

(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
2020 年 9 月 10 日 (10.09.2020)



(10) 国际公布号
WO 2020/177263 A1

(51) 国际专利分类号:
H04L 12/801 (2013.01)

(21) 国际申请号: PCT/CN2019/096675

(22) 国际申请日: 2019 年 7 月 19 日 (19.07.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910156869.8 2019 年 3 月 1 日 (01.03.2019) CN

(71) 申请人: 烽火通信科技股份有限公司 (FIBERHOME TELECOMMUNICATION TECHNOLOGIES CO., LTD) [CN/CN]; 中国湖北省武汉市东湖高新技术开发区高新四路 6 号, Hubei 430000 (CN)。武汉飞思灵微电子技术有限公司 (WUHAN FISILINK MICROELECTRONICS TECHNOLOGY CO., LTD.) [CN/CN]; 中国湖北省武汉市东湖开发区关东工业园烽火路光通信大楼产业大楼二楼, Hubei 430000 (CN)。

(72) 发明人: 陈永洲 (CHEN, Yongzhou); 中国湖北省武汉市东湖高新技术开发区高新四路 6 号, Hubei 430000 (CN)。周万涛 (ZHOU, Wantao); 中国湖北省武汉市东湖高新技术开发区高新四路 6 号, Hubei 430000 (CN)。李仁杰 (LI, Renjie); 中国湖北省武汉市东湖高新技术开发区高新四路 6 号, Hubei 430000 (CN)。

(74) 代理人: 武汉智权专利代理事务所 (特殊普通合伙) (WUHAN ZHI QUAN PATENT AGENCY); 中国湖北省武汉市东湖高新技术开发区珞喻路 727 号星光无限 4 栋 21 层 2103 室, Hubei 430000 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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,

(54) Title: TRAFFIC MANAGEMENT METHOD AND SYSTEM AND FABRIC NETWORK PROCESSOR

(54) 发明名称: 一种流量管理方法、系统及织状网络处理器

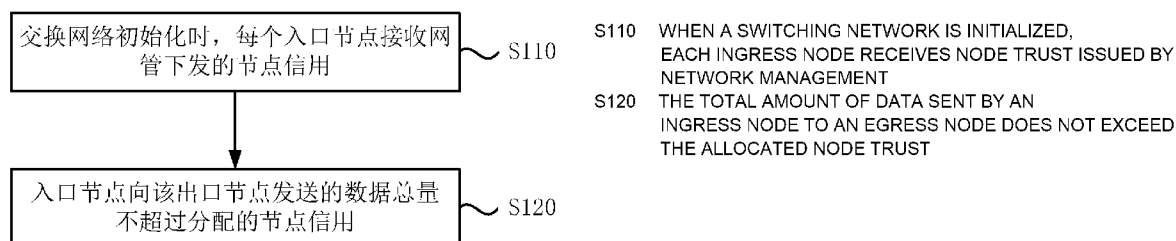


图 1

(57) Abstract: The present invention relates to the technical field of communications, and disclosed are a traffic management method and system and a fabric network processor, which are applied to switching networks. The traffic management method comprises that: the total amount of data sent by an ingress node to an egress node does not exceed the node trust allocated thereto, and the sum of the node trust of a plurality of ingress nodes which send data to egress nodes does not exceed the total egress traffic of the egress nodes. In the present invention, the traffic sent to a fabric network processor is controlled by means of node trust, thereby avoiding the circumstance in which the fabric network processor is unable to work normally due to too much traffic.

(57) 摘要: 本发明公开了一种流量管理方法、系统及织状网络处理器, 应用于交换网络, 涉及通信技术领域。流量管理方法包括: 入口节点向出口节点发送的数据总量不超过其分配到的节点信用, 且向出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量。本发明通过节点信用控制发往织状网络处理器的流量, 避免织状网络处理器因流量过大而不能正常工作。

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)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

一种流量管理方法、系统及织状网络处理器

技术领域

本发明涉及通信技术领域，具体是涉及一种流量管理方法、系统及织状网络处理器。

背景技术

在通信领域，基于以太网报文的通信占主流地位，路由器和以太网织状交换单元等数据通信设备都需要采用芯片的包处理模块对以太网包或 IP 包（Internet Protocol）进行处理。许多数据网络使用基于织状（Fabric）架构的交换网络来传输数据，例如在超大数据带宽交换处理系统中，每个织状网络处理器（Fabric Network Processor, FNP）为一个节点，织状网络处理器 FNP 之间通过多个织状交换单元（Fabric Switch Element, FSE）交换数据。在每个节点上，入口流量管理器（Ingress Traffic Manager）决定所有入口流量管理队列（Ingress Traffic Management Queue）是否有资格发送信元到织状交换单元 FSE，继而发送到对应的出口流量管理队列（Egress Traffic Management Queue）。

目前，通用的交换网络的流量管理（Traffic Management）方法是服务质量（Quality of Service, QoS）+层次化服务质量（Hierarchy Quality of Service, HQoS），入口与出口之间通过流状态信息（Flow Status Message）和信用信息（Credit Message）对虚拟输出队列（Virtual Output Queues, VOQs）进行调度。这种输出队列和业务类型的 pull 调度模式能够根据端口速率灵活地进行整形（shaping）和调度（scheduling）。

但是，实际应用中存在由于网络系统流量不规则而导致织状网络处理器 FNP 崩溃的情况。例如，业务盘上电复位或者突然插拔业务芯片的时候，发往某个或某些织状网络处理器 FNP 的数据包过多，造成上述织状网络处理器 FNP 因流量过大而不能正常工作。

发明内容

针对现有技术中存在的缺陷，本发明实施例的目的在于提供一种流量管理方法、系统及织状网络处理器，通过节点信用控制发往织状网络处理器的流量，避免织状网络处理器因流量过大而不能正常工作。

第一方面，本发明实施例提供一种流量管理方法，应用于交换网络；

入口节点向出口节点发送的数据总量不超过其分配到的节点信用，且向所述出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量。

结合第一方面，在第一种可选的实现方式中，所述节点信用是网管下发的或者由所述出口节点周期性地重新分配的。

在第一种可选的实现方式中，在每个所述周期中，所述入口节点实时计算节点信用余额，其中，节点信用余额=所述节点信用 - 已发送的数据数量；当所述节点信用余额达到阈值时，停止向所述出口节点发送数据，以及向外发送节点信用请求。

在第一种可选的实现方式中，所述入口节点包括入口流量管理器，入口流量管理器用于配置输出队列和队列信用；

当所述节点信用大于零，队列信用大于或者等于零，且输出队列为非空时，向所述出口节点发送数据。

第二方面，本发明实施例提供一种流量管理方法，应用于交换网

络；

出口节点周期性地将出口总流量划分为多个节点信用，并一一分配给向其发送数据的多个入口节点。

结合第二方面第一种可选的实现方式，所述出口节点根据拥塞信息和/或所述入口节点的节点信用请求，增加、减少或者维持所述节点信用；

所述拥塞信息来自所述交换网络的织状交换单元和/或所述出口节点的出口包缓存器。

第三方面，本发明实施例提供一种流量管理系统，其设于交换网络的入口节点上，每个入口节点均包括入口包缓存器，所述流量管理系统包括：

入口节点信用计数器，其用于存储分配给该入口节点的节点信用，其中，向所述出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量；

入口管理逻辑单元，其用于控制所述入口包缓存器向所述出口节点发送的数据总量不超过所述节点信用。

结合第三方面，在第一种可选的实现方式中，所述节点信用是网管下发的或者由所述出口节点周期性地重新分配的。

在第一种可选的实现方式中，在每个所述周期中，所述入口管理逻辑单元用于实时计算节点信用余额，其中，节点信用余额=所述节点信用 - 已发送的数据数量；当所述节点信用余额达到阈值时，向外发送节点信用请求。

结合第三方面，在第二种可选的实现方式中，所述入口节点还包括入口流量管理器，入口流量管理器用于配置输出队列和队列信用；或者，所述入口管理逻辑单元还用于配置输出队列和队列信用；

所述入口管理逻辑单元还用于控制所述入口包缓存器发送数据，包括：当所述节点信用大于零，队列信用大于或者等于零，且输出队列为非空时，向所述出口节点发送数据。

第四方面，本发明实施例提供一种流量管理系统，其设于交换网络的出口节点上，每个出口节点均包括出口包缓存器，所述流量管理系统包括：

出口管理逻辑单元，其用于统计从每个入口节点接收到的数据数量，以及周期性地将出口总流量划分为多个节点信用，并一一分配给向其发送数据的多个入口节点；

出口节点信用计数器，其用于存储从每个入口节点接收到的数据数量。

结合第四方面，在第一种可选的实现方式中，所述出口管理逻辑单元还用于根据拥塞信息和/或所述入口节点的节点信用请求，增加、减少或者维持所述节点信用；

所述拥塞信息来自织状交换单元或者所述出口包缓存器。

第五方面，本发明实施例提供一种织状网络处理器，其设于交换网络中，所述织状网络处理器包括入口包缓存器和出口包缓存器，所述织状网络处理器还包括上述第三方面的流量管理系统，以及上述第四方面的流量管理系统。

与现有技术相比，本发明实施例提供一种流量管理方法，应用于交换网络，入口节点向出口节点发送的数据总量不超过其分配到的节点信用，且向出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量，通过节点信用控制发往织状网络处理器的流量，避免织状网络处理器因流量过大而不能正常工作。

附图说明

图 1 是本发明实施例流量管理方法流程图；

图 2 是本发明另一实施例流量管理方法流程图；

图 3 是本发明另一实施例流量管理方法中，入口节点和出口节点之间节点信用的功能示意图；

图 4 是本发明实施例入口节点的流量管理系统的应用示意图；

图 5 是本发明另一实施例出口节点流量管理系统的应用示意图；

图 6 是图 4 和图 5 的流量管理系统的应用示意总图，其中包含出口节点以及向该出口节点发送数据的一个入口节点；

图 7 是一种现有的基于 Fabric 的交换网络架构；

图 8 是本发明实施例在织状网络处理器 FNP 中的应用示意图；

图 9 是本发明实施例在织状网络处理器 FNP 的另一个具体应用示意图。

具体实施方式

下面结合附图及具体实施例对本发明作进一步的详细描述。

本发明实施例提供一种流量管理方法，可以应用于分布式交换网络，也可以应用于集中式交换网络，流量管理方法包括：入口节点向出口节点发送的数据总量不超过其分配到的节点信用，且向出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量。

具体的，节点信用（Node Credit，NC）表示可发送的数据总量，是对端织状网络处理器 FNP 允许本端织状网络处理器 FNP 向对端织状网络处理器 FNP 的各端口发送的数据流量的总和。

本端织状网络处理器 FNP 向对端织状网络处理器 FNP 的各端口发送的数据可以是信元、数据包或者指定的数据单位，不作限定，其中，指定的数据单位包括具有指定比特数的码块，例如 66B 比特等。

根据上述数据的具体发送形式，入口节点向出口节点发送的数据总量相应地是信元总数、数据包总数、数据字节总数或者指定的数据单位总数。

以下的本发明实施例均以分布式交换网络为例进行说明。

参见图 1 所示，在入口节点，流量管理方法包括：

S110 交换网络初始化时，每个入口节点接收网管下发的节点信用。

具体的，在初始化之前，网管将一个出口节点的出口总流量划分为多个节点信用，初始化时一一分配给向该出口节点发送数据的多个入口节点，因此，向该出口节点发送数据的每个入口节点都得到为其分配的节点信用。

网管将一个出口节点的出口总流量划分为多个节点信用时，可以采用平均分配的方式，使得每个入口节点的节点信用相同，也可以根据实际应用采用不平均分配的方式，使得不同入口节点的节点信用不同，不作限定。

S120 入口节点向该出口节点发送的数据总量不超过分配的节点信用。

目前业内通用的分布式交换网络的流量管理方法是 QOS+HQOS，但是，当分布式交换网络上电复位或者拔插分布式交换网络的织状网络处理器 FNP 时，分布式交换网络进行初始化，多个入口节点的虚拟输出队列 VOQs 都向出口节点发送数据，导致网络系统的流量不规则，以及织状交换单元和/或出口节点崩溃的现象。

在本实施方式中，节点信用是网管下发的，在其他的实施方式中，节点信用也可以通过其他途径下发给入口节点，不作限定。

通过本实施例流量管理方法，在交换网络初始化时，配置入口节

点的节点信用，通过节点信用达到对各节点之间的流量控制，避免流量过多导致织状交换单元和/或出口节点瘫痪。

参见图 2 所示的本发明另一实施例流量管理方法流程图，交换网络初始化时，节点信用是网管下发的，初始化后，出口节点周期性地重新分配节点信用。

流量管理方法包括：

S210 交换网络初始化时，每个入口节点接收网管下发的节点信用。此时，该节点信用为初始值。

S220 入口节点向该出口节点发送的数据总量不超过分配的节点信用。

S230 初始化后，出口节点周期性地将出口总流量划分为多个节点信用，并一一分配给向其发送数据的多个入口节点。

具体的，该重新分配的节点信用可以与步骤 S210 中网管下发的节点信用的初始值相同，也可以不同，不作限定。

出口节点将出口总流量划分为多个节点信用的周期也称为老化时间（EligTime）。老化时间可以根据出口节点的带宽等参数确定，例如对 Fabric 分布式架构进行系统仿真计算得到老化时间。

S240 入口节点接收出口节点周期性地重新分配的节点信用。

S250 在每个周期中，当节点信用大于零，队列信用大于或者不超过零，且输出队列为非空时，入口节点向出口节点发送数据。

参见图 3 所示，入口节点 Node X（入口织状网络处理器 Ingress FNP）通过织状交换单元 FSE 向二个出口节点 Node Y（出口织状网络处理器 Egress FNP）和 Node Z（出口织状网络处理器 Egress FNP）发送数据。在初始化时，入口节点 Node X 从中央处理器（Central Processing Unit, CPU）接收下发的节点信用初始值（NCCIFloor），

节点信用初始值 (NCCIFloor) 包括出口节点 Node Y 分配给入口节点 Node X 的节点信用初始值 (NCCIFloor_Y), 以及出口节点 Node Z 分配给入口节点 Node X 的节点信用初始值 (NCCIFloor_Z)。其中, NCCIFloor_Y 和 NCCIFloor_Z 可以相等, 也可以不相等, 不作限定。

作为一个示例, 在初始化时, 入口节点 Node X 的两个输出队列向出口节点 Node Y 所发送的数据总量不超过节点信用初始值 NCCIFloor_Y, 节点信用 NCCIFloor_Y 是分配给入口节点 Node X 的节点信用, 用于控制入口节点 Node X 向出口节点 Node Y 发送的数据总量。入口节点 Node X 统计节点信用余额 NCCI[Y]。同样的, 节点信用初始值 NCCIFloor_Z 是分配给入口节点 Node X 的节点信用, 用于控制入口节点 Node X 向出口节点 Node Z 发送的数据总量。入口节点 Node X 统计节点信用余额 NCCI[Z]。

初始化后, 出口节点周期性地将出口总流量划分为多个节点信用, 并一一分配给向其发送数据的多个入口节点, 此时的节点信用可以等于或者不等于节点信用初始值, 不作限定。

目前业内通用的分布式交换网络的入口节点均包括独立的入口流量管理器, 入口流量管理器用于配置输出队列和队列信用, 例如虚拟输出队列 VOQs 基于队列信用 (Queue Credit, QC) 进行流量管理。

本实施例采用节点信用 NC 机制和队列信用 QC 机制相结合的技术方案, 可以避免流量过多导致织状交换单元和/或出口节点瘫痪。

进一步的, 在步骤 S250 之后, 流量管理方法还包括:

S260 入口节点实时计算节点信用余额, 其中, 节点信用余额=节点信用 - 已发送的数据数量。当节点信用余额达到阈值时, 停止向出口节点发送数据, 以及向外发送节点信用请求。

具体的, 当节点信用余额达到阈值时, 停止向出口节点发送数据,

以及向外发送节点信用请求的执行顺序不作限定。可以先停止向出口节点发送数据，然后向外发送节点信用请求；也可以先向外发送节点信用请求，然后停止向出口节点发送数据；或者，停止向出口节点发送数据，同时，向外发送节点信用请求。

S270 出口节点根据拥塞信息（Congestion Message, CM）和/或入口节点的节点信用请求，增加、减少或者维持节点信用。

拥塞信息来自交换网络的织状交换单元 FSE 和/或出口节点的出口包缓存器。拥塞信息的触发条件包括：由于交换网络的负载不均衡，或者超过时延等因素，导致织状交换单元 FSE 的包缓存器和/或出口节点的出口包缓存器中的流量超过预设水线。

具体的，增加的节点信用=当前的节点信用+节点信用减少值（NodeDec）。

减少的节点信用=当前的节点信用-节点信用增加值（NodeInc）。

其中，节点信用减少值 NodeDes 和节点信用增加值 NodeInc 都可以根据出口节点的带宽和老化时间等参数确定，例如通过对 Fabric 分布式架构进行系统仿真得到，并作为初始化配置在软件开发工具包（Software Development Kit, SDK）中，下发到出口节点的芯片中。

出口节点根据拥塞信息和/或入口节点的节点信用请求，可以根据实际情况，例如综合入口节点的业务情况、优先级等因素，采用以下方式之一或者组合对节点信用进行调整：增加、减少或者维持节点信用，不作限定。无论如何调整，在每个周期内出口节点向所有入口节点分配的节点信用总和不超过该出口节点的出口总流量。

本实施例可以避免流量过多导致织状交换单元和/或出口节点瘫痪，也避免由于流量过少导致织状交换单元和/或出口节点长时间处于非全速状态，并可以根据实际节点之间的流量进行可变粒度的、动

态的调整，实现业务流量均匀。

参见图 4 所示的一种流量管理系统的应用示意图，用于实现前述实施例的方法，其设于交换网络的入口节点上，即图 4 中的入口流量管理系统 100，每个入口节点均包括入口包缓存器 130（Packet Buffer, PB），入口流量管理系统 100 包括入口节点信用计数器（Node Credit Counter Ingress, NCCI）120 和入口管理逻辑单元 110。

入口节点信用计数器（NCCI）120 用于存储分配给该入口节点的节点信用，其中，向出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量。

入口管理逻辑单元 110 用于控制入口包缓存器向出口节点发送的数据总量不超过节点信用。

具体的，入口管理逻辑单元 110 用于接收分配给该入口节点的节点信用，并转发给入口节点信用计数器（NCCI）120；从入口包缓存器 130 发出的数据数量达到节点信用时，向入口包缓存器 130 发送停止发送数据的反压信息，禁止入口包缓存器 130 向分配该节点信用的出口节点输出数据，入口包缓存器 130 缓存接收的数据，当接收的数据超过入口包缓存器 130 的缓存大小时，则丢弃。

具体的，节点信用是网管下发的或者由出口节点周期性地重新分配的。出口节点周期性地重新分配节点信用时，入口管理逻辑单元 110 用于周期性地接收分配给该入口节点的节点信用，并转发给入口节点信用计数器（NCCI）120。

进一步的，在每个周期中，入口管理逻辑单元 110 用于实时计算节点信用余额，其中，节点信用余额=节点信用 - 已发送的数据数量；当节点信用余额达到阈值时，向入口包缓存器发送停止发送数据的反压信息，以及向外发送节点信用请求。

具体的, 根据实际应用设置阈值的大小, 不作限定, 例如本发明实施例中设置阈值=0。

出口节点将出口总流量划分为多个节点信用的周期也称为老化时间 (EligTime)。老化时间可以根据出口节点的带宽、以及入口和出口包缓存器的大小等参数确定, 例如对 Fabric 分布式架构进行系统仿真计算得到老化时间。

在一种实施方式中, 入口节点还可以包括入口流量管理器, 入口流量管理器用于配置输出队列 (Queue) 和队列信用 QC。入口管理逻辑单元 110 控制节点信用, 并从入口流量管理器获取对队列信用以及输出队列结果。在另一种实施方式中, 入口管理逻辑单元 110 除了控制节点信用, 还用于配置输出队列和队列信用。

入口管理逻辑单元 110 控制入口包缓存器 130 向出口节点发送数据, 包括: 当节点信用大于零, 队列信用大于或者等于零, 且输出队列为非空时, 向出口节点发送数据。

入口管理逻辑单元 110 接收出口节点周期性地重新分配的节点信用, 并转发给入口节点信用计数器 (NCCI) 120。入口管理逻辑单元 110 从入口包缓存器 130 获取消耗的节点信用, 并在每个周期中实时计算节点信用余额。当节点信用余额达到阈值时, 向入口包缓存器发送停止发送数据的反压信息, 通过织状交换单元 FSE 向出口节点发送节点信用请求。

当入口管理逻辑单元 110 以信元方式向出口节点发送数据时, 每个信元包括数据字节数为 N_C , 节点信用= 信元个数*每个信元的数据字节数 N_C 。

如果从入口包缓存器 130 发出 K 字节 (bytes) 的信元, 则节点信用余额 $NCCI[Y]$ = 发送信元前的节点信用余额 $NCCI[Y]$ -发送的信

元的字节数量 K 。如果 $NCCI[Y]=0$ ，则发送节点信用请求。

参见图 5 所示的一种流量管理系统的应用示意图，其设于交换网络的出口节点上，即图 5 中的出口流量管理系统 200，每个出口节点均包括出口包缓存器 230 (Packet Buffer, PB)，出口流量管理系统 200 包括出口管理逻辑单元 210 和出口节点信用计数器 220。

出口管理逻辑单元 210 用于周期性地将出口总流量划分为多个节点信用，并通过织状交换单元 FSE 一一分配给向其发送数据的多个入口节点。

出口节点信用计数器 220 用于存储从每个入口节点接收到的数据数量。

进一步的，出口管理逻辑单元还用于根据拥塞信息和/或入口节点的节点信用请求，增加、减少或者维持节点信用；其中，拥塞信息来自织状交换单元 FSE 和/或出口包缓存器 230。

具体的，出口管理逻辑单元 210 用于统计从每个入口节点接收到的信元的字节数量，并转发给出口节点信用计数器 220。如果从入口节点 NodeX 收到 K 字节的信元 (cell)，则收到的信元的字节数量增加 K ，即 $CredCell[X]=+K$ 。

对于任意一个入口节点，如果出口管理逻辑单元 210 接收到拥塞信息，则减少的节点信用 $CredCells$ =当前的节点信用 $CredCells$ -节点信用减少值 $NodeDes$ 。

如果出口管理逻辑单元 210 收到节点信用请求，则增加的节点信用 $CredCells$ =当前的节点信用 $CredCells$ +节点信用增加值 $NodeInc$ 。

否则，节点信用 $CredCells$ 不变。

其中，节点信用减少值 $NodeDes$ 和节点信用增加值 $NodeInc$ 都可以根据出口节点的带宽和老化时间等参数确定，例如通过对 Fabric

分布式架构进行系统仿真得到，并作为初始化配置在 SDK 中，下发到出口节点的芯片中。

出口节点根据拥塞信息和/或入口节点的节点信用请求，可以根据实际情况，例如综合入口节点的业务情况、优先级等因素，采用以下方式之一或者组合对节点信用进行调整：增加、减少或者维持节点信用，不作限定。无论如何调整，在每个周期内出口节点向所有入口节点分配的节点信用总和不超过该出口节点的出口总流量。

本发明实施例还提供一种织状网络处理器 FNP，其设于交换网络中，织状网络处理器包括入口包缓存器 130 和出口包缓存器 230，织状网络处理器还包括图 4 中的入口流量管理系统 100，以及图 5 中的出口流量管理系统 200。

图 4 和图 5 所示的流量管理系统一起，显示一个出口节点以及向该出口节点发送数据的一个入口节点的流量管理过程，如图 6 所示。

作为一种示例，例如图 7 所示的基于 Fabric 的交换网络架构，该交换网络架构包括多个节点，每个节点具有一个织状网络处理器 FNP（A.1 至 A.M，其中， $M \geq 1$ ），织状网络处理器 FNP 之间通过 N 个织状交换单元 FSE（C.1 至 C.N，其中， $N \geq 1$ ）交换数据。每个织状网络处理器 FNP 既是入口节点，也是出口节点。

为了便于说明，下文以本端织状网络处理器 FNP 与对端织状网络处理器 FNP 之间交换信元为例进行说明，两个织状网络处理器 FNP 结构相同。

当本端织状网络处理器 FNP 作为入口节点时，对端织状网络处理器 FNP 作为出口节点。

当本端织状网络处理器 FNP 作为出口节点时，那么对端织状网络处理器 FNP 就作为入口节点。

图 8 所示的本端织状网络处理器 FNP 包括入口包缓存器 130 和出口包缓存器 230。

本端织状网络处理器 FNP 还包括入口节点信用管理模块 (Node Credit Manager Ingress, NCMI) 100a 和出口节点信用管理模块 (Node Credit Manager Egress, NCME) 200a, 分别用于处理该本地织状网络处理器 FNP 的节点信用 NC。

其中, NCMI100a 的主要功能包括: 本端入口节点接收对端出口节点分配的节点信用, 向入口包缓存器 130 发送停止发送数据的反压信息, 以及发送节点信用请求到对端出口节点。

NCME200a 的主要功能包括: 本端出口节点分配节点信用到对端入口节点, 接收来自织状交换单元 FSE 和/或出口包缓存器 230 的拥塞信息, 以及接收对端出口节点的节点信用请求。

NCMI100a 例如图 4 中的入口流量管理系统 100, 每个入口管理逻辑单元 110 用于从入口包缓存器 130 获取消耗的节点信用, 记录一个老化时间 (EligTime) 内本端入口节点发往对端出口节点的信元数量。图 4 中的入口流量管理系统 100 仅显示了一个入口节点信用计数器 120 (NCCI), 在 NCMI100a 中, 入口节点信用计数器 120 (NCCI) 的数量为该网络交换系统中所有对端出口节点的数量。

如果节点信用 $NCCI[EgressFNP] > 0$, 则本端入口节点可以发送单播报文 (PktUc cells) 到对应的对端出口节点。例如, 如果本端入口节点 NodeX 发送 K 字节 bytes 的信元 (cell) 到对端出口节点 NodeY, 则入口管理逻辑单元 110 记录节点信用余额 $NCCI[Y] = \text{发送信元前的节点信用余额 } NCCI[Y] - \text{发送的信元的字节数量 } K$ 。在前述的老化时间 (EligTime) 内, 如果节点信用余额 $NCCI[EgressFNP]$ 减为零, 则将节点信用请求信号 (Node Credit Request, NCR) [EgressFNP] 置

为 1，并发送到织状交换单元（FSE）的织状接口（入口处）510（FI Ingress），然后转发给对端出口节点。同时，发送反压信息给入口包缓存器 130，停止发送信元。

如果织状交叉单元 FSE 上的包缓存器的缓存超过水线导致数据拥塞，或者，各线路上延时导致出口节点出现数据拥塞，则将一路反压信息通过 NCMI100a 发送到入口包缓存器 130，停止向外发送数据包，等到流量正常后再继续向外发送数据包。

NCME200a 如图 5 中的出口流量管理系统 200，用于将分配给对端入口节点的节点信用 NC 发送到织状接口（入口处）510（FI Ingress），并转发给对端入口节点。

NCME200a 统计对端入口节点的节点信元，每当出口包缓存器 230（Packet Buffer, PB）收到 K 字节信元的数据报文，则统计收到的信元的字节数量 $CredCell[X]=+K$ 。

如果在前述的老化时间（EligTime）内，NCME200a 收到来自织状交叉盘 FSE 的 FSE 拥塞信息或者本端出口缓存器 230 发送过来的包缓存拥塞信息，即将拥塞信息置为 1，当下一个老化时间，发送给对端入口节点的节点信用则修改为 $CreditCells-NodeDec$ 。

如果 NCME200a 收到对端入口节点发送过来的节点信用请求信号（Node Credit Request, NCR），则下一个老化时间，发送给对端入口节点的节点信用则修改为 $CreditCells+NodeInc$ ；如果 NCME200a 没有收到拥塞信息，也没有收到 NCR 信号，则下个老化时间，发送给对端入口节点的节点信用则仍为 $CreditCells$ 。

每个老化时间后，NCME200a 重新发送给各对端入口节点相应的节点信用（ $CredCells[IngressFNP]$ ），并清空本端的节点信用 $CredCells$ 计数。而对端入口节点会把收到的节点信用（ $CredCells[EgressFNP]$ ）

存储在对端的入口节点信用计数器（NCCI[EgressFNP]）。

图 9 所示为本端织状网络处理器 FNP 的一个具体示例，在图 8 所示的本端织状网络处理器 FNP 的基础上，本端织状网络处理器 FNP 还可以包括组装模块 300 和切片模块 400，组装模块 300 用于将本端织状网络处理器 FNP 的所有控制信息都组装成数据包发到织状交换单元 FSE，数据包的大小根据实际应用情况确定，例如 256 Bytes 字节的标准数据包，切片模块 400 用于将织状交换单元 FSE 下发的所有标准数据包切成控制信息给本端织状网络处理器 FNP，其中，控制信息包括节点信用和节点信用请求，以及队列信用和队列信用请求。

组装模块 300 向织状接口（入口处）510（FI Ingress）发送发往入口的控制信息，并从织状接口（入口处）510（FI Ingress）接收控制信息反压。

切片模块 400 向织状接口（出口处）520（FI Egress）发送控制信息反压响应，从织状接口（出口处）520（FI Egress）接收发往出口的控制信息（MsgDest）。

在本端入口节点，切片模块 400 从织状接口（出口处）520（FI Egress）接收发往出口的控制信息（MsgDest），其中携带对端出口节点分配的节点信用，将节点信用发送到 NCMI100a。

切片模块 400 从织状接口（出口处）520（FI Egress）接收发往出口的控制信息（MsgDest），其中，发往出口的控制信息（MsgDest）携带对端入口节点的节点信用请求和/或 FSE 拥塞信息，将对端入口节点的节点信用请求和/或 FSE 拥塞信息发送到 NCME200a。

在上述基于队列信用的流量管理之外，本实施例也可以将节点信用（NC）机制和队列信用（QC）机制相结合，例如前述实施例的步骤 S250，可以避免流量过多导致织状交换单元和/或出口节点瘫痪。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行计算机程序指令时，全部或部分地产生按照本申请实施例的流程或功能。计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(Digital Subscriber Line, DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。计算机可读存储介质可以是计算机能够读取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。可用介质可以是磁性介质，(例如，软盘、硬盘、磁带)、光介质(例如，数字通用光盘(Digital Video Disc, DVD))或者半导体介质(例如，固态硬盘(Solid State Disk, SSD))等。

本发明不局限于上述实施方式，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也视为本发明的保护范围之内。本说明书中未作详细描述的内容属于本领域专业技术人员公知的现有技术。

权 利 要 求 书

1. 一种流量管理方法，应用于交换网络，其特征在于：

入口节点向出口节点发送的数据总量不超过其分配到的节点信用，且向所述出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量。

2. 如权利要求 1 所述的流量管理方法，其特征在于：

所述节点信用是网管下发的或者由所述出口节点周期性地重新分配的。

3. 如权利要求 2 所述的流量管理方法，其特征在于：

在每个所述周期中，所述入口节点实时计算节点信用余额，其中， $\text{节点信用余额} = \text{所述节点信用} - \text{已发送的数据数量}$ ；当所述节点信用余额达到阈值时，停止向所述出口节点发送数据，以及向外发送节点信用请求。

4. 如权利要求 1 至 3 任一项所述的流量管理方法，其特征在于：

所述入口节点配置输出队列和队列信用；

当所述节点信用大于零，队列信用大于或者等于零，且输出队列为非空时，向所述出口节点发送数据。

5. 一种流量管理方法，应用于交换网络，其特征在于：

出口节点周期性地将出口总流量划分为多个节点信用，并一一分配给向其发送数据的多个入口节点。

6. 如权利要求 5 所述的流量管理方法，其特征在于：

所述出口节点根据拥塞信息和/或所述入口节点的节点信用请求，增加、减少或者维持所述节点信用；

所述拥塞信息来自所述交换网络的织状交换单元和/或所述出口节点的出口包缓存器。

7. 一种流量管理系统，其设于交换网络的入口节点上，每个入口节点均包括入口包缓存器，其特征在于，所述流量管理系统包括：

入口节点信用计数器，其用于存储分配给该入口节点的节点信用，其中，向所述出口节点发送数据的多个入口节点的节点信用总和不超过出口节点的出口总流量；

入口管理逻辑单元，其用于控制所述入口包缓存器向所述出口节点发送的数据总量不超过所述节点信用。

8. 如权利要求 7 所述的流量管理系统，其特征在于：所述节点信用是网管下发的或者由所述出口节点周期性地重新分配的。

9. 如权利要求 8 所述的流量管理系统，其特征在于：

在每个所述周期中，所述入口管理逻辑单元用于实时计算节点信用余额，其中，节点信用余额=所述节点信用 - 已发送的数据数量；当所述节点信用余额达到阈值时，向所述入口包缓存器发送停止发送数据的反压信息，以及向外发送节点信用请求。

10. 如权利要求 7 所述的流量管理系统，其特征在于：

所述入口节点还包括入口流量管理器，入口流量管理器用于配置输出队列和队列信用；或者，所述入口管理逻辑单元还用于配置输出队列和队列信用；

所述入口管理逻辑单元还用于控制所述入口包缓存器发送数据，包括：当所述节点信用大于零，队列信用大于或者等于零，且输出队列为非空时，向所述出口节点发送数据。

11. 一种流量管理系统，其设于交换网络的出口节点上，每个出口节点均包括出口包缓存器，其特征在于，所述流量管理系统包括：

出口管理逻辑单元，其用于统计从每个入口节点接收到的数据数量，以及周期性地将出口总流量划分为多个节点信用，并一一分配给

向其发送数据的多个入口节点；

出口节点信用计数器，其用于存储从每个入口节点接收到的数据数量。

12. 如权利要求 11 所述的流量管理系统，其特征在于：

所述出口管理逻辑单元还用于根据拥塞信息和/或所述入口节点的节点信用请求，增加、减少或者维持所述节点信用；

所述拥塞信息来自织状交换单元或者所述出口包缓存器。

13. 一种织状网络处理器，其设于交换网络中，所述织状网络处理器包括入口包缓存器和出口包缓存器，其特征在于：

所述织状网络处理器还包括如权利要求 7 至 10 任一项所述的流量管理系统，以及如权利要求 10 或 11 所述的流量管理系统。

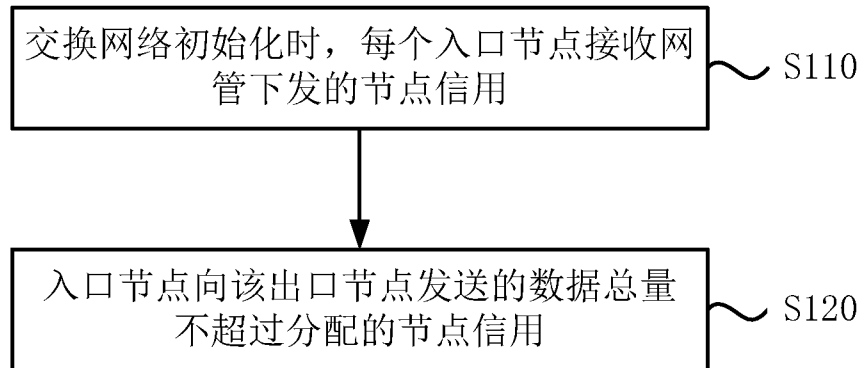


图 1

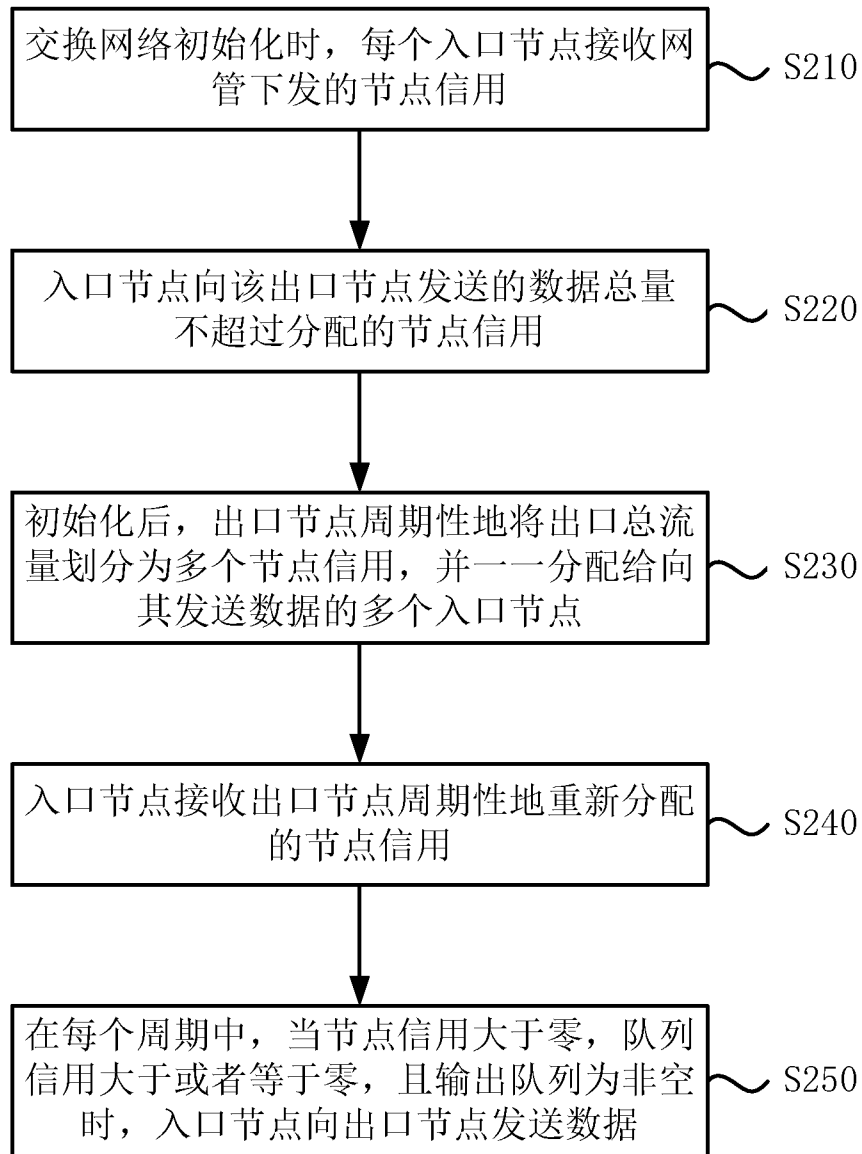


图 2

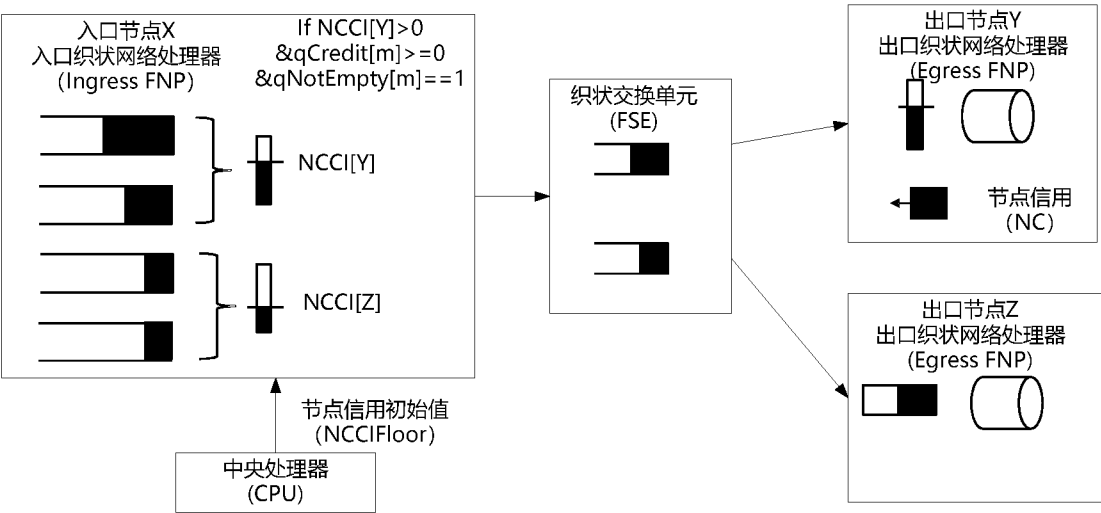


图 3

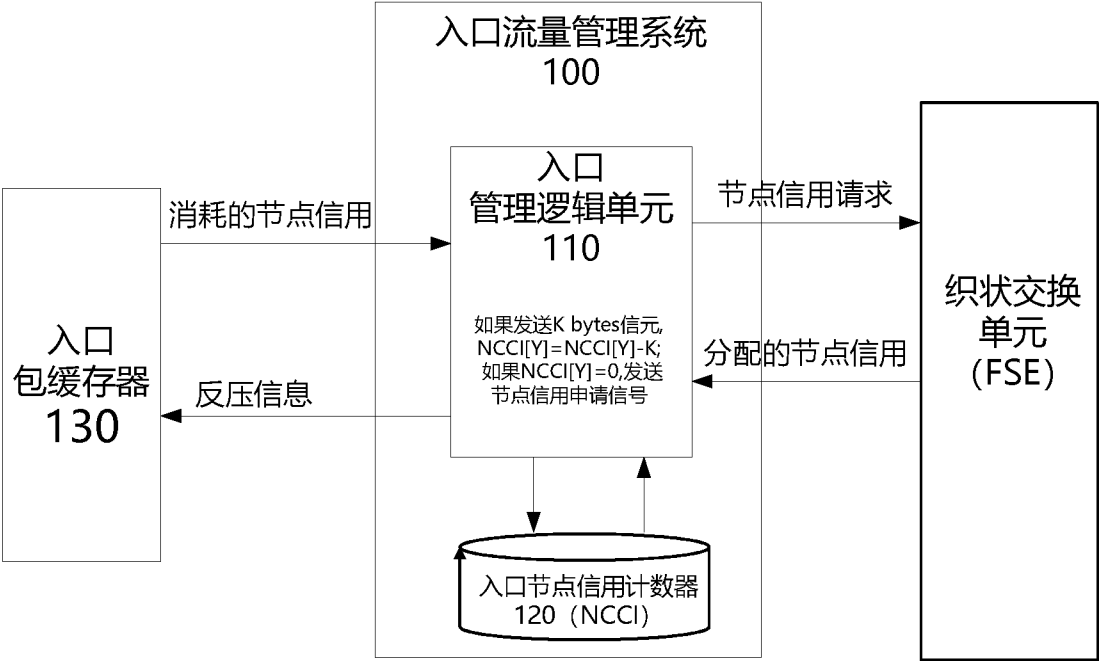


图 4

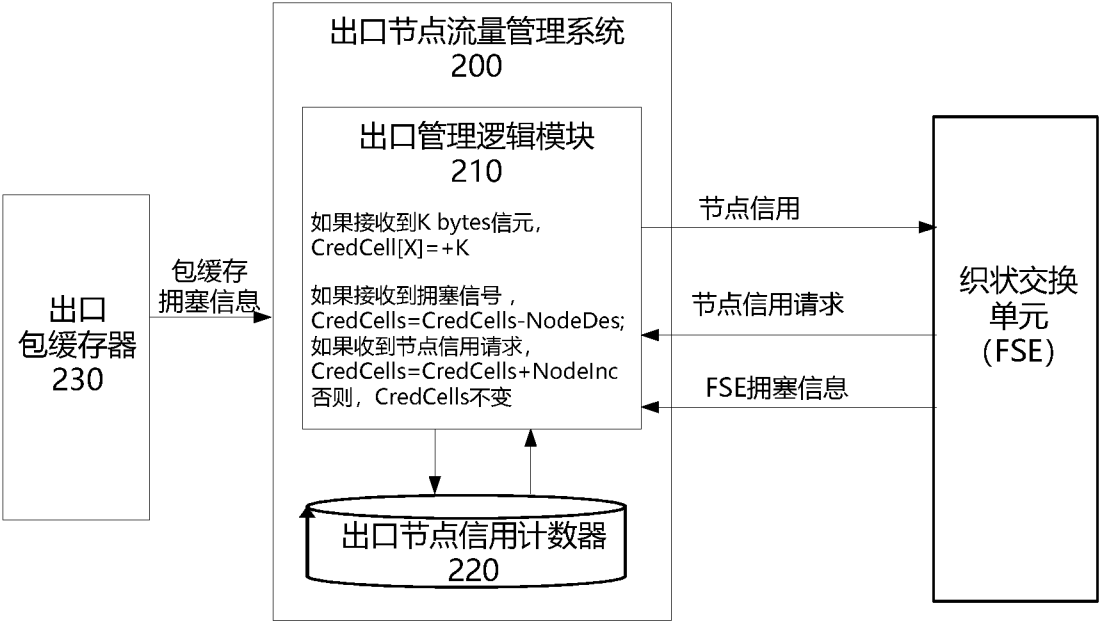


图 5

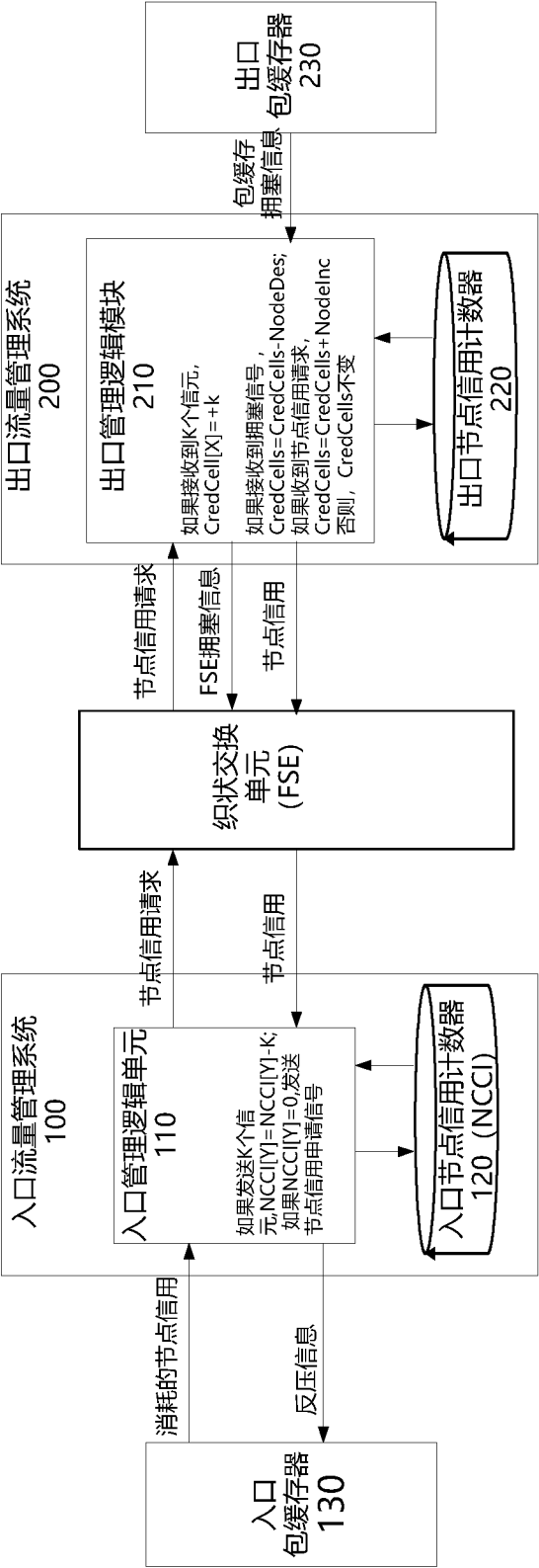


图6

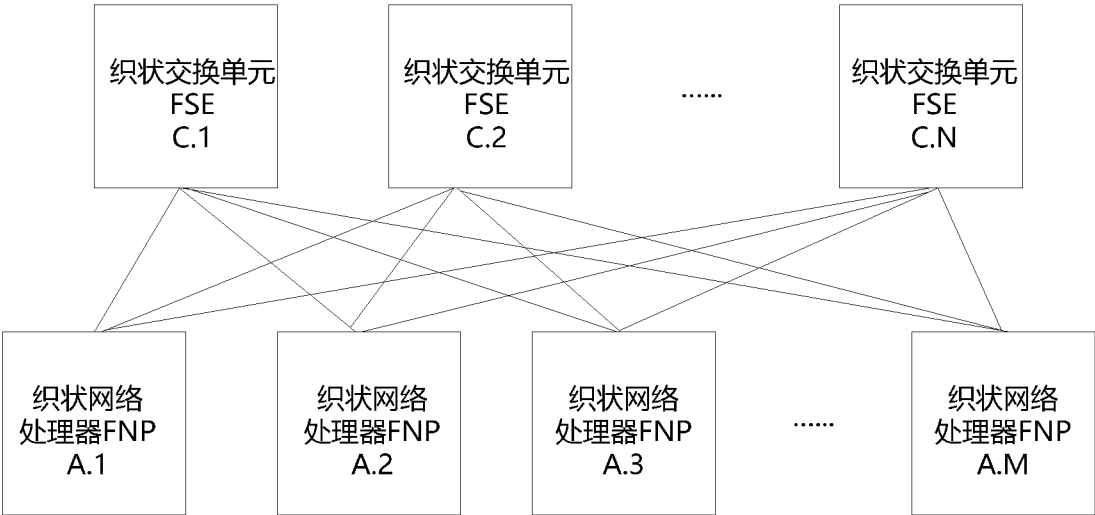


图 7

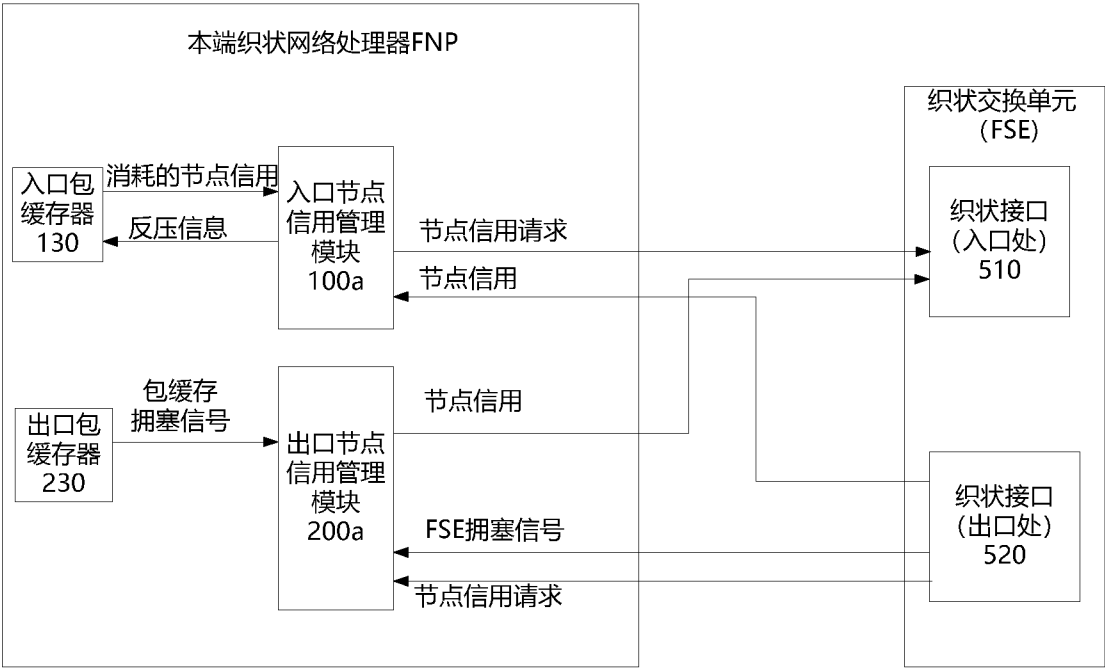


图 8

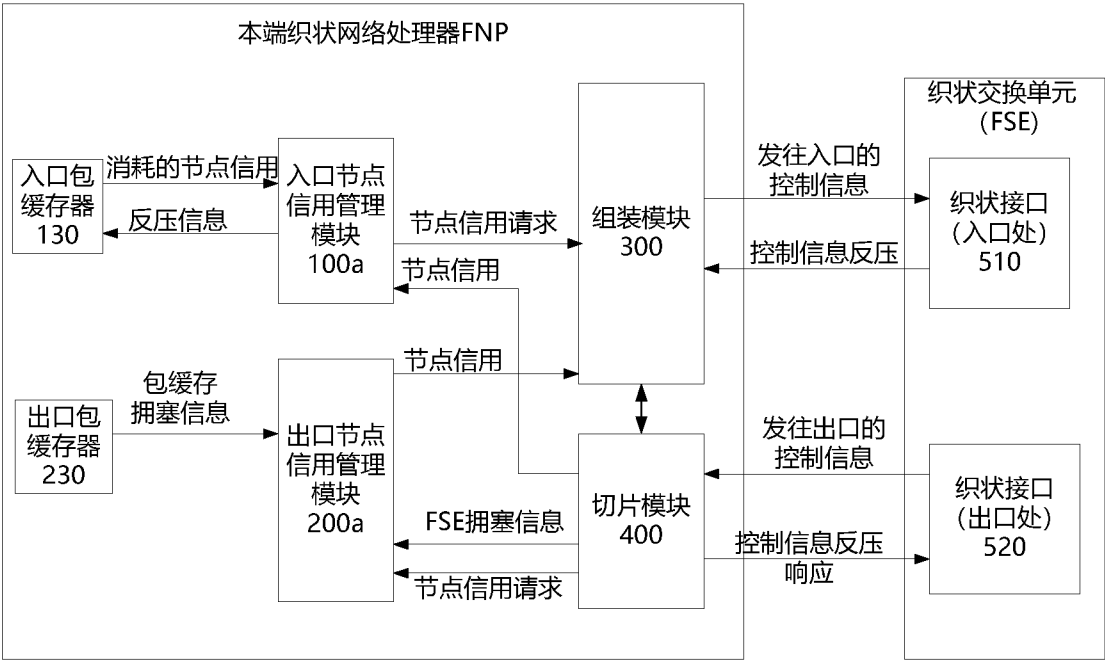


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/096675

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/801(2013.01)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)

CNPAT; CNKI; WPI; EPODOC; IEEE: 流量, 控制, 管理, 交换机, 节点, 信用, 总, 分配, 入口, 输入, 出口, 输出, 多个, 织状, 交织, traffic, manage, control, switch, node, credit, sum, allocate, outlet, inlet, multiple, woven, fabric

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109787910 A (FIBERHOME TELECOMMUNICATION TECHNOLOGIES CO., LTD. et al.) 21 May 2019 (2019-05-21) claims 1-13	1-13
X	CN 101123587 A (H3C TECHNOLOGIES CO., LTD.) 13 February 2008 (2008-02-13) description, page 7, paragraph 4 to page 13, last paragraph	1-13
A	CN 101018156 A (HUAWEI TECHNOLOGIES CO., LTD.) 15 August 2007 (2007-08-15) entire document	1-13
A	US 2007291643 A1 (CISCO TECHNOLOGY, INC.) 20 December 2007 (2007-12-20) entire document	1-13
A	US 2014269294 A1 (CISCO TECHNOLOGY, INC.) 18 September 2014 (2014-09-18) entire document	1-13
A	CN 102487358 A (JUNIPER NETWORKS INC.) 06 June 2012 (2012-06-06) entire document	1-13



Further documents are listed in the continuation of Box C.



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

05 November 2019

Date of mailing of the international search report

28 November 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/096675

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109787910	A	21 May 2019	None			
CN	101123587	A	13 February 2008	None			
CN	101018156	A	15 August 2007	None			
US	2007291643	A1	20 December 2007	AT	548830	T	15 March 2012
				EP	2027672	A4	23 December 2009
				CN	101467393	A	24 June 2009
				EP	2027672	A2	25 February 2009
				WO	2007145845	A2	21 December 2007
US	2014269294	A1	18 September 2014	CN	105052095	A	11 November 2015
				EP	2974183	A1	20 January 2016
				WO	2014158246	A1	02 October 2014
CN	102487358	A	06 June 2012	CN	105323185	A	10 February 2016
				EP	3297234	A1	21 March 2018
				US	2017257328	A1	07 September 2017
				EP	2461529	A2	06 June 2012
				US	2012140626	A1	07 June 2012

国际检索报告

国际申请号

PCT/CN2019/096675

A. 主题的分类 H04L 12/801(2013.01)i 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																							
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) H04L 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)) CNPAT;CNKI;WPI;EPDOC;IEEE:流量, 控制, 管理, 交换机, 节点, 信用, 总, 分配, 入口, 输入, 出口, 输出, 多个, 织状, 交织, traffic, manage, control, switch, node, credit, sum, allocate, outlet, inlet, multiple, woven, fabric																							
C. 相关文件 <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>PX</td> <td>CN 109787910 A (烽火通信科技股份有限公司等) 2019年 5月 21日 (2019 - 05 - 21) 权利要求1-13</td> <td>1-13</td> </tr> <tr> <td>X</td> <td>CN 101123587 A (杭州华三通信技术有限公司) 2008年 2月 13日 (2008 - 02 - 13) 说明书第7页第4段至第13页最后1段</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>CN 101018156 A (华为技术有限公司) 2007年 8月 15日 (2007 - 08 - 15) 全文</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>US 2007291643 A1 (CISCO TECHNOLOGY, INC.) 2007年 12月 20日 (2007 - 12 - 20) 全文</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>US 2014269294 A1 (CISCO TECHNOLOGY, INC.) 2014年 9月 18日 (2014 - 09 - 18) 全文</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>CN 102487358 A (丛林网络公司) 2012年 6月 6日 (2012 - 06 - 06) 全文</td> <td>1-13</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 109787910 A (烽火通信科技股份有限公司等) 2019年 5月 21日 (2019 - 05 - 21) 权利要求1-13	1-13	X	CN 101123587 A (杭州华三通信技术有限公司) 2008年 2月 13日 (2008 - 02 - 13) 说明书第7页第4段至第13页最后1段	1-13	A	CN 101018156 A (华为技术有限公司) 2007年 8月 15日 (2007 - 08 - 15) 全文	1-13	A	US 2007291643 A1 (CISCO TECHNOLOGY, INC.) 2007年 12月 20日 (2007 - 12 - 20) 全文	1-13	A	US 2014269294 A1 (CISCO TECHNOLOGY, INC.) 2014年 9月 18日 (2014 - 09 - 18) 全文	1-13	A	CN 102487358 A (丛林网络公司) 2012年 6月 6日 (2012 - 06 - 06) 全文	1-13
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 109787910 A (烽火通信科技股份有限公司等) 2019年 5月 21日 (2019 - 05 - 21) 权利要求1-13	1-13																					
X	CN 101123587 A (杭州华三通信技术有限公司) 2008年 2月 13日 (2008 - 02 - 13) 说明书第7页第4段至第13页最后1段	1-13																					
A	CN 101018156 A (华为技术有限公司) 2007年 8月 15日 (2007 - 08 - 15) 全文	1-13																					
A	US 2007291643 A1 (CISCO TECHNOLOGY, INC.) 2007年 12月 20日 (2007 - 12 - 20) 全文	1-13																					
A	US 2014269294 A1 (CISCO TECHNOLOGY, INC.) 2014年 9月 18日 (2014 - 09 - 18) 全文	1-13																					
A	CN 102487358 A (丛林网络公司) 2012年 6月 6日 (2012 - 06 - 06) 全文	1-13																					
☐ 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(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																							
国际检索实际完成的日期 2019年 11月 5日		国际检索报告邮寄日期 2019年 11月 28日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局(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		授权官员 张倩 电话号码 86-(10)-53961574																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/096675

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	109787910	A	2019年 5月 21日	无			
CN	101123587	A	2008年 2月 13日	无			
CN	101018156	A	2007年 8月 15日	无			
US	2007291643	A1	2007年 12月 20日	AT	548830	T	2012年 3月 15日
				EP	2027672	A4	2009年 12月 23日
				CN	101467393	A	2009年 6月 24日
				EP	2027672	A2	2009年 2月 25日
				WO	2007145845	A2	2007年 12月 21日
US	2014269294	A1	2014年 9月 18日	CN	105052095	A	2015年 11月 11日
				EP	2974183	A1	2016年 1月 20日
				WO	2014158246	A1	2014年 10月 2日
CN	102487358	A	2012年 6月 6日	CN	105323185	A	2016年 2月 10日
				EP	3297234	A1	2018年 3月 21日
				US	2017257328	A1	2017年 9月 7日
				EP	2461529	A2	2012年 6月 6日
				US	2012140626	A1	2012年 6月 7日

表 PCT/ISA/210 (同族专利附件) (2015年1月)