

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04L 12/56 (2006.01)

H04L 5/14 (2006.01)



[12] 发明专利申请公开说明书

[21] 申请号 200510087168.1

[43] 公开日 2006年2月1日

[11] 公开号 CN 1728687A

[22] 申请日 2005.7.27

[21] 申请号 200510087168.1

[30] 优先权

[32] 2004.7.27 [33] US [31] 10/898,989

[71] 申请人 阿尔卡特公司

地址 法国巴黎

[72] 发明人 M·佐兰诺维奇 B·麦克布赖德

P·拉比诺维奇

[74] 专利代理机构 北京市中咨律师事务所

代理人 杨晓光 于 静

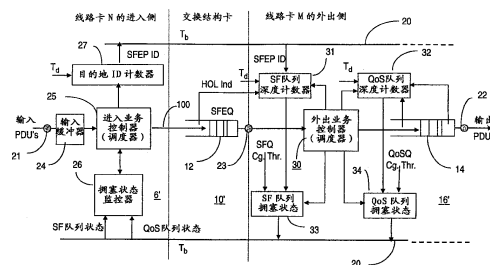
权利要求书 5 页 说明书 11 页 附图 3 页

[54] 发明名称

用于闭环、带外背压机制的方法和装置

[57] 摘要

背压机制使用 TDM 背压总线，为每个端口卡分配了时隙。在其时隙期间，每个进入卡/端口在所述总线上写其在上一个数据更新周期期间传送到外出端口的分组的数量。所述外出端口读取该信息，并用其计算交换结构输出队列的当前深度。除了从进入卡接收的信息之外，外出卡还对接收自适当的交换结构端口的分组的数量进行计数，并基于此估计/计算所述交换结构输出队列的当前深度。通过所述队列深度与分别的阈值的比较，为所有队列计算拥塞状态。每个外出卡使用所述拥塞状态以产生到进入卡的背压信号。进入卡使用所述信号来决定是否向目的地外出卡/端口发送业务。



1. 一种在通信节点使用的背压机制，所述通信节点用于沿着数据路径发送协议数据单元（PDU），所述数据路径是通过交换结构（SF）外出端口在进入端口和外出端口之间被建立的，所述背压机制包括：

背压总线，用于在所述进入端口和所述外出端口之间建立带外双向连接；

外出逻辑，用于估计表征沿所述数据路径的拥塞特性的拥塞状态指令；和

进入逻辑，用于基于所述拥塞状态指令来控制所述 PDU 沿所述数据路径的传输，

其中，所述外出逻辑通过所述背压总线，向所述进入逻辑传送所述拥塞状态指示符，而所述进入逻辑通过所述背压总线向所述外出逻辑传送业务统计。

2. 根据权利要求 1 的背压机制，其中，所述外出端口维持 QoS 队列，以基于所述 PDU 的 QoS，按照到达所述外出端口的顺序来保持所述 PDU，并且所述 SF 外出端口维持 SF 队列，以按照到达所述 SF 外出端口的顺序来保持所述 PDU。

3. 根据权利要求 1 的背压机制，其中，所述数据路径的更新周期是 T_d ，并且所述背压总线的更新周期 $T_b < T_d$ 。

4. 根据权利要求 2 的背压机制，其中，所述拥塞状态指令包括 QoS 队列拥塞状态和 SF 队列拥塞状态，所述 QoS 队列拥塞状态提供了所述 QoS 队列的拥塞的测量，所述 SF 队列拥塞状态提供了所述 SF 队列的拥塞的测量。

5. 根据权利要求 4 的背压机制，其中，所述外出逻辑包括：

外出业务控制器，用于从所述 SF 外出端口向所述 QoS 队列传送所述 PDU；

QoS 深度计数器，用于确定所述 QoS 队列的当前深度（D）；和

QoS 队列拥塞状态单元,用于比较所述当前深度 D 和 QoS 拥塞阈值,并产生所述 QoS 队列拥塞状态。

6. 根据权利要求 4 的背压机制,其中,所述外出逻辑包括:

SF 队列深度计数器,用于确定所述 SF 队列的当前深度 (D_{SF}); 和 SF 队列拥塞状态单元,用于比较所述当前深度 D_{SF} 和 SF 拥塞阈值,并产生所述 SF 队列拥塞状态。

7. 根据权利要求 1 的背压机制,其中,所述进入逻辑包括:

拥塞状态监控器,用于保持通过所述背压总线所接收的所述拥塞状态指示;

进入业务控制器,用于识别所述 PDU 的所述 SF 队列和所述 QoS 队列,并根据所述拥塞状态指令,调度所述 PDU 沿所述数据路径的传输; 和

目的地 ID 计数器,用于产生所述业务统计,并沿所述背压总线向所述外出端口传送所述业务统计。

8. 根据权利要求 1 的背压机制,其中,所述业务统计包括在上一个数据更新周期 T_d 期间,从所述进入端口沿所述数据路径传送的 PDU 的计数。

9. 根据权利要求 8 的背压机制,其中,根据所述 SF 外出端口的标识来评估所述计数。

10. 根据权利要求 1 的背压机制,其中,所述背压总线利用固定长度的背压帧,根据时分复用 (TDM) 协议操作。

11. 根据权利要求 10 的背压机制,其中,在所述背压帧内所分配的时隙 T_{be} 期间,所述外出端口在所述背压总线上写所述拥塞状态指令,而所述进入端口从所述时隙 T_{be} 读取所述拥塞状态指令。

12. 根据权利要求 10 的背压机制,其中,在所述背压帧内所分配的时隙 T_{bi} 期间,所述进入端口在所述背压总线上写所述业务统计,而所述外出端口从所述时隙 T_{bi} 读取该业务统计。

13. 根据权利要求 10 的背压机制,其中,在所述背压帧内所分配的

时隙 T_{bsf} 期间，所述 SF 外出端口在所述背压总线上写附加拥塞状态数据，而所述外出端口从所述时隙 T_{bsf} 读取所述附加拥塞状态数据。

14. 一种用于带外背压总线的帧，该带外背压总线用于提供沿数据路径的闭环，所述数据路径是通过通信节点的交换结构 (SF) 外出端口、在进入端口和外出端口之间被建立的，所述帧包括：

外出端口时隙 T_{be} ，用于承载拥塞状态指令，该指令针对数据更新间隔 T_d 提供了所述数据路径的拥塞状态；和

进入端口时隙 T_{bi} ，用于承载 PDU 的计数，所述 PDU 是在所述数据更新间隔 T_d 期间向所述 SF 外出端口传送的。

15. 根据权利要求 14 的帧格式，还包括 SF 时隙 T_{bsf} ，用于承载附加状态数据，该数据是关于针对所述数据更新间隔 T_d 在所述 SF 外出端口的拥塞的。

16. 根据权利要求 14 的帧格式，其中，所述拥塞状态指令包括 QoS 队列拥塞状态和 SF 队列拥塞状态，所述 QoS 队列拥塞状态提供了所述外出端口的拥塞的测量，所述 SF 队列拥塞状态提供了所述 SF 外出端口的拥塞的测量。

17. 一种用于改善沿数据路径的业务吞吐量的方法，所述数据路径是在通信节点处通过交换结构 (SF) 外出端口在进入端口和外出端口之间被建立的，所述方法包括：

- a) 估计表征沿所述数据路径的拥塞特性的拥塞状态指令；
- b) 利用背压总线在所述进入端口和所述外出端口之间建立带外双向连接；
- c) 通过所述背压总线向所述进入端口传送所述拥塞状态指示符，并通过所述背压总线向所述外出端口传送业务统计；和
- d) 基于所述拥塞状态指令，控制所述 PDU 沿所述数据路径的传输。

18. 根据权利要求 17 的方法，其中，所述拥塞状态指令包括 QoS 队列拥塞状态和 SF 队列拥塞状态，所述 QoS 队列拥塞状态提供了所述外出端口的拥塞的测量，所述 SF 队列拥塞状态提供了所述外出端口的拥塞的

测量。

19. 根据权利要求 18 的方法，其中，所述步骤 (a) 包括：

a1) 基于所述 PDU 的 QoS，按照到达所述外出端口的顺序，将所述 PDU 置于 QoS 队列中；

a2) 确定所述 QoS 队列的深度 D ；

a3) 比较所述深度 D 和 QoS 拥塞阈值，用来产生所述 QoS 队列拥塞状态。

20. 根据权利要求 19 的方法，其中，所述步骤 (a) 还包括：

a4) 按照到达所述 SF 外出端口的顺序，将所述 PDU 置于 SF 中；

a5) 确定所述 SF 队列的深度 D_{SF} ；

a6) 比较所述深度 D_{SF} 和 SF 拥塞阈值，用来产生所述 SF 队列拥塞状态。

21. 根据权利要求 19 的方法，其中，所述步骤 a2) 还包括：

确定上一个数据更新周期期间，被置于所述 QoS 队列中的 PDU 的数量；

确定上一个数据更新周期期间，从所述 QoS 队列所传送的 PDU 的数量；和

通过将置于所述 QoS 队列中的 PDU 的数量与从所述 QoS 队列所传送的 PDU 的数量的差值，与所述深度 D 相加，来确定针对当前数据更新周期的被更新的深度 D' 。

22. 根据权利要求 20 的方法，其中，所述步骤 a5) 包括：

根据所述业务统计来建立 PDU 计数，该计数指出了在上一个数据更新周期期间被置于所述 SF 队列中的 PDU 的数量；

确定上一个数据更新周期期间，从所述 SF 队列所传送的 PDU 的数量；和

通过将置于所述 SF 队列中的 PDU 的数量与从所述 SF 队列所传送的 PDU 的数量的差值，与所述深度 D_{SF} 相加，来确定针对当前数据更新周期的被更新的深度 D_{SF}' 。

23. 根据权利要求 18 的方法, 其中, 所述步骤 (b) 包括:
通过更新周期 T_b 的固定长度帧, 为所述总线提供 TDM 协议;
在所述帧内分配外出端口字段 T_{be} 、进入端口字段 T_{bi} 和 SF 端口字段 T_{bsf} ; 和
在所述分别的字段内, 为每个所述进入、外出和 SF 端口分配时隙,
其中, 当所述端口之一在所述背压总线上写时, 所有剩余的端口可以从所述背压总线读取。

24. 根据权利要求 23 的方法, 其中, 所述步骤 (c) 包括:
在所述外出端口, 在所述外出端口字段 T_{be} 中写入所述 QoS 队列拥塞状态和所述 SF 队列拥塞状态, 并从所述进入端口字段 T_{bi} 读取所述业务统计; 和

在所述进入端口, 在所述进入端口字段 T_{bi} 中写入所述业务统计, 并从所述外出端口字段 T_{be} 读取所述 QoS 队列拥塞状态和所述 SF 队列拥塞状态。

25. 根据权利要求 23 的方法, 其中, 所述步骤 (c) 还包括:
在所述 SF 外出端口, 在所述 SF 端口字段 T_{bsf} 上写关于所述 SF 队列的附加拥塞状态数据;

在所述外出端口, 从所述 SF 端口字段 T_{bsf} 读取所述附加拥塞数据;
和

根据所述附加拥塞数据, 调整所述 SF 队列拥塞状态。

26. 根据权利要求 18 的方法, 其中, 所述步骤 (d) 包括:
在所述进入端口接收所述 QoS 队列拥塞状态和所述 SF 队列拥塞状态;

确定所述 QoS 队列是否拥塞;
如果所述 QoS 队列不拥塞, 则确定所述 SF 队列是否拥塞;
如果所述 SF 队列不拥塞, 则沿所述数据路径, 从所述进入端口向所述外出端口传送所述 PDU; 和

如果任何所述 SF 队列和所述 QoS 队列拥塞, 则丢弃所述 PDU。

用于闭环、带外背压机制的方法和装置

技术领域

本发明涉及通信网络,并且特别涉及用于交换机/路由器系统的闭环背压机制。

背景技术

通信速度的当前需求和现代数据网络的容量需要较高性能的网络结构,该结构是高度可升级的,并且能够有效地处理大量多播和单播业务。期望这些网络支持多媒体应用并能够动态地重新配置,以便保证针对约定的服务质量(QoS)的预定带宽量。例如集成的语音、视频和数据应用的新兴服务,增大了这样的需求:网络提供商基于用户在特定时间的特定需求,对每个用户保证某个可变的QoS。该需求只能通过逐步将新的硬件和软件开发集成到网络中来达到,所述开发用于处理随着网络发展和客户要求新服务进展而出现的特定问题。

通信应用的不断增长的种类需要节点结构使用广泛选择的线路和交换板设计,其能够实现灵活的高性能的数字、模拟或混合的交换系统。当接收端口不能按照向其传送信息的速度消化该信息时或者当业务的优先权改变时,就出现了问题。当积聚在交换机或路由器中的交叉点处的数据量超过数据存储的容量时,数据被丢弃,因而迫使目的地端口请求数据重新传输。处理所述拥塞问题的解决方案是通过在每个可能的交叉点提供足够的数据存储,使用“强力攻击(brute-force)”的方法。然而,该方案相当昂贵,还需要构成所述网络交换机和/或路由器的板上的额外不动产(real-estate)容量。

常见的是向数据源预分配接收机的缓冲器信用,并通知相应的发送器

能够发送多少数据。根据在外出端口 (egress port) 或者在节点内部检测到拥塞, 所述接收机扣留缓冲器信用, 迫使相应的发送器降低发送分组的速度, 或者完全停止传送。该过程被称为“背压 (backpressure)”拥塞控制。通常, 背压机制提供了交换机/路由器的外出和进入线路卡之间的通信, 力图最小化数据丢失、线头阻塞和时延。通过有效管理所述交换结构 (switch fabric) 和外出线路卡处的拥塞级别, 以及通过向所述进入线路卡及时地传播所述信息, 达到了背压拥塞控制。该通信通常通过所述交换结构卡 (带内) 来承载。

带内背压机制假设所述背压路径的所有元件能够支持该机制。然而, 许多当前的交换结构实现不支持任何种类的带内背压机制。在这些情况下, 当需要升级交换机/路由器以实现具有 QoS 保证的新应用时, 需要使用能够提供相同能力的可选解决方案, 而不是替换整个交换机/路由器。

发明内容

本发明的目的是提供闭环背压机制, 该机制能够完全地或部分地减轻现有技术拥塞控制机制的缺点。

因此, 本发明为通信节点提供了背压机制, 所述节点沿着数据路径传送协议数据单元 (PDU), 所述数据路径是通过交换结构 (SF) 外出端口, 在进入端口 (ingress port) 和外出端口之间被建立的。该背压机制包括背压总线, 用来在所述进入端口和外出端口之间建立带外双向连接; 还包括外出逻辑, 用来估计表征沿所述数据路径的拥塞特性的拥塞状态指令; 还包括进入逻辑, 用来基于所述拥塞状态指令, 控制沿所述数据路径的 PDU 的传输, 其中, 所述外出逻辑通过所述背压总线向所述进入逻辑传送所述拥塞状态指示符, 而该进入逻辑通过所述背压总线向该外出逻辑传送业务统计。

本发明还涉及用于带外背压总线的帧, 用来提供沿着通过通信节点的交换结构 (SF) 外出端口、在进入端口和外出端口之间所建立的数据路径的闭环, 该帧包括: 用来承载拥塞状态指令的外出端口时隙 T_{be} , 该拥塞

状态指令针对数据更新间隔 T_d 提供数据路径的拥塞状态；以及用来承载 PDU 的计数的进入端口时隙 T_{bi} ，该 PDU 是在数据更新间隔 T_d 期间被传送到 SF 外出端口的。

本发明还涉及用于改善沿数据路径的业务吞吐量的方法，所述数据路径是通过交换结构 (SF) 外出端口，在进入端口和外出端口之间的通信节点处被建立的。该方法包括 (a) 估计表征沿所述数据路径的拥塞特性的拥塞状态指令；(b) 利用背压总线，在进入端口和外出端口之间建立带外双向连接；(c) 通过所述背压总线，向进入端口传送所述拥塞状态指示符，并通过该背压总线向外出端口传送业务统计；和 (d) 基于所述拥塞状态指令，控制沿所述数据路径的 PDU 传输。

本发明的机制尤其适合需要实时发送和交换数字化数据信元的通信系统。

有利地，根据本发明的闭环背压机制，不需要用复杂昂贵的、支持 QoS 保证的交换结构卡来替换现有交换机/路由器上的交换结构卡。也就是说，根据本发明的机制即使在所述交换结构卡不能提供时也能够提供服务区分；这能被简单地实现，而不需要替换所述交换结构卡。这就延长了交换机/路由器的使用期，使网络提供商能够通过通过对现有设备的较小投资获得更多的收入。

建造新系统时也应用了类似的考虑。这样，不必使用/设计昂贵而复杂的新交换结构卡来向用户提供更多的 QoS 性能，所述系统能够使用简单且不昂贵的现有交换结构卡，并通过实现本发明的机制，仍能达到约定的 QoS 性能保证。

附图说明

结合附图，根据下面对优选实施例的更详细描述，本发明的前述和其它目的、特征和优点将是显而易见的，其中：

图 1 是示出了带外背压机制的结构的框图；

图 2 是根据本发明的一个实施例的背压机制的框图；和

图 3 示出了背压总线 and 特定节点配置的关联逻辑的实现。

具体实施方式

交换结构的拥塞导致了较大的时延、分组丢失和阻塞。假设不存在针对所述外出或者交换结构数据路径的带内机制，以在合理的时间帧内对所述进入数据路径指出拥塞状态，带外背压机制成为处理拥塞的实际解决方案。本发明的背压机制是利用背压总线、基于外出和进入线路卡之间的闭环、带外通信的。本发明对于所述交换结构不能保证所需 QoS 的节点尤其地有用。

图 1 示出了所述带外背压机制的结构，所说明的节点具有进入接口 5、外出接口 15 以及接口 5 和 15 之间所连接的交换结构 (SF) 卡 10。每个进入和外出接口包括表示为 6-1、6-2、6-3、6-4 和分别地 16-1、16-2、16-3、16-4 的多个线路卡。每个输出卡 6 为每个外出端口维持逻辑队列 14 (例如八个队列代表八个 QoS 级别)，该队列将协议数据单元 (PDU) 发送到该外出端口，并具有相同的 QoS 保证。例如，对四个外出端口和八个 QoS 级别，外出卡 16 维持 $4 \times 8 = 32$ 个计数器。这些队列在此被称为“QoS 队列”或者“逻辑队列”。所述交换结构卡 10 也为每个交换结构外出端口 22 维持外出队列 12，用来安排交换后接收自不同 SF 进入端口的 PDU。这些队列在此被称为“SF 队列”。所述 PDU 按照到达的顺序被安排在队列 12、14 中，并且如果队列变的拥塞 (所述缓冲器变满)，则该 PDU 将被丢弃。

为了示出所述线路卡事实上具有多个端口，图 1 所示的业务使用了双线。图 1 也以例子的方式示出了通过所述交换机/路由器的 PDU 的路由，下文称之为“数据路径” 100。该例子中，在进入卡 6-2 上的进入端口 21 和外出卡 16-1 上的外出端口 22 之间建立了数据路径 100。这样，进入端口 21 从所述网络中的源接收 PDU，确定卡 10 上的 SF 外出端口 23 的地址，并向交换结构 10 传输该 PDU。该交换结构 10 向 SF 外出端口 23 发送所述 PDU，其中该 PDU 被置于所述 SF 队列 12 内。当轮流到它时，该 PDU 从 SF 外出端口 23 被传输到卡 16-1，其中该 PDU 基于其 QoS，被置于外出

端口 22 的适当的 QoS 队列 14 中。因此，当轮流到它时，外出端口 22 将该 PDU 传输到所述网络内的目的地。下面，所述数据（业务）更新周期表示为 T_d 。

图 1 还用粗线示出了背压总线 20，其提供了所述进入和外出线路卡之间的双向通信。本发明的背压机制针对改善的时延性能和有限的阻塞状况，管理外出队列的大小，通过提供控制 PDU 丢弃的进入线路卡来减少过度的排队。优选地通过背板连接器（backplane connector），总线 20 与参与所述背压机制的所有线路卡相连接。该结构使本发明的交换机元件能够在例如“闭环”ATM 系统中，以最大化通过所述交换机的数据吞吐量，而不丢失信元。

为了总线容量的最佳使用，背压总线 20 使用了 TDM（时分复用）协议。TDM 解决方案避免使用开销比特，并避免使用相关的硬件和软件来进行开销比特处理，也避免使用总线仲裁机制。此处所述背压总线的更新周期表示为 T_b 。

在所述例子中，所有的进入接口卡、外出接口卡和交换结构卡通过总线 20 相互连接。应当指出，如后面所讨论的，也可设想所述交换结构没有连接到总线 20 的变型。根据所提供的 QoS 提供的级别和哪个用户能够使用该服务，也可以使少于全部的进入和外出卡在所述总线上被连接。如上所述，图 1 所说明的解决方案可以被用于这样的节点：由于其交换结构使用了不能进行 QoS 区分的、早期的或不昂贵的卡（板）变型，因而不能保证需要的服务质量。

参与所述背压机制的每个线路卡 6、16 在一个背压帧 T_b 期间，从所述背压总线 20 读取或在该背压总线上写；在所述帧中的时隙被分配，以便当一个卡在总线上写时，其它卡能够读取。背压总线帧格式和速度是专有的，并能针对分别的节点配置而被定制。背压更新周期 T_b 的最大长度必须足够短，以便外出端口能够足够快地对进入端口施加背压，以避免所述交换结构外出缓冲器和外出卡缓冲器的溢出。

图 2 示出了根据本发明的一个实施例的背压机制的更详细的框图。它

是通过 SF 外出端口 23 从输入端口 21 被发送到输出端口 22 的 PDU 的数据路径 100。图 2 还示出了参与所述业务传输和背压机制的进入、外出和交换结构卡上的单元，在此被称为进入逻辑 6'、SF 逻辑 10' 和分别的外出逻辑 16'。

给每个进入卡提供了进入业务控制器 25，其知道所有外出端口 23 的 SF 队列 12 的拥塞状态，以及所有外出端口 22 的所有逻辑队列 14 的拥塞状态，如状态监控器 26 的直观显示。控制器 25 使用进入数据路径 PDU 传输机制，由此其从背压总线 20 读取指出队列 12 和 14 的拥塞状态的状态比特，并相应地更新由监控器 26 维持的分别的状态变量计数器。只要进入端口 22 准备好传输 PDU，状态监控器 26 就检查 SF 队列 12 和该 PDU 的外出队列 14 的状态。每个外出队列 12、14 的状态可以用一个比特来指出，例如对于“发送下一个 PDU”消息使用逻辑“1”，并且对于“停止、队列拥塞”消息使用逻辑“0”。因此根据该状态信息，继续或者停止所述 PDU 的传输。

进入业务控制器 25 首先基于 PDU 地址，识别所述 SF 外出端口 23 和外出端口 22。然后，控制器 25 在外出端口 22 处识别对应将被发送的 PDU 的 QoS 的 QoS 队列 14，并基于该队列的拥塞状态变量来检查分别的队列 14 的状态比特。如果队列 14 的状态比特指出分别的 QoS 队列 14 不拥塞，则第二级比较检查分别的 SF 队列 12 的拥塞比特。如果两个条件都满足，则通过 SF 队列 12、SF 外出端口 23 和外出队列 14，沿着从进入端口 21 到外出端口 22 的数据路径传送所述 PDU。如果这两个队列任何一个拥塞，则不能发送该 PDU，并且控制器 25 寻找下一个 PDU，并按照类似的方法进行。

进入逻辑 6' 也使用进入背压总线使用机制，由此分别的进入端口向外出端口传送业务统计，该业务统计用来计算队列 12、14 的深度。所述统计在该例子中包括在先前的数据路径更新周期 T_d 期间，从进入端口 21 向外出端口 22 发送的 PDU 的数量。在图 2 的实施例 中，对每一被发送的 PDU，进入逻辑 6' 写入向其发送该 PDU 的外出端口的唯一标识，如所述目的地

ID 计数器 27 通常所示。该信息所需要的比特数量取决于所述背压系统中的现用端口的数量。保持这些统计的目的地 ID 计数器 27 在时间 T_d 被复位。如果在上一个数据更新周期没有发送 PDU，则进入端口写零。

在总线 20 上广播所述比特（业务统计）。外出逻辑 16' 读取该统计，外出端口识别与其 ID 相匹配的比特，并使用该标识来增加 PDU 计数，其意味着 PDU 已经由进入端口按照其路径被传送。外出端口 22 的标识符被记在计数器 27 中，直到下一次端口 21 在背压总线 20 上进行写操作。实际上，由于在一个更新周期 T_d 期间能发送最多两个分组，因此针对每个进入端口需要存储最多两个端口 ID。

总而言之，进入逻辑 6' 在分别的总线时隙内向/从总线 20 写和读比特。每个进入逻辑 6' 使用背压总线 20，以向外出逻辑 16' 传送统计（为在先前的更新周期内发送的每个分组，在总线上写入目的地交换结构外出端口/队列 ID），并从所述外出逻辑接收拥塞状态指令（从总线读取针对所有 SF 和 QoS 队列的状态比特）。

根据所述目的地地址，所述交换结构逻辑 10' 接收所述 PDU，并将其置于分别的 SF 队列 12 中。如果该队列已经溢出，则所述交换机被阻塞并且所述 PDU 被丢弃，这导致了低质量的传输。为了减少这种情况的发生，从所述外出卡 16 在总线 20 上控制队列 12 的深度。

因此，外出逻辑 16' 知道由所述 SF 外出队列 12 接收的 PDU 的总数，在此用 PDU_{RxSF} 表示，还知道在上一个更新周期 T_d 内从分别的 SF 外出端口 23 所传输的 PDU 的数量，在此用 PDU_{TxSF} 表示。SF 队列深度计数器 31 计算差值 $D_{SF} = PDU_{RxSF} - PDU_{TxSF}$ ，并将其与所述 SF 外出队列 12 的当前深度相加，或者从该 SF 外出队列 12 的当前深度中减去，以通过与上一个更新周期内的队列相加或相减的 PDU 数量来更新队列 12 的当前深度。每个进入卡的计数器 31 和交换结构传输队列 12 一样多。

当能够在所述背压总线上报告动作（即能在所述总线上写）的进入端口传送了 PDU 时，通过只要从 SF 外出端口 23 接收了 PDU 就递减所述队列深度计数器 31，来确定所述差值 $D_{SF} = PDU_{RxSF} - PDU_{TxSF}$ （队列 12 的

深度)。只要控制器 30 观察到进入端口 21 向所述交换结构外出端口之一发送了 PDU (当其从总线上读取了由所述进入卡传输的分别的 SF 外出端口的 ID), 就递增计数器 31。

最好, 外出业务控制器 30 知道每个 QoS 队列 14 接收的 PDU 的总数, 在此表示为 PDU_{Rx} , 还知道在上一个更新周期 T_d 内从分别的外出端口 22 所传输的 PDU 的数量, 在此表示为 PDU_{Tx} 。QoS 队列深度计数器 32 维持差值 $D = PDU_{Rx} - PDU_{Tx}$, 并将其与所述外出队列 14 的当前深度相加, 或者从该外出队列 14 的当前深度中减去, 以通过与上一个更新周期内的队列相加或相减的 PDU 数量来更新队列 14 的当前深度。

当 PDU 到达分别的外出卡端口时, 通过递增所述队列深度计数器 32, 来确定差值 $D = PDU_{Rx} - PDU_{Tx}$ (队列 14 的深度)。当所述 PDU 离开所述输出端口队列 14 时, 递减计数器 32。另外, 只要外出业务控制器 30 观察到交换结构外出端口不在使用中, 就将计数器 32 复位为零。

图 2 的实施例使用了两个阈值, 以确定任何两个进入/外出端口之间所建立的每个数据路径的拥塞状态指令。即, 其针对队列 14 使用下文称为“QoS 队列拥塞阈值”的第一拥塞阈值, 并且针对 SF 队列 12 使用下文称为“SF 队列拥塞阈值”的第二拥塞阈值。SF 外出队列拥塞状态单元 33 通过比较所述 SF 拥塞阈值和分别的队列的深度 D_{SF} , 计算 SF 外出队列的拥塞状态。QoS 队列拥塞状态单元 34 通过比较 QoS 阈值和分别的外出队列的深度 D , 计算 QoS 队列 14 的拥塞状态。

每次 PDU 离开分别的队列 12 时, 控制器 30 基于深度 D_{SF} 以分别的队列的当前状态对拥塞状态单元 33 进行更新。类似地, 每次 PDU 离开分别的队列 14 时, 控制器 30 基于深度 D 以分别的队列的当前状态对拥塞状态单元 34 进行更新。

所述外出逻辑 16' 向所有进入端口广播所述状态比特。所述进入端口使用外出卡通过背压总线 20 所发送的信号, 来更新每个队列 12、14 的状态变量。然后, 每次所述进入卡发送 PDU 以确定沿数据路径 100 的分别的队列是否拥塞时, 就检查分别的目的地端口的状态值。

总而言之，所述外出逻辑 16' 在每个背压总线周期 T_b ，更新或不更新队列 12 和 14 的拥塞状态比特，并在总线 20 上向所有的进入端口广播该拥塞状态指令。在所述进入卡，进入逻辑 6' 接收所述队列状态，并基于该状态重新开始或者停止沿分别数据路径的 PDU 的传输。同样，所述外出逻辑 16' 在每个背压总线周期 T_b 从所述进入逻辑 6' 接收业务统计（从所述进入端口传输的 PDU 的数量），并使用该统计来更新分别的队列 12、14 的当前深度。

所述背压机制能够通过或者不通过来自所述交换结构卡 10 的拥塞指令来操作背压；图 2 示出的实施例中，所述交换结构 10 发送指出 SF 外出队列 12 中的拥塞状态的比特。由于所述附加拥塞状态数据被用于估计 SF 队列的拥塞状态，所以仅所述外出逻辑 16' 使用所述比特。例如，交换结构卡 10 在总线 20 上传输的消息可以通知外出逻辑 16' 是否超过了所述线头（HOL）阈值。如果所述数据更新周期 T_d 大大长于所述背压总线更新周期 T_b ，则在多个总线更新周期内不改变该消息。如果需要，所述比特也可以被用来检查和调整在外出卡处计算的传输队列深度的估计值。在某些情况下，一旦在超过所述 HOL 阈值时设置了状态比特之一，该状态比特可以在大大长于所述数据更新周期 T_d 的周期内保持不变。在该时间内，外出逻辑完全依赖于所述估计的队列深度。

为了解决所述问题，所述外出逻辑可以使用另一个特性来调整所述队列深度值。例如，外出逻辑 16' 能够计算一个时间周期内从所述交换结构外出端口 23 接收的字节的数量，其取决于链路的速度。如果该数量为零，则假定队列 12 为空。如果当前队列深度值不为零，则在该点将其复位为零。

图 3 示出了针对特定交换机配置的所述背压机制的实际实现的例子。所述线路卡 6、16 可以是语音（模拟）或/和数据卡，其在背板 7 上与所述背压总线 20 相连接。语音卡仅向数据卡、交换机管理卡 10 或控制处理器卡传送 PDU；语音卡之间存在最少的通信或没有通信。

在工作的例子中，所述配置包括八个数据卡和九个语音卡，所述数据卡具有到所述交换机管理卡 10' 的 2 x 100M 端口，所述语音卡具有到所述

卡 10' 的 1x100M 端口。每个卡最多有四个端口。在所述实施例中，利用例如 11、11' 处所示的 FPGA 来实现所述背压总线逻辑，该 FPGA 在每个线路卡上通过总线 19、19' 与网络处理器 13、13' 相连接。在所述网络处理器和所述背压总线具有不同的互连规格时，使用所述 FPGA 来互连所述网络处理器 13、13' 和背压总线。所述背压网络处理器可以是例如 IXP1200' 并且所述总线 19、19' 可以是 IX 总线。所述网络处理器 13、13' 为分别的卡的所有端口维持所述进入和外出逻辑 6、16'。

交换机管理卡 10' 在所述实施例中包括交叉设备 (crossbar device) SF (例如 SMC 48300) 和四个 SF-1 设备 (例如 SMC 48310)，其可以通过接口 17 向 FPGA 18 提供附加拥塞状态数据，该 FGPA 18 还在总线 20 上被连接。

图 3 的插图也以例子的方式示出了所述配置的背压总线帧的格式。该帧包括下面描述的专用于由所述外出逻辑传输的比特的 398-比特字段 1、专用于由所述进入逻辑传输的比特的 290-比特字段 2，以及专用于所述交换结构的 32-比特字段 3。在该例子中，所述背压帧的总长度是 561 比特，这就导致了 52.91 Mbps 的总线速度。

外出端口字段 1 (T_{be}) 使所述外出逻辑能够向所述进入逻辑广播状态变量。每个外出数据卡在每个背压帧 T_b 写 38 个比特。这 38 个比特中的六个专用于两个 SF 外出队列 12 的拥塞状态 (每队列 3 比特)。其余的 32 比特专用于更新外出端口队列 14 的拥塞状态 (四个端口/卡，八个 QoS 队列/端口)，每个 QoS 队列一个比特。每个语音卡 (在该实施例中九个语音卡) 在每个背压帧写一个比特，但是在该情况中，无论哪一个队列 12、14 拥塞，该比特被设置为逻辑 "0"，以对所有分别的进入卡施加背压。同样地，在该例子中 $T_{be}=389$ 比特。

如上所述，每个进入端口针对其在上一个更新周期期间所传送的每一个分组，写入该外出端口的唯一的标识符。一个更新周期内一个进入端口能够传送最多两个分组。因此，每个进入端口在时隙 T_{bi} 期间在所述背压总线帧内写入十个比特。全部地，29 个进入端口每一个总线帧使用 290 个

比特 ($T_{bi}=290$ 比特), 如在该例子中的字段 (时隙) 2 所示。在背压总线 20 上广播所述比特, 并且假定每个外出卡读取指定给该外出卡的比特, 用来计算分别的队列深度和拥塞状态变量。

字段 3 专用于所述 SF 端口, 如图 2 的实施例所示, 该 SF 端口可以在所述背压总线上写。这样, 只有所述外出端口访问所述时隙, 以便接收关于 SF 外出队列 12 的附加拥塞状态数据。在该例子中, 字段 3 为 $T_{bsf}=5$ 比特, 以能够识别唯一的 $32 (2^5)$ 个外出端口。

为了清楚地说明本发明, 尽管在此参考特定的硬件和软件实现描述了本发明, 然而, 将如同附加的权力要求所定义的本发明应用于不同通信环境的变型, 对于本领域的普通技术人员而言将是显而易见的。

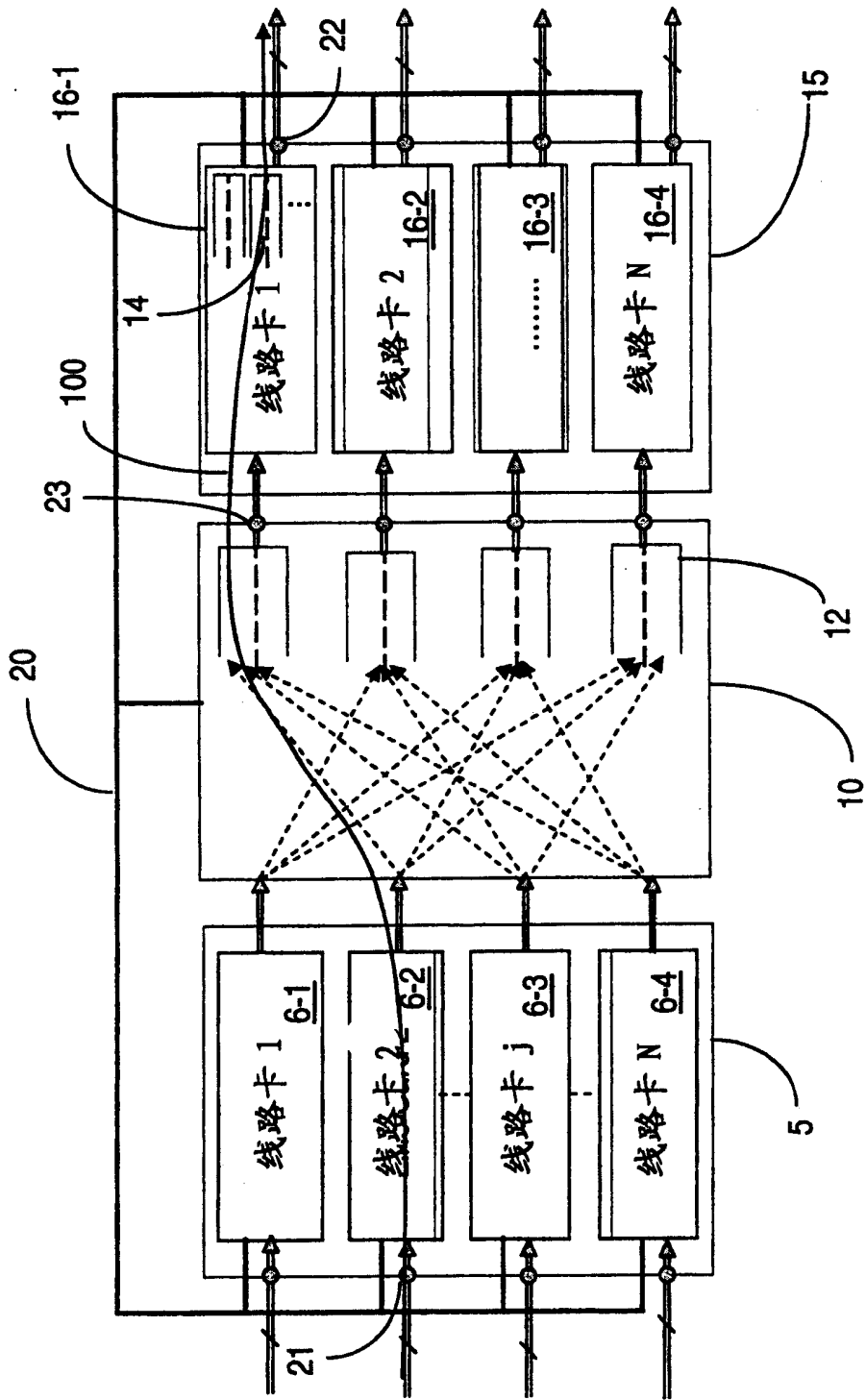


图 1

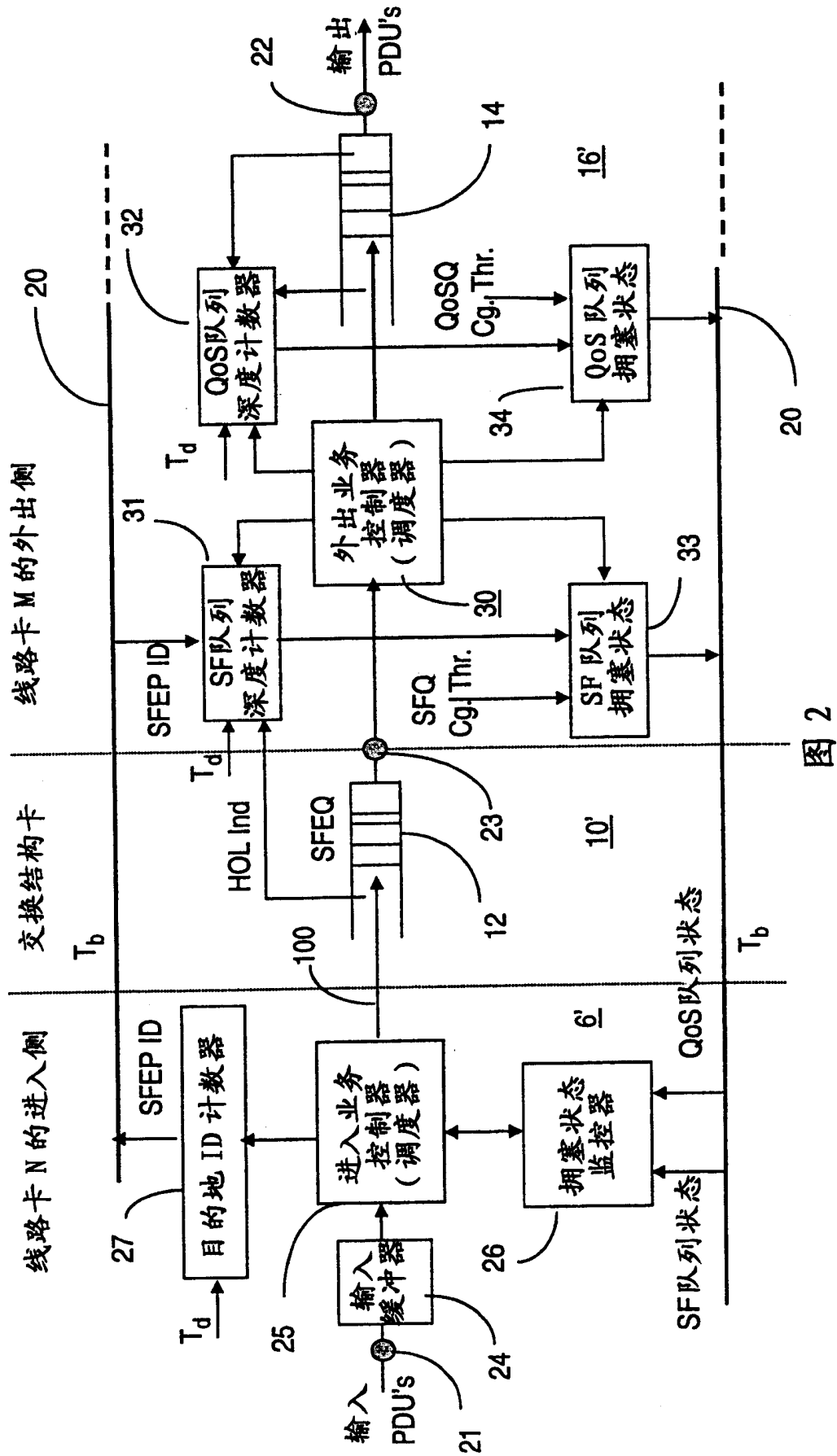


图 2

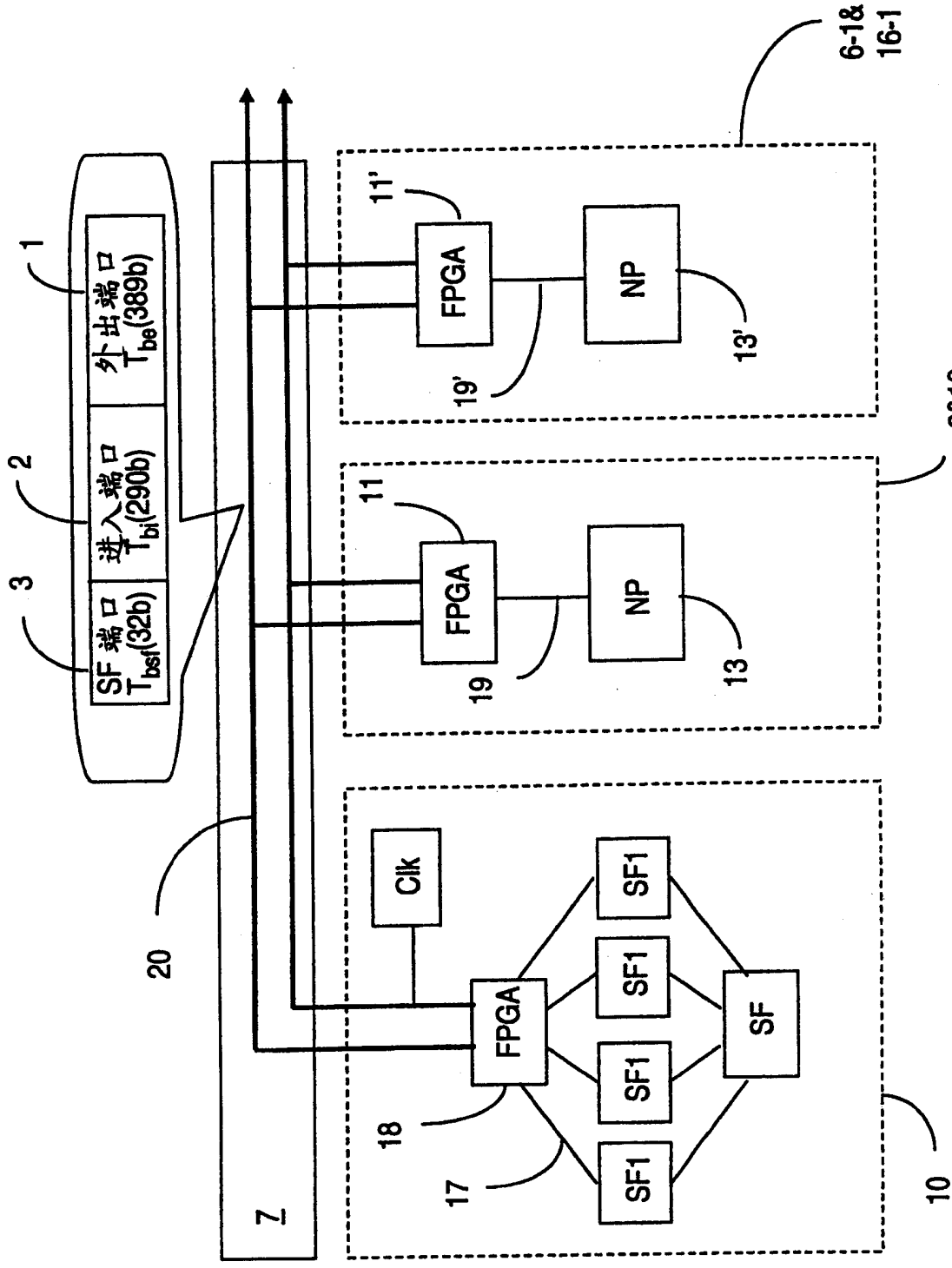


图 3