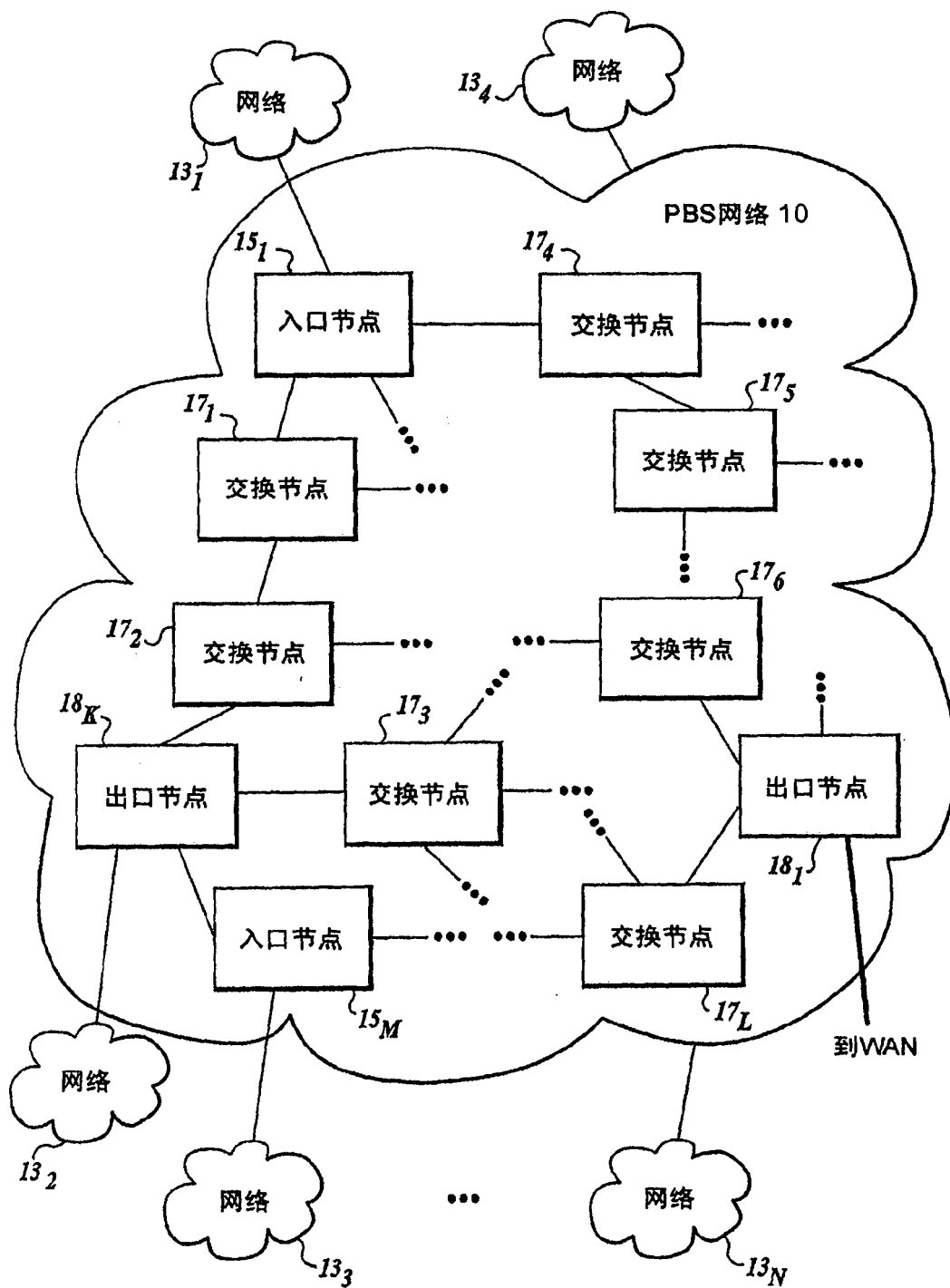


图 1

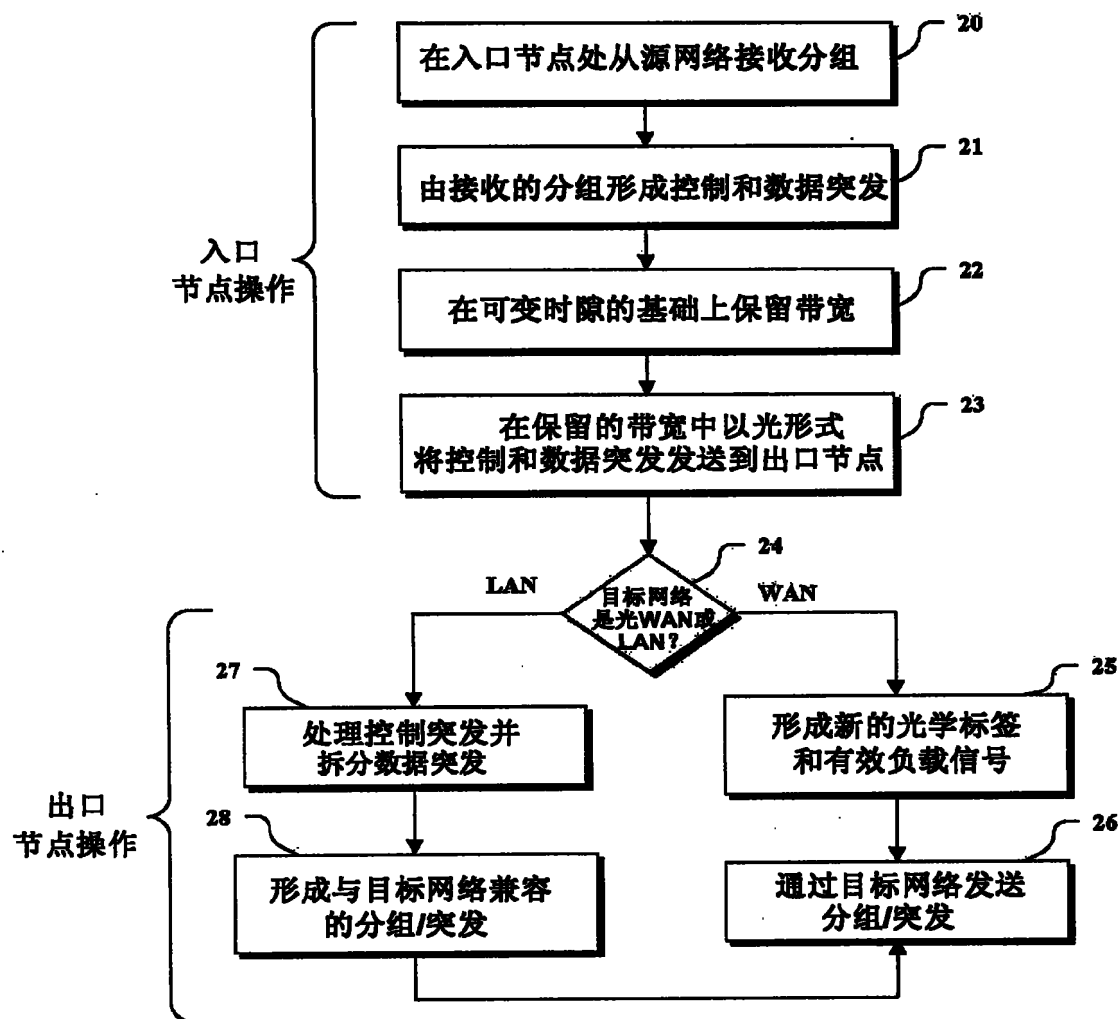


图 2

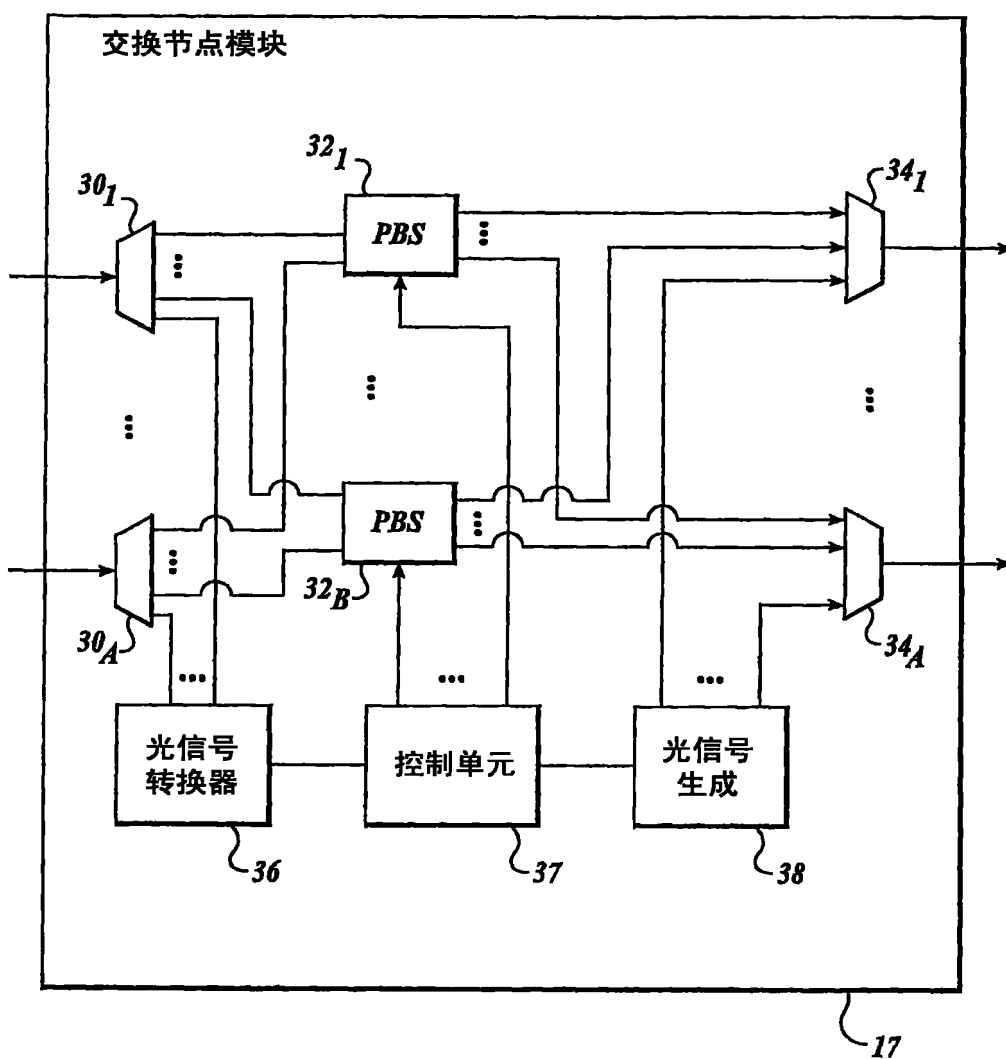


图 3

光数据突发格式

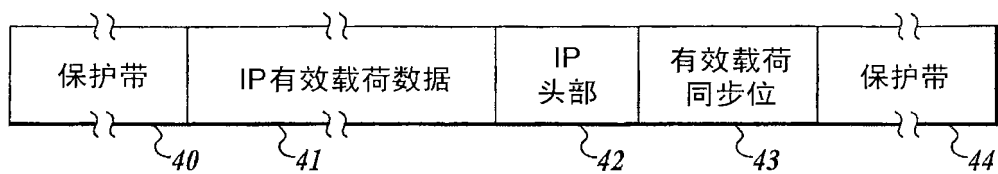


图 4A

光控制突发格式

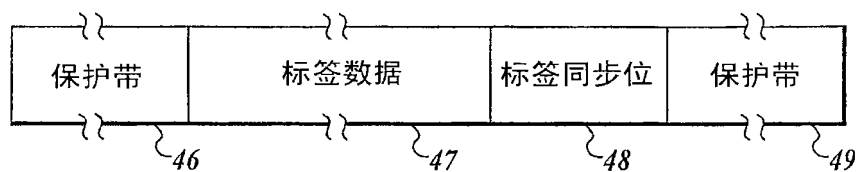


图 4B

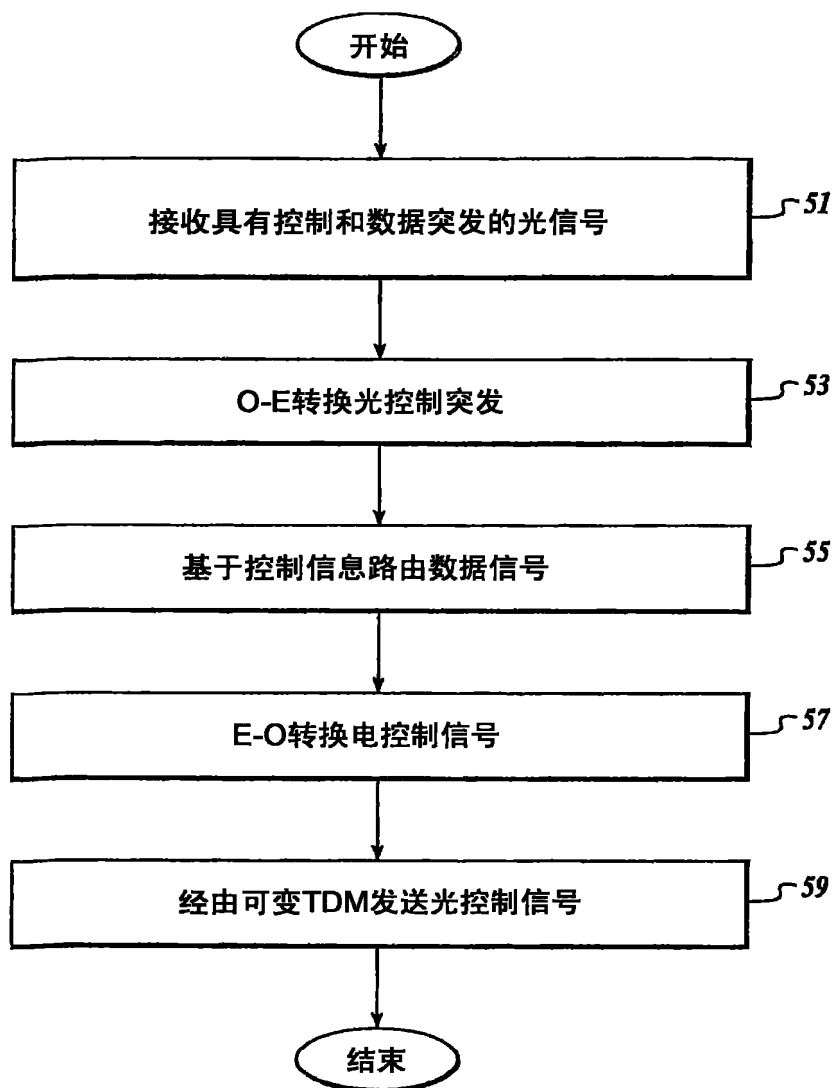


图 5

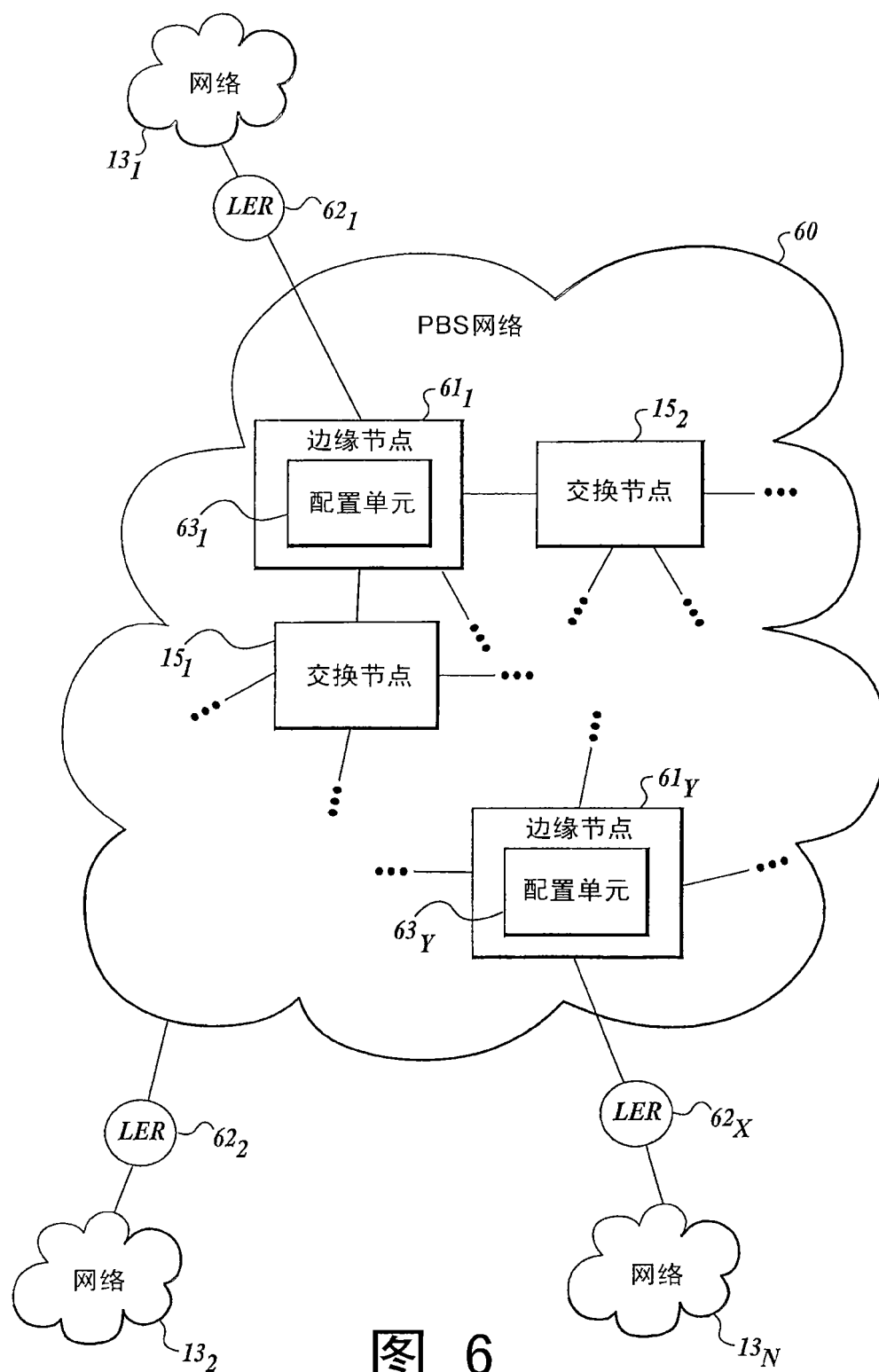


图 6



图 6A

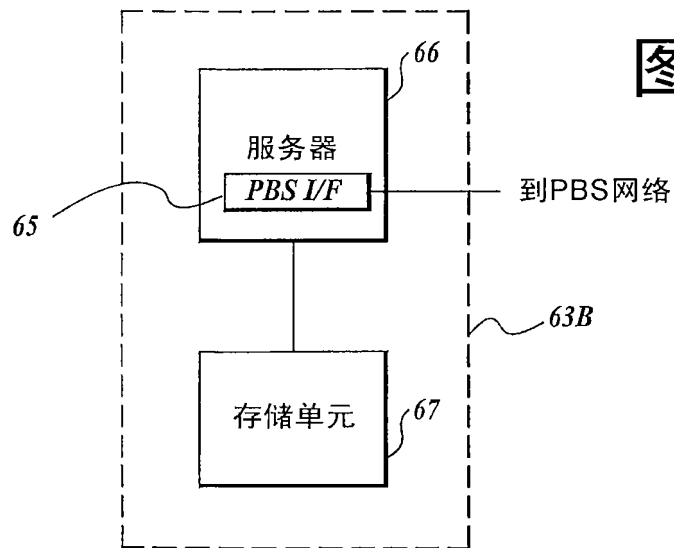


图 6B

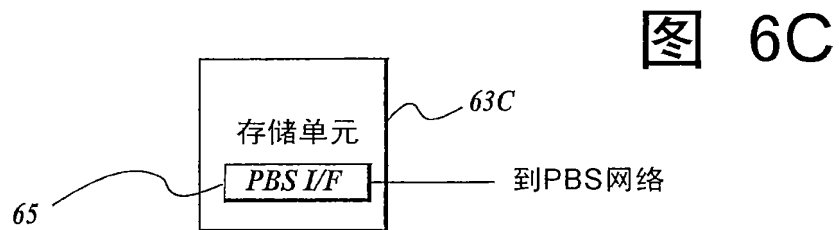


图 6C

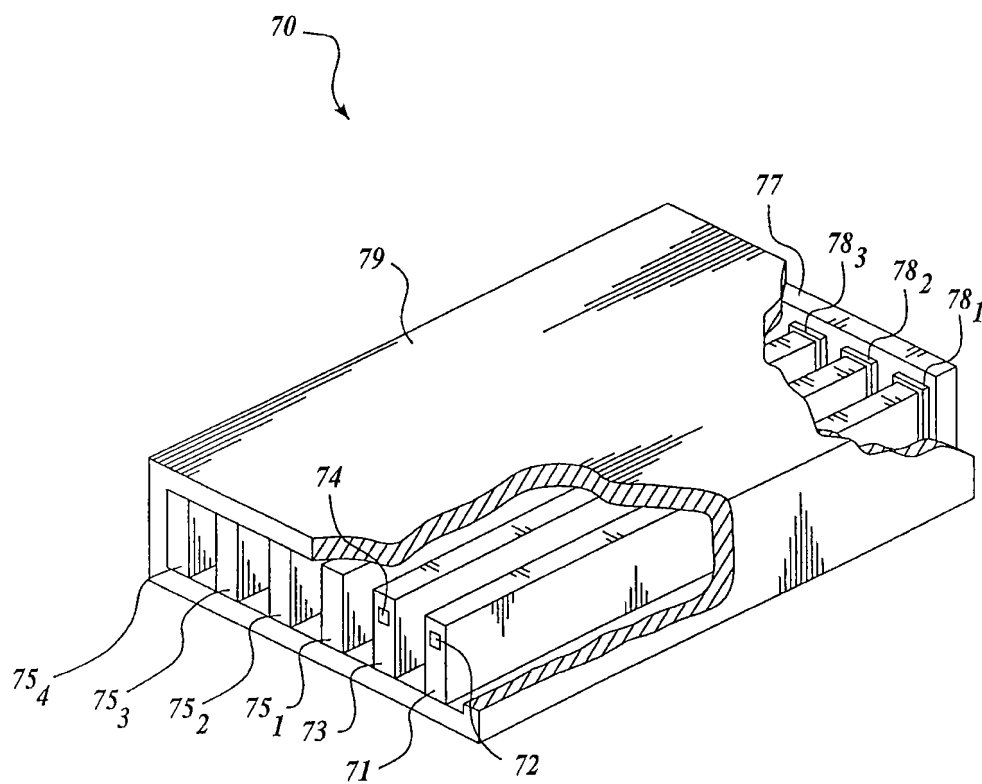
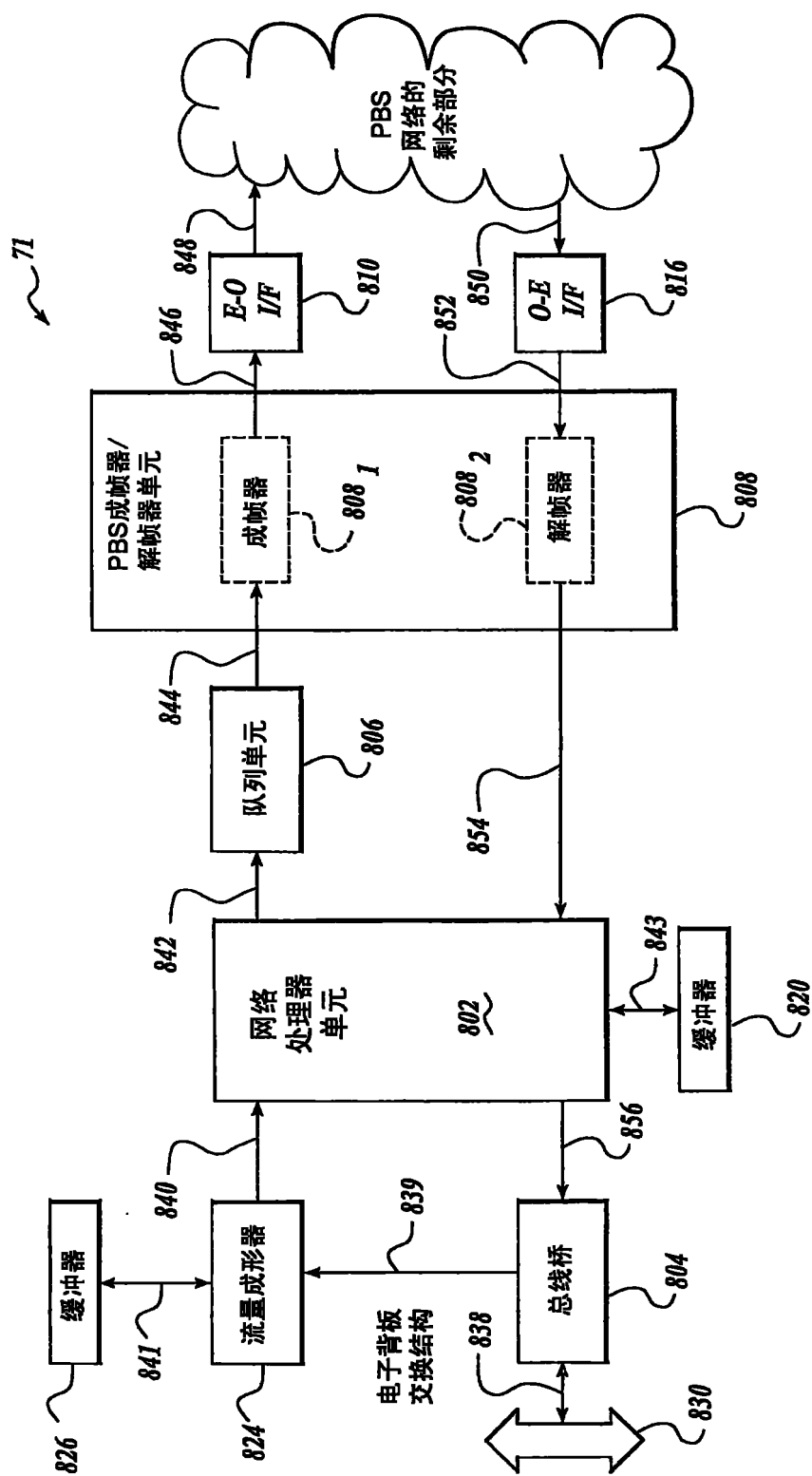


图 7



8
[A]

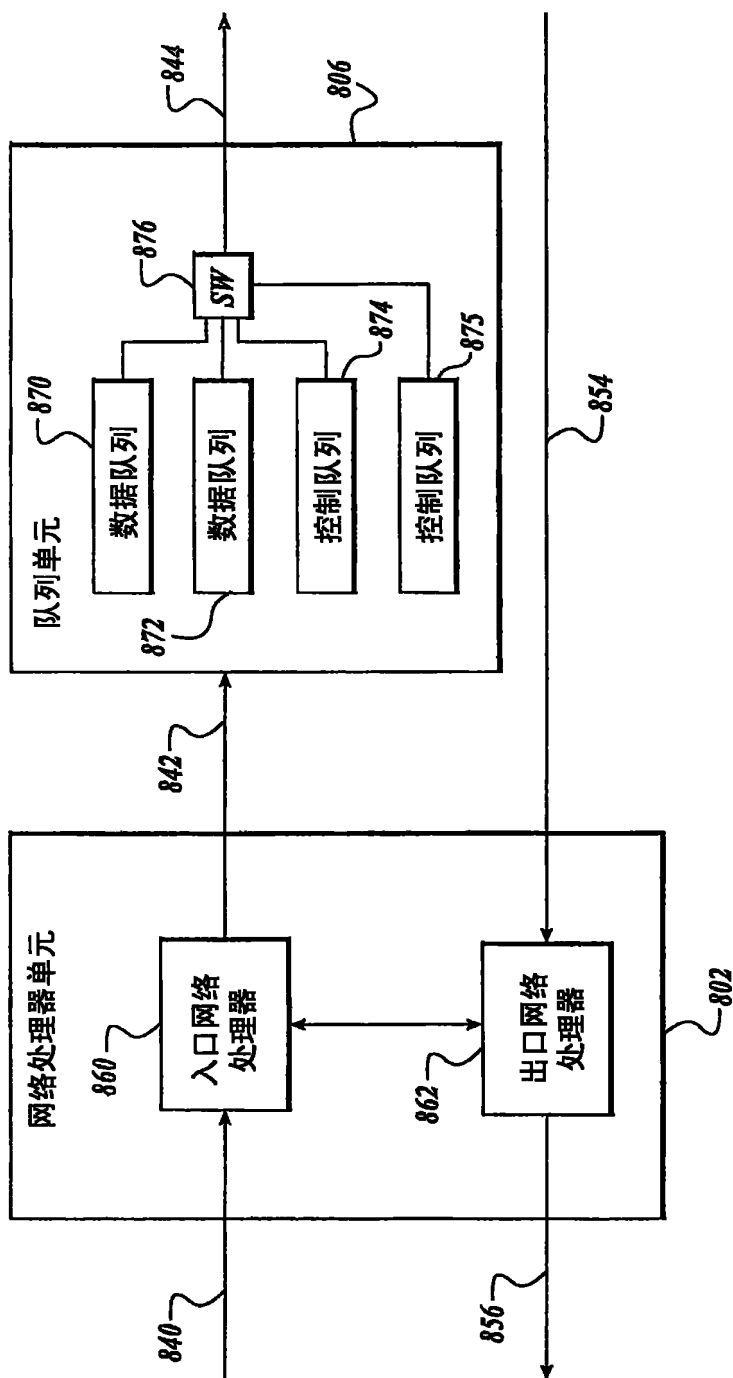


图 8A

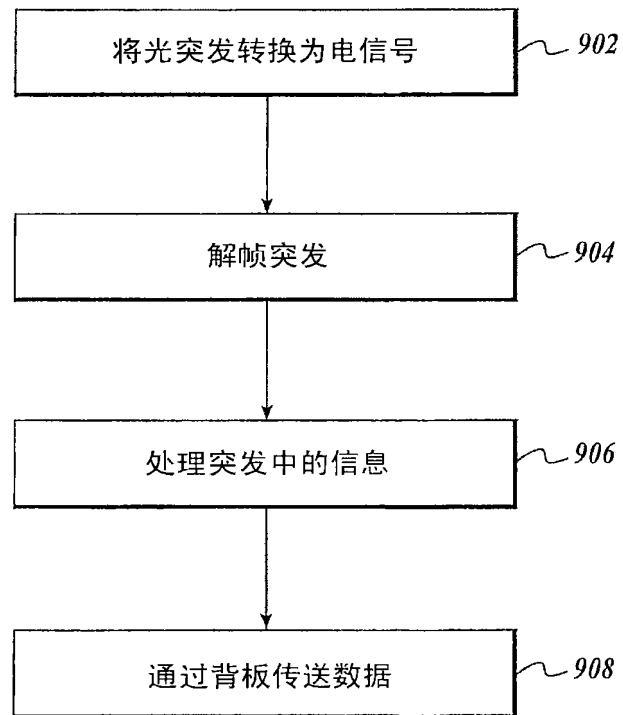


图 9

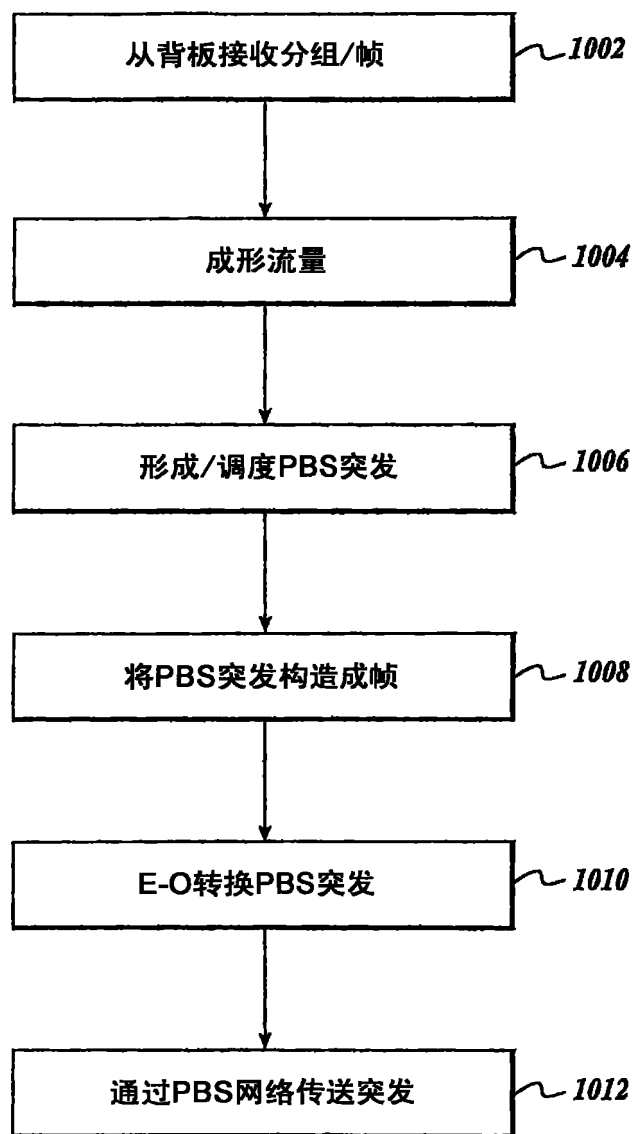


图 10