

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日
2008年6月26日 (26.06.2008)

PCT

(10) 国际公布号
WO 2008/074191 A1

(51) 国际专利分类号:

H04L 12/56 (2006.01)

[CN/CN]; 中国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, Guangdong 518057 (CN)。

(21) 国际申请号:

PCT/CN2006/003526

(74) 代理人: 北京康信知识产权代理有限公司 (KANGXIN PARTNERS, P. C.); 中国北京市海淀区知春路甲48号盈都大厦A座16层余刚, Beijing 100098 (CN)。

(22) 国际申请日:

2006年12月21日 (21.12.2006)

(25) 申请语言:

中文

(26) 公布语言:

中文

(71) 申请人 (对除美国外的所有指定国): 中兴通讯股份有限公司 (ZTE CORPORATION) [CN/CN]; 中国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, Guangdong 518057 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

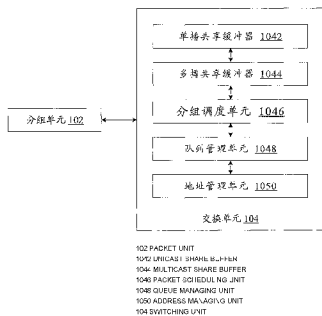
(72) 发明人; 及

(75) 发明人/申请人 (仅对美国): 杨祖发 (YANG, Zufa)

[见续页]

(54) Title: A PACKET SWITCHING SYSTEM AND A METHOD THEREOF

(54) 发明名称: 分组交换装置及方法



(57) Abstract: An equipment for packet switching and a method thereof, the packet switching equipment includes: a packet unit, for dividing or filling the input data into the fixed length packet; a switching unit, for dividing at least one priority for the unicast packet and the multicast packet respectively which are from the packet unit, and respectively assigning a queue to the unicast packet and the multicast packet of each priority, then scheduling it according to the priority of the queue. The unicast packet queue which is assigned to the highest priority and the multicast packet queue which is assigned to the highest priority are used to carrying the TDM service or the packet service, and the unicast packet queue which is assigned to the other priority and the multicast packet queue which is assigned to the other priority are used to carrying the packet service. The invention is also applied to the TDM service and the packet service switching.

(57) 摘要:

本发明公开了一种分组交换装置及方法, 其中, 该分组交换装置: 分组单元, 用于将输入数据划分或填充为定长的分组; 以及交换单元, 用于将来自分组单元的单播分组和多播分组分别划分为至少一个优先级并为每个优先级的单播分组和多播分组分别分配一个队列, 并且根据队列的优先级进行调度, 其中, 分配给最高优先级的单播分组的队列和分配给最高优先级的多播分组的队列用于承载 TDM 业务或分组业务, 分配给其它优先级的单播分组的队列和分配给其它优先级的多播分组的队列用于承载分组业务。本发明同时适用于 TDM 业务和分组业务的交换。

WO 2008/074191 A1



(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR),

OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告。

分组交换装置及方法

技术领域

本发明涉及通信领域，更具体地涉及一种分组交换装置及方法。

背景技术

5 时分复用（Time Division Multiplexing，简称 TDM）网络基于固定时隙交叉业务，同一业务的数据占用等间隔的定长时隙转发，适应传送对时延敏感的业务，如实时语音、高清晰视频信息等。分组交换网络基于存贮转发和统计复用交换业务，同一业务相邻分组间隔不确定，适应传送对时延不敏感的业务，如电子邮件、文件、禁止图像等。两种网络特性不同，承载的业务差异性较大，
10 同时在电信领域均得到了广泛应用。

 近年来，人们一直在探索简化电信网络的结构，降低网络的建设成本，提高现有网络资源的利用率，并相继推出了支持多业务支持平台或多业务交换平台（Multi-Service Provisioning Platform or Multi-Service Switching Platform，简称 MSPP）和边缘到边缘伪线仿真（Pseudo Wire Emulation Edge-to-Edge，简称 PWE3）等功能的网络设备。前者采用统一的数据封装格式同时将分组业务和 TDM 业务通过同步数字系列（Synchronous digital hierarchy，简称 SDH）或同步光网络（Synchronous optical network，简称 SONET）或准同步数字体系（Plesiochronous Digital Hierarchy，简称 PDH）或光传送体系（Optical Transport Hierarchy，简称 OTH）网络传送，后者采用边缘节点仿真技术同时将分组业务
15 和 TDM 业务通过分组交换网络传送。
20

 伴随着因特网（Internet）的高速发展，在电信网络传送的业务中，分组业务的占有率远远超过 TDM 业务，并仍在进一步加大差距。与之相适应，电信网络大部分的业务交换设备采用分组交换方式。

 TDM 业务常常要求按照字节间插方式实现交换，如通过 SDH 或 SONET
25 网络传送的 E1、T1 等 TDM 业务，不同业务数据通过字节间插复用后传送。如果在 TDM 交换设备的线卡上，将输入方向的每时隙由单字节扩展到多字节，

TDM 同步时隙交叉部件按多字节时隙进行交叉，在线卡的输出方向重新恢复为字节间插方式，可以实现与字节交叉方式 TDM 交换设备等价的功能。

5 通常的分组交换部件采用共享存贮转发和基于输出排队的方式实现。所有端口共享统一的存贮器将引起业务对交换设备内部缓冲资源的竞争，多条业务统计复用到同一端口时将引起业务对端口带宽资源的竞争。解决竞争的有效办法就是使用队列管理缓冲区，限制各业务对共享缓冲区的占用，并根据某种策略，如基于优先级调度端口排队的分组。

10 分组的可变长特性必然引入不确定的缓冲时延。在分组的交换过程中，资源竞争的出现也将引入不确定的转发时延，因此，传统的分组交换方式不适合传送 TDM 业务，分组交换网络和 TDM 交换网络还将会在相当长时期内并存。

由上可见，需要一种新的适合传送 TDM 业务的分组交换方式，从而使得 TDM 业务和分组业务都可以通过该种新的分组交换方式进行传送。

发明内容

15 鉴于上述问题，本发明提出了一种分组交换装置和方法，以同时适应 TDM 业务和分组业务的交换需求。

20 根据本发明的一方面，提出了一种分组交换装置，包括：分组单元 102，用于将输入数据划分或填充为定长的分组；以及交换单元 104，用于将来自分组单元的单播分组和多播分组分别划分为至少一个优先级并为每个优先级的单播分组和多播分组分别分配一个队列，并且根据队列的优先级进行调度，其中，分配给最高优先级的单播分组的队列和分配给最高优先级的多播分组的队列用于承载 TDM 业务或分组业务，分配给其它优先级的单播分组的队列和分配给其它优先级的多播分组的队列用于承载分组业务。

在上述分组交换装置中，队列用于缓存分组的地址，采用先入先出方式进行操作。

25 在上述分组交换装置中，交换单元包括：单播共享缓冲器，用于存储来自分组单元的单播分组；多播共享缓冲器，用于存储来自分组单元的多播分组；分组调度单元，用于根据队列的优先级从队列中调度分组；队列管理单元，用

于分别为每个端口的单播分组队列和多播分组队列分配共享缓冲空间，并将为最高优先级的单播分组队列分配的单播共享缓冲空间用于 TDM 单播业务或单播分组业务，将为最高优先级的多播分组队列分配的多播共享缓冲空间用于 TDM 多播业务或多播分组业务，以及在单播分组和多播分组被转发送指定端口后，释放单播分组和多播分组对应的共享缓冲空间；以及地址管理单元，用于管理单播共享缓冲器和多播共享缓冲期的空闲存储地址。

在上述分组交换装置中，单播共享缓冲器和多播共享缓冲器的写入速度不低于交换单元的所有输入端口的分组输入速率之和，单播共享缓冲器和多播共享缓冲器的读出速度不低于交换单元的所有输出端口的分组输出速度之和。

10 在上述分组交换装置中，队列管理单元在多播共享缓冲器中将多播分组的存储地址复制到多个指定端口的相应多播队列，在达到预定时间时，释放多播分组存储地址。

在上述分组交换装置中，在占用的单播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在占用的多播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

在上述分组交换装置中，在单播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在多播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

在上述分组交换装置中，队列调度单元通过以下方式调度相同优先级的单播业务和多播业务：优先调度单播业务，优先调度多播业务，随机调度。

在上述分组交换装置中，队列调度单元优先调度最高优先级队列中的分组。

25 根据本发明的另一方面，提供了一种分组交换方法，其包括以下步骤：S1202，分组单元将输入数据划分或填充为定长的分组；以及 S1204，交换单元将来自分组单元的单播分组和多播分组分别划分为至少一个优先级并为每个

优先级的单播分组和多播分组分别分配一个队列，并根据队列的优先级调度分组，其中，分配给最高优先级的单播分组的队列和分配给最高优先级的多播分组的队列用于承载 TDM 业务或分组业务，分配给其它优先级的单播分组的队列和分配给其它优先级的多播分组的队列用于承载分组业务。

- 5 在上述分组交换方法中，队列用于缓存分组的地址，采用先入先出方式进行操作。

- 在上述分组交换方法中，步骤 S1204 包括：S12042，通过单播共享缓冲器存储来自分组单元的单播分组并将为最高优先级的单播分组队列分配的单播共享缓冲空间用于 TDM 单播业务或单播分组业务，以及通过多播共享缓冲器存储来自分组单元的多播分组并将为最高优先级的多播分组队列分配的多播共享缓冲空间用于 TDM 多播业务或多播分组业务；以及 S12044，分组调度单元，用于根据队列的优先级从队列中调度分组；以及 S12046，在单播分组和多播分组被转发送指定端口后，释放单播分组和多播分组对应的共享缓冲空间。

- 15 在上述分组交换方法中，单播共享缓冲器和多播共享缓冲器的写入速度不低于交换单元的所有输入端口的分组输入速率之和，单播共享缓冲器和多播共享缓冲器的读出速度不低于交换单元的所有输出端口的分组输出速度之和。

在上述分组交换方法中，多播共享缓冲器将多播分组的存储地址复制到多个指定端口的相应多播队列，在达到预定时间时，释放多播分组存储地址。

- 20 在上述分组交换方法中，在占用的单播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在占用的多播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

- 在上述分组交换方法中，在单播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在多播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

在上述分组交换方法中,通过以下方式调度相同优先级的单播业务和多播业务: 优先调度单播业务, 优先调度多播业务, 随机调度。

本发明保证了最高优先级业务优先占用全局共享缓冲空间,并将最高优先级业务分组用于转发 TDM 业务或分组业务, 将其他优先级分组仅用于转发分
5 组业务, 从而使得本发明的分组交换装置和方法能够同时适应 TDM 业务和分
组业务的交换。

附图说明

此处所说明的附图用来提供对本发明的进一步理解, 构成本申请的一部分, 本发明的示意性实施例及其说明用于解释本发明, 并不构成对本发明的不
10 当限定。在附图中:

图 1 是根据本发明实施例的分组交换装置的功能框图;

图 2 是根据本发明实施例的分组交换装置的功能结构图;

图 3 是根据本发明实施例的单播共享缓冲器门限设置图;

图 4 是根据本发明实施例的多播共享缓冲器门限设置图;

15 图 5 是根据本发明实施例的单播优先级队列结构图;

图 6 是根据本发明实施例的单播优先级队列门限设置图;

图 7 是根据本发明实施例的单播优先级队列到单播共享缓冲器的映射图;

图 8 是根据本发明实施例的多播优先级队列结构图;

图 9 是根据本发明实施例的多播优先级队列门限设置图;

20 图 10 是本发明实施例的多播优先级队列到多播共享缓冲器的映射图;

图 11 是根据本发明实施例的端口队列调度时隙分布图; 以及

图 12 是根据本发明实施例的分组交换方法的流程图。

具体实施方式

下面参考附图，详细说明本发明的具体实施方式。

参考图 1，说明根据本发明实施例的分组交换装置。如图 1 所示，该分组交换装置包括：分组单元 102，用于将输入数据划分或填充为定长的分组；以及交换单元 104，用于将来自分组单元的单播分组和多播分组分别划分为至少一个优先级并为每个优先级的单播分组和多播分组分别分配一个队列，并且根据队列的优先级进行调度，其中，分配给最高优先级的单播分组的队列和分配给最高优先级的多播分组的队列用于承载 TDM 业务或分组业务，分配给其它优先级的单播分组的队列和分配给其它优先级的多播分组的队列用于承载分组业务。

其中，队列用于缓存分组的地址，采用先入先出方式进行操作。

其中，交换单元 104 包括：单播共享缓冲器 1042，用于存储来自分组单元的单播分组；多播共享缓冲器 1044，用于存储来自分组单元的多播分组；分组调度单元 1046，用于根据队列的优先级从队列中调度分组；队列管理单元 1048，用于分别为每个端口的单播分组队列和多播分组队列分配共享缓冲空间，并将为最高优先级的单播分组队列分配的单播共享缓冲空间用于 TDM 单播业务或单播分组业务，将为最高优先级的多播分组队列分配的多播共享缓冲空间用于 TDM 多播业务或多播分组业务，以及在单播分组和多播分组被转发指定端口后，释放单播分组和多播分组对应的共享缓冲空间；以及地址管理单元 1050，用于管理单播共享缓冲器和多播共享缓冲器的空闲存储地址。

其中，单播共享缓冲器和多播共享缓冲器的写入速度不低于交换单元的所有输入端口的分组输入速率之和，单播共享缓冲器和多播共享缓冲器的读出速度不低于交换单元的所有输出端口的分组输出速度之和。

其中，队列管理单元在多播共享缓冲器中将多播分组的存储地址复制到多个指定端口的相应多播队列，在达到预定时间时，释放多播分组存储地址。

其中，在占用的单播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在占用的多播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

- 5 其中，在单播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在多播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

- 10 其中，队列调度单元通过以下方式调度相同优先级的单播业务和多播业务：优先调度单播业务，优先调度多播业务，随机调度。

其中，队列调度单元优先调度最高优先级队列中的分组。

- 15 具体地，如图 2 所示，在实际网络中，根据本发明实施例的分组交换装置包括以下部分：单播共享缓冲器（Unicast Share Buffer，简称 USB）、多播共享缓冲器（Multicast Share Buffer，简称 MSB）、单播优先级队列（Unicast Priority Queue，简称 UPQ）、多播优先级队列（Multicast Priority Queue，简称 MPQ）、分组调度器（Packet Scheduler，简称 PSch）、队列管理（Queue Management，简称 QM）、及地址管理（Address Management，简称 AM）。

- 20 在下面的描述中，我们假设根据本发明实施例的分组交换装置分别将单播业务和多播业务划分为了 4 个优先级别。单播业务和多播业务的 4 优先级分别命名为 Pri0（Priority 0，优先级 0）、Pri1、Pri2、Pri3。其中，Pri0 优先级最高，Pri3 优先级最低。每端口每 Pri0 单播业务和每 Pri0 多播业务可以选择 TDM 属性。当选择 TDM 属性时，指示相应分组承载的是 TDM 业务，否则，指示相应分组承载的是分组业务。每端口每 Pri1、Pri2、Pri3 单播业务和每端口每 Pri1、Pri2、Pri3 多播业务没有 TDM 属性，只能转发分组业务。

- 25 其中，USB 被分组交换装置的交换部件的所有端口单播业务共享使用，采用读写操作分离的双口存贮器结构，以分组为单位存贮业务数据，写入速率不低于交换部件所有输入端口最大分组输入速率之和，读出速率不低于交换部件所有输出端口最大分组输出速率之和。

QM 采用门限值控制单播业务对 USB 的占用，如图 3 所示，USB 高门限（USB High Threshold，简称 USB_HT）是 USB 的总容量。USB 低门限（USB Low Threshold，USB_LT）是 USB 已经分配给交换部件所有单播优先级业务队列和单播 TDM 业务队列的容量。USB_LT 必须小于 USB_HT。从 USB_LT 到 USB_HT 之间的缓冲容量保留给 TDM 单播业务使用。当对 USB 的占用达到 USB_LT 时，禁止继续缓冲任何分组业务类型的单播分组。只要对 USB 的占用没有达到 USB_HT，总是允许缓冲任何单播 TDM 业务分组。

MSB 为交换部件所有端口共享使用，采用读写操作分离的双口存储器结构，以分组为单位存储业务数据，写入速率不低于交换部件所有输入端口最大分组输入速率之和，读出速率不低于交换部件所有输出端口最大分组输出速率之和。

QM 采用门限值控制多播业务对 MSB 的占用，如图 4 所示，MSB 高门限（MSB High Threshold，简称 MSB_HT）是 MSB 的总容量。MSB 低门限（MSB Low Threshold，简称 MSB_LT）是 MSB 已经分配给交换部件所有多播优先级业务队列和多播 TDM 业务队列的容量。MSB_LT 必须小于 MSB_HT。从 MSB_LT 到 MSB_HT 之间的缓冲容量保留给 TDM 多播业务使用。当对 MSB 的占用达到 MSB_LT 时，禁止继续缓冲任何分组业务类型的多播分组。只要对 MSB 的占用没有达到 MSB_HT，总是允许缓冲任何多播 TDM 业务分组。

每端口包含的 4 个单播优先级队列分别为单播优先级队列 0（Unicast Priority Queue 0，简称 UPQ0）、UPQ1、UPQ2、UPQ3。这些 UPQ 结构相同，采用 FIFO 操作方式，如图 5 所示。UPQ 单元存储特定单播优先级业务分组的检索地址，而不是特定优先级单播业务分组所包含的数据。单播分组存储地址总是从 UPQ 头部（UPQ Head，简称 UPQ_H）输出，从 UPQ 尾部（UPQ Tail，简称 UPQ_T）的下一个单元加入。

QM 使用门限值限制每端口每 UPQ 对 USB 的占用，如图 6 所示，每端口每 UPQ 包含 UPQ 低门限（UPQ Lower Threshold，简称 UPQ_LT）和 UPQ 高门限（UPQ High Threshold，简称 UPQ_HT）。UPQ_HT 是特定 UPQ 的存储空间。UPQ_LT 用于控制特定 UPQ 对应的单播业务对 USB 的占用，当特定 UPQ 的分组排队长度低于 UPQ_LT 时，允许 USB 继续缓冲对应的单播业务分组。对于每 UPQ，设置的 UPQ_LT 不能大于 UPQ_HT。所有 UPQ 的 UPQ_LT 之和

等于 USB_LT 等价的单播分组存储地址数。

任何 UPQ 当前排队的单播分组并不是连续地存储在 USB 中,如图 7 所示,属于同一 UPQ 的单播分组可以分布在 USB 的任何位置。

5 每端口包含的 4 个多播优先级队列分别为多播优先级队列 0 (Multicast Priority Queue 0, 简称 MPQ0)、MPQ1、MPQ2、MPQ3。这些 MPQ 结构相同,采用 FIFO 操作方式,如图 8 所示。MPQ 单元存储特定多播优先级业务分组的检索地址,而不是特定优先级多播业务分组所包含的数据。多播分组存储地址总是从 MPQ 头部 (MPQ Head, 简称 MPQ_H) 输出,从 MPQ 尾部 (MPQ Tail, 简称 MPQ_T) 的下一个单元加入。

10 QM 使用门限值限制每端口每 MPQ 对 MSB 的占用,如图 9 所示,每端口每 MPQ 包含 MPQ 低门限 (MPQ Lower Threshold, 简称 MPQ_LT) 和 MPQ 高门限 (MPQ High Threshold, 简称 MPQ_HT)。MPQ_HT 是特定 MPQ 的存储空间。MPQ_LT 用于控制特定 MPQ 对应的多播业务对 MSB 的占用,当特定 MPQ 的分组排队长度低于 MPQ_LT 时,允许 MSB 继续缓冲对应的多播业务分
15 组。对于每 MPQ,设置的 MPQ_LT 不能大于 MPQ_HT。所有 MPQ 的 MPQ_LT 之和等于 MSB_LT 等价的多播分组存储地址数。

QM 使用队列输入许可 (Queue Input Grant, 简称 QIG) 信号通知外界输入分组,每 UPQ 和每 MPQ 都存在对应的 QIG 信号。对于单播分组业务,任何 UPQ 排队长度低于该 UPQ 的 UPQ_LT 时,QM 将对应用于该 UPQ 的 QIG 信号设置成真,否则设置成假;对于多播分组业务,任何 MPQ 排队长度低于该
20 MPQ 的 MPQ_LT 时,QM 将对应用于该 MPQ 的 QIG 信号设置成真,否则设置成假。对于单播 TDM 业务,如果当前对 USB 的占用低于 USB_HT,QM 将对应用于该 UPQ 的 QIG 信号设置成真,否则设置成假;对于多播 TDM 业务,如果当前对 MSB 的占用低于 MSB_HT 时,QM 将对应用于该 MPQ 的 QIG 信号设置成真,否则设置成假。
25

任何 MPQ 当前排队的单播分组并不是连续地存储在 MSB 中,如图 10 所示,属于同一 MPQ 的多播分组可以分布在 MSB 的任意位置。从图 10 可以看出,对于任何多播分组,QM 仅在 MSB 保留一份拷贝,将多播分组的存储地址复制到多个指定端口的相应多播队列。如“多播分组 y”需要多播到端口 i、

端口 j、端口 k，相应的存贮地址被分别复制到 MPQi、MPQj、MPQk。

PSch 按照定长的调度时隙 (Schedule Time Slot, 简称 STS) 从 UPQs 或 MPQs 调度业务分组。每端口包含一个 PSch, 负责调度该端口的 4 个 UPQs 和 4 个 MPQs。各端口的 PSch 独立工作。所有 PSch 每秒钟能够处理的 STS 总数
5 等价于交换部件每秒钟能够转发的最大分组数, 对应于交换部件的最大分组交换容量。PSch 基于优先级调度分组, 结合图 2 和图 11, 调度过程描述如下:
尽管特定端口的 STS 并不邻接, 但相邻 STS 之间的时间间隔相同, 且是 STS 的整数倍, PSch 只有在每 STS 期间才能执行调度操作, 每 STS 最多能够从对应端口的 4 个 UPQs 和 4 个 MPQs 中调度一个分组; 每端口的 PSch 存在 8 条
10 队列输出许可 (Queue Output Grant, 简称 QOG) 输入信号, 分别控制该端口 4 个 UPQs 和 4 个 MPQs 的调度; 特定端口上的相同 Pri 的 UPQ 和 MPQ 具有相同的调度优先级, PSch 可优先调度 UPQ, 或优先调度 MPQ, 也可以随机选择; 当特定的 PSch 执行调度时, 首先调度 Pri0 队列的分组, 如果 Pri0 队列非空, 且对应的 QOG 信号状态为真, 则 PSch 从相应的 Pri0 队列中输出一个分组,
15 之后, 进入等待状态, 等待下一个 STS 的到来; 否则, 如果 Pri1 队列非空, 且对应的 QOG 信号状态为真, 则 PSch 从相应的 Pri1 队列中输出一个分组, 之后, 进入等待状态, 等待下一个 STS 的到来; 依此类推, 直到 Pri3 队列被调度为止; 如果当前 STS 期间没有合适的分组输出, 则 PSch 直接进入等待状态, 等待下一个 STS 的到来; 特定 PSch 每次从选择的 UPQ 输出一个单播分组,
20 立即释放该单播分组的存贮地址; 特定 PSch 每次从选择的 MPQ 输出一个多播分组, 如果对应的端口是该多播分组最后一个多播端口, 则立即释放该多播分组的存贮地址, 否则, 继续保持该多播分组的存贮地址; 当特定 PSch 输出一个分组时, 同时输出一个多播输出指示 (Multicast Output Indicating, 简称 MCOI) 信号, 当 MCOI 信号为真时, 指示当前从 MSB 输出分组多播分组, 否
25 则, 指示从 USB 输出单播分组。

AM 负责管理 USB 和 MSB 的空闲存贮地址。当特定业务需要缓冲一个单播或多播输入分组时, AM 为该业务对应的队列分配一个 USB 或 MSB 的空闲存贮地址。每 PSch 释放一个单播或多播存贮地址, AM 立即回收该存贮地址。

通过以上描述可知, 根据本发明实施例的分组交换装置执行的分组交换方法如图 12 所示, 包括以下步骤: S1202, 分组单元将输入数据划分或填充为定长的分组; 以及 S1204, 交换单元将来自分组单元的单播分组和多播分组分别
30

划分为至少一个优先级并为每个优先级的单播分组和多播分组分别分配一个队列，并根据队列的优先级调度分组，其中，分配给最高优先级的单播分组的队列和分配给最高优先级的多播分组的队列用于承载 TDM 业务或分组业务，分配给其它优先级的单播分组的队列和分配给其它优先级的多播分组的队列用于承载分组业务。

其中，队列用于缓存分组的地址，采用先入先出方式进行操作。步骤 S1204 包括：S12042，通过单播共享缓冲器存储来自分组单元的单播分组并将为最高优先级的单播分组队列分配的单播共享缓冲空间用于 TDM 单播业务或单播分组业务，以及通过多播共享缓冲器存储来自分组单元的多播分组并将为最高优先级的多播分组队列分配的多播共享缓冲空间用于 TDM 多播业务或多播分组业务；以及 S12044，分组调度单元，用于根据队列的优先级从队列中调度分组；以及 S12046，在单播分组和多播分组被转发送指定端口后，释放单播分组和多播分组对应的共享缓冲空间。

其中，单播共享缓冲器和多播共享缓冲器的写入速度不低于交换单元的所有输入端口的分组输入速率之和，单播共享缓冲器和多播共享缓冲器的读出速度不低于交换单元的所有输出端口的分组输出速度之和。

其中，多播共享缓冲器将多播分组的存储地址复制到多个指定端口的相应多播队列，在达到预定时间时，释放多播分组存储地址。

其中，在占用的单播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在占用的多播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

其中，在单播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及在多播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

其中，通过以下方式调度相同优先级的单播业务和多播业务：优先调度单播业务，优先调度多播业务，随机调度。

综上所述,本发明的核心思想是:为最高优先级业务保留一定的全局缓冲空间,基于全局共享缓冲门限缓冲 TDM 业务的分组,基于每业务队列门限缓冲分组业务的分组。按照优先级执行端口业务的调度,优先转发最高优先级业务。使用独立的单播共享缓冲器和多播共享缓冲器,并为每端口设置独立的单播业务队列和多播业务队列,避免两者之间的相互影响。

具体地,本发明实现该核心思想的过程主要包括以下步骤:步骤一,根据本发明的分组交换装置的分组部件按照定长分组交换业务,其中,从用户接口接入的任何业务数据必须分割或填充为特定长度的定长分组后才能发送到分组交换设备的交换部件。根据本发明的分组交换装置中的交换部件将单播业务划分为一个或多个优先级别,并为每端口每优先级单播业务分配一个队列;将多播业务也划分为一个或多个优先级别,并为每端口每优先级多播业务分配一个队列。其中,队列用于排队缓冲分组的存贮地址,按照先进先出(First Input First Output,简称 FIFO)方式操作。交换部件包含单播共享缓冲器、多播共享缓冲器、分组调度、队列管理、地址管理等部件。

步骤二,单播共享缓冲器存贮单播分组;多播共享缓冲器存贮多播分组。两个缓冲器都是双端口存贮结构,一个端口用于写入分组,另一个端口用于读出分组,禁止同时读写相同的存贮地址。共享单播缓冲器和共享多播缓冲器的写入速率不低于交换设备所有输入端口的分组输入速率之和,读出速度不低于交换设备所有输出端口的分组输出速率之和,保证不会出现输入和输出堵塞。

步骤三,队列管理部件为每端口每优先级单播队列分配特定的单播共享缓冲器空间,并保留部分单播共享缓冲空间用于 TDM 单播业务;为每端口每优先级多播队列分配特定的多播共享缓冲器空间,并保留部分多播共享缓冲空间用于 TDM 多播业务。

步骤四,对于任何多播分组,队列管理部件仅在多播共享缓冲器中保留一份拷贝,将多播分组的存贮地址复制到多个指定端口的相应多播队列。队列管理部件为任何多播分组设置一个计数器,控制多播分组存贮地址的释放。仅当多播分组已经转发到所有指定的端口之后,队列管理部件才能够释放该多播分组对应的存贮空间。

步骤五,所有最高优先级的单播和多播分组具有可选的 TDM 属性,如果

当前分组包含 TDM 属性，指示当前分组承载的是 TDM 业务，否则，指示当前分组承载的是分组业务。

5 步骤六，对于单播分组业务，如果占用的单播共享缓冲空间低于分配值，允许缓冲当前输入的单播分组，否则，丢弃当前输入的单播分组。对于多播分组业务，如果占用的多播共享缓冲空间低于分配值，允许缓冲当前输入的多播分组，否则，丢弃当前输入的多播分组。

10 步骤七，对于单播 TDM 业务，如果单播共享缓冲器存在空闲存贮空间，允许缓冲当前输入的单播分组，否则，丢弃当前输入的单播分组。对于多播 TDM 业务，如果多播共享缓冲器存在空闲存贮空间，允许缓冲当前输入的 TDM 多播分组，否则，丢弃当前输入的 TDM 多播分组。

步骤八，分组调度部件根据优先级调度业务，最先调度最高优先级业务，最后调度最低优先级业务。相同优先级的单播业务和多播业务属于同一调度级别，可选择优先调度单播业务或多播业务，或者随机选择。

15 步骤九，地址管理部件管理单播共享缓冲器和多播共享缓冲器的空闲存贮地址。每存贮地址可以缓冲一个分组。

20 本发明使用了独立的单播共享缓冲器和多播共享缓冲器，并为每端口的每单播优先级业务，和每端口每多播优先级业务设置独立的队列，当交换部件转发无连接的分组业务时，如以太网媒介访问控制（Media Access Control，简称 MAC）帧，可以将未知目的地址的 MAC 帧转换为定长的多播分组转发，从而，避免了地址学习过程对单播业务的影响。另外，本发明根据特定的队列门限缓冲分组业务，根据单播共享缓冲器门限和多播共享缓冲器门限缓冲 TDM 业务，并优先调度 TDM 业务，在保证 TDM 业务转发实效性的同时，还能保证 TDM 业务的转发不受分组业务拥塞的影响。

25 以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权利要求书

1. 一种分组交换装置，其特征在于包括：

分组单元，用于将输入数据划分或填充为定长的分组；以及

- 5 交换单元，用于将来自所述分组单元的单播分组和多播分组分别划分为至少一个优先级并为每个优先级的单播分组和多播分组分别分配一个队列，并且根据队列的优先级进行调度，其中，分配给最高优先级的单播分组的队列和分配给最高优先级的多播分组的队列用于承载 TDM 业务或分组业务，分配给其它优先级的单播分组的队列和分配给其它优先级的多播分组的队列用于承载所述分组业务。

- 10 2. 根据权利要求 1 所述的分组交换装置，其特征在于，所述队列用于缓存所述分组的地址，采用先入先出方式进行操作。

3. 根据权利要求 1 所述的分组交换装置，其特征在于，所述交换单元包括：

单播共享缓冲器，用于存储来自所述分组单元的单播分组；

多播共享缓冲器，用于存储来自所述分组单元的多播分组；

- 15 分组调度单元，用于根据所述队列的优先级从所述队列中调度所述分组；

- 队列管理单元，用于分别为每个端口的单播分组队列和多播分组队列分配共享缓冲空间，并将为最高优先级的单播分组队列分配的共享缓冲空间用于 TDM 单播业务或单播分组业务，将为最高优先级的多播分组队列分配的共享缓冲空间用于 TDM 多播业务或多播分组业务，以及在所述单播分组和所述多播分组被转发送指定端口后，释放所述单播分组和所述多播分组对应的共享缓冲空间；以及
- 20

地址管理单元，用于管理所述单播共享缓冲器和所述多播共享缓冲器的空闲存储地址。

25

4. 根据权利要求 3 所述的分组交换装置, 其特征在于, 所述单播共享缓冲器和所述多播共享缓冲器的写入速度不低于所述交换单元的所有输入端口的分组输入速率之和, 所述单播共享缓冲器和所述多播共享缓冲器的读出速度不低于所述交换单元的所有输出端口的分组输出速度之和。
5. 根据权利要求 3 所述的分组交换装置, 其特征在于, 所述队列管理单元在所述多播共享缓冲器中将多播分组的存储地址复制到多个指定端口的相应多播队列, 在达到预定时间时, 释放所述多播分组存储地址。
6. 根据权利要求 4 所述的分组交换装置, 其特征在于:
在占用的单播共享缓冲空间低于分配值的情况下, 允许缓冲当前输入的单播分组, 否则丢弃当前输入的单播分组; 以及
在占用的多播共享缓冲空间低于分配值的情况下, 允许缓冲当前输入的多播分组, 否则丢弃当前输入的多播分组。
7. 根据权利要求 4 所述的分组交换装置, 其特征在于:
在所述单播共享缓冲器存在空闲存储空间的情况下, 允许缓冲当前输入的单播分组, 否则丢弃当前输入的单播分组; 以及
在所述多播共享缓冲器存在空闲存储空间的情况下, 允许缓冲当前输入的多播分组, 否则丢弃当前输入的多播分组。
8. 根据权利要求 3 所述的分组交换装置, 其特征在于, 所述队列调度单元通过以下方式调度相同优先级的单播业务和多播业务: 优先调度单播业务, 优先调度多播业务, 随机调度。
9. 根据权利要求 3 所述的分组交换装置, 其特征在于, 所述队列调度单元优先调度最高优先级队列中的分组。
10. 一种分组交换方法, 其特征在于包括:
S1202, 分组单元将输入数据划分或填充为定长的分组; 以及
S1204, 交换单元将来自所述分组单元的单播分组和多播分组分别划分为至少一个优先级并为每个优先级的单播分组和多播分组分别分配

一个队列，并根据所述队列的优先级调度所述分组，其中，分配给最高优先级的单播分组的队列和分配给最高优先级的多播分组的队列用于承载 TDM 业务或分组业务，分配给其它优先级的单播分组的队列和分配给其它优先级的多播分组的队列用于承载所述分组业务。

5 11. 根据权利要求 10 所述的分组交换方法，其特征在于，所述队列用于缓存所述分组的地址，采用先入先出方式进行操作。

12. 根据权利要求 10 所述的分组交换方法，其特征在于，所述步骤 S1204 包括：

10 S12042，通过单播共享缓冲器存储来自所述分组单元的单播分组并将为最高优先级的单播分组队列分配的单播共享缓冲空间用于 TDM 单播业务或单播分组业务，以及通过多播共享缓冲器存储来自所述分组单元的多播分组并将为最高优先级的多播分组队列分配的多播共享缓冲空间用于 TDM 多播业务或多播分组业务；以及

15 S12044，分组调度单元，用于根据所述队列的优先级从所述队列中调度所述分组；以及

S12046，在所述单播分组和所述多播分组被转发送指定端口后，释放所述单播分组和所述多播分组对应的共享缓冲空间。

13. 根据权利要求 12 所述的分组交换方法，其特征在于，所述单播共享缓冲器和所述多播共享缓冲器的写入速度不低于所述交换单元的所有输入端口的分组输入速率之和，所述单播共享缓冲器和所述多播共享缓冲器的读出速度不低于所述交换单元的所有输出端口的分组输出速度之和。

20

14. 根据权利要求 12 所述的分组交换方法，其特征在于，所述多播共享缓冲器将多播分组的存储地址复制到多个指定端口的相应多播队列，在达到预定时间时，释放所述多播分组存储地址。

25 15. 根据权利要求 12 所述的分组交换方法，其特征在于：

在占用的单播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及

在占用的多播共享缓冲空间低于分配值的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

16. 根据权利要求 12 所述的分组交换方法，其特征在于：

5 在所述单播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的单播分组，否则丢弃当前输入的单播分组；以及

在所述多播共享缓冲器存在空闲存储空间的情况下，允许缓冲当前输入的多播分组，否则丢弃当前输入的多播分组。

17. 根据权利要求 12 所述的分组交换方法，其特征在于，通过以下方式调度
10 相同优先级的单播业务和多播业务：优先调度单播业务，优先调度多播业务，随机调度。

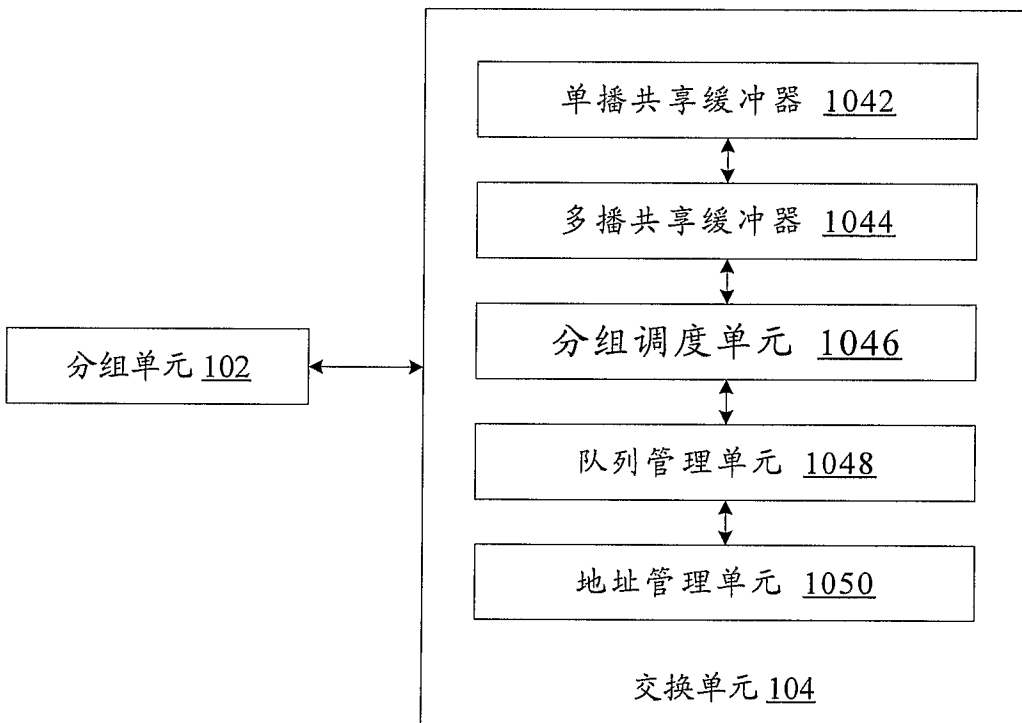


图 1

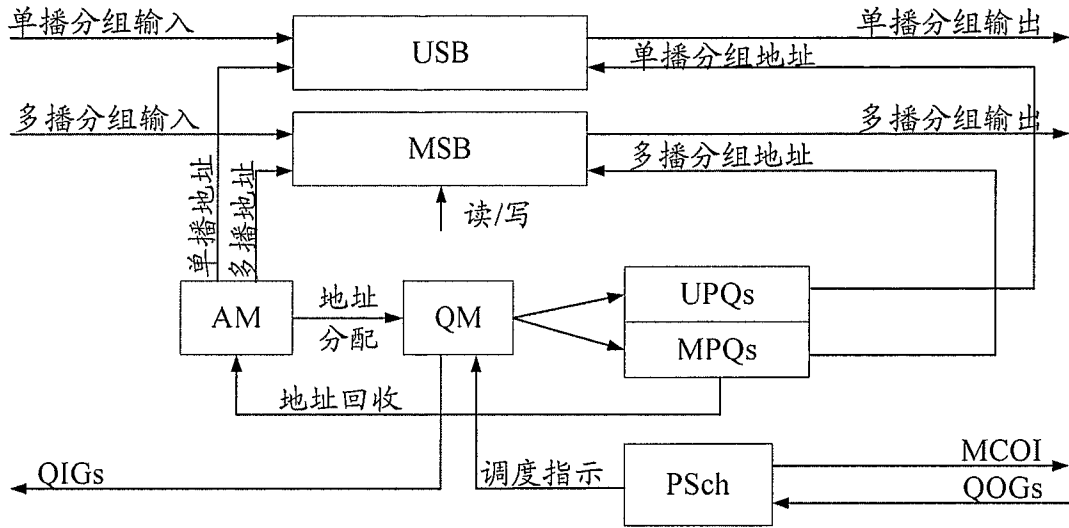


图 2

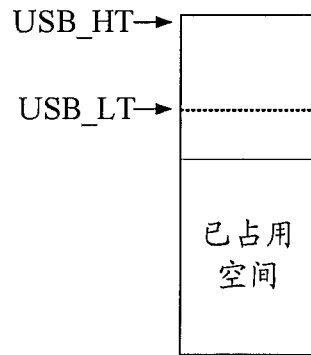


图 3

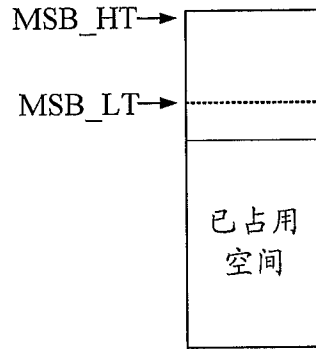


图 4

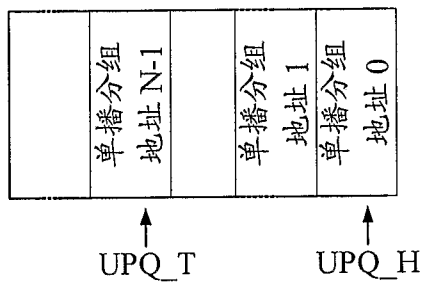


图 5

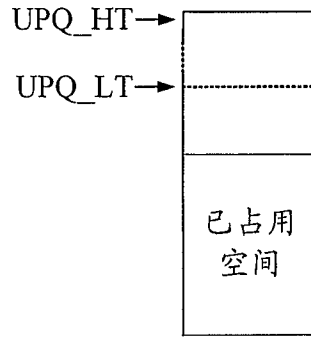


图 6

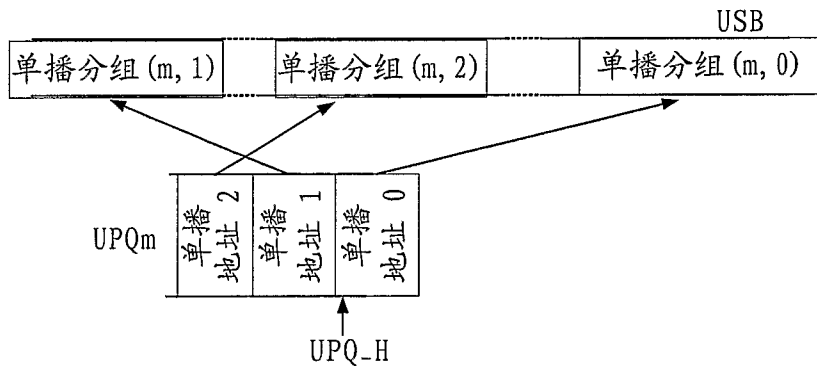


图 7

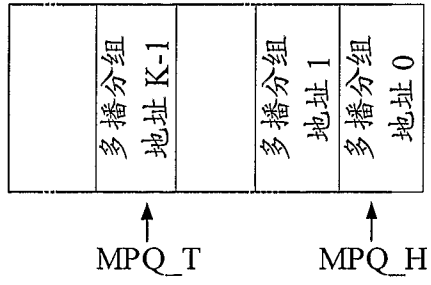


图 8

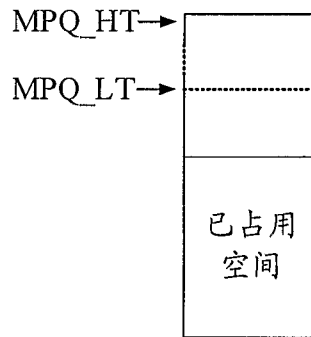


图 9

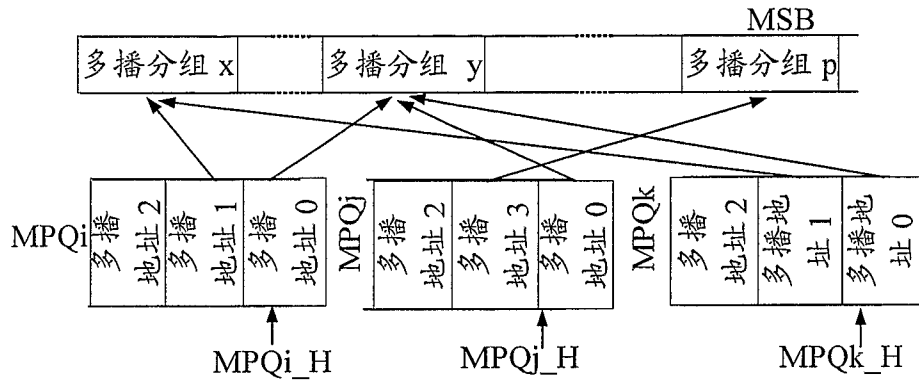


图 10

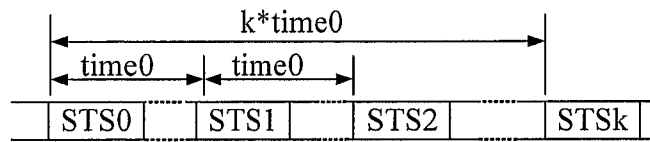


图 11

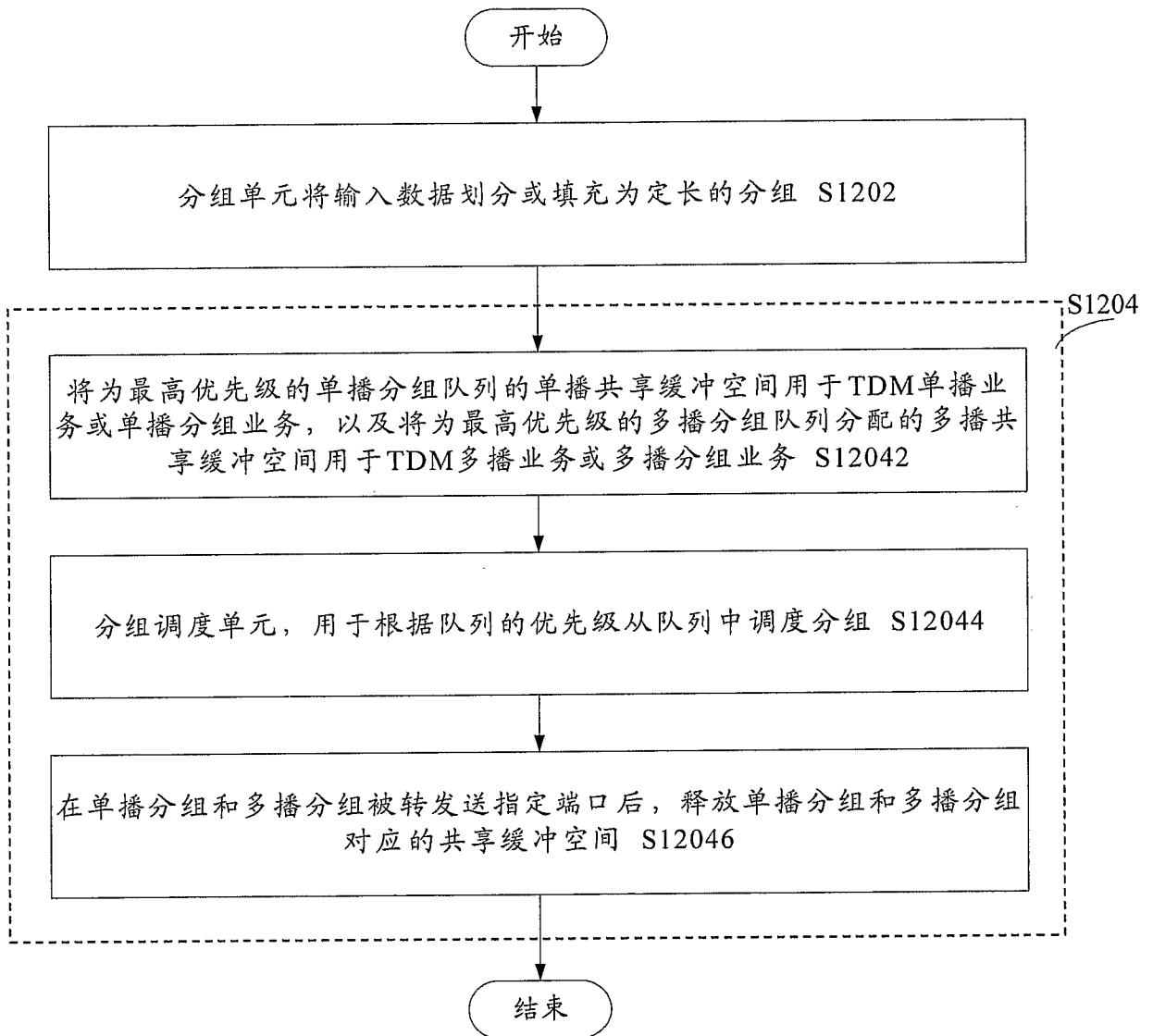



图 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2006/003526

A. CLASSIFICATION OF SUBJECT MATTER H04L 12/56 (2006.01) i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC:H04L H04Q H04M Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) style="text-align: center;">WPI, EPODOC, PAJ, CNPAT, CNKI: PACKET?, PRIORIT+, TDM, TIME, DIVISION, QUEUE?, SEQUENCE, MULTICAST+		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US20040213149A1 (ALCATEL) 28 Oct. 2004 (28.10.2004) the whole document	1-17
A	US20050094645A1 (ALCATEL) 05 May 2005 (05.05.2005) the whole document	1-17
A	WO2004066570A1 (FUJITSU LTD) 28 Aug. 2004 (05.08.2004) the whole document	1-17
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search 12 Sep. 2007 (12.09.2007)	Date of mailing of the international search report 27 SEP. 2007 (27.09.2007)	
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer  Telephone No. (86-10)62086068	

INTERNATIONAL SEARCH REPORT

Information patent family members

Search request No.
PCT/CN2006/003526

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US20040213149A1	28.10.2004	EP1471673A1	27.10.2004
		CN1540943A	27.10.2004
		EP1471673B1	02.08.2006
		DE60307230E	14.09.2006
		AT335323T	15.08.2006
US20050094645A1	05.05.2005	EP1528729A2	04.05.2005
		CN1614956A	11.05.2005
WO2004066570A1	05.08.2004	AU2003303790A1	13.08.2004
		TW595170A	21.06.2004
		US2005129044A1	16.06.2005
		TW200414720A	01.08.2004
		JP2004567109T	18.05.2006
		KR20050050110A	27.05.2005
		CN1729655A	01.02.2006
TW595170BB	21.06.2004		

国际检索报告

国际申请号
PCT/CN2006/003526

A. 主题的分类

H04L 12/56 (2006.01) i

按照国际专利分类表(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: H04L H04Q H04M

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

WPI, EPODOC, PAJ, CNPAT, CNKI:

PACKET?, PRIORIT+, TDM, TIME, DIVISION, QUEUE?, SEQUENCE, MULTICAST+, 分组, 包, 优先级, 队列, 序列, 多播

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	US20040213149A1 (阿尔卡特公司) 28.10 月 2004 年 (28.10.2004) 全文	1-17
A	US20050094645A1 (阿尔卡特公司) 05.5 月 2005 年 (05.05.2005) 全文	1-17
A	WO2004066570A1 (富士通株式会社) 05.8 月 2004 年 (05.08.2004) 全文	1-17

其余文件在 C 栏的续页中列出。

见同族专利附件。

* 引用文件的具体类型:

- “A” 认为不特别相关的表示了现有技术一般状态的文件
- “E” 在国际申请日的当天或之后公布的在先申请或专利
- “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件
- “O” 涉及口头公开、使用、展览或其他方式公开的文件
- “P” 公布日先于国际申请日但迟于所要求的优先权日的文件

- “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件
- “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性
- “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性
- “&” 同族专利的文件

国际检索实际完成的日期
12.9 月 2007 (12.09.2007)

国际检索报告邮寄日期
27. 9 月 2007 (27.09.2007)

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路 6 号 100088
传真号: (86-10)62019451

受权官员



电话号码: (86-10)62086068

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2006/003526

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
US20040213149A1	28.10.2004	EP1471673A1	27.10.2004
		CN1540943A	27.10.2004
		EP1471673B1	02.08.2006
		DE60307230E	14.09.2006
		AT335323T	15.08.2006
US20050094645A1	05.05.2005	EP1528729A2	04.05.2005
		CN1614956A	11.05.2005
WO2004066570A1	05.08.2004	AU2003303790A1	13.08.2004
		TW595170A	21.06.2004
		US2005129044A1	16.06.2005
		TW200414720A	01.08.2004
		JP2004567109T	18.05.2006
		KR20050050110A	27.05.2005
		CN1729655A	01.02.2006
		TW595170BB	21.06.2004