[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04L 12/56 (2006.01)

[12] 发明专利说明书

专利号 ZL 02819147.1

[45] 授权公告日 2008 年 4 月 23 日

[11] 授权公告号 CN 100384174C

[22] 申请日 2002.9.24 [21] 申请号 02819147.1

[30] 优先权

[32] 2001. 9.27 [33] US [31] 09/963,487

[86] 国际申请 PCT/CA2002/001451 2002.9.24

[87] 国际公布 WO2003/028288 英 2003.4.3

[85] 进入国家阶段日期 2004.3.29

[73] 专利权人 4198638 加拿大有限公司

地址 加拿大魁北克

[72] 发明人 史蒂夫·罗尚 理查德·S·诺曼 罗宾·布瓦万

[56] 参考文献

CA 2292828 A1 2001.6.22

WO 01/05107 A1 2001.1.18

US 6252848 B1 2001.6.26

审查员 赵 晨

[74] 专利代理机构 中原信达知识产权代理有限责任公司

代理人 顾红霞 钟 强

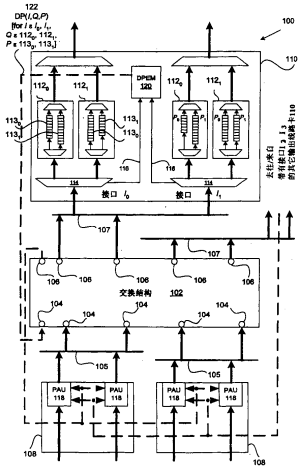
权利要求书 10 页 说明书 21 页 附图 2 页

[54] 发明名称

在包交换装置中用于避免拥堵的方法和系统

[57] 摘要

一种用于调节通过一种装置比如带有交换结构的路由器的数据包流的方法、装置和计算机可读的存储介质。 为设在装置的出口级的多个队列中的每一个队列收集拥堵信息，比如带宽利用率统计数字。 根据带宽利用率的统计数字，为每个队列计算一个“丢弃概率”。 这个信息被周期性地或者按照其它的一个控制时间周期，比如当丢弃概率发生巨大变化时，传输到入口级。 根据数据包要到达的队列以及该队列的丢弃概率，入口级将数据可控制地传输到交换结构或者不进行传输。 在这种方式中，甚至在拥堵有机会发生以前就已经避免了拥堵的出现。 这将提高带宽的利用率，因为在入口级被丢弃的数据包就不会占用装置中的其它资源。



1. 一种通过一个装置调节数据包流的方法，该装置具有带有至少一个输入端口和至少一个输出端口的交换结构，连接于至少一个输入端口的控制实体，用来调节流到所述输入端口的数据包，和连接于至少一个输出端口的多个出口队列，用来临时存储从输出端口接收的数据包，所述方法包括：

获得关于在出口队列接收的数据包的带宽利用率信息，其中，获得所述带宽利用率信息包括确定在每一个所述出口队列接收的数据包消耗的带宽量；

从带宽利用率相信和在每一个所述出口队列接收的数据包消耗的带宽量，确定与每个出口队列有关的丢弃概率

将与每个出口队列有关的丢弃概率提供到所述控制实体，用来通过所述控制实体选择性地传输数据包到交换结构的至少一个输入端口。

2. 如权利要求 1 所述的方法，还包括：

对于控制实体收到的每个数据包，在有关所述至少一个出口队列的带宽利用率信息的基础上将收到数据包传输或者不传输到入口实体。

3. 如权利要求 2 所述的方法，其特征在于，不将收到数据包传输到入口实体包括丢弃收到的数据包。

4. 如权利要求 2 所述的方法，其特征在于，不将收到数据包传输到入口实体包括在一个存储位置存储数据包。

5. 如权利要求 2 所述的方法，其特征在于，不将收到数据包传输到入口实体包括沿着一个备用线路重新传输数据包。

6. 如权利要求 1 所述的方法，还包括：

为控制实体收到的每个数据包产生一个数量值；

将数量值与丢弃概率进行比较；以及

在所述比较步骤结果的基础上，将收到的数据包传输或者不传输到入口实体。

7. 如权利要求 6 所述的方法，其特征在于，数量值是一个随机的数字。

8. 如权利要求 1 所述的方法，其特征在于，出口实体包括多个出口队列，并且，拥堵信息还包括每个出口队列的占用率。

9. 如权利要求 9 所述的方法，其特征在于，出口实体包括多个出口队列，并且，拥堵信息包括每个出口队列占用率的可变性。

10. 如权利要求 1 所述的方法，还包括

从有关所述出口队列中特定的一个的带宽利用率信息中确定用于所述特定出口队列的丢弃概率；

其中将有关所述至少一个出口队列的带宽利用率信息提供到控制实体包括将用于所述至少一个出口队列中的每一个的丢弃概率提供到控制实体。

11. 如权利要求 1 所述的方法，其特征在于，获得关于所述至少一个出口队列的带宽利用率信息包括从至少一个业务流管理实体接收所述带宽利用率信息，业务流管理实体设在所述至少一个出口队列和至少一个输出端口之间。

12. 如权利要求 10 所述的方法，其特征在于，每个数据包或者由多个业务流字节或者由非业务流字节组成，并且，获得关于所述至少一个出口队列的带宽利用率信息还包括为所述至少一个输出端口中的

每一个特定端口确定连接到该特定输出端口的每个出口队列每单位时间收到的业务流字节平均值。

13. 如权利要求 12 所述的方法，其中，从带宽利用率信息，为特定一个出口队列确定丢弃概率，包括：

为所述特定出口队列确定一个分配业务流带宽；

将所述特定出口队列的收到业务流字节平均值与该出口队列的分配业务流带宽进行比较；以及

如果所述特定出口队列的收到业务流字节平均值比该出口队列的分配业务流带宽大，增加该出口队列的丢弃概率；

如果所述特定出口队列的收到业务流字节平均值比该出口队列的分配业务流带宽小，降低该出口队列的丢弃概率。

14. 如权利要求 12 所述的方法，其特征在于，从有关所述特定出口队列的带宽利用率信息中为所述特定的出口队列确定丢弃概率包括：

为所述特定出口队列确定一个分配业务流带宽；

将所述特定出口队列的收到业务流字节平均值与该出口队列的分配业务流带宽进行比较；以及

如果所述特定出口队列的收到业务流字节平均值比该出口队列的分配业务流带宽大，就将该出口队列的丢弃概率设为该出口队列以前的丢弃概率的时间平均值与一个正的增加值之和；

如果所述特定出口队列的收到业务流字节平均值比该出口队列的分配业务流带宽小，就将该出口队列的丢弃概率设为该出口队列以前的丢弃概率的时间平均值与一个负的增加值之和。

15. 如权利要求 12 所述的方法，其特征在于，为所述特定的出口队列确定丢弃概率包括：

（a）将临时收到业务流字节平均值设定为收到业务流字节的平均数；

（b）将临时丢弃概率设定为所述特定出口队列以前丢弃概率值的时间平均值；

（c）为所述特定出口队列确定一个分配业务流带宽；

（d）将临时收到业务流字节平均值与所述特定出口队列的分配业务流带宽进行比较；

（e）如果临时收到业务流字节平均值比所述特定出口队列的分配业务流带宽大，就给临时丢弃概率加上一个正的概率增加值，并且给临时收到业务流字节平均值增加一个负的带宽增加值；

（f）如果临时收到业务流字节平均值比所述特定出口队列的分配业务流带宽小，就给临时丢弃概率加上一个负的概率增加值，并且给临时收到业务流字节平均值增加一个正的带宽增加值；

（g）将所述特定出口队列的丢弃概率设定为临时丢弃概率。

16. 如权利要求 15 所述的方法，还包括执行步骤（d），(e)和(f)一个预定的次数。

17. 如权利要求 15 所述的方法，还包括执行步骤（d），(e)和(f)，直到临时收到业务流字节平均值在所述特定出口队列分配业务流带宽所需要的范围内。

18. 如权利要求 17 所述的方法，还包括测量所述特定出口队列的深度并执行步骤（d），(e)和(f)，直到所述特定出口队列的深度在所需的范围内。

19. 如权利要求 18 所述的方法，还包括测量所述特定出口队列深度的可变性，并执行步骤（d），(e)和(f)，直到所述特定出口队列深度的可变性在所需要的范围内。

20. 如权利要求 15 所术的方法，还包括执行步骤（d），(e)和(f)，直到所述特定出口队列的临时丢弃概率收敛到一个需要的精确度。

21. 如权利要求 15 所述的方法，其特征在于，为所述特定出口队列确定一个分配业务流带宽包括：

如果所述特定出口队列的丢弃概率为零，则确定在出口队列将收到的业务流字节平均值；以及

如果所述特定出口队列的丢弃概率为零，而且如果出口队列将收到的业务流字节平均值大于该队列的分配业务流带宽，则为该出口队列的分配业务流带宽加上一个正的增加值；

如果所述特定出口队列的丢弃概率为零，而且如果出口队列将收到的业务流字节平均值小于该队列的分配业务流带宽，则为该出口队列的分配业务流带宽加上一个负的增加值。

22. 如权利要求 15 所述的方法，还包括：

为所有连接到所述特定输出端口的出口队列确定可用的业务流带宽；

为所有连接到所述特定输出端口的出口队列确定分配的总业务流带宽；

其中，只有当分配给连接到所述特定输出端口的所有出口队列的总业务流带宽小于连接到该输出端口的所有出口队列的可用业务流带宽时，才执行给所述特定出口队列分配业务流带宽增加一个正的增加值这个步骤。

23 如权利要求 22 所述的方法，其特征在于，为连接到输出端口的所有出口队列确定一个可用业务流带宽包括：

确定带宽梯度，该带宽梯度表明了与所述特定输出端口相连的所有出口队列的可用业务流带宽增加或者降低的速率；

按照与梯度带宽的函数关系增加或者降低与所述特定输出端口相连的所有出口队列的可用业务流带宽。

24. 如权利要求 23 所述的方法，其特征在于，获得与所述至少一

个出口队列相关的带宽利用率信息还包括：为所述至少一个输出端口中的每一个特定端口确定从所述特定输出端口单位时间收到的非业务流字节平均值，并且还在于，为连接到所述特定输出端口的所有出口队列确定一个可用业务流带宽还包括：

为连接到所述特定输出端口的所有出口队列确定一个总链路容量；

将最大可用业务流带宽设定为所述总链路容量与所述非业务流字节平均值之差；

其中，连接到所述特定输出端口的所有出口队列的可用业务流带宽不会超过最大可用业务流带宽。

25. 如权利要求 24 所述的方法，其特征在于，如果所述特定出口队列的丢弃概率为零，确定在该出口队列将收到的业务流字节平均值还包括：计算该出口队列每个时间单位收到业务流字节平均值和该出口队列以前丢弃概率值的时间平均值的一个函数。

26. 如权利要求 25 所述的方法，其特征在于，所述函数是以下两项的商：（i）所述特定出口队列每个时间单位收到的业务流字节平均值，（ii）所述特定出口队列以前丢弃概率值的时间平均值与整体时间的差。

27. 如权利要求 15 所述的方法，还包括：

如果所述特定出口队列的丢弃概率为零，确定在该出口队列将收到的业务流字节平均值；以及

执行步骤（d）,(e),(f)至少两次；

其中，如果出口队列为零，正的宽带增加量是所述特定出口队列将收到的业务流平均数的第一部分，所述第一部分随着步骤（f）的执行将减少；并且

其中，如果出口队列为零，负的宽带增加量是所述特定出口队列将收到的业务流平均数的第二部分，所述第二部分随着步骤（e）的执

行将减少。

28. 如权利要求 27 所述的方法，其特征在于，正的概率增加值具有随着步骤（e）的执行而减少的一个值，并且，负的概率增加值具有随着步骤（f）的执行而减少的一个值。

29. 如权利要求 23 所述的方法，其特征在于，获得与所述至少一个出口队列有关的带宽利用率信息包括为所述至少一个输出端口中的每个特定输出端口确定从该输出端口收到连续数据包之间的平均空闲时间。

30. 如权利要求 29 所述的方法，其特征在于，确定一个带宽梯度包括：

将从所述特定输出端口收到的连续信息包之间的平均空闲时间与一个第一阈值进行比较；以及

如果从所述特定输出端口收到的连续信息包之间的平均空闲时间低于第一阈值，设定带宽梯度来为连接到所述特定输出端口的所有出口队列指示可用业务流带宽中的一个第一降低速率。

31. 如权利要求 30 所述的方法，还包括：

将从所述特定输出端口收到的连续信息包之间的平均空闲时间与一个小于第一阈值的第二阈值进行比较；以及

如果从所述特定输出端口收到的连续信息包之间的平均空闲时间低于第二阈值，设定带宽梯度来为连接到所述特定输出端口的所有出口队列指示可用业务流带宽中的一个第二降低速率，其中，所述第二降低速率比所述第一降低速率大。

32. 如权利要求 31 所述的方法，还包括：

将从所述特定输出端口收到的连续信息包之间的平均空闲时间与一个第三阈值进行比较；以及

如果从所述特定输出端口收到的连续信息包之间的平均空闲时间高于第三阈值，设定带宽梯度来为连接到所述特定输出端口的所有出口队列指示可用业务流带宽中的一个增加速率。

33. 如权利要求 32 所述的方法，还包括：

确定多个出口队列中存储器利用率的程度；以及

编程将第一、第二和第三阈值中至少一个表示成所述存储器利用率程度的函数。

34. 如权利要求 10 所述的方法，其特征在于，所述交换结构的至少一个输出端口为多个输出端口，并且，多个输出端口中的每个端口都分别与多个出口队列中的一个队列相连接。

35. 如权利要求 10 所述的方法，其特征在于，所述交换结构的所述至少一个输出端口中的至少一个分别与多个出口队列中的一个队列相连接。

36. 如权利要求 10 所述的方法，其特征在于，将用于所述至少一个出口队列中每一个的丢弃概率提供给控制实体是在可编程的基础上来进行操作。

37. 如权利要求 10 所述的方法，还包括：

记录用于所述至少一个出口队列中每一个在选择时间的丢弃概率；

探测是否有一个预定幅度的变化在用于至少一个出口队列的丢弃概率中出现；

其中，只有探测到有至少一个预定幅度的变化在用于一个特定出口队列的丢弃概率中出现，才能执行将用于所述特定出口队列的丢弃概率提供给控制实体。

38. 如权利要求 10 所述的方法，还包括：

记录用于所述至少一个出口队列中每一个在选择时间的丢弃概率；

探测是否至少有一个预定幅度的变化在用于至少一个出口队列的丢弃概率中出现；

其中，在下面两种情况下执行将用于一个特定出口队列的丢弃概率提供给控制实体的步骤（i）探测到有至少一个预定幅度的变化在用于所述特定出口队列的丢弃概率中出现；或者（ii）经过预定的时间后，不管是否探测到有至少一个预定幅度的变化在用于所述特定出口队列的丢弃概率中出现。

39. 如权利要求 10 所述的方法，还包括：

对于每个收到数据包，控制实体确定收到数据包前往哪个出口队列，并在数据包要传输到用于那个出口队列的丢弃概率的基础上，确定将收到的数据包传输还是不传输到交换结构。

40. 如权利要求 39 所述的方法，其特征在于，在数据包要传输到用于那个出口队列的丢弃概率的基础上，将收到的数据包传输到交换结构还是不进行传输包括：

为收到的数据包产生一个随机数；

将随机数与收到数据包要传输到用于那个出口队列的丢弃概率就比较；以及

在比较的基础上将收到的数据包传输或者不传输到交换结构。

41. 如权利要求 40 所述的方法，其特征在于，不传输一个收到的数据包包括丢弃数据包。

42. 如权利要求 40 所述的方法，其特征在于，不传输一个受到的数据包包括将数据包标记为可丢弃数据包。

43. 如权利要求 40 所述的方法，其特征在于，不传输一个收到的数据包包括在一个存储位置存储收到的数据包并将数据包标记为可丢弃数据包，并且，传输一个数据包包括只传输没有被标记为可丢弃的数据包。

44. 如权利要求 43 所述的方法，其特征在于，不传输一个收到的数据包还包括：

确定在出口队列中是否存在拥堵减轻的情况；以及

如果在出口队列中存在拥堵减轻的情况，就确定是否存储位置要被用来存储其它数据包，如果不是，就不再标记数据包为可丢弃数据包。

45. 一个用来在一个装置中使用的丢弃概率计算模块，该装置设有（i）一个处理结构，设有至少一个输入端口和至少一个输出端口；（ii）一个控制实体，与所述至少一个输入端口相连，用来调节流入到该端口的数据包；以及（iii）多个出口队列，与所述至少一个输出端口相连，用来临时存储从出口收到的数据包，所述丢弃概率计算模块包括：

一个分配处理实体，用于为每个出口队列确定一个分配的业务流带宽；以及

一个与分配处理实体相连的概率处理实体，所述概率处理实体适合于从分配处理实体中为每个出口队列接收分配业务流带宽，还适合于从一个外部实体中为每个出口队列接收业务流字节的平均值，概率处理实体可用于：

将每个出口队列的收到业务流字节平均值与所述特定出口队列的分配业务流带宽进行比较；

根据所述特定出口队列收到业务流字节的平均值是比该出口队列的分配业务流带宽更大还是更小，将出口队列的丢弃概率设定为该出口队列以前丢弃概率值的时间平均值与一个正的或者负的增加值的和。

在包交换装置中用于避免拥堵的方法和系统

发明领域

本发明涉及在包交换装置中避免拥堵的方法和系统，尤其是使用数据包丢弃技术来避免拥堵的方法和系统。

发明背景

通过一个包交换装置（比如，一个路由器）来支持的所有链路的集成链路带宽通常比装置的总交换能力高。这就引起位于包交换装置入口（交换前）和出口（交换后）级的缓冲器中出现拥堵。拥堵甚至会在无阻塞交换结构（switch fabrics）上出现。通常可以使用缓冲来控制拥堵。然而，缓冲会造成数据包传输的延迟，因此，为了避免过度延迟，就需要开发更复杂的技术来解决拥堵。

一种用于控制拥堵的方案是尾部丢弃法，藉此，允许将数据包填充到出口缓冲器中，如果数据包来自交换结构而同时缓冲器容量已满，就将数据包丢弃。然而，这种方法会造成多路数据流丢失数据包。高层协议会通过终止数据流并重新传输丢失信息的方法来对这种类型的数据包丢失做出反应。虽然拥堵本身被终止了，而最终的结果就是包交换装置那种非常令人讨厌的减速—加速—减速等等的工作方式。

在另一种传统的拥堵控制算法中，出口级对每一个数据包采取一个行动，这种行动可以是使数据包排队等待，也可以是将数据包丢弃。这种类型算法的一个例子是随机初期丢弃（RED）算法。RED（随机初期丢弃）算法的具体例子包括 RED（随机初期丢弃）算法（该算法在 1998 年 4 月的 Internet Request For Comments(RFC)2309 中进行了描述，特此引用），以及 BLUE 算法（该算法在 BLUE：A New Class of Active Queue Management Algorithms 的第 1-26 页进行了描述，Wu-

chang Feng 等著，特此引用）。数据包应该丢弃还是排队等待的决定通过监控特定链路上出口缓冲器已充满的程度，并相应地给数据包产生一个丢弃概率来进行确定。如果给数据包产生的随机数字低于丢弃概率，数据包就被丢弃；否则将数据包放入出口缓冲器中。在这种方法中，出口缓冲器的拥堵可以通过出口缓冲器采取的行动来进行控制。

但是，由于延迟不丢弃的数据包增加的不稳定信号和等待时间以及发送要丢弃的数据包都需要大量地提供交换结构。因此，通过装置出口丢弃数据包或者使数据包排队等候的每一个行动（比如，一旦交换资源已经被用来交换数据包），这些最终丢弃的数据包将消耗掉包交换装置整个入口和交换级的资源。明显地，以测量拥堵水平为基础来做出决定，被使用装置的存储器和/或交换资源利用程度不可避免地效率低下。

因此，在工业上有必要开发一种能够限制拥堵，同时又能够在像路由器这样的包交换装置中有更大的资源利用效率的机制。

发明简述

本发明提供了一种用于在包交换装置的入口级调节数据包流量的方法。特别地，将为装置出口级的多个队列中的每一个获得带宽利用率信息。根据带宽利用率信息，为每个队列计算“丢弃概率”。这个信息周期性地或者在其它控制的时间段传输到入口级，比如当丢弃概率出现极大变化时。然后入口级就能够根据数据包要到达的队列以及那个队列的丢弃概率，继续进行数据包到交换结构的可控传输或者停止进行传输。在这样的方式中，甚至在拥堵有机会发生以前就被避免了。这就提高了带宽的利用率，因为在入口级丢弃的数据包不会占据装置中的其它资源。

因此，本发明可以概括为利用一种调节通过一个装置的数据包流

的方法，该装置设有包括多个输入端口和多个输出端口的交换结构；一个控制实体，该控制实体连接到输入端口用于调节到入口的数据包流；以及多个出口队列，该出口队列连接到输出端口用于临时存储从出口收到的数据包。该方法包括获得有关出口队列收到的数据包的带宽利用率信息；从带宽利用率信息中确定各出口队列的丢弃概率；向控制实体提供各出口队列的丢弃概率，供控制实体用来有选择性的将数据包传输到交换结构的输入端口。

在一个特定的实施例中，获得有关出口队列收到数据包的带宽利用率信息包括为每一个输出端口确定从该输出端口收到连续数据包之间的平均空闲时间；为每一个输出端口确定连接到该输出端口的每个出口队列单位时间收到业务流字节数的平均值，以及为每个输出端口确定单位时间从该输出端口收到非业务流字节数的平均值。

在一个特定实施例中，通过为某一出口队列确定分配业务流带宽，以及将该出口队列收到的业务流字节数平均值与该出口队列分配的业务流带宽进行比较，就可以确定一个出口队列的丢弃概率。如果该出口队列收到的业务流字节数平均值比该出口队列分配到的业务流带宽更大（更小），则该出口队列的丢弃概率被设定为该出口队列以前丢弃概率值的时间平均值与一个正（负）增加值之和。

在一个特定的实施例中，可以为每个出口队列和数据包优先级的组合的丢弃概率进行计算。

在一个特定的实施例中，本发明的方法可以实现为计算机可读储存介质上的指令序列。

根据第二个主要的方面，本方法可以概括为丢弃概率计算模块，该模块包括一个分配处理实体，用于为每个出口队列和一个与分配处理实体相连的概率处理实体确定一个分配的业务流带宽，概率处理实

体适合于从分配处理实体为每个出口队列接收分配的业务流带宽，还适合于从外部装置中为每个出口队列接收已收到业务流字节的平均值。

概率处理实体可以将每个出口队列收到业务流字节的平均数与该出口队列分配的业务流宽带进行比较，并根据该出口队列收到的业务流字节平均值比该出口队列分配业务流带宽是大还是小，将某一出口队列的丢弃概率设定为该出口队列以前丢弃概率值的时间平均值与或正或负的增加值之和。

根据第三个主要的方面，本方法可以概括为设有交换结构的一种装置，该交换结构设有多个输入端口和多个输出端口，该交换结构适合于在该装置的输入端口和输出端口之间交换数据包。该装置还包括多个出口队列，出口队列与交换结构输出端口中相应的一个相连，每个出口队列都适合于（i）临时存储从相应交换结构出口收到的数据包，（ii）在出口队列收到数据包的基础上确定带宽使用信息。

所述装置还包括一个与出口队列相连的丢弃概率计算模块，该丢弃概率计算模块适合于以带宽使用率信息为基础确定每个出口队列的丢弃概率。所述装置还包括一个与交换结构输入端口和丢弃概率计算模块相连的数据包接收部件，数据包接受部件适合于（i）接收传输到交换结构输出端口的数据包；（ii）确认与每个收到的数据包相关的出口队列；（iii）根据与每个收到数据包相关的出口队列的丢弃概率，将收到数据包传输或者不传输到交换结构的一个输入端口。

根据另一主要方面，本发明可以概括为一种调节通过一个装置的数据包流量的方法，该装置设有一个入口实体；一个出口实体；一个处理结构，该处理结构设在入口实体和出口实体之间；以及一个控制实体，该控制装置用于在将数据包传输到入口实体之前对数据包进行处理。该方法包括获得与出口实体的收到数据包相关的拥堵信息，并

将拥堵信息提供给控制实体，供控制实体用来在数据包传输到入口实体之前对数据包进行处理。

通过对本发明优选实施例的说明和附图的审查，本领域一般技术人员对本发明的这些和其它方面以及特点将更加清楚明了。

附图简述

其中：

图 1 是设有本发明实施例的随机数据包丢弃功能的一个包交换装置的框图；

图 2 是图 1 装置中丢弃概率计算模块实施例的框图。

优选实施例详述

参考图 1，图中显示了本发明实施例的包交换装置，该装置以入口级获得的拥堵信息为基础，通过在入口级传输或者不传输数据包来避免信息拥堵的发生。

在一个实施例中，包交换装置 100 是一种多级路由器，而且出口级可以是多级路由器的中间级或者最后级。能在包交换网络中与邻近的装置（比如，路由器或者节点）相连的装置 100 包括一个交换结构 102，交换结构 102 设有多个输入端口 104 和多个输出端口 106。交换结构 102 的输入端口 104 与一组输入线路卡 108 相连，交换结构 102 的输出端口 106 与一组输出线路卡 110 相连。在线路卡为双向线路的实施例中，输入线路卡 108 和输出线路卡 110 可以没有差别。同样，在多级路由器中，输出线路卡 110 将作为路由选择的下一级。

输入线路卡 108 适合于从外部实体中接收数据包流，并将它们提供给交换结构 102 的输入端口 104。每个输入线路卡 108 可以通过一个总线结构 105 与交换结构 102 的一个、两个或者多个输入端口 104 相连。输出线路卡 110 适合于从交换结构 102 的输出接口 106 接收数

据包流并将它们转发到网络下游临近的路由器或者节点上。每个输出线路卡 110 都设有一个、两个或者多个物理接口，与交换器 102 的输出端口 106 中的每一个相对应。每个输出线路卡 110 上的接口通过输出线路卡 110 能共同使用的总线结构 107 与交换结构 102 上相应的输出端口 106 相连。在图示实施例中，有四个标记为 I_0、I_1、I_2、I_3 的接口，两个接口位于两个输出线路卡 110 中的每一个上面。然而，显而易见还能够采用其他布置方式，本发明不受任何接口、输出线路卡 110 数量或者接口在输出线路卡 110 上的布置方式的限制。

典型地，数据包由首部和有效负载构成，而且还与优先级关联（或者优先权或服务级别）。首部通常确定数据包的优先级（如果可用）以及数据包发送的目的地节点。收到一个特定数据包的输入线路卡 108 将它的目的地节点转换成交换结构 102 的目的地输出端口，比如 106x，并且将这个目的地输出端口 106x 的标识插入到数据包的首部。利用在数据包的首部所指定的目的地输出端口 106x 的标识，数据包就通过适当的总线 105 被提供给交换结构 102 的一个输入端口 104。交换结构 102 负责确保数据包确实出现在要求的目的地输出端口 106x。具有一个（通过输出线路卡的总线 107）连接到目的地输出端口 106x 的接口的输出线路卡 110 从数据包的首部将目的地输出端口 106x 的标识去掉，以数据包最初的形式将其转发到网络上临近的路由器或者节点上。在这种方式中，数据包被移动到更靠近它的目的地节点。在多级路由器的情况下，输出线路卡就代表了路由选择的下一级。

在一个实施例中，同那种其输出线路卡 110 能够传输到装置 100 外面的交换结构相比，交换结构 102 能够以更高的速率向每个输出线路卡 110 提供信息。这是为了允许输出线路卡 110 来缓冲超出的带宽，并就如何路由缓冲数据包做出明智的决定。在另外的实施例中，严格的讲，如果装置 100 中数据包发送的速率与能够通过一个与输出线路卡 110 相连的外部装置进行控制的速率匹配或者比这个速率小，出口实体就没有必要是队列。

不管在装置 100 出口是否需要缓冲能力，在出口线路卡 110 上的接口都包含一个出口实体。在一个实施例中，每个输出线路卡 110 上的每个接口都与一个、两个或者多个构成出口实体部件的出口队列 112 相连。出口队列被用来临时存储通过适当总线 107 从交换结构 102 的相应输出端口 106 发送来的数据包流中的数据包。出口队列 112 可以说是作为与它们相连的物理接口充当虚拟接口或者虚拟端口。应当提及的是在多级路由器中，相对于路由选择的后面级，出口队列 112 实际上是入口队列。

在图示实施例中，每个接口上有两个出口队列 112，分别标记为 Q_0，Q_1。因此，接口 I_0 与它自己的出口队列 Q_0，Q_1，而接口 I_1 与它自己的出口队列 Q_0，Q_1 等等相连。但是，可以理解的是还可以有其它布置方式，而且本发明并不局限于每个接口上出口队列 112 的任何特定数量。此外，在数据包有不同优先级的情况下（比如，P_0 和 P_1），出口队列 112 可以在优先级的基础上划分为相应的一组子队列 113。应当认识到的是子队列 113 并不要求全体都是一样的深度，对于每个出口队列 112 也不需要具有相同的子队列 113 数目。

根据本发明的一个实施例，还为每个接口设有一个出口业务流管理器（ETM）114。每个出口业务流管理器 114 包括适当的电路、软件和/或者控制逻辑用于从每个收到的首部去掉目的地输出标识以及用于确定将收到的数据包放到哪个接口、哪个出口队列 112 和子队列 113。显然可以理解的是在本发明的其它实施例中，每个输出线路卡 110 设有一个输出业务流管理器，这样，每个出口业务流管理器 114 将直接连接到与总线 107 相对应的一个通路中。

向哪个接口发送一个收到的数据包，要在收到数据包首部中指定的信息（比如目的地节点）的基础上做出决定。向哪个出口队列 112 插入收到的数据包，要在收到数据包首部中指定的信息（比如虚拟端

口标识符）的基础上做出决定。向哪个子队列 113 插入收到的数据包，要在收到数据包首部中指定的信息（比如优先级）的基础上做出决定。

为了说明的目的，假定每个接口实际设有一个出口业务流管理器 114，每个这样的出口业务流管理器 114 都额外设有电路、软件和/或控制逻辑，用于监控从交换结构 102 的相应输出端口 106 收到的数据包目的地和数量。在这个信息的基础上，出口业务流管理器 114 产生表明拥堵的信息。

拥堵信息可能包括带宽利用率信息，比如：

接口 I_i 上收到数据包之间的平均空闲时间(记做 AIT（I_i）- 平均空闲时间)；

接口 I_i 收到非业务流字节平均值(记做 ARNB（I_i）- 收到非业务流字节平均值)；

与接口 I_i 关联的每个出口队列 $Q_\varepsilon\{112_0, 112_1\}$的每个优先级 $P_\varepsilon\{113_0, 131_1\}$收到业务流字节平均值(记做 ARTB（$I_i$，Q，P）- 平均收到业务流字节)。

可选地，拥堵信息还可以包括对每个出口队列 112 深度的测量值或者每个队列可变性的测量值。为了说明的目的，假定拥堵信息是带宽利用率信息，该信息由位于一个给定出口线路卡 110 上的每个出口业务流管理器 114 产生，这样的信息被提供给公共丢弃概率计算模块（DPEM）120，该公共丢弃概率计算模块有输出线路卡 110 共用。（另外一个方法是，为输出线路卡 110 上的每个出口业务流管理器 114 设有一个单独的 DPEM 120）。

位于给定输出线路卡 110 上的 DPEM 120 包括电路、软件和/或控制逻辑，用于为每个与位于给定出口线路卡 110 上的每个接口关联的出口队列 112 和子队列 113 计算丢弃概率。因此，根据设有 DPEM 120 的出口线路卡上的接口数量，每个 DPEM 120 将为一个、两个或者多

个接口计算丢弃概率。为了便于标记，接口 I_i，队列 112_q 和子队列 113_P 的丢弃概率将被记做 DP（I_i, 112_q, 113_p）。

每个 DPEM 120 通过一个控制链路 122 与每个输入线路卡 108 中的一个或者多个数据包接收单元(PAUs)118 相连。从一个给定 DPEM 120 连接到输入线路卡 108 的控制链路携带有每个队列和优先级的组合的丢弃概率，所述每个组合是对于与包含 DPEM 120 的输出线路卡 110 关联的每个接口可能的组合。由于这个工作由位于每个输出线路卡 110 中的每个 DPEM 120 完成，所以位于每个输入线路卡 108 中的每个 PAU 118 就可以为每个可能的接口、队列和子队列组合获取丢弃概率。由特定 DPEM 120 传输的丢弃概率就可以以广播信息的形式进行发送。交换结构 102 实际上被用做一个通道，用来将丢弃概率 DP（I，Q，P）从 DPEM 120 运送到 PAUs 118 的。

现在考虑位于一个特定输入线路卡 108 上的 PAU 118，这个装置被用来在数据包传输到交换结构 102 的相应输入端口 104 之前处理数据包流。PAU 118 的功能之一就是以每个数据包的接口和输入队列（如果适用，还有优先级）为基础，以及与接口、输出队列（如果适用，还有优先级）的组合相联系的丢弃概率为基础，通过使用数据包的随机丢弃功能来避免信息拥堵。虽然用来说明的实施例中交换结构 102 的每个输入端口 104 显示了一个 PAU 118，但是在某些实施例中，给每个输入线路卡 108 都设一个 PAU 118，或者设置唯一的一个布置在输入线路卡 108 中的 PAU 118，这样是非常有利的。

用来处理一个特定数据包流的 PAU 118 设有合适的电路、软件和/或控制逻辑，用于确定每个收到数据包目的地输出端口。这个目的地输出端口与输出线路卡 110 之一上的一个接口（比如，I_i）相对应。此外，PAU 118 还包括合适的电路、软件和/或控制逻辑，用于确定输出业务流管理器 114 要将收到的数据包插进去的输出队列 112（比如，112_q），所述输出业务流管理器 114 关联于与目的地输出端口相对应

的接口。此外，如果一个收到的数据包具有几个优先级中的一个，那么，PAU 118 还包括合适的电路、软件和/或控制逻辑，用于确定数据包的优先级以及要确定数据包要插入到的子队列（比如，113_p）。基于这种信息并且基于丢弃概率 DP（I_i, 112_q,113_p）， PAU 118 做出是要丢弃收到的数据包还是继续向交换结构 102 传输的决定。

为了做出是否丢弃收到的具有接口 I_i、输出队列 112q 和子队列 113p 特征的数据包的决定，PAU 118 包括合适的电路、软件和/或控制逻辑，用于为收到的数据包产生一个随机如数字 R 并将 R 与丢弃概率 DP（I_i, 112_q,113_p）进行比较。比如说，如果 R 低于 DP（I_i, 112_q,113_p），那么数据包就被丢弃，否则数据包就被发送到交换结构 102 的相应输入端口 104。另外一种方法是，如果随机数字 R 比丢弃概率 DP（I_i, 112_q,113_p）高，就将数据包丢弃。术语“随机数字”在这种意义上就是指包括一个由伪随机或者拟随机技术产生的数字。

在它决定确实要将收到数据包转发到交换结构 102 的情况下，PAU 118 包括合适的电路、软件和/或控制逻辑，用于将目的地输出端口的标识插入到数据包首部，并将数据包转发到交换结构 102 的相应输入端口 104，这时数据包的首部规定了目的地输出端口的标识。但是，在它决定要丢弃收到数据包的情况下，数据包不会传输到交换结构 102，而将完全地被丢弃，或者发送到一个用于存放丢弃数据包的单独的存储库。非常具有优势的是，不传输的数据包不会消耗交换结构 102 中的资源或者 PAU 118 拥堵管理装备中的资源，提高了资源的利用率。

现在来描述通过与接口 I_0 关联的输出业务流管理器 114 来产生带宽利用值（比如，AIT（I_0），ARBN（I_0）和 ARBT（I_0，Q，P））。第一，关于 AIT（I_0）值，它指示接口 I_0 的总带宽利用率。如果使用了长度固定的数据包，那么总带宽利用率就可以通过计算每秒钟到达出口业务流管理器 114 的数据包数量来直接进行测量。在这种情况下，

就不需要计算平均空闲时间。但是，如果使用了可变长度数据包，优选地，总宽带利用率通过计算连续收到数据包时非传输时间间隔的平均值来进行间接测量。这个时间是指数据包与数据包之间的平均空闲时间。

如果一个字中的专用比特被用来显示那个字是否在它所在的数据包中距离最后一个字有一定数目的字，那么就便于实现测量数据包之间的平均空闲时间的方法。这样一种用于用信号指示数据包即将结束的技术在授予给 Norman 等人、申请号为 09/870,776 的美国专利申请中进行了描述，该专利申请日期为 2001 年 5 月 31 日，转让给本发明的受让人，特此引用。

与接口 I_0 关联的出口业务流管理器 114 也为 Q_ε $\{112_0, 112_1\}$，P_ε $\{113_0, 113_1\}$产生 ARTB（I_0，Q，P）值，该值是为接口 I_0 表示发送到每个出口队列和子队列组合的业务流字节数平均值。一个业务流字节从属于一个必须满足特定的面向用户的质量服务标准的数据包。换言之，业务流字节属于必须避免信息拥堵的数据包。为了计算每个 ARTB（I_0，Q，P）值，出口业务流管理器 114 包括合适的电路、软件和/或控制逻辑，用于分析每个到来数据包的首部，并且从首部的信息中确定让数据包前往哪个出口队列 112，以及数据包优先级。

另外，与接口 I_0 关联的出口业务流管理器 114 也产生 ARNB（I_0）值，该值表示了在接口 I_0 收到的非业务流字节数的平均值。一个非业务流字节属于没有附带面向用户服务质量标准的数据包中的字节。为了计算 ARNB（I_0）值，出口业务流管理器 114 包括合适的电路、软件和/或控制逻辑，用于分析每个到来数据包的首部，并且从首部的信息中确定数据包是业务流数据包还是非业务流数据包。显而易见的是，对于每个数据包，只在计算 ARNB（I_0）值和 ARTB（I_0，Q，P）值的过程中对数据包部首进行一次分析。

参考图 2，现在详细描述适合为每个有效的 I,Q,P 组合计算丢弃概率 DP（I_i, 112_q,113_p）的丢弃概率计算模块（DPEM）120 的例子。特别地，图示的与一个出口线路卡 110 关联的 DPEM 120 包括：一个总计（aggregation）处理实体 208，一个利用率处理实体 210，一个分配处理实体 220，一个概率处理实体 230，以及一个外推处理实体 240。

总计处理实体 208 从与接口 I_0 关联的出口业务流管理器 114 接收 ARNB（I_0）和 AIT（I_0）值，从与接口 I_1 关联的出口业务流管理器 114 接收 ARNB（I_1）和 AIT（I_1）值。总计处理实体 208 根据其输入确定收到的非业务流字节（标记为 ARNB）的总计平均值，以及带宽梯度（标记为 BWGR）。ARNB 和 BWGR 值被提供给利用率处理实体 210。ARNB 值的计算可以通过加上由于 i=0 和 i=1 而增加的 ARNB（I_i）值来得到。BWGR 值的计算可以通过以下方式完成。

为了得到一个总平均空闲时间，对每个接口测量的平均空闲时间 AIT（I_0），AIT（I_1）求平均值。然后将总平均空闲时间与一组预设阈值进行比较。在一个实施例中，每个接口的总平均空闲时间首先与临界最小平均阈值时间（记作 T_1）进行比较。如果它小于 T_1，那么这种情况就表明交换结构 102 中带宽的超临界利用。因此，设定带宽梯度值（BWGR）来表明在入口侧迫切需要缩小带宽。

然而，如果总平均空闲时间不小于 T_1，那么就让它与预设最小平均阈值（标记为 T_2）进行比较。如果总平均空闲时间比 T_2 小，那么这种情况表明了非临界性拥堵，非临界性拥堵消耗了交换结构 102 的缓冲空间。因此，设定带宽梯度值（BWGR）表明在入口侧需要适当的缩小带宽。

如果总平均空闲时间不比 T_2 小，那么就将它与预设最大平均阈值（标记为 T_3）进行比较。如果总平均空闲时间比 T_3 大，那么这种情况就表明交换结构 102 中带宽利用率不足。因此，设定带宽（BWGR）

来表明在入口侧需要增加带宽。

最后，如果总平均空闲时间介于 T_2 和 T_3 之间，那么这种情况表明交换结构 102 中的带宽的使用不需要进行补偿。因此，设定带宽梯度值（BWGR）表明在入口侧既不需要增加带宽，也不需要减少带宽。

应当指出的是，阈值 T1、T2 和 T3 可以在各种参数如带宽利用率，还有队列深度和带宽的可变性（猝发性）的基础上进行动态调节。

利用率处理实体 220 从总计处理实体 208 中收到 BWGR 值（比如，带宽梯度）和 ARNB 值（比如，平均收到非业务流字节）。以其输入为基础，利用率处理实体 210 为业务流数据包确定一个总利用带宽，然后把总利用带宽以 BWAVAIL 值（可利用带宽）的形式提供给分配处理实体 220。BWAV 值的计算按照以下步骤进行：

利用率处理实体对所有数据包（业务流数据包和非业务流数据包）可以利用的总带宽保留内部记录，记做 AGG_AVAIL。如果 BWGR 值表明在入口侧的带宽需要增加，于是 AGG_AVAIL 通过预设步长值来增加入口侧带宽，直到最大可利用总带宽；如果 BGWR 值表明入口侧的带宽需要缩小，那么 AGG_AVAIL 就通过预设步长值缩小入口侧带宽，降到一个最小可利用总带宽；如果 BGWR 值表明出口侧的带宽既不需要增加也不需要缩小，那么 AGG_AVAIL 保持不变；如果 BWGR 值表明在入口侧迫切需要缩小带宽，那么 AGG_AVAIL 设定到一个预设（低）值。

其次，从 AGG_AVAILA 的结果中减去 ARNB 值就得到 BWAVAIL 值，BWAVAIL 值是只有业务流数据包可以利用的总带宽。在本发明的一个实施例中，增加和减少操作中的步长值都可以设为最小可利用总带宽的百分率。因为缩小带宽比增加带宽更重要，所以增加操作的步长值要低于缩小操作中的步长值。

分配处理实体220除了从利用率处理实体210中以BWAVAIL值的形式收到业务流数据包可利用总带宽以外，还收到平均字节数的显示值，如果丢弃概率为零，就应当为输出线路卡上的每个输出队列112接收平均字节数显示值。这个信息从外推处理实体240中以多个ARTBDP0（I，Q）值（比如，如果丢弃概率为0的平均收到业务流字节）的形式进行接收，其中Iε $\{I_0, I_1\}$和Qε$\{112_0, 112_1\}$。随后，在外推处理实体240内容中将对每个ARTBDP0（I，Q）值的计算进行详细描述。

以其输入为基础，分配处理实体220为每个有效的I,Q组合的业务流字节分配带宽。这个计算结果以给I和Q的组合分配带宽值的形式（记作BWALLOC（I，Q）—分配的带宽）提供给概率处理实体230。

BWALLOC（I，Q）值的计算按照以下方式进行：分配处理实体220首先确定为每个I和Q的组合分配的带宽是否满足要求。这可以通过将以前的BWALLOC（I，Q）值与相应的ARTBDP0（I，Q）值进行比较来完成。因此，将分配的带宽与I和Q组合能够收到的最大可能带宽进行比较。

如果BWALLOC（I，Q）值超过ARTBDP0（I，Q）值，那么就要缩小BWALLOC（I，Q）值，比如，缩小一个固定的量，或者以一个依赖于BWALLOC（I，Q）值与ARTBDP0（I，Q）值之差的因子来缩小BWALLOC（I，Q）值。另一方面，如果BWALLOC（I，Q）值小于ARTBDP0（I，Q）值，那么BWALLOC（I，Q）的值就要增加，比如，增加一个固定的量，或者以一个依赖于BWALLOC（I，Q）值与ARTBDP0（I，Q）值之差的因子来增加BWALLOC（I，Q）值。BWALLOC（I，Q）值的增加或者缩小值被提供给概率处理实体230。

应当指出的是，另外一个实施例中，“不改变”的结果将应用到

特定的 BWALLOC（I，Q）值中，该实施例同样在本发明的范围内。还应当指出的是，当从概率处理实体 210 上收到数据时，为了确保在同一线路卡上，检查 BWALLOC（I，Q）在整个 I 和 Q 范围内的总和不会超过该线路卡的 BWAVAIL 值是非常有利的。

概率处理实体 230，除了从分配处理实体 220 接收（$I_\varepsilon\{I_0, I_1\}$和 $Q_\varepsilon\{112_0, 112_1\}$）的 BWALLOC（I，Q）值，还从外推处理实体 240 接收（$I_\varepsilon\{I_0, I_1\}$和 $Q_\varepsilon\{112_0, 112_1\}$以及 $P_\varepsilon\{113_0, 113_1\}$）的 ARTBDP0（I，Q，P）值，从与接口 I_0 关联的出口业务流控制器 114 接收（$Q_\varepsilon\{112_0, 112_1\}$以及 $P_\varepsilon\{113_0, 113_1\}$的）ARTB（$I_0$，Q，P）值，从与接口 I_1 关联的出口业务流管理器 114 接收（$Q_\varepsilon\{112_0, 112_1\}$以及 $P_\varepsilon\{113_0, 113_1\}$的）ARTB（$I_1$，Q，P）值。

以其输入为基础，概率处理实体 230 为每个有效的 I,Q,P 组合产生丢弃概率 DP（I，Q）P），在这种情况下为 $I_\varepsilon\{I_0, I_1\}$和 $Q_\varepsilon\{112_0, 112_1\}$以及 $P_\varepsilon\{113_0, 113_1\}$产生。为给定 I 值（比如 i）和 Q 值(比如 q)对所有的 P 值计算丢弃概率 DP（I，Q，P）可以按照以下方式进行：

首先，对所有的 P 值计算 ARTB（I,q,p）的总和。这就得到一个数量，该数量表示了与接口 I_i 关联的出口队列 112q 收到业务流字节的总平均数量，可以记做 TARTB（i,q）。为了确定是否分配了比需要的更大的带宽，将这个数量与 BWALLOC（i,q）进行比较。因为当分配的带宽与实际使用的带宽相匹配的时候，就会产生最优的资源利用效率，两个参数的差提供了一个误差的信号，要使这个差别变为零。在这个阶段，可以采取一种简单的方法和一种更加复杂的方法。下面先描述简单的方法，然后描述更加复杂的方法。

在分配带宽比总平均使用带宽大的情况下，在简单方法中，丢弃概率 DP（i,q,P）将对一个或者多个 P 降低（取决于使用了 intserv 还是 diffserv 模型），使得在入口处丢弃的数据包更少，最终造成了

TARTB（i,q）的增加。相反地，如果平均总带宽小于实际使用带宽，丢弃概率 DP(i,q,P)将对一个或者多个 P 增加(取决于使用了 intserv 还是 diffserv 模型），使得大量数据包在入口处被丢弃，最终造成了 TARTB（i,q）的减少。用于增加丢弃概率 DP（i,q,P）的幅度将不同于缩小的幅度。

重复以上过程，直到分配的带宽处于总平均使用带宽的一定范围内。有利的是，这种方法避免了一定程度的信息拥堵。但是，收敛将花费相对较长的时间。这是因为在改变丢弃概率和业务流字节数的平均值发生相应改变之间的大量时间都浪费了。此外，如果丢弃概率在业务流字节数的平均值被检测到之前就改变了，那么就可能“超过”最终丢弃概率，而最终丢弃概率将允许分配的带宽位于总平均使用带宽的一定范围之内。为了减少收敛的时间，可以求助于一种更复杂的方法。

在更复杂的方法中，每个 P 的丢弃概率的净增加/减少量是一个迭代过程的结果，它取决于（i）丢弃概率（记作 ADP（I,q,P）并被提供给外推处理实体 240）的时间平均值；（ii）一个临时丢弃概率（记作 $DP_{temp}(i,q,P)$;(iii)收到业务流临时平均字节数（记作 $ARTB_{temp}(i,q,P)$）。

在初始化时，临时丢弃概率 $DP_{temp}(i,q,P)$设定为以前的 DP (i,q,P)值，收到业务流临时平均字节数 $ARTB_{temp}(i,q,P)$ 设定为以前的收到业务流字节 ARTB (i,q,P) 的平均数。与前面一样，这个迭代过程的起点是通过比较分配带宽 BWALLOC(i,q)和使用的平均总带宽 TARTB(i,q)来确定是否需要增加或者降低丢弃概率。取决于是否需要增加或者减少丢弃概率，来对一个或者多个 P 的临时丢弃概率值做出相应的改变。

在这一点上，收到业务流字节临时平均字节数 $ARTB_{temp}(i,q,P)$被改变了，但是在相反的方向上。因此，临时丢弃概率的增加与收到业

务流平均临时平均字节数相对应，而临时丢弃概率的降低与收到业务流平均临时平均字节数的增加相对应。利用每个收到业务流平均临时字节数的新值，将对所有 P 的 $ARTB_{temp}(i,q,P)$值全部相加，就计算出了使用的临时平均总带宽 $TARTB_{temp}(i,q)$。将 $TARTB_{temp}(i,q)$的值与 BWALLOC(i,q)值进行比较，比较的结果就表明了分配带宽是否比期望的总平均使用带宽大。

改变一个或者多个 P 的临时丢弃概率 $DP_{temp}(i,q,P)$的步骤、重新计算 $ARTB_{temp}(i,q, P)$对整个 P 的值的步骤以及计算 $TARTB_{temp}(i,q)$值的步骤都可以多次重复。在一个实施例中，一直重复这个步骤直到 $TARTB_{temp}(i,q)$值处于所需要的 BWALLOC(i,q)值范围内。一个可选方案是，将这个步骤重复固定的次数或者直到临时丢弃概率实现收敛。在经过需要的重复次数以后，在任何情况下，每个丢弃概率 DP(i,q,P)都设定为相应临时丢弃概率 $DP_{temp}(i,q,P)$的当前值，并将丢弃概率提供给输入线路卡中的 PAUs。在这种方法中，就可以预测由于丢弃概率的变化而造成的带宽利用率的变化，以便更快地得到所需要的带宽利用率。

对 $DP_{temp}(i,q,P)$和 $ARTB_{temp}(i,q,P)$的值可以使用不同的初始步长，初始步长可以是一个固定值。对于 $ARTB_{temp}(i,q,P)$值，初始步长可以是一个取决于的值，在每次（或者第 N 次）迭代中以 2 的幂次方减小。在每次（或者每第 N 次）迭代过程中，相对于它们以前的值来为 $DP_{temp}(i,q,P)$和 $ARTB_{temp}(i,q,P)$ 的增加或降低来减少步长，这样是很有利的。举个不作为限制的例子，减少量可以是对数的（比如：2 的幂次方）。

应当认识到的是，通过将丢弃概率不仅作为带宽利用率的函数，还作为其它参数的函数关系，所述参数是比如出口队列 112 和/或子队列的深度、发送到每个接口出口队列和/或子队列的单个数据流的带宽变化性（猝发性），就能够推动本领域一般技术人员提高概率处理实

体 230 的性能。 一个数据流的猝发性可以看作是那个数据流带宽的导数。

外推处理实体 240 除了从概率处理实体 230 接收 ADP（I,Q,P）值以外，还从出口业务流管理器 114 接收 ARTB（I,Q,P）值，即每个有效的 I,Q,P 组合的平均收到字节数。如果丢弃概率为零，外推处理实体 240 就以其输入为基础为每个有效的 I,Q,P 组合计算收到业务流字节的平均数。这些值以 ARTBDP0（I,Q,P）值的形式被提供给概率处理实体 230。

ARTBDP0（I,Q,P）值的计算按照以下方式进行：给定每个有效的 I,Q,P 组合的收到业务流字节数的平均值，并且给定相同的 I,Q,P 组合的平均丢弃概率，如果丢弃概率为零，就通过外推计算收到业务流字节数的平均值。特别地，ARTBDP0（I,Q,P）= ARTB（I,Q,P）/（1-ADP（I,Q,P））。如果更经常地收到 ARTB（I,Q,P）值而非 ADP（I,Q,P）值，那么一旦收到 ARTB（I,Q,P）值，就要读取相应的 ADP（I,Q,P）值并且计算相应的 ARTBDP0（I,Q,P）。

另外，外推处理实体 240 还对那些与一个共用接口和出口队列相关联的所有优先级的 ARTBDP0（I,Q,P）值求总和，得到给定 I,Q 组合的接收字节平均数以及该 I,Q 组合效率最低情况下收到的字节。外推处理实体 240 以该 I,Q 组合的 ARTBDP0（I,Q）值的形式将这个信息提供给分配处理实体 220。因此，ARTBDP0（I,Q）= （ARTBDP0（I,Q,P））p ,其中 （X）p 表示一系列 X（P）在整个 P 值范围内总和。

在一些实施例中，这样是有利的：限制将 DP（I,Q,P）值传输到 PAUs 118 的速度以限制通过交换结构 02 的非业务流字节，同时还限制丢弃概率改变后由于更长的反应时间所引起的不稳定性。对于一个给定的（I,Q,P）三元组的新丢弃概率，不论何时当它从原先的值改变

一个超过预设绝对值或者相对值的量时都可以被发送出去，而不用以更高的速率进行更新。这个预设值是可编程的。这个预设值对同一接口的一个输出端口到下一个输出端口或者从一个出口队列 112 到下一个队列可以各不相同，或者对于同一接口 I 和出口队列 Q 组合，从一个子队列 113 到下一个子队列也不相同。在另外的实施例中，只要同一（I,Q）组合的所有丢弃概率中的一个变化超出了预设绝对值或者相对值，全部丢弃概率就会被发送出去。在丢弃概率没有实质性差别的情况下设有一个定时器，使得丢弃概率值至少按照每秒中的预设时间次数将概率值发送到 PAUs 118。

此外，根据本发明的一个实施例，为每个输出端口、与该输出端口关联的出口队列、以及优先级的组合单独产生一个丢弃概率。在其它的实施例中，一个数据包的优先级没有考虑到是丢弃还是转发一个数据包的决定中去，因此，一个单一的丢弃概率将与每个有效的输出端口和出口队列的组合关联。

而且，前面已经提到过，在决定不传输收到的数据包的情况下，数据包可能完全从存储器中丢弃或者被发送到一个存储丢弃数据包的单独存储库。在另外的实施例中，没有传输到处理结构的数据包可以沿着一个备用通道重新发送。

在又一实施例中，要丢弃的数据包被“标记”为“可丢弃数据”，直到需要使用该数据包在存储器中占用的空间时，才将该数据包丢弃。在这种方法中，如果导致数据包被打上“标记”的拥堵减轻，数据包会取消标记并继续发送到交换结构。“标记”的数据包的特性可以规定在数据包的首部。在再一实施例中，被标记的数据包仍然可通过交换结构来传输，但是标记可以用做发给高层应用程序的信号，必须迅速降低某个数据流。

显而易见，虽然上述说明提到了“丢弃”概率，但这并不意味着

数据包就真正根据这种概率被丢弃了。一个例子是不传输到交换结构102 的包被存储在一个单独的存储器中或者记录下来供以后参考。因此，术语“丢弃概率”可以看作是指非传输概率的更宽泛的概念。根据数据包发送到的出口队列的“丢弃”概率，PAU 118 做出传输或者不传输收到数据包的决定。

本领域一般技术人员应理解的是前面所述关于路径接收单元118、丢弃概率计算模块 120、利用率处理实体 210、分配处理实体 220、概率处理实体 230 和外推处理实体 240 功能性的全部或部分可以作为预编程的硬件或者固件元件（比如，专用集成电路（ASICs）,电可擦可编程只读存储器（EEPROMs），等等），或者其它相关元件来执行完成。

在本发明的其它实施例中，前面所述关于路径接收单元 118、丢弃概率计算模块 120、利用率处理实体 210、分配处理实体 220、概率处理实体 230 和外推处理实体 240 功能性的全部或部分可以作为软件来实现，所述软件由一系列通过数字计算机来执行程序指令构成，数字计算机包括一个处理器和与通信总线相连的一个存储器。这种存储器包括数据和程序指令。为了执行规范中描述的功能块，处理器适合于处理数据和程序指令。为此在图中对操作进行了描述。

程序指令能被存储在一个固定、有明确形状并且计算机系统可读的媒体中，（如，可移动磁盘，CD-ROM，ROM，或者硬盘），或者将程序指令远程存储，但通过由传输媒体连接到网络的调制解调器或者其它接口设备（如通道转接器）可以传输到计算机系统。传输媒体可以是有明确形状的媒体（如光纤或者模拟传输线）或者是使用无线技术完成的媒体（如微波、红外线或者其他传输方法）。

本领域一般技术人员显然能够理解，程序指令可以用许多编程语言写成，和很多计算机体系结构或者操作系统一起使用。比如，有些

实施例可以在过程编程语言（如“C 语言”）或者面向对象的编程语言（比如“C++”或“JAVA”）中完成。

本发明的实施例已经进行了描述和说明，本领域的技术人员显然能够理解，可以对本发明进行各种修改和变化可以对本发明做出修改、变化而不脱离本发明的范围，如同权利要求书要求的那样。

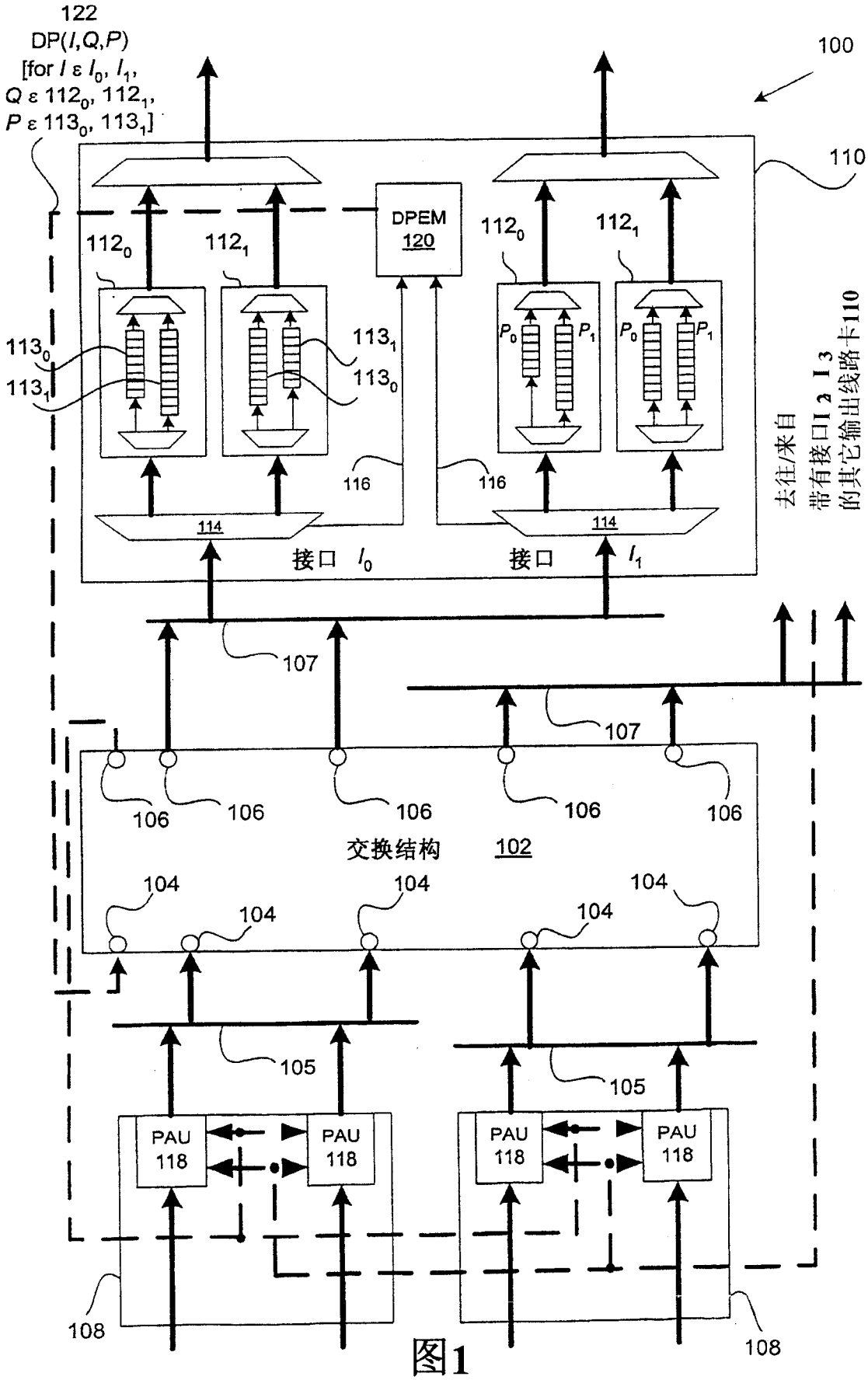

122
DP(I,Q,P)
[for I ε I0, I1,
Q ε 1120, 1121,
P ε 1130, 1131]
100
110
DPEM
120
1120
1121
1130
1131
P0
P1
116
114
接口 I0
接口 I1
去往/来自
带有接口I0 I1 I2 I3
的其它输出线路卡110
107
106
交换结构 102
104
105
PAU
118
108

图1

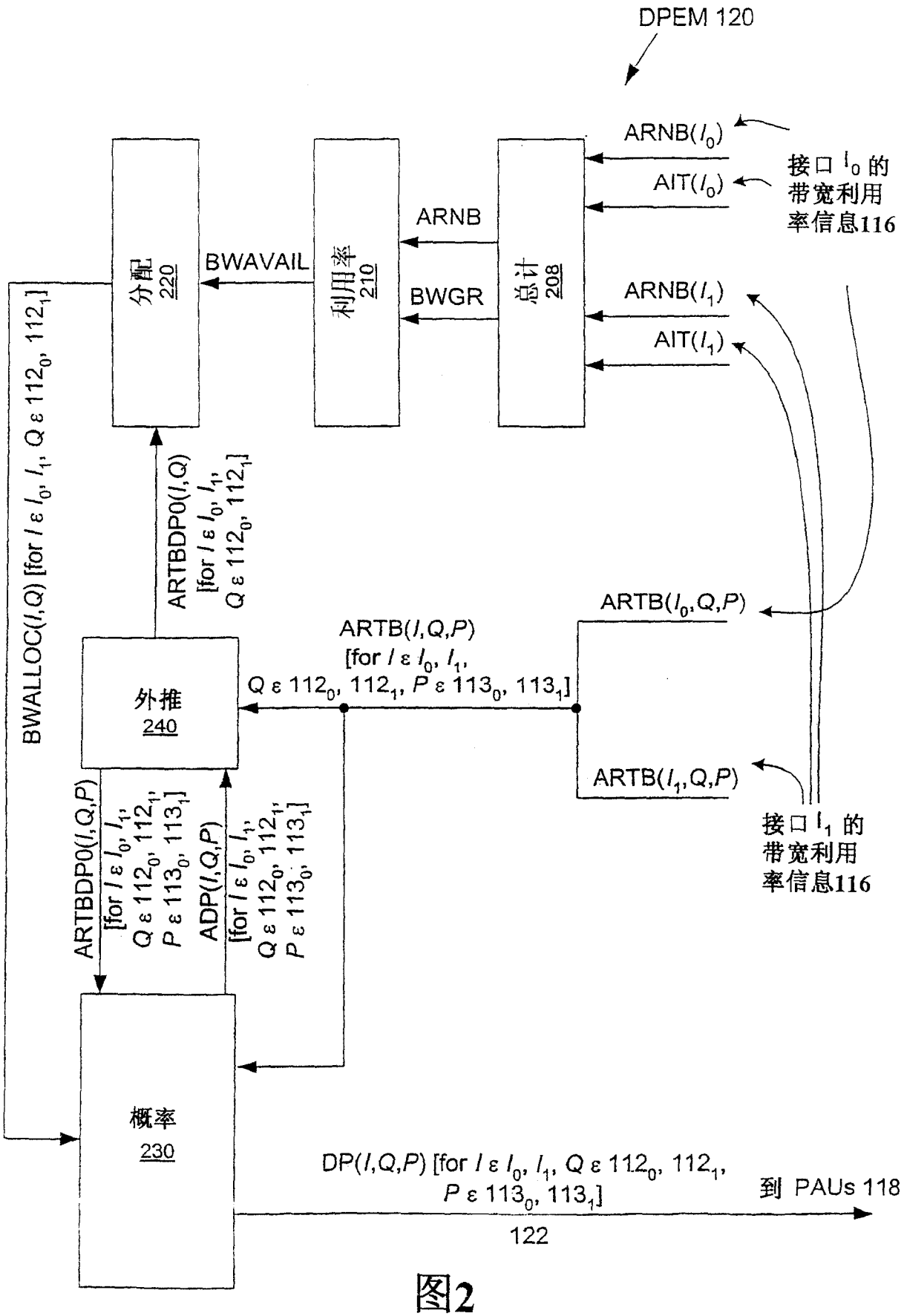

DPEM 120
ARNB(I_0)
AIT(I_0)
接口 I_0 的带宽利用率信息116
ARNB
BWGR
总计 208
利用率 210
BWAVAIL
分配 220
ARNB(I_1)
AIT(I_1)
BWALLOC(I,Q) [for $I \varepsilon I_0, I_1$, $Q \varepsilon 112_0, 112_1$]
ARTBDP0(I,Q) [for $I \varepsilon I_0, I_1$, $Q \varepsilon 112_0, 112_1$]
ARTB(I_0,Q,P)
ARTB(I,Q,P) [for $I \varepsilon I_0, I_1$, $Q \varepsilon 112_0, 112_1$, $P \varepsilon 113_0, 113_1$]
外推 240
ARTB(I_1,Q,P)
接口 I_1 的带宽利用率信息116
ARTBDP0(I,Q,P) [for $I \varepsilon I_0, I_1$, $Q \varepsilon 112_0, 112_1$, $P \varepsilon 113_0, 113_1$]
ADP(I,Q,P) [for $I \varepsilon I_0, I_1$, $Q \varepsilon 112_0, 112_1$, $P \varepsilon 113_0, 113_1$]
概率 230
DP(I,Q,P) [for $I \varepsilon I_0, I_1$, $Q \varepsilon 112_0, 112_1$, $P \varepsilon 113_0, 113_1$]
到 PAUs 118
122

图2