

(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
2016年6月2日 (02.06.2016)

(10) 国际公布号
WO 2016/082169 A1

- (51) 国际专利分类号: H04L 12/911 (2013.01) H04L 29/12 (2006.01)
- (21) 国际申请号: PCT/CN2014/092421
- (22) 国际申请日: 2014年11月28日 (28.11.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 陶怡栋 (TAO, Yidong); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。何睿 (HE, Rui); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。董晓文 (DONG, Xiaowen); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (74) 代理人: 北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (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, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

(54) Title: MEMORY ACCESS METHOD, SWITCH AND MULTI-PROCESSOR SYSTEM

(54) 发明名称: 内存访问方法、交换机及多处理器系统

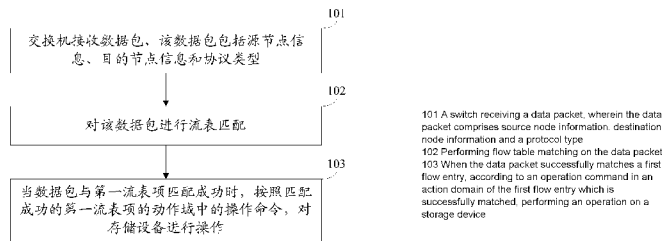


图 2 /Fig 2

(57) Abstract: Provided are a memory access method, a switch and a multi-processor system. The memory access method comprises: a switch receiving a data packet; performing flow table matching on the data packet, wherein the flow table comprises at least one flow entry, the flow entry comprises a matching domain and an action domain, the at least one flow entry comprises a first flow entry, the matching domain of the first flow entry is used for matching source node information, destination node information and a protocol type in the data packet, and the action domain of the first flow entry is used for indicating an operation command for a storage device built in the switch; and when the data packet successfully matches the first flow entry, according to the operation command in the action domain of the first flow entry which is successfully matched, performing an operation on the storage device. By building a storage device in a switch which is accessed by a processor, the present invention can reduce or even avoid accessing a far-end memory by a processor in a multi-processor system, thus reducing a memory access time delay.

(57) 摘要: 本发明实施例提供了一种内存访问方法、交换机及多处理器系统。该内存访问方法包括: 交换机接收数据包; 对数据包进行流表匹配, 所述流表包括至少一个流表项, 流表项包括匹配域和动作域, 至少一个流表项包括第一流表项, 第一流表项的匹配域用于匹配数据包中的源节点信息、目的节点信息和协议类型, 第一流表项的动作域用于指示对交换机内置的存储设备的操作命令; 当数据包与第一流表项匹配成功时, 按照匹配成功的第一流表项的动作域中的操作命令, 对所述存储设备进行操作。本发明通过在交换机中内置存储设备, 处理器对其进行访问, 可以减少甚至避免多处理器系统中处理器访问远端内存, 进而减少了内存访问延时。

WO 2016/082169 A1

说明书

内存访问方法、交换机及多处理器系统

技术领域

5 本发明涉及计算机技术领域，特别涉及一种内存访问方法、交换机及多处理器系统。

背景技术

10 随着人类对计算机速度和计算规模需求不断提高，多处理器系统应运而生。在多处理器系统中，多个处理器通过互连网络进行通信。互连网络通常由多个交换机构成，该互连网络连接负责计算的处理器，同时也可以连接负责存储的内存。当处理器需要访问内存时，请求经过互连网络转发至内存，然而，随着处理器和内存的数量越来越多，互连网络的规模也越来越大，处理器访问远端内存时的访问延时也随之增大，从而导致系统性能的下降。

15 现有技术中提出了一种减少处理器访问远端内存（即连接在交换机端口上的内存）时的访问延时的方法，在该方法中，互连网络中的交换机内均带有缓存（Cache）功能，从而可以缓存一部分内存的数据。当处理器需要访问的数据存在于交换机中时，可以直接由该交换机中的缓存返回数据，而不必访问远端的内存，进而减少了访问延时。

20 在实现本发明的过程中，发明人发现现有技术至少存在以下问题：

各个交换机中都带有 Cache，各个 Cache 中缓存的数据可能包括共享数据，即被多个处理器使用的数据，当某个交换机的 Cache 内的共享数据被修改且该共享数据在其他的交换机的 Cache 中存有副本时，如果其他的交换机中的副本不能得到及时的修改，这时该数据若被其他处理器访问就会出现错误。因此，
25 为了避免出现处理器错误，必须保证 Cache 内数据的一致性，而 Cache 一致性的维护通常十分复杂。

发明内容

30 为了解决现有技术中采用交换机内的 Cache 减少访问延时时存在的 Cache 一致性维护困难的问题，本发明实施例提供了一种内存访问方法、交换机及多

处理器系统。所述技术方案如下：

第一方面，本发明实施例提供了一种内存访问方法，所述方法包括：

交换机接收数据包，所述数据包包括源节点信息、目的节点信息和协议类型，所述协议类型用于指示所述数据包的类型；

5 对所述数据包进行流表匹配，所述流表包括至少一个流表项，所述流表项包括匹配域和动作域，所述至少一个流表项包括第一流表项，所述第一流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第一流表项的动作域用于指示对所述交换机内置的存储设备的操作命令；

10 当所述数据包与所述第一流表项匹配成功时，按照匹配成功的所述第一流表项的动作域中的操作命令，对所述存储设备进行操作。

在第一方面的第一种可能的实现方式中，所述按照匹配成功的第一流表项的动作域中的操作命令，对所述存储设备进行操作，包括：

15 当匹配成功的第一流表项的动作域中的所述操作命令为读操作命令时，从所述存储设备中读取数据，并将读取的所述数据返回给所述源节点信息对应的节点；

当匹配成功的第一流表项的动作域中的所述操作命令为写操作命令时，将所述数据包中的数据写入所述存储设备中。

20 在第一方面的第二种可能的实施方式中，所述至少一个流表项还包括第二流表项，所述第二流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第二流表项的动作域用于指示对所述数据包中的数据

进一步地，在该第二种可能的实施方式中，所述方法还包括：

25 当所述数据包与所述第二流表项匹配成功时，按照匹配成功的所述第二流表项的动作域中的操作命令，对所述数据包中的数据进行计算处理，得到计算结果；

将所述计算结果发送给所述数据包中的所述源节点信息对应的节点。

在第一方面的第三种可能的实施方式中，所述方法还可以包括：

接收控制器发送的流表配置消息，所述流表配置消息用于为所述交换机配置所述流表项；

30 根据所述流表配置消息配置所述流表项。

第二方面，本发明实施例提供了一种交换机，所述交换机包括：

第一接收模块，用于接收数据包，所述数据包包括源节点信息、目的节点信息和协议类型，所述协议类型用于指示所述数据包的类型；

5 匹配模块，用于对所述第一接收模块接收的所述数据包进行流表匹配，所述流表包括至少一个流表项，所述流表项包括匹配域和动作域，所述至少一个流表项包括第一流表项，所述第一流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第一流表项的动作域用于指示对所述交换机内置的存储设备的操作命令；

10 操作模块，用于当所述数据包与所述第一流表项匹配成功时，按照匹配成功的所述第一流表项的动作域中的操作命令，对所述存储设备进行操作。

在第二方面的第一种可能的实施方式中，所述操作模块包括：

读取单元，用于当所述操作命令为读操作命令时，从所述存储设备中读取数据；

15 发送单元，用于将所述读取单元读取的所述数据返回给所述源节点信息对应的节点；

写入单元，用于当所述操作命令为写操作命令时，将所述数据包中的数据写入所述存储设备中。

20 在第二方面的第二种可能的实施方式中，所述至少一个流表项还包括第二流表项，所述第二流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第二流表项的动作域用于指示对所述数据包中的数据

进一步地，在该第二种可能的实施方式中，所述交换机还包括：

25 处理模块，用于当所述数据包与所述第二流表项匹配成功时，按照匹配成功的所述第二流表项的动作域中的操作命令，对所述数据包中的数据进行计算处理，得到计算结果；

发送模块，用于将所述处理模块得到的所述计算结果发送给所述数据包中的所述源节点信息对应的节点。

在第二方面的第三种可能的实施方式中，所述交换机还包括：

30 第二接收模块，用于接收控制器发送的流表配置消息，所述流表配置消息用于为所述交换机配置所述流表项；

配置模块，用于根据所述第二接收模块收到的所述流表配置消息配置所述

流表项。

第三方面，本发明实施例提供了一种交换机，所述交换机包括：处理器、存储器、总线和通信接口；所述存储器用于存储计算机执行指令，所述处理器与所述存储器通过所述总线连接，当所述计算机运行时，所述处理器执行所述
5 存储器存储的所述计算机执行指令，以使所述交换机执行前述第一方面提供的方法。

第四方面，本发明实施例提供了一种交换机，所述交换机包括：

10 输入端口，用于接收数据包，所述数据包包括源节点信息、目的节点信息和协议类型，所述协议类型用于指示所述数据包的类型；

存储器，用于存储流表，所述流表包括至少一个流表项，所述流表项包括匹配域和动作域，所述至少一个流表项包括第一流表项，所述第一流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第一流表项的动作域用于指示对所述交换机内置的存储设备的操作命令；

15 存储设备，用于存储数据；

查表逻辑电路，用于采用所述存储器存储的流表对所述输入端口接收的数据包进行流表匹配；

操作逻辑电路，用于当所述数据包与所述第一流表项匹配成功时，按照匹配成功的所述第一流表项的动作域中的操作命令，对所述存储设备进行操作；

20 纵横开关总线，用于为所述操作逻辑电路传输的数据包选择输出端口；

输出端口，用于发送所述纵横开关总线传输的数据包。

第五方面，本发明实施例提供了一种多处理器系统，所述多处理器系统包括：多个处理器和互连网络，所述多个处理器通过所述互连网络通信连接，所
25 述互连网络包括多个交换机，所述交换机包括前述第二方面或第三方面或第四方面提供的交换机。

在第五方面的第一种可能的实施方式中，所述多处理器系统还包括多个外部存储设备，所述多个外部存储设备通过所述互连网络与所述多个处理器通信连接。

30 在第五方面的第二种可能的实施方式中，所述多处理器系统为片上系统。
本发明实施例提供的技术方案的有益效果是：通过在交换机中内置存储设

备，并对接收到的数据包进行流表匹配，当数据包与流表中的第一流表项匹配成功时，根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作，从而减少甚至避免了对处理器对远端内存的访问，可以减少内存访问的延时，同时，由于采用各交换机内的存储设备单独存储数据，不存在在其他交换机的存储设备中存在副本的情况，因此不存在需要维护 Cache 一致性的问题，实现简单。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 是多处理器系统的结构示意图；

图 2 是本发明实施例一提供的内存访问方法的流程图；

15 图 3 是本发明实施例二提供的内存访问方法的流程图；

图 4 是本发明实施例二提供的内存访问方法的访问实例示意图；

图 5 是本发明实施例三提供的交换机的结构框图；

图 6 是本发明实施例四提供的交换机的结构框图；

图 7 是本发明实施例五提供的交换机的结构框图；

20 图 8 是本实施例五提供的交换机的具体实现结构图；

图 9 是本发明实施例六提供的交换机的硬件结构图；

图 10 是本发明实施例七提供的多处理器系统的结构框图。

具体实施方式

25 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

本发明实施例提供了一种内存访问方法、交换机和多处理器系统。下面先结合图 1 介绍多处理器系统的网络架构。

参见图 1，在多处理器系统中，多个处理器 2 通过互连网络 1 相互连接，互连网络包括多个交换机，这些交换机负责转发处理器 2 之间的通信数据。同时，该多处理器系统还可以包括多个独立的存储设备 3，存储设备 3 通过该互

连网络 1 与处理器 2 连接，因此，该互连网络 1 中的交换机还负责转发处理器 2 对存储设备 3 的访问请求以及存储设备 3 向处理器 2 返回的响应消息等。

容易知道，以上多处理器系统的网络架构仅为举例，并不作为对本发明实施例的限制，例如，该处理器系统中也可以不包括存储设备 3。

5

实施例一

本发明实施例提供了一种内存访问方法，适用于前述多处理器系统，该方法可以由交换机执行，在具体实现中，交换机可以为开放流交换机（OpenFlow Switch，简称 OFS），也可以为其他具有匹配功能的交换机。如图 2 所示，该方法包括：

步骤 101：交换机接收数据包，该数据包包括源节点信息、目的节点信息和协议类型。

其中，源节点信息可以为源节点标识、源节点的 MAC 地址等；目的节点信息可以为目的节点标识、目的节点的 MAC 地址等；协议类型用于指示数据包的类型，例如读请求数据包、写请求数据包、计算请求数据包等。

在本发明实施例中，交换机中内置有存储设备，该存储设备可以为静态随机存取存储器（Static Random Access Memory，简称 SRAM）、动态随机存取存储器（Dynamic Random Access Memory，简称 DRAM）等。

容易知道，OFS 中设有流表，流表包括至少一个流表项，每个流表项都包括匹配域和动作域。流表中的流表项通常包括转发流表项，该转发流表项用于确定数据包的转发出口。流表中的流表项还可以包括数据包修改流表项，该数据包修改流表项用于确定对数据包内的信息进行修改，例如修改数据包的头域等。

步骤 102：对该数据包进行流表匹配。

如前所述，交换机中的流表可以包括多种表项，而在本发明实施例中，流表项还包括第一流表项，第一流表项的匹配域用于匹配数据包中的源节点信息、目的节点信息和协议类型，第一流表项的动作域用于指示对交换机内置的存储设备的操作命令。该操作命令包括但不限于读操作命令、写操作命令。

容易知道，第一流表项的动作域中的操作命令与其匹配域中的协议类型通常是相互对应的，例如，如果第一流表项的匹配域中的协议类型用于指示读请求数据包，则该第一流表项的动作域中的操作命令为对交换机内置的存储设备

的读操作命令。

在实现时，各种流表项（例如前述第一流表项、转发流表项、数据包修改流表项）可以配置在一个流表中，也可以按照流表项的功能，将不同的流表项配置在不同的流表中。

5 各种流表通常设置在交换机的三态内容寻址存储器（Ternary Content Addressable Memory，简称 TCAM）或低延迟动态随机存取存储器（Reduced Latency Dynamic Random Access Memory，简称 RLDRAM）中。

容易知道，交换机中的流表项可以由开放流控制器（OpenFlow Controller，简称 OFC）配置。

10 步骤 103：当数据包与第一流表项匹配成功时，按照匹配成功的第一流表项的动作域中的操作命令，对存储设备进行操作。

具体地，该步骤 103 可以包括：

当匹配成功的第一流表项的动作域中的操作命令为读操作命令时，从存储设备中读取数据，并将读取的数据返回给源节点信息对应的节点；

15 当匹配成功的第一流表项的动作域中的操作命令为写操作命令时，将数据包中的数据写入存储设备中。

本发明实施例通过在交换机中内置存储设备，并对接收到的数据包进行流表匹配，当数据包与流表中的第一流表项匹配成功时，根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作，从而减少甚至避免了对处理器对远端内存的访问，可以减少内存访问的延时，同时，由于采用各交换机内的存储设备单独存储数据，不存在在其他交换机的存储设备中存在副本的情况，因此不存在需要维护 Cache 一致性的问题，实现简单。并且，当第一流表项匹配成功时，由于各个交换机直接对自身内部的存储设备进行访问，所以对于一个访问请求而言，只需要出入互连网络一次（即接收访问请求和返回
25 响应消息），从而可以节省网络资源。

实施例二

本发明实施例提供了一种内存访问方法，可以应用于前述多处理器系统。其中，交换机中设有流表，流表中包括至少一条流表项，每条流表项均包括匹
30 配域和动作域，其中，匹配域用于匹配交换机接收到的数据包中的信息，动作域用于指示当流表项匹配成功时对数据包进行何种操作。通常，流表中的流表

项通常包括转发流表项，该转发流表项用于确定数据包的转发出口。流表中的流表项还可以包括数据包修改流表项，该数据包修改流表项用于确定对数据包内的信息进行修改，例如修改数据包的头域等。同时，交换机中还内置有存储设备（也可称内存），例如 SRAM 或 DRAM 等。在实现时，交换机可以为 OFS，
5 也可以为其他具有匹配功能的交换机。

下面结合图 3 对本实施例的方法进行详细说明，如图 3 所示，该方法包括：

步骤 201：交换机接收处理器发送的数据包，该数据包包括源节点信息、目的节点信息和协议类型。

其中，源节点信息可以为源节点标识、源节点的 MAC 地址等；目的节点
10 信息可以为目的节点标识、目的节点的 MAC 地址等；协议类型用于指示数据包的类型，例如读请求数据包、写请求数据包、计算请求数据包等。

该数据包中还可以包括操作地址，例如读写地址。可以理解地，当数据包为写数据包时，该数据包中还包括待写入的数据。

步骤 202：交换机对该数据包进行流表匹配。当数据包与第一流表项匹配
15 成功时，执行步骤 203；当数据包与第二流表项匹配成功时，执行步骤 204 和步骤 205。

如前所述，交换机中的流表可以包括多种表项，需要强调的是，在本发明
20 实施例中，流表项还包括第一流表项，第一流表项的匹配域用于匹配数据包中的源节点信息、目的节点信息和协议类型，第一流表项的动作域用于指示对交换机内置的存储设备的操作命令。该操作命令包括但不限于读操作命令、写操作命令。

容易知道，第一流表项的动作域中的操作命令与其匹配域中的协议类型
25 通常是相互对应的，例如，如果第一流表项的匹配域中的协议类型用于指示读请求数据包，则该第一流表项的动作域中的操作命令为对交换机内置的存储设备的读操作命令。

此外，在本实施例中，流表项还可以包括第二流表项，第二流表项的匹配
30 域用于匹配数据包中的源节点信息、目的节点信息和协议类型，第二流表项的动作域用于指示对数据包中的数据进行处理的操作命令，该计算处理包括但不限于循环冗余校验（Cyclic Redundancy Check，简称 CRC）或快速傅立叶变换（Fast Fourier Transformation，简称 FFT）。

在实现时，各种流表项（例如前述第一流表项、转发流表项、数据包修改

流表项)可以配置在一个流表中,也可以按照流表项的功能,将不同的流表项配置在不同的流表中,例如,将第一流表项配置在一个流表中,将转发流表项配置在一个流表中,本发明实施例对此不做限制。

各种流表可以设置在交换机的 TCAM 或 RLDRAM 中。

5 容易知道,该步骤 202 中,会对收到的数据包与每一条流表项进行匹配。

进一步地,在具体实现中,一个交换机中的流表还可以根据交换机的端口来设置,例如,可以交换机的一个端口对应一组流表,每组流表可以包括一个流表(该流表中包括所有种类的流表项),每组流表也可以包括多个流表(该多个流表分别包括不同种类的流表项),从一个端口接收的数据包仅在该端口
10 对应的流表中进行匹配;又例如,可以交换机中仅配置一组流表,所有端口接收到的数据包均在该组流表中进行匹配。

步骤 203: 交换机按照匹配成功的第一流表项的动作域中的操作命令,对存储设备进行操作。

具体地,该步骤 203 可以包括:

15 当匹配成功的第一流表项的动作域中的操作命令为读操作命令时,从存储设备中读取数据,并将读取的数据返回给源节点信息对应的节点;

当匹配成功的第一流表项的动作域中的操作命令为写操作命令时,将数据包中的数据写入存储设备中。

下面结合图 4 对本实施例的具体应用场景以及交换机的具体操作流程进行
20 详细说明。

处理器 04 和处理器 42 共同运行一应用程序,两者交互,确定处理器 42 运行该应用程序需要的数据需要从处理器 04 获取。该应用程序向控制器申请共享存储空间(该共享存储空间即为交换机内置的存储设备的存储空间),以便于处理器 04 能将处理器 42 需要的数据写入共享存储空间,供处理器 42
25 访问。该控制器在系统建立时,即获知所有交换机中的存储设备,并对所有交换机中的存储设备进行管理。

控制器为该应用程序分配了位于交换机 11 上的存储空间,会将分配的存储空间的地址告知处理器 04 和处理器 42(例如,会发送存储空间分配信息,存储空间分配信息可以包括交换机标识、以及存储空间地址),并向交换机
30 11 下发用于配置第一流表项的流表配置信息,配置的第一流表项至少包括两条,一条的匹配域包括源节点标识(04)、目的节点标识(11)和协议类型(WR),

对应的动作域为写操作命令 (MEM WR), 另一条的匹配域包括源节点标识 (xx)、目的节点标识 (11) 和协议类型 (RD), 对应的动作域为读操作命令 (MEM RD)。

此外, 控制器还会向其他交换机 (例如交换机 00、01 等) 发送用于配置
5 转发流表项的流表配置信息, 该转发流表项用于指示将处理器 04 和处理器 42 发送给交换机 11 的数据包转发给交换机 11。

处理器 04 获得控制器分配的存储空间后, 如图 4 所示, 处理器 04 发送写
请求数据包②至互连网络中的交换机 00, 该写请求数据包包括源节点标识
(04)、目的节点标识 (11)、协议类型 (WR, 表示写请求)、写入地址 (00)
10 和待写入的数据 (xx), 该待写入的数据 (xx) 即为处理器 42 运行该应用程序
需要的数据。交换机 00 将该写请求数据包转发至交换机 01, 交换机 01 将该写
请求数据包转发至交换机 11 (即目的节点)。容易知道, 交换机 00 和交换机
01 收到该写请求数据包后, 对其进行流表匹配, 根据匹配成功的转发流表项中
的动作域对写请求数据包进行转发。

如前所述, 交换机 11 中的流表①包括两条第一流表项, 其中第一条流表
项的匹配域包括源节点标识 (04)、目的节点标识 (11) 和协议类型 (WR),
对应的动作域为写操作命令 (MEM WR), 其中第二条流表项的匹配域包括源
节点标识 (xx)、目的节点标识 (11) 和协议类型 (RD), 对应的动作域为读
操作命令 (MEM RD)。交换机 11 对写请求数据包进行流表匹配后, 该写请求
20 数据包可以与第一条流表项匹配成功, 因此, 对交换机 11 中的存储设备执行
写操作命令, 将待写入的数据 xx 写入交换机 11 中的存储设备中的地址 00。

类似地, 处理器 42 发送读请求数据包③至互连网络中的交换机 02, 该读
请求数据包包括源节点标识 (42)、目的节点标识 (11)、协议类型 (RD, 表
示读请求) 和读取地址 (00), 交换机 02 将该读请求数据包转发至交换机 12,
25 交换机 12 将该读请求数据包转发至交换机 11 (即目的节点)。容易知道, 交换
机 02 和交换机 12 收到该读请求数据包后, 对其进行流表匹配, 根据匹配成功
的转发流表项中的动作域对读请求数据包进行转发。

交换机 11 对读请求数据包进行流表匹配后, 该读请求数据包可以与第二
条流表项匹配成功, 因此, 对交换机 11 中的存储设备执行读操作命令, 从交
30 换机 11 中的存储设备中的地址 00 读取数据 xx。并生成响应数据包④, 该响应
数据包包括源节点标识 (11), 目的节点标识 42, 协议类型 (RD RLY, 表示读

响应), 读取地址 (00) 和读出的数据 (xx), 该响应数据包经由交换机 12、交换机 02 返回给处理器 42。

从该读请求数据包的处理流程可以看出, 与现有的访问外部存储设备的读请求流程相比 (处理器发送的读请求数据包先进入互连网络, 然后由互连网络
5 转发至外部存储设备, 然后外部存储设备将读响应数据包发送至互连网络, 由互连网络转发至处理器, 一个访问过程需要有数据包进出互连网络各两次), 本实施例的一个访问过程只需要数据包进出互连网络一次, 所以本发明实施例的内存访问方法可以节省网络资源。

需要说明的是, 前述应用场景仅为举例, 并不以此为限, 本发明实施例还
10 适用于以下场景: 控制器先为应用分配了共享存储空间, 然后将相应的应用分配给处理器处理, 这时分配到处理器上的应用事先就已经知道有共享存储空间可以使用, 从而能够直接对该共享存储空间进行访问。

步骤 204: 交换机按照匹配成功的第二流表项的动作域中的操作命令, 对数据包中的数据进行计算处理, 得到计算结果。

15 在实现时, 可以在交换机中设置计算模块, 也可以设置一个专门的计算设备, 交换机将该数据发送给该专门的计算设备, 并接收计算设备返回的计算结果。

步骤 205: 交换机将计算结果发送给处理器。

该处理器即为数据包中的源节点信息对应的节点。

20 在本实施例中, 即将计算结果发送给步骤 201 中的处理器。

容易知道, 交换机中的流表项可以由 OFC 配置, 所以该方法还可以包括: 接收 OFC 发送的流表配置消息, 该流表配置消息用于为该交换机配置流表项;

根据该流表配置消息配置流表项。

25 本发明实施例通过在交换机中内置存储设备, 并对接收到的数据包进行流表匹配, 当数据包与流表中的第一流表项匹配成功时, 根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作, 从而减少甚至避免了对处理器对远端内存的访问, 可以减少内存访问的延时, 同时, 由于采用各交换机内的存储设备单独存储数据, 不存在在其他交换机的存储设备中存在副本
30 的情况, 因此不存在需要维护 Cache 一致性的问题, 实现简单。并且, 当第一流表项匹配成功时, 由于各个交换机直接对自身内部的存储设备进行访问, 所

以对于一个访问请求而言，只需要出入互连网络一次（即接收访问请求和返回响应消息），从而可以节省网络资源。此外，本实施例中，交换机中还配置有第二流表项，当数据包与第二流表项匹配时，可以对数据包中的数据进行计算处理，增强了网络的硬件计算能力。

5

实施例三

本发明实施例提供了一种交换机，其中设有流表，流表中包括至少一条流表项，每条流表项均包括匹配域和动作域，其中，匹配域用于匹配交换机接收到的数据包中的信息，动作域用于指示当流表项匹配成功时对数据包进行何种操作。通常，流表中的流表项通常包括转发流表项，该转发流表项用于确定数据包的转发出口。流表中的流表项还可以包括数据包修改流表项，该数据包修改流表项用于确定对数据包内的信息进行修改，例如修改数据包的头域等。同时，交换机中还内置有存储设备，例如 SRAM 或 DRAM 等。在实现时，交换机可以为 OFS。

15 参见图 5，该交换机包括：第一接收模块 301、匹配模块 302 和操作模块 303。

其中，第一接收模块 301 用于接收数据包，该数据包包括源节点信息、目的节点信息和协议类型，该协议类型用于指示数据包的类型。其中，源节点信息可以为源节点标识、源节点的 MAC 地址等；目的节点信息可以为目的节点标识、目的节点的 MAC 地址等；协议类型用于指示数据包的类型，例如读请求数据包、写请求数据包、计算请求数据包等。

25 匹配模块 302 用于对第一接收模块 301 接收的数据包进行流表匹配，该流表包括至少一个流表项，该流表项包括匹配域和动作域，至少一个流表项包括第一流表项，第一流表项的匹配域用于匹配数据包中的源节点信息、目的节点信息和协议类型，第一流表项的动作域用于指示对交换机内置的存储设备的操作命令。该操作命令包括但不限于读操作命令、写操作命令。

容易知道，第一流表项的动作域中的操作命令与其匹配域中的协议类型通常是相互对应的，例如，如果第一流表项的匹配域中的协议类型用于指示读请求数据包，则该第一流表项的动作域中的操作命令为对交换机内置的存储设备的读操作命令。

30 在实现时，各种流表项（例如前述第一流表项、转发流表项、数据包修改

流表项)可以配置在一个流表中,也可以按照流表项的功能,将不同的流表项配置在不同的流表中。

各种流表通常设置在交换机的三态内容寻址存储器(Ternary Content Addressable Memory,简称 TCAM)或低延迟动态随机存取存储器(Reduced Latency Dynamic Random Access Memory,简称 RLDRAM)中。

操作模块 303 用于当数据包与第一流表项匹配成功时,按照匹配成功的第一流表项的动作域中的操作命令,对存储设备进行操作。

本发明实施例通过在交换机中内置存储设备,并对接收到的数据包进行流表匹配,当数据包与流表中的第一流表项匹配成功时,根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作,从而减少甚至避免了对处理器对远端内存的访问,可以减少内存访问的延时,同时,由于采用各交换机内的存储设备单独存储数据,不存在在其他交换机的存储设备中存在副本的情况,因此不存在需要维护 Cache 一致性的问题,实现简单。并且,当第一流表项匹配成功时,由于各个交换机直接对自身内部的存储设备进行访问,所以对于一个访问请求而言,只需要出入互连网络一次(即接收访问请求和返回响应消息),从而可以节省网络资源。

实施例四

本发明实施例提供了一种交换机,其中设有流表,流表中包括至少一条流表项,每条流表项均包括匹配域和动作域,其中,匹配域用于匹配交换机接收到的数据包中的信息,动作域用于指示当流表项匹配成功时对数据包进行何种操作。通常,流表中的流表项通常包括转发流表项,该转发流表项用于确定数据包的转发出口。流表中的流表项还可以包括数据包修改流表项,该数据包修改流表项用于确定对数据包内的信息进行修改,例如修改数据包的头域等。同时,交换机中还内置有存储设备,例如 SRAM 或 DRAM 等。在实现时,交换机可以为 OFS,也可以为其他具有匹配功能的交换机。

参见图 6,该交换机包括:第一接收模块 401、匹配模块 402 和操作模块 403。

其中,第一接收模块 401 用于接收数据包,该数据包包括源节点信息、目的节点信息和协议类型,该协议类型用于指示数据包的类型。匹配模块 402 用于对第一接收模块 401 接收的数据包进行流表匹配,该流表包括至少一个流表

项，该流表项包括匹配域和动作域，至少一个流表项包括第一流表项，第一流表项的匹配域用于匹配数据包中的源节点信息、目的节点信息和协议类型，第一流表项的动作域用于指示对交换机内置的存储设备的操作命令。操作模块 403 用于当数据包与第一流表项匹配成功时，按照匹配成功的第一流表项的动作域中的操作命令，对存储设备进行操作。

其中，源节点信息可以为源节点标识、源节点的 MAC 地址等；目的节点信息可以为目的节点标识、目的节点的 MAC 地址等；协议类型用于指示数据包的类型，例如读请求数据包、写请求数据包、计算请求数据包等。

该数据包中还可以包括操作地址，例如读写地址。可以理解地，当数据包为写数据包时，该数据包中还包括待写入的数据。

在本实施例的一种实现方式中，该操作模块 402 可以包括：

读取单元，用于当第一流表项中的操作命令为读操作命令时，从存储设备中读取数据；

发送单元，用于将读取单元读取的数据返回给源节点信息对应的节点；

写入单元，用于当第一流表项中的操作命令为写操作命令时，将数据包中的数据写入存储设备中。

第一流表项中的操作命令包括但不限于读操作命令、写操作命令。容易知道，第一流表项的动作域中的操作命令与其匹配域中的协议类型通常是相互对应的，例如，如果第一流表项的匹配域中的协议类型用于指示读请求数据包，则该第一流表项的动作域中的操作命令为对交换机内置的存储设备的读操作命令。

在本实施例中，交换机中的流表项还包括第二流表项，第二流表项的匹配域用于匹配数据包中的源节点信息、目的节点信息和协议类型，第二流表项的动作域用于指示对数据包中的数据计算处理的操作命令。该计算处理包括但不限于 CRC 或 FFT。

在实现时，各种流表项（例如前述第一流表项、转发流表项、数据包修改流表项）可以配置在一个流表中，也可以按照流表项的功能，将不同的流表项配置在不同的流表中，例如，将第一流表项配置在一个流表中，将转发流表项配置在一个流表中，本发明实施例对此不做限制。

各种流表可以设置在交换机的 TCAM 或 RLDRAM 中。

容易知道，匹配模块 402 会对交换机收到的数据包与交换机中的每一条流

表项进行匹配。

进一步地，在具体实现中，一个交换机中的流表还可以根据交换机的端口来设置，例如，可以交换机的一个端口对应一组流表，每组流表可以包括一个流表（该流表中包括所有种类的流表项），每组流表也可以包括多个流表（该
5 多个流表分别包括不同种类的流表项），从一个端口接收的数据包仅在该端口对应的流表中进行匹配；又例如，可以交换机中仅配置一组流表，所有端口接收到的数据包均在该组流表中进行匹配。

相应地，该交换机还可以包括：

处理模块 404，用于当数据包与第二流表项匹配成功时，按照匹配成功的
10 第二流表项的动作域中的操作命令，对数据包中的数据进行计算处理，得到计算结果；

发送模块 405，用于将处理模块 404 得到的计算结果发送给数据包中的源节点信息对应的节点，例如处理器。

在实现时，可以在交换机中设置计算模块，也可以设置一个专门的计算设
15 备，交换机将该数据发送给该专门的计算设备，并接收计算设备返回的计算结果。

可选地，该交换机还可以包括：

第二接收模块 406，用于接收控制器发送的流表配置消息，流表配置消息用于为交换机配置流表项；

20 配置模块 407，用于根据第二接收模块 406 收到的流表配置消息配置流表项。

本发明实施例通过在交换机中内置存储设备，并对接收到的数据包进行流表匹配，当数据包与流表中的第一流表项匹配成功时，根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作，从而减少甚至避免了对处理器对远端内存的访问，可以减少内存访问的延时，同时，由于采用各交换机内的存储设备单独存储数据，不存在在其他交换机的存储设备中存在副本的情况，因此不存在需要维护 Cache 一致性的问题，实现简单。并且，当第一流表项匹配成功时，由于各个交换机直接对自身内部的存储设备进行访问，所以对于一个访问请求而言，只需要出入互连网络一次（即接收访问请求和返回
25 响应消息），从而可以节省网络资源。此外，本实施例中，交换机中还配置有
30 第二流表项，当数据包与第二流表项匹配时，可以对数据包中的数据进行计算

处理，增强了网络的硬件计算能力。

实施例五

本发明实施例提供了一种交换机，参见图 7，该交换机包括处理器 501、
5 存储器 502、总线 503 和通信接口 504。其中，存储器 502 用于存储计算机执行指令，处理器 501 与存储器 502 通过总线 503 连接，当所述计算机运行时，处理器 501 执行存储器 502 存储的所述计算机执行指令，以使所述交换机执行实施例一或实施例二中交换机所执行的方法。

该交换机还包括内置的存储设备，该内置的存储设备可以为 SRAM 或
10 DRAM 等。该存储设备可以为存储器 502，也可以为独立于存储器 502 的存储设备。

在具体实现中，该交换机的硬件结构可以如图 8 所示，包括：输入端口 601、
处理器 602、内置的存储设备 603、存储器 604、纵横开关 (CrossBar) 总线 605、
输出端口 606。

15 其中，输入端口 601 用于接收数据包，该数据包包括源节点信息、目的节点信息和协议类型。

存储器 604 用于存储流表。流表包括至少一个流表项，每个流表项都包括
匹配域和动作域。流表中的流表项通常包括转发流表项，该转发流表项用于确定数据包的转发出口。流表中的流表项还可以包括数据包修改流表项，该数据
20 包修改流表项用于确定对数据包内的信息进行修改，例如修改数据包的头域等。

在本发明实施例中，流表项还包括第一流表项，第一流表项的匹配域用于
匹配数据包中的源节点信息、目的节点信息和协议类型，第一流表项的动作域
用于指示对交换机内置的存储设备的操作命令。该操作命令包括但不限于读操
25 作命令、写操作命令。

容易知道，第一流表项的动作域中的操作命令与其匹配域中的协议类型通常
是相互对应的，例如，如果第一流表项的匹配域中的协议类型用于指示读请
求数据包，则该第一流表项的动作域中的操作命令为对交换机内置的存储设备
的读操作命令。

30 在实现时，各种流表项（例如前述第一流表项、转发流表项、数据包修改流表项）可以配置在一个流表中，也可以按照流表项的功能，将不同的流表项

配置在不同的流表中。

处理器 602 用于采用存储器 604 中的流表，对输入端口的数据包进行流表匹配，并且，当数据包与流表中的第一流表项匹配成功时，按照匹配成功的第一流表项的动作域中的操作命令，对内置的存储设备 603 进行操作，例如，读
5 操作和写操作。

操作完成后，将响应数据包通过 CrossBar 总线 605 输出到输出端口 606，从而反馈给请求设备。

处理器 602 还用于当数据包与第二流表项匹配成功时，按照匹配成功的第二流表项的动作域中的操作命令，对数据包中的数据进行计算处理，得到计算
10 结果，并将该计算结果通过 CrossBar 总线 605 和输出端口 606 发送给请求设备。

本发明实施例通过在交换机中内置存储设备，并对接收到的数据包进行流表匹配，当数据包与流表中的第一流表项匹配成功时，根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作，从而减少甚至避免了对处理器对远端内存的访问，可以减少内存访问的延时，同时，由于采用各交
15 换机内的存储设备单独存储数据，不存在在其他交换机的存储设备中存在副本的情况，因此不存在需要维护 Cache 一致性的问题，实现简单。并且，当第一流表项匹配成功时，由于各个交换机直接对自身内部的存储设备进行访问，所以对于一个访问请求而言，只需要出入互连网络一次（即接收访问请求和返回响应消息），从而可以节省网络资源。此外，本实施例中，交换机中还配置有
20 第二流表项，当数据包与第二流表项匹配时，可以对数据包中的数据进行计算处理，增强了网络的硬件计算能力。

实施例六

本发明实施例提供了一种交换机，参见图 9，该交换机包括：输入端口 701、
25 存储器 702、查表逻辑电路 703、存储设备 704、操作逻辑电路 705、CrossBar 总线 706、输出端口 707。

其中，输入端口 701 用于接收数据包，该数据包包括源节点信息、目的节点信息和协议类型。

存储器 702 用于存储流表。流表包括至少一个流表项，每个流表项都包括
30 匹配域和动作域。流表中的流表项通常包括转发流表项，该转发流表项用于确定数据包的转发出口。流表中的流表项还可以包括数据包修改流表项，该数据

包修改流表项用于确定对数据包内的信息进行修改，例如修改数据包的头域等。

在本发明实施例中，流表项还包括第一流表项，第一流表项的匹配域用于匹配数据包中的源节点信息、目的节点信息和协议类型，第一流表项的动作域
5 用于指示对交换机内置的存储设备的操作命令。该操作命令包括但不限于读操作命令、写操作命令。

容易知道，第一流表项的动作域中的操作命令与其匹配域中的协议类型通常是相互对应的，例如，如果第一流表项的匹配域中的协议类型用于指示读请求数据包，则该第一流表项的动作域中的操作命令为对交换机内置的存储设备
10 的读操作命令。

在实现时，各种流表项（例如前述第一流表项、转发流表项、数据包修改流表项）可以配置在一个流表中，也可以按照流表项的功能，将不同的流表项配置在不同的流表中。

查表逻辑电路 703 用于采用存储器 702 中的流表，对输入端口的数据包进行
15 流表匹配，操作逻辑电路 705 用于根据查表逻辑电路 703 的输出结果，按照匹配成功的第一流表项的动作域中的操作命令，对内置的存储设备 704 进行操作，例如，读操作和写操作。

操作完成后，将响应数据包通过 CrossBar 总线 706 输出到输出端口 707，从而反馈给请求设备。

操作逻辑电路 705 还用于当数据包与第二流表项匹配成功时，按照匹配成功
20 的第二流表项的动作域中的操作命令，对数据包中的数据进行计算处理，得到计算结果，并将该计算结果通过 CrossBar 总线 706 和输出端口 707 发送给请求设备。

其中，存储器 702 可以为 TCAM 或 RDRAM，存储设备 704 可以为 SRAM
25 或 DRAM 等。查表逻辑电路 703 和操作逻辑电路 705 的具体电路为本领域技术的常规电路，故在此省略详细结构描述。

本发明实施例通过在交换机中内置存储设备，并对接收到的数据包进行流表匹配，当数据包与流表中的第一流表项匹配成功时，根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作，从而减少甚至避免了
30 对处理器对远端内存的访问，可以减少内存访问的延时，同时，由于采用各交换机内的存储设备单独存储数据，不存在在其他交换机的存储设备中存在副本

的情况，因此不存在需要维护 Cache 一致性的问题，实现简单。并且，当第一流表项匹配成功时，由于各个交换机直接对自身内部的存储设备进行访问，所以对于一个访问请求而言，只需要出入互连网络一次（即接收访问请求和返回响应消息），从而可以节省网络资源。此外，本实施例中，交换机中还配置有第二流表项，当数据包与第二流表项匹配时，可以对数据包中的数据进行计算处理，增强了网络的硬件计算能力。

实施例七

本发明实施例提供了一种多处理器系统，参见图 10，该系统包括：多个处理器 801 和互连网络 800，多个处理器 801 通过互连网络 800 通信连接，互连网络 800 包括多个交换机 802，交换机 802 为实施例四、五或六提供的交换机。

在本实施例的一种实现方式中，该多处理器系统还可以包括多个外部存储设备 803，多个外部存储设备 803 通过互连网络 800 与多个处理器 801 通信连接。

在本实施例的另一种实现方式中，该多处理器系统可以为片上系统（System on Chip，简称 SoC）。容易知道，交换机 802 也可以为独立的通信设备。

本发明实施例通过在交换机中内置存储设备，并对接收到的数据包进行流表匹配，当数据包与流表中的第一流表项匹配成功时，根据第一流表项的动作域中的操作命令直接对交换机内置的存储设备进行操作，从而减少甚至避免了对处理器对远端内存的访问，可以减少内存访问的延时，同时，由于采用各交换机内的存储设备单独存储数据，不存在在其他交换机的存储设备中存在副本的情况，因此不存在需要维护 Cache 一致性的问题，实现简单。并且，当第一流表项匹配成功时，由于各个交换机直接对自身内部的存储设备进行访问，所以对于一个访问请求而言，只需要出入互连网络一次（即接收访问请求和返回响应消息），从而可以节省网络资源。此外，本实施例中，交换机中还配置有第二流表项，当数据包与第二流表项匹配时，可以对数据包中的数据进行计算处理，增强了网络的硬件计算能力。

需要说明的是：上述实施例提供的交换机根据收到的数据包进行内存访问时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而

将上述功能分配由不同的功能模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的交换机与内存管理方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

5

上述本发明实施例序号仅仅为了描述，不代表实施例的优劣。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

10

以上所述仅为本发明的较佳实施例，并不用以限制本发明，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、一种内存访问方法，其特征在于，所述方法包括：

5 交换机接收数据包，所述数据包包括源节点信息、目的节点信息和协议类型，所述协议类型用于指示所述数据包的类型；

对所述数据包进行流表匹配，所述流表包括至少一个流表项，所述流表项包括匹配域和动作域，所述至少一个流表项包括第一流表项，所述第一流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第一流表项的动作域用于指示对所述交换机内置的存储设备的操作命令；

10 当所述数据包与所述第一流表项匹配成功时，按照匹配成功的所述第一流表项的动作域中的操作命令，对所述存储设备进行操作。

2、根据权利要求1所述的方法，其特征在于，所述按照匹配成功的第一流表项的动作域中的操作命令，对所述存储设备进行操作，包括：

15 当匹配成功的第一流表项的动作域中的所述操作命令为读操作命令时，从所述存储设备中读取数据，并将读取的所述数据返回给所述源节点信息对应的节点；

当匹配成功的第一流表项的动作域中的所述操作命令为写操作命令时，将所述数据包中的数据写入所述存储设备中。

20

3、根据权利要求1或2所述的方法，其特征在于，所述至少一个流表项还包括第二流表项，所述第二流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第二流表项的动作域用于指示对所述数据包中的数据计算处理的操作命令。

25

4、根据权利要求3所述的方法，其特征在于，所述方法还包括：

当所述数据包与所述第二流表项匹配成功时，按照匹配成功的所述第二流表项的动作域中的操作命令，对所述数据包中的数据进行计算处理，得到计算结果；

30 将所述计算结果发送给所述数据包中的所述源节点信息对应的节点。

5、根据权利要求 1-4 任一项所述的方法，其特征在于，所述方法还包括：
接收控制器发送的流表配置消息，所述流表配置消息用于为所述交换机配置所述流表项；

根据所述流表配置消息配置所述流表项。

5

6、一种交换机，其特征在于，所述交换机包括：

第一接收模块，用于接收数据包，所述数据包包括源节点信息、目的节点信息和协议类型，所述协议类型用于指示所述数据包的类型；

10 匹配模块，用于对所述第一接收模块接收的所述数据包进行流表匹配，所述流表包括至少一个流表项，所述流表项包括匹配域和动作域，所述至少一个流表项包括第一流表项，所述第一流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第一流表项的动作域用于指示对所述交换机内置的存储设备的操作命令；

15 操作模块，用于当所述数据包与所述第一流表项匹配成功时，按照匹配成功的所述第一流表项的动作域中的操作命令，对所述存储设备进行操作。

7、根据权利要求 6 所述的交换机，其特征在于，所述操作模块包括：

读取单元，用于当所述操作命令为读操作命令时，从所述存储设备中读取数据；

20 发送单元，用于将所述读取单元读取的所述数据返回给所述源节点信息对应的节点；

写入单元，用于当所述操作命令为写操作命令时，将所述数据包中的数据写入所述存储设备中。

25 8、根据权利要求 6 或 7 所述的交换机，其特征在于，所述至少一个流表项还包括第二流表项，所述第二流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第二流表项的动作域用于指示对所述数据包中的数据计算处理的操作命令。

30 9、根据权利要求 8 所述的交换机，其特征在于，所述交换机还包括：
处理模块，用于当所述数据包与所述第二流表项匹配成功时，按照匹配成

功的所述第二流表项的动作域中的操作命令，对所述数据包中的数据进行处理，得到计算结果；

发送模块，用于将所述处理模块得到的所述计算结果发送给所述数据包中的所述源节点信息对应的节点。

5

10、根据权利要求 6-9 任一项所述的交换机，其特征在于，所述交换机还包括：

第二接收模块，用于接收控制器发送的流表配置消息，所述流表配置消息用于为所述交换机配置所述流表项；

10 配置模块，用于根据所述第二接收模块收到的所述流表配置消息配置所述流表项。

11、一种交换机，其特征在于，所述交换机包括：处理器、存储器、总线和通信接口；所述存储器用于存储计算机执行指令，所述处理器与所述存储器通过所述总线连接，当所述计算机运行时，所述处理器执行所述存储器存储的所述计算机执行指令，以使所述交换机执行如权利要求 1~5 任意一项所述的方法。

12、一种交换机，其特征在于，所述交换机包括：
20 输入端口，用于接收数据包，所述数据包包括源节点信息、目的节点信息和协议类型，所述协议类型用于指示所述数据包的类型；

存储器，用于存储流表，所述流表包括至少一个流表项，所述流表项包括匹配域和动作域，所述至少一个流表项包括第一流表项，所述第一流表项的匹配域用于匹配所述数据包中的源节点信息、目的节点信息和协议类型，所述第
25 一流表项的动作域用于指示对所述交换机内置的存储设备的操作命令；

存储设备，用于存储数据；

查表逻辑电路，用于采用所述存储器存储的流表对所述输入端口接收的数据包进行流表匹配；

操作逻辑电路，用于当所述数据包与所述第一流表项匹配成功时，按照匹
30 配成功的所述第一流表项的动作域中的操作命令，对所述存储设备进行操作；

纵横开关总线，用于为所述操作逻辑电路传输的数据包选择输出端口；

输出端口，用于发送所述纵横开关总线传输的数据包。

13、一种多处理器系统，所述多处理器系统包括：多个处理器和互连网络，所述多个处理器通过所述互连网络通信连接，所述互连网络包括多个交换机，
5 其特征在于，所述交换机包括为权利要求 6-12 任一项所述的交换机。

14、根据权利要求 13 所述的多处理器系统，其特征在于，所述多处理器系统还包括多个外部存储设备，所述多个外部存储设备通过所述互连网络与所述多个处理器通信连接。

10

15、根据权利要求 13 或 14 所述的多处理器系统，其特征在于，所述多处理器系统为片上系统。

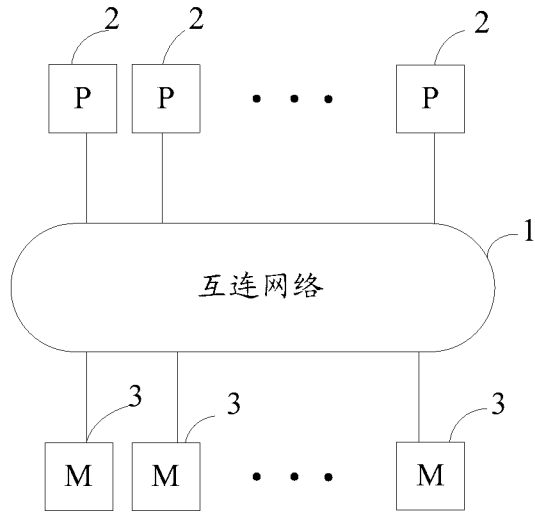


图 1

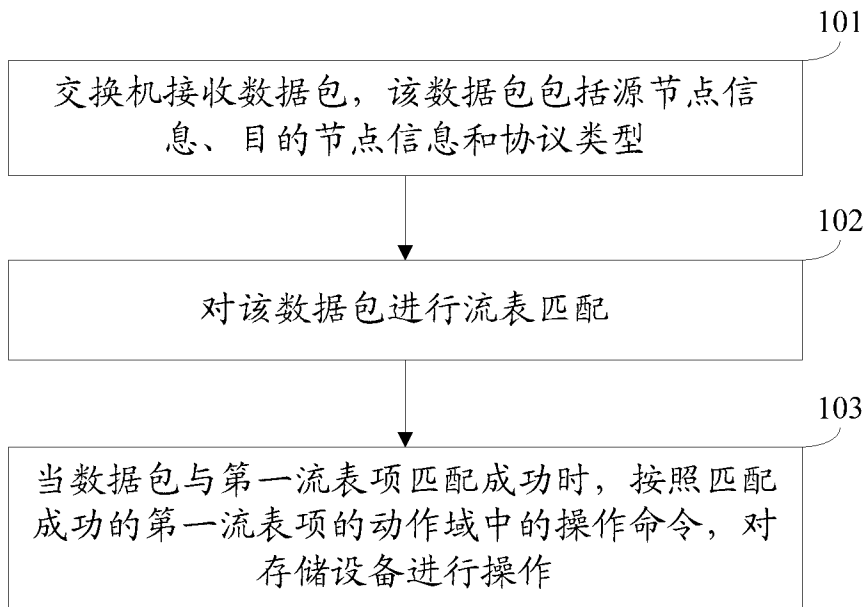


图 2

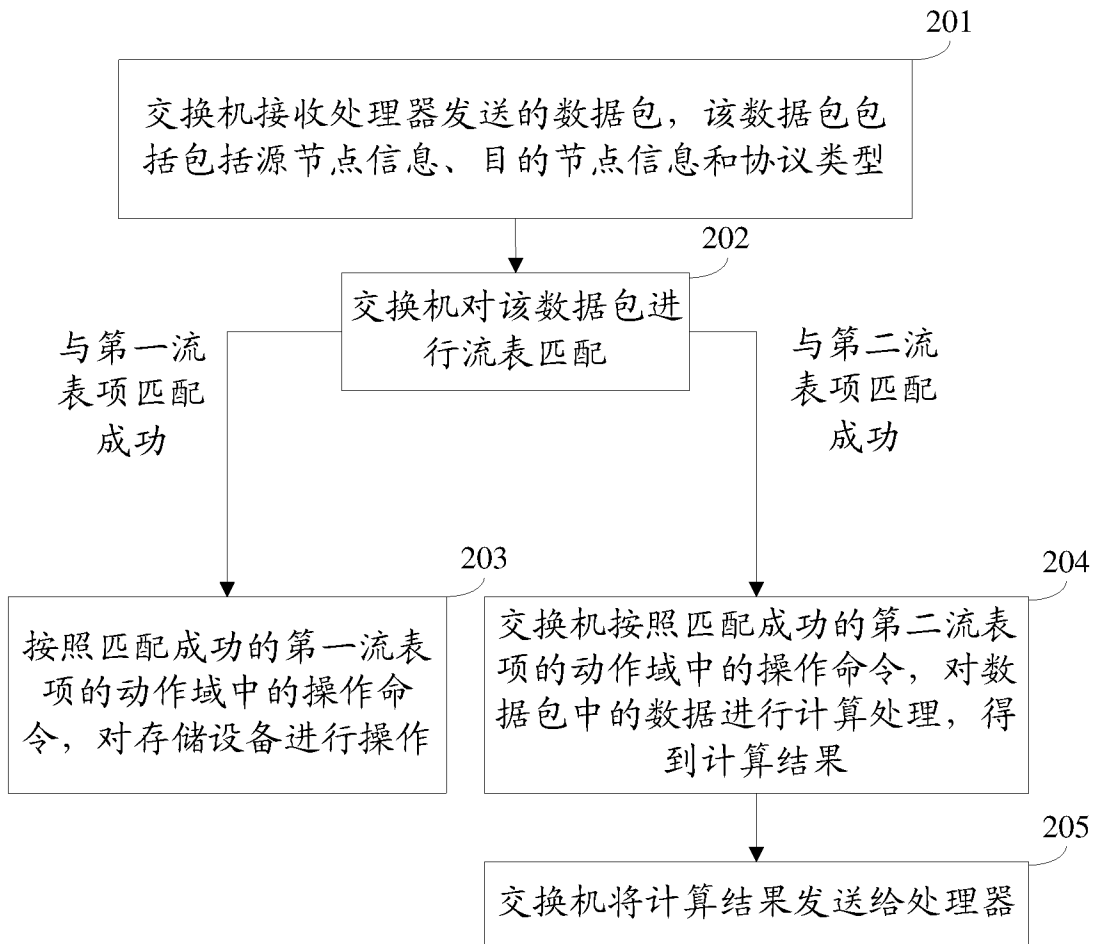


图 3

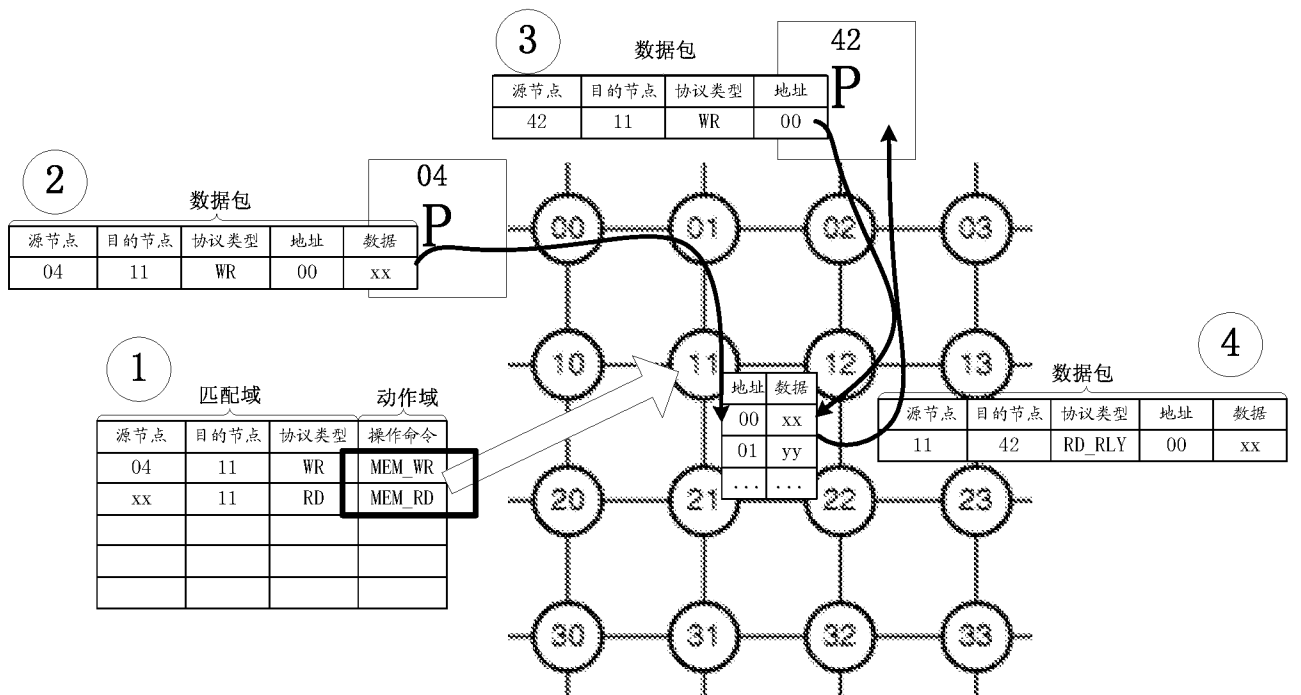


图 4

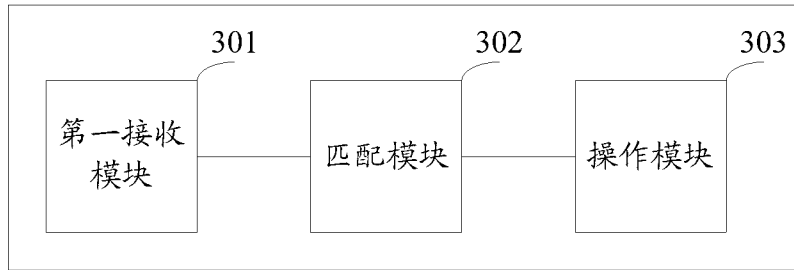


图 5

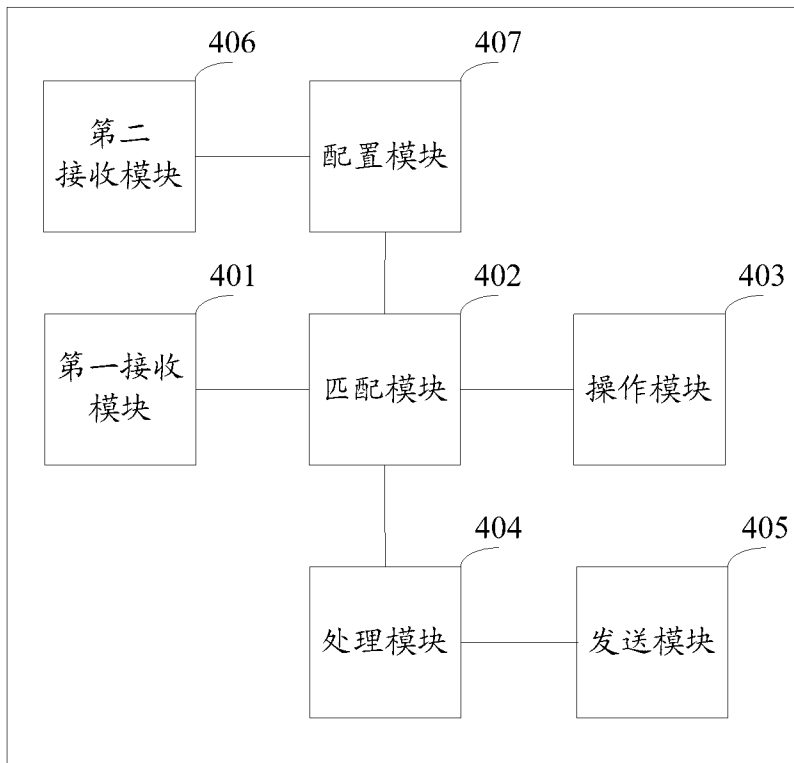


图 6

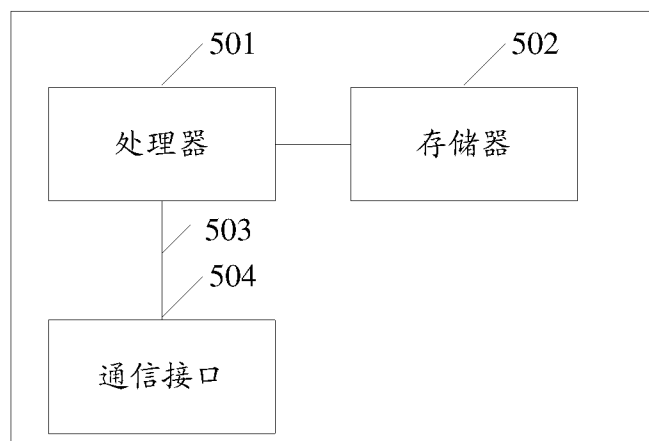


图 7

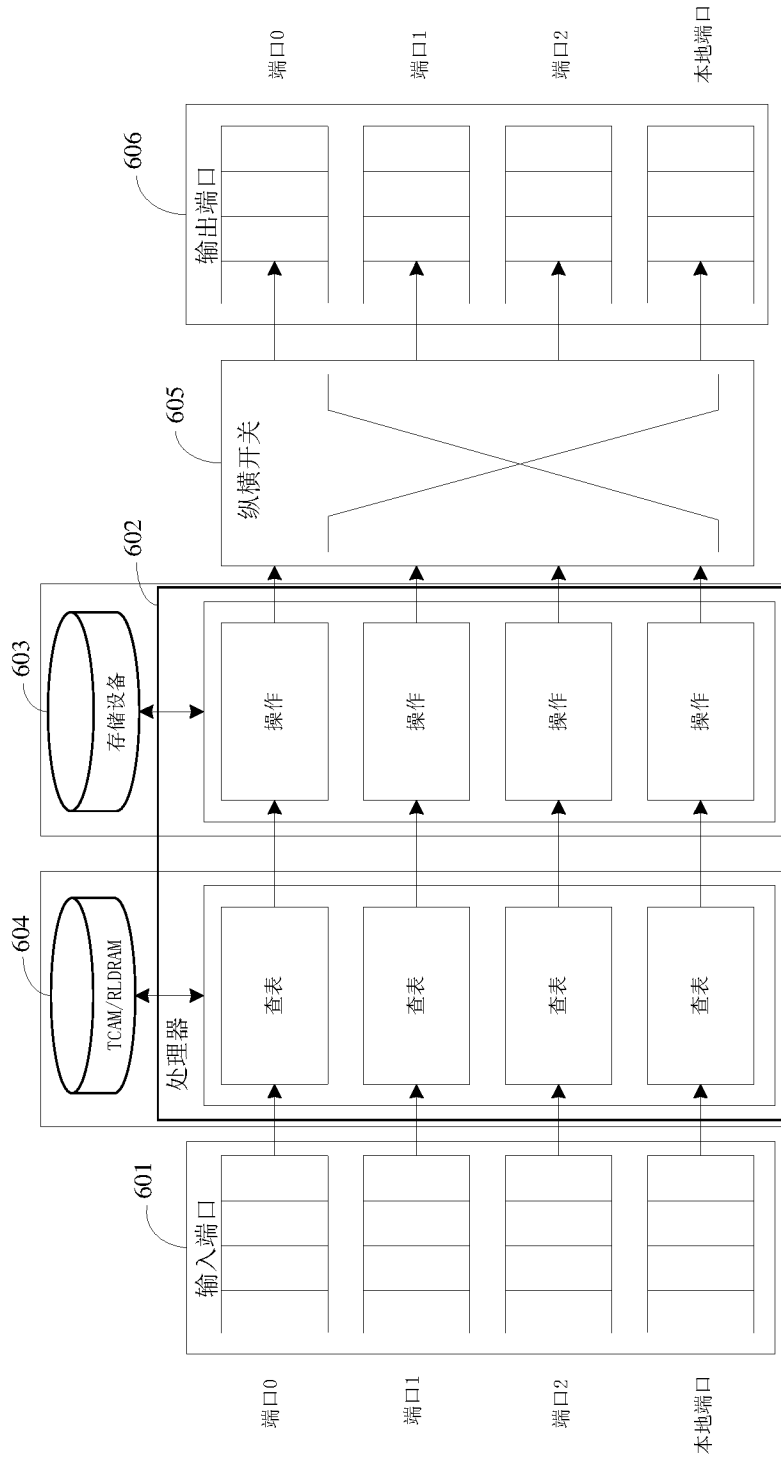


图8

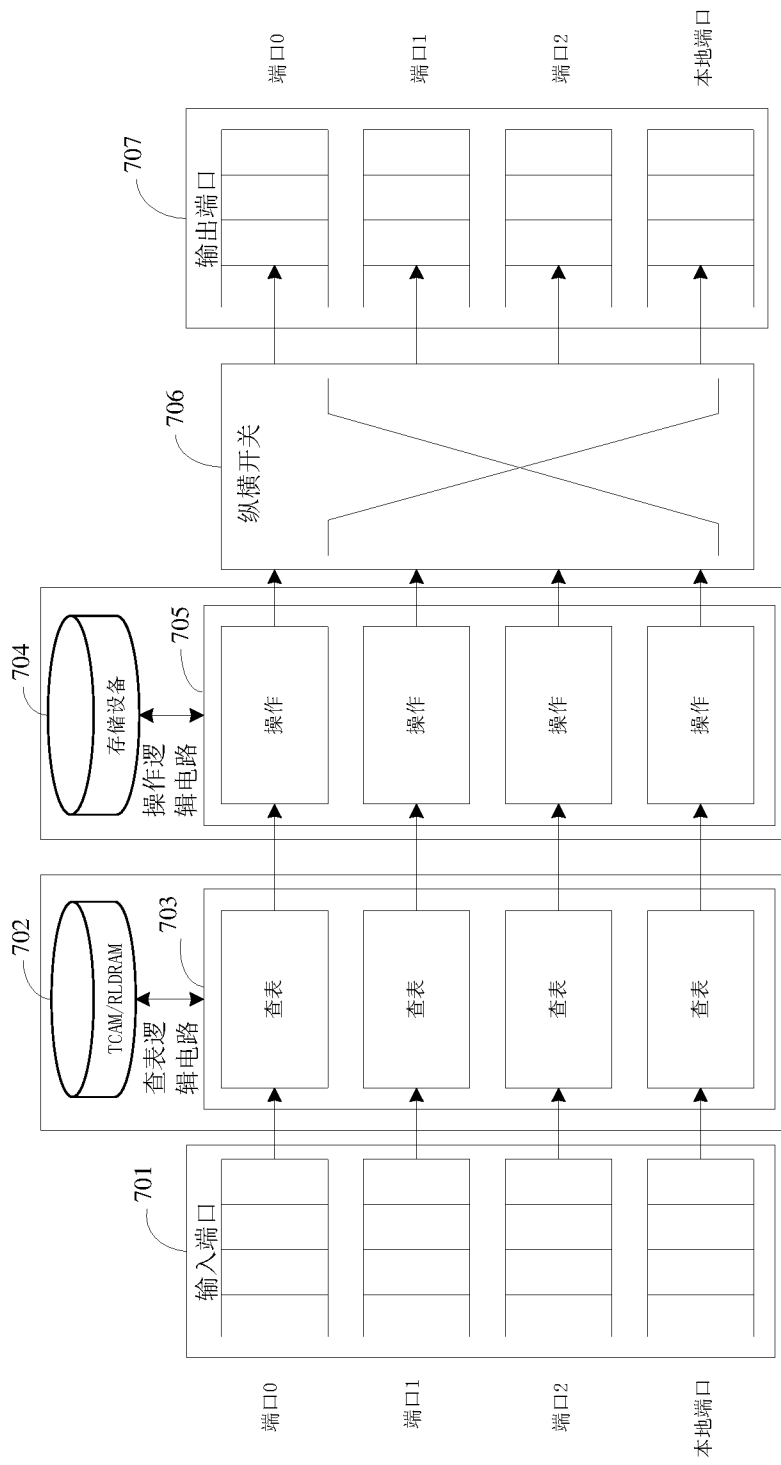


图9

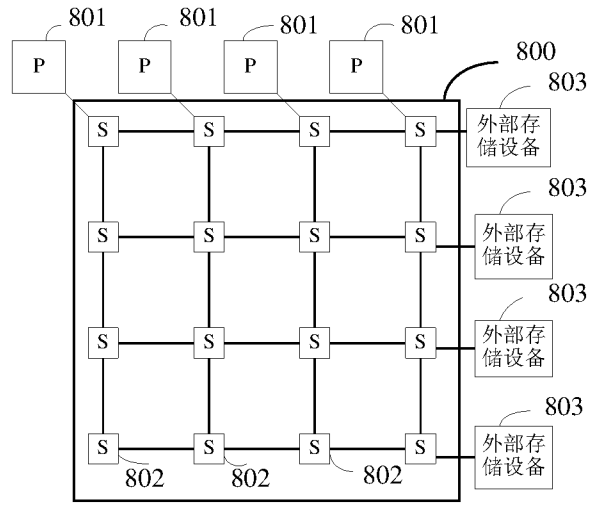


图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/092421

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/911 (2013.01) i; H04L 29/12 (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)

H04L

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)

CNABS; CNTXT; CNKI; VEN: SWITCH, MATCH+, FLOW TABLE, STOR+, MEMORY

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103501280 A (UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA) 08 January 2014 (08.01.2014) abstract, claims 1-8	1-15
A	CN 102946325 A (ZTE CORPORATION) 27 February 2013 (27.02.2013) the whole document	1-15
A	CN 103036653 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 10 April 2013 (10.04.2013) the whole document	1-15
A	CN 103401784 A (HUAWEI TECHNOLOGIES CO., LTD.) 20 November 2013 (20.11.2013) the whole document	1-15
A	WO 2013133227 A1 (NEC CORPORATION et al.) 12 September 2013 (12.09.2013) the whole document	1-15

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&” document member of the same patent family</p>
---	---

Date of the actual completion of the international search
20 August 2015

Date of mailing of the international search report
02 September 2015

Name and mailing address of the ISA
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No. (86-10) 62019451

Authorized officer
DING, Ling
Telephone No. (86-10) 62411483

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2014/092421

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103501280 A	08 January 2014	None	
CN 102946325 A	27 February 2013	CN 102946325 B	03 June 2015
CN 103036653 A	10 April 2013	None	
CN 103401784 A	20 November 2013	None	
WO 2013133227 A1	12 September 2013	CN 104185972 A	03 December 2014
		US 2015009828 A1	08 January 2015
		EP 2824877 A1	14 January 2015

国际检索报告

国际申请号

PCT/CN2014/092421

<p>A. 主题的分类</p> <p>H04L 12/911(2013.01)i; H04L 29/12(2006.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>H04L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNABS;CNTXT;CNKI;VEN: 交换机, 流表, 匹配, 储存, 存储, 内存; SWITCH, MATCH+, FLOW TABLE, STOR+, MEMORY</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 103501280 A (电子科技大学等) 2014年 1月 8日 (2014 - 01 - 08) 摘要, 权利要求1-8</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>CN 102946325 A (中兴通讯股份有限公司) 2013年 2月 27日 (2013 - 02 - 27) 全文</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>CN 103036653 A (华中科技大学) 2013年 4月 10日 (2013 - 04 - 10) 全文</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>CN 103401784 A (华为技术有限公司) 2013年 11月 20日 (2013 - 11 - 20) 全文</td> <td>1-15</td> </tr> <tr> <td>A</td> <td>WO 2013133227 A1 (NEC CORP等) 2013年 9月 12日 (2013 - 09 - 12) 全文</td> <td>1-15</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 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(如具体说明的) “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</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 103501280 A (电子科技大学等) 2014年 1月 8日 (2014 - 01 - 08) 摘要, 权利要求1-8	1-15	A	CN 102946325 A (中兴通讯股份有限公司) 2013年 2月 27日 (2013 - 02 - 27) 全文	1-15	A	CN 103036653 A (华中科技大学) 2013年 4月 10日 (2013 - 04 - 10) 全文	1-15	A	CN 103401784 A (华为技术有限公司) 2013年 11月 20日 (2013 - 11 - 20) 全文	1-15	A	WO 2013133227 A1 (NEC CORP等) 2013年 9月 12日 (2013 - 09 - 12) 全文	1-15
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
A	CN 103501280 A (电子科技大学等) 2014年 1月 8日 (2014 - 01 - 08) 摘要, 权利要求1-8	1-15																		
A	CN 102946325 A (中兴通讯股份有限公司) 2013年 2月 27日 (2013 - 02 - 27) 全文	1-15																		
A	CN 103036653 A (华中科技大学) 2013年 4月 10日 (2013 - 04 - 10) 全文	1-15																		
A	CN 103401784 A (华为技术有限公司) 2013年 11月 20日 (2013 - 11 - 20) 全文	1-15																		
A	WO 2013133227 A1 (NEC CORP等) 2013年 9月 12日 (2013 - 09 - 12) 全文	1-15																		
<p>国际检索实际完成的日期</p> <p>2015年 8月 20日</p>	<p>国际检索报告邮寄日期</p> <p>2015年 9月 2日</p>																			
<p>ISA/CN的名称和邮寄地址</p> <p>中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国</p> <p>传真号 (86-10)62019451</p>	<p>受权官员</p> <p>丁玲</p> <p>电话号码 (86-10)62411483</p>																			

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/092421

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103501280	A	2014年 1月 8日	无			
CN	102946325	A	2013年 2月 27日	CN	102946325	B	2015年 6月 3日
CN	103036653	A	2013年 4月 10日	无			
CN	103401784	A	2013年 11月 20日	无			
WO	2013133227	A1	2013年 9月 12日	CN	104185972	A	2014年 12月 3日
				US	2015009828	A1	2015年 1月 8日
				EP	2824877	A1	2015年 1月 14日

表 PCT/ISA/210 (同族专利附件) (2009年7月)