



(12) 发明专利

(10) 授权公告号 CN 101483593 B

(45) 授权公告日 2011. 10. 26

(21) 申请号 200910006948. 7

审查员 熊金安

(22) 申请日 2009. 02. 13

(73) 专利权人 中兴通讯股份有限公司

地址 518057 广东省深圳市南山区高新技术
产业园科技南路中兴通讯大厦法律部

(72) 发明人 潘庭山

(74) 专利代理机构 北京安信方达知识产权代理
有限公司 11262

代理人 龙洪 霍育栋

(51) Int. Cl.

H04L 12/56 (2006. 01)

(56) 对比文件

US 6363077 S1, 2002. 03. 26, 全文.

CN 1780252 A, 2006. 05. 31, 全文.

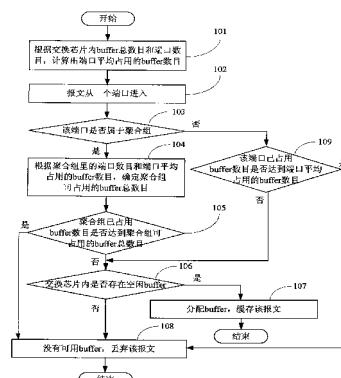
权利要求书 2 页 说明书 5 页 附图 2 页

(54) 发明名称

一种交换设备中基于聚合链路分配缓存的方
法及装置

(57) 摘要

一种交换设备中基于聚合链路分配缓存的方
法及装置, 该方法包括: 当报文从链路聚合组内
的一端口进入, 交换设备将该聚合组内所有端口
可占用的缓存总和作为聚合组可占用的缓存共享
给该聚合组内的各个端口, 如判断该聚合组内已
占用的缓存数目尚未达到该聚合组可占用的缓存
的总数目, 则分配相应容量的缓存存储报文, 否则
丢弃报文。本发明对传统的静态、动态缓存分配方
法进行了改进, 通过将聚合链路内所有端口的缓
存总和共享给各端口, 进一步提高了交换机链路
聚合组的转发性能, 同时还不会对其它普通端口
或者其它聚合组造成任何影响。



1. 一种交换设备中基于聚合链路分配缓存的方法,包括:当报文从链路聚合组内的一端口进入,交换设备将所述聚合组内所有端口可占用的缓存总和作为聚合组可占用的缓存共享给所述聚合组内的各个端口,如判断所述聚合组内已占用的缓存数目尚未达到所述聚合组可占用的缓存的总数目,则分配相应容量的缓存存储所述报文,否则丢弃所述报文。

2. 按照权利要求1所述的方法,其特征在于,根据所述聚合组内的端口数目和所述交换设备内端口平均占用的缓存数目,计算所述聚合组可占用的缓存的总数目=所述交换设备内端口平均占用的缓存数目×所述聚合组内的端口数目;所述交换设备内端口平均占用的缓存数目=所述交换设备内缓存的总数目/所述交换设备内的端口数目。

3. 按照权利要求1所述的方法,其特征在于,根据对聚合组内各个端口当前可占用的缓存的数目求和,计算所述聚合组可占用的缓存的总数目;所述聚合组内各个端口当前可占用的缓存的数目=所述交换设备当前剩余的缓存数目×所述聚合组内端口当前的活跃程度。

4. 按照权利要求2所述的方法,其特征在于,当报文从除所述聚合组内的端口以外的普通端口进入,则

计算确定所述交换设备内端口平均占用的缓存数目,或者根据交换设备当前剩余的缓存数目和所述普通端口当前的活跃程度,计算所述普通端口当前可占用的缓存数目;并且,当所述普通端口已占用的缓存数目尚未达到所述交换设备内端口平均占用的缓存数目,或者当所述普通端口已占用的缓存数目尚未达到所述普通端口当前可占用的缓存数目,则分配相应容量的缓存存储所述报文,否则丢弃所述报文。

5. 按照权利要求1至4任一项所述的方法,其特征在于,在分配相应容量的缓存之前还包括:查询所述交换设备内是否尚存在空闲缓存,如存在则分配所述缓存,否则丢弃所述报文。

6. 一种交换设备中基于聚合链路分配缓存的装置,其特征在于,所述装置包括依次连接的端口管理单元、缓存分配单元以及缓存单元;其中:

所述端口管理单元,用于管理交换设备内的所有端口,并当对进入报文的端口区分为是链路聚合组端口,则将所述聚合组内所有端口可占用的缓存总和作为聚合组可占用的缓存共享给所述聚合组内的各个端口;当通过所述缓存分配单元查询到,所述聚合组内已占用的缓存数目尚未达到所述聚合组可占用的缓存的总数目,则向所述缓存分配单元申请分配相应容量的缓存存储所述报文,否则丢弃所述报文;

所述缓存分配单元,用于根据所述端口管理单元的申请,将分配的所述缓存提供给所述端口管理单元;

所述缓存单元,用于提供存储报文的所述缓存。

7. 按照权利要求6所述的装置,其特征在于,所述装置还包括分别与所述端口管理单元和所述缓存分配单元连接的

计算单元,用于根据从所述端口管理单元获取的所述交换设备内的端口数目,以及从所述缓存分配单元获取的所述交换设备内的缓存的总数目,计算所述交换设备内端口平均占用的缓存数目=所述交换设备内的缓存的总数目/所述交换设备内的端口数目;然后根据从所述端口管理单元获取的聚合组内的端口数目和计算的所述交换设备内端口平均占用的缓存数目,计算所述聚合组可占用的缓存的总数目=所述交换设备内端口平均占用的

缓存数目 × 所述聚合组内的端口数目，并将计算结果输出给所述端口管理单元。

8. 按照权利要求 7 所述的装置，其特征在于，

所述计算单元，或用于根据从所述缓存分配单元获取的所述交换设备当前剩余的缓存数目，以及从所述端口管理单元获取的所述聚合组内端口当前的活跃程度，计算所述聚合组内各端口当前可占用的缓存的数目＝所述交换设备当前剩余的缓存数目 × 所述聚合组内端口当前的活跃程度，然后对所述聚合组内各端口当前可占用的缓存的数目求和，计算所述聚合组可占用的缓存的总数目，并将计算结果输出给所述端口管理单元。

9. 按照权利要求 7 所述的装置，其特征在于，

所述端口管理单元，当对进入报文的端口区分为是除所述聚合组内的端口以外的普通端口，则根据从所述计算单元输入的所述交换设备内端口平均占用的缓存数目，或指示所述计算单元输出计算的普通端口当前可占用的缓存数目，当通过所述缓存分配单元查询到，所述普通端口已占用的缓存数目尚未达到所述交换设备内端口平均占用的缓存数目，或者所述普通端口已占用的缓存数目尚未达到所述普通端口当前可占用的缓存数目，则向所述缓存分配单元申请分配相应容量的缓存存储所述报文，否则丢弃所述报文；

所述计算单元，还用于根据所述端口管理单元的指示，通过从所述缓存分配单元获取的所述交换设备当前剩余的缓存数目，以及从所述端口管理单元获取的所述普通端口当前的活跃程度，计算获得所述普通端口当前可占用的缓存数目，并将计算结果输出给所述端口管理单元。

10. 按照权利要求 6 至 9 任一项所述的装置，其特征在于，

所述缓存分配单元，还用于在分配缓存之前，查询所述交换设备是否尚存在空闲缓存，存在则分配所述缓存，否则向所述端口管理单元返回分配失败的信息；

所述端口管理单元，在收到所述分配失败的信息后，丢弃所述报文。

一种交换设备中基于聚合链路分配缓存的方法及装置

技术领域

[0001] 本发明涉及交换机等网络设备分配资源的方法，尤其涉及交换芯片中基于聚合链路分配缓存的方法及装置。

背景技术

[0002] 交换芯片内部存在一定数量的缓存(buffer)，当报文从交换机的源端口进入交换机后，如果有可用的buffer，那么报文会被缓存在交换机的buffer里面，直到交换机将报文从所有的目的端口发出后，该buffer才会被释放。

[0003] 交换芯片中分配buffer的方法大致有三种：第一种是交换芯片的buffer被其所有端口全局共享，只要有空闲buffer，所有端口都可以去抢占；第二种是静态buffer分配，即通过端口设置静态值来设置平均可以占用的buffer数目；第三种是动态buffer分配，即端口可以占用的buffer和交换芯片剩余buffer成线性关系，也就是说端口可以占用的buffer随着交换芯片剩余buffer减少而减少。

[0004] 以上第二种、第三种方法虽然比起第一种方法具有更好效果和更优的性能，但二者都没有考虑到一种情况：两台交换机通过链路聚合控制协议(LACP, Link Aggregation Control Protocol)，可以把多个端口聚合成为一个聚合组(trunk)。在这种情况下，如果实现buffer在端口聚合组内的共享，则能够进一步提高链路聚合组的报文转发性能，而如果仍然依上述方法基于端口分配而不基于聚合组分配，则会不利于链路聚合组的转发性能的进一步提高。

发明内容

[0005] 本发明所要解决的技术问题是提供一种交换设备中基于聚合链路分配缓存的方法及装置，能够进一步提高链路聚合组端口的转发性能。

[0006] 为了解决上述技术问题，本发明提供了一种交换设备中基于聚合链路分配缓存的方法，包括：当报文从链路聚合组内的一端口进入，交换设备将该聚合组内所有端口可占用的缓存总和作为聚合组可占用的缓存共享给该聚合组内的各个端口，如判断该聚合组内已占用的缓存数目尚未达到该聚合组可占用的缓存的总数目，则分配相应容量的缓存存储报文，否则丢弃报文。

[0007] 进一步地，根据聚合组内的端口数目和交换设备内端口平均占用的缓存数目，计算该聚合组可占用的缓存的总数目=交换设备内端口平均占用的缓存数目×该聚合组内的端口数目；交换设备内端口平均占用的缓存数目=交换设备内缓存的总数目/交换设备内的端口数目。

[0008] 进一步地，根据对聚合组内各个端口当前可占用的缓存的数目求和，计算该聚合组可占用的缓存的总数目；该聚合组内各个端口当前可占用的缓存的数目=交换设备当前剩余的缓存数目×该聚合组内端口当前的活跃程度。

[0009] 进一步地，当报文从除聚合组内的端口以外的普通端口进入，则

[0010] 计算确定交换设备内端口平均占用的缓存数目,或者根据交换设备当前剩余的缓存数目和该普通端口当前的活跃程度,计算该普通端口当前可占用的缓存数目;并且,当该普通端口已占用的缓存数目尚未达到交换设备内端口平均占用的缓存数目,或者当该普通端口已占用的缓存数目尚未达到该普通端口当前可占用的缓存数目,则分配相应容量的缓存存储报文,否则丢弃报文。

[0011] 进一步地,在分配相应容量的缓存之前还包括:查询交换设备内是否尚存在空闲缓存,如存在则分配缓存,否则丢弃报文。

[0012] 为了解决上述技术问题,本发明提供了一种交换设备中基于聚合链路分配缓存的装置,包括依次连接的端口管理单元、缓存分配单元以及缓存单元;其中:

[0013] 端口管理单元,用于管理交换设备内的所有端口,并当对进入报文的端口区分为是链路聚合组端口,则将聚合组内所有端口可占用的缓存总和作为 聚合组可占用的缓存共享给聚合组内的各个端口;当通过缓存分配单元查询到,该聚合组内已占用的缓存数目尚未达到该聚合组可占用的缓存的总数目,则向缓存分配单元申请分配相应容量的缓存存储报文,否则丢弃报文;

[0014] 缓存分配单元,用于根据端口管理单元的申请,将分配的缓存提供给端口管理单元;

[0015] 缓存单元,用于提供存储报文的缓存。

[0016] 进一步地,本发明的装置还包括分别与端口管理单元和缓存分配单元连接的计算单元,用于根据从端口管理单元获取的交换设备内的端口数目,以及从缓存分配单元获取的交换设备内的缓存的总数目,计算交换设备内端口平均占用的缓存数目=交换设备内的缓存的总数目 / 交换设备内的端口数目;然后根据从端口管理单元获取的聚合组内的端口数目和计算的交换设备内端口平均占用的缓存数目,计算该聚合组可占用的缓存的总数目=交换设备内端口平均占用的缓存数目 × 该聚合组内的端口数目,并将计算结果输出给端口管理单元。

[0017] 进一步地,

[0018] 计算单元,或用于根据从缓存分配单元获取的交换设备当前剩余的缓存数目,以及从端口管理单元获取的聚合组内端口当前的活跃程度,计算该聚合组内各端口当前可占用的缓存的数目=交换设备当前剩余的缓存数目 × 该聚合组内端口当前的活跃程度,然后对该聚合组内各端口当前可占用的缓存的数目求和,计算该聚合组可占用的缓存的总数目,并将计算结果输出给端口管理单元。

[0019] 进一步地,

[0020] 端口管理单元,当对进入报文的端口区分为是除所述聚合组内的端口以外的普通端口,则根据从计算单元输入的交换设备内端口平均占用的缓存数目,或指示计算单元输出计算的普通端口当前可占用的缓存数目,当通过缓存分配单元查询到,该普通端口已占用的缓存数目尚未达到交换设备内端口平均占用的缓存数目,或者该普通端口已占用的缓存数目尚未达到该普通端口当前可占用的缓存数目,则向缓存分配单元申请分配相应容量的缓存存储 报文,否则丢弃报文;

[0021] 计算单元,还用于根据端口管理单元的指示,通过从缓存分配单元获取的交换设备当前剩余的缓存数目,以及从端口管理单元获取的该普通端口当前的活跃程度,计算获

得该普通端口当前可占用的缓存数目，并将计算结果输出给端口管理单元。

[0022] 进一步地，

[0023] 缓存分配单元，还用于在分配缓存之前，查询交换设备是否尚存在空闲缓存，存在则分配缓存，否则向端口管理单元返回分配失败的信息；

[0024] 端口管理单元，在收到该分配失败的信息后，丢弃报文。

[0025] 本发明对传统的静态、动态缓存分配方法进行了改进，通过将链路聚合组内所有端口的缓存总和共享给聚合组内的各端口，实现基于聚合链路的缓存分配，由此进一步提高了交换机链路聚合组的转发性能，同时还不会对其它普通端口或者其它聚合组造成任何影响。

附图说明

[0026] 图 1 是本发明交换设备中基于聚合链路分配缓存的方法实施例流程图；

[0027] 图 2 是本发明基于聚合链路分配缓存的装置实施例结构框图。

具体实施方式

[0028] 本发明引入了一种交换设备中基于聚合链路分配缓存的方法及装置，其发明构思是，交换设备将链路聚合组里所有端口可占用的缓存共享给聚合组内的每一端口；当报文从交换设备聚合组内的端口进入，且交换设备判断聚合组里已占用的缓存数目尚未达到聚合组共享缓存的总数目，则分配缓存存储报文，否则丢弃报文。

[0029] 以下结合附图和优选实施例，对本发明的上述构思展开进行详细阐述，以解释清楚本发明的技术方案。为了阐述方便，以下实施例仅在现有的第二种方法的基础上面进行阐述。实际上，本发明的技术方案对于现有的第二种、第三种 buffer 分配方法均适用。

[0030] 请参见图 1，本发明交换设备中基于聚合链路分配缓存的方法包括以下步骤：

[0031] 步骤 101：根据交换芯片内 buffer 总数目和端口数目，计算出端口平均占用的 buffer 数目=交换芯片内 buffer 总数目 / 端口数目；

[0032] 实际上，如果是动态分配 buffer，则步骤 101 是根据交换芯片当前剩余的 buffer 数目和端口当前的活跃程度（接收报文的频度），计算端口当前可占用的 buffer 数目=交换芯片当前剩余的 buffer 数目 × 端口当前的活跃程度，即如果端口当前的活跃程度高，则在当前剩余的 buffer 中可占用的 buffer 数目就大，反之亦然。

[0033] 步骤 102、103：报文从端口进入，如果判断为链路聚合组里的端口，执行步骤 104，如果判断是除链路聚合组里的端口外的普通端口，执行步骤 109；

[0034] 步骤 104 ~ 108：根据聚合组里的端口数目和端口平均占用的 buffer 数目，确定聚合组可占用的 buffer 总数目=端口平均占用的 buffer 数目 × 聚合组里的端口数目；判断聚合组内（所有端口）已占用的 buffer 数目是否达到聚合组可占用的 buffer 总数目，达到则丢弃报文，没有达到则继续查询交换芯片内是否尚存在空闲缓存，存在则分配 buffer 并缓存报文，否则丢弃报文；结束流程。

[0035] 如果是动态分配 buffer，则步骤 104 是根据步骤 101 计算的聚合组内各端口当前可占用的 buffer 数目，确定的聚合组可占用的 buffer 总数目，等于聚合组内各端口当前可占用的 buffer 数目之和。

[0036] 步骤 109 :判断该端口已占用 buffer 数目是否达到端口平均占用的 buffer 数目, 达到则执行步骤 108, 即丢弃报文 ; 没有达到则执行步骤 106, 即继续判断交换芯片内是否尚存在空闲缓存, 存在则获取 buffer 并缓存报文, 否则丢弃报文。

[0037] 上述步骤 106 作为可选步骤, 用于保证聚合组里的端口或普通端口在分 配的 buffer 中缓存报文时, 不会与其它端口产生 buffer 资源冲突, 亦即保证不会对其他端口产生影响。

[0038] 下面就上述方法实施例给出一个应用示例来帮助进一步理解本发明的技术方案。

[0039] 假设交换机共有 2000 个 buffer, 每个 buffer 256 字节 ; 交换机共有 20 个物理端口, 端口 1、端口 2、端口 3 及端口 4 都在链路聚合组 1 里面。

[0040] 假设报文从聚合组 1 进入, 则基于聚合链路分配缓存的方法就按如下步骤执行 :

[0041] 步骤 1 :计算端口平均占用的 buffer 数目 = $2000/20 = 100$, 即每个端口平均可以占用 100 个 buffer ;

[0042] 步骤 2 :如果报文从端口 1 ~ 端口 4 的任意一个端口进入, 则交换机判断该端口属于链路聚合组里的端口, 先计算链路聚合组可占用的 buffer 总数目 = 端口平均占用的 buffer 数目 × 聚合组里的端口数目 = $100 \times 4 = 400$, 即链路聚合组可占用 400 个 buffer ;

[0043] 步骤 3 :判断端口 1 ~ 端口 4 当前已占用的 buffer 总数目是否达到 400, 如果未达到则交换机分配缓存并存储报文, 否则丢弃从聚合组 1 进来的报文。

[0044] 如图 2 所示, 本发明根据图 1 所示的方法相应地为交换机提供的一种基于聚合链路分配缓存的装置的实施例, 该装置 200 包括依次连接的计算单元 210、端口管理单元 220、缓存分配单元 230 以及缓存单元 240 ; 其中 :

[0045] 计算单元 210, 还与缓存分配单元 230 连接, 用于通过缓存分配单元 230 获取交换芯片内 buffer 总数目, 再根据从端口管理单元 220 获取的交换芯片内端口数目, 计算出端口平均占用的 buffer 数目 ; 根据从端口管理单元 220 获取的聚合组里的端口数目和计算出的端口平均占用的 buffer 数目 N, 计算聚合组可占用的 buffer 总数目 M ; 将计算结果输出给端口管理单元 220。

[0046] $N = \text{交换芯片内 buffer 总数目} / \text{端口数目}$;

[0047] $M = \text{端口平均占用的 buffer 数目} \times \text{聚合组里的端口数目}$ 。

[0048] 当然, 如果是采用动态分配 buffer 的方法, 计算单元 210 则可分别计算端口当前可占用的 buffer 数目和聚合组可占用的 buffer 总数目, 具体方法如前所述, 故此不再赘述。

[0049] 端口管理单元 220, 用于管理交换芯片内的所有端口, 确定端口数目, 并区分进入报文的端口是链路聚合组端口还是非链路聚合组端口 ; 若区分为链路聚合组端口, 则根据记录的该聚合组已占用的 buffer 数目 m 和输入 M, 判断 m 是否小于 M, 小于则向缓存分配单元 230 申请分配 buffer, 并将报文缓存如缓存分配单元 230 分配的 buffer 中, 否则丢弃该报文 ; 若区分为非链路聚合组端口, 则根据记录的该端口已占用的 buffer 数目 n 和输入 N, 判断 n 是否小于 N, 小于则根据报文容量向缓存分配单元 230 申请分配 buffer, 并将报文缓存如缓存分配单元 230 分配的 buffer 中, 否则丢弃该报文 ; 在目的端口提出发送报文的请求时, 从缓存分配单元 230 取出相应的报文输出给该目的端口, 并向缓存分配单元 230 请求释放 buffer。

[0050] 图 2 所示的计算单元 210, 可合并在端口管理单元 220 中。

[0051] 缓存分配单元 230, 用于根据端口管理单元 220 的申请分配相应容量的 buffer, 并根据端口管理单元 220 的请求释放相应的 buffer。

[0052] 缓存单元 240, 用于提供存储报文的 buffer。

[0053] 本发明与现有技术相比较, 引入了一种基于聚合链路的缓存分配方法, 能够实现对链路聚合组里的端口 buffer 在聚合组内的共享, 从而提高链路聚合组的报文转发性能。

[0054] 当然, 本发明还可以有其他多种实施例, 在不背离本发明精神及其实质的情况下, 熟悉本领域的技术人员可根据本发明作出各种相应的改变和变形, 但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

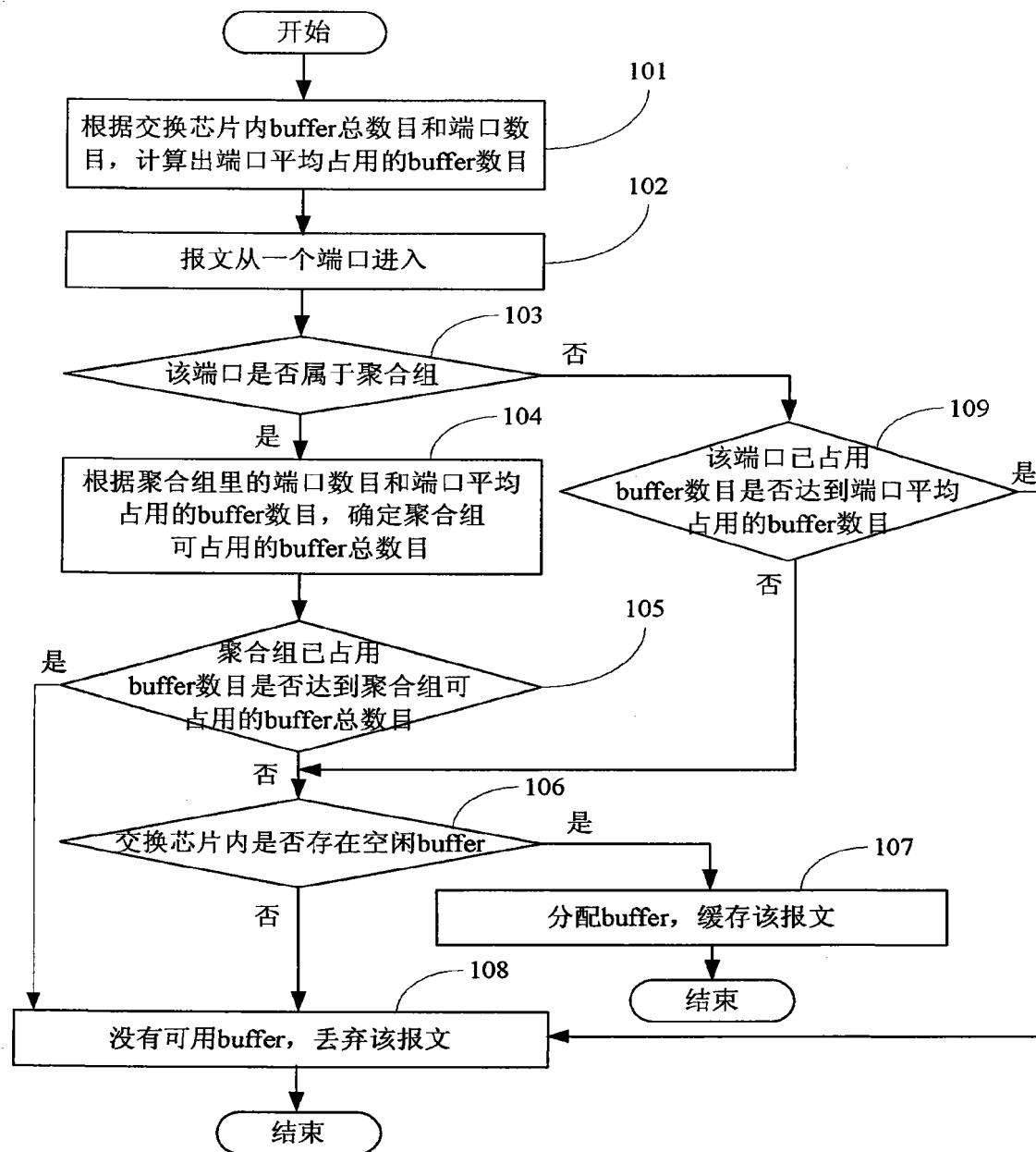


图 1

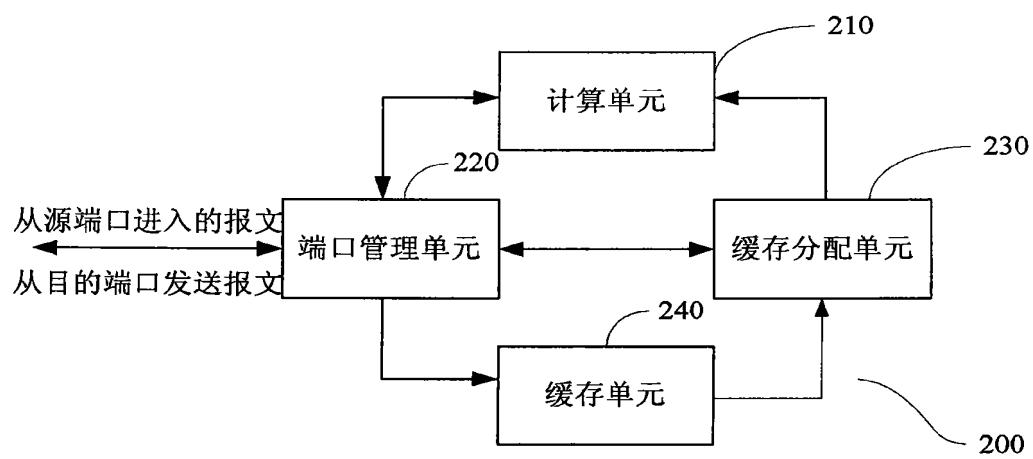


图 2