(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 5 月 7 日 (07.05.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/087523 A1

(51) 国际专利分类号:
H04L 29/08 (2006.01)

(21) 国际申请号: PCT/CN2018/113789

(22) 国际申请日: 2018 年 11 月 2 日 (02.11.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 阿里巴巴集团控股有限公司 (**ALIBABA GROUP HOLDING LIMITED**) [—/CN]; 开曼群岛英属开曼群岛大开曼资本大厦一座四层847号邮箱, Grand Cayman (KY)。

(72) 发明人: 马可 (**MA, Ke**); 美国华盛顿州默瑟岛东南57街9105号, Washington 98040 (US)。 欧阳晋 (**OUYANG, Jin**); 中国北京市朝阳区望京东园四区9号楼阿里中心A座, Beijing 100102 (CN)。

(74) 代理人: 北京清源汇知识产权代理事务所（特殊普通合伙） (**BEIJING TSINGYUANHUI INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区苏州街一号3层362, Beijing 100080 (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, 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,

(54) **Title:** NETWORK COMMUNICATION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

(54) **发明名称**: 网络通信的方法、装置及电子设备

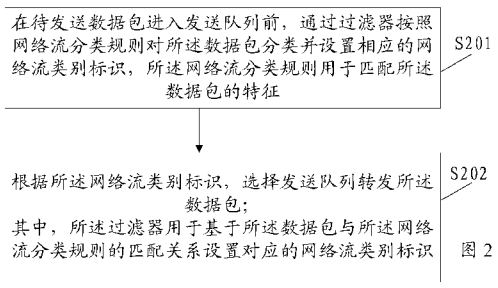

图 2

S201 CLASSIFY, BEFORE DATA PACKETS TO BE SENT ENTER SENDING QUEUES, THE DATA PACKETS BY MEANS OF A FILTER ACCORDING TO A NETWORK FLOW CLASSIFICATION RULE AND SET CORRESPONDING NETWORK FLOW CLASS IDENTIFIERS, THE NETWORK FLOW CLASSIFICATION RULE BEING USED FOR MATCHING THE FEATURES OF THE DATA PACKETS

S202 SELECT, ACCORDING TO THE NETWORK FLOW CLASS IDENTIFIERS, SENDING QUEUES TO FORWARD THE DATA PACKETS, WHEREIN THE FILTER IS USED FOR SETTING CORRESPONDING NETWORK FLOW CLASS IDENTIFIERS ACCORDING TO A MATCHING RELATIONSHIP BETWEEN THE DATA PACKETS AND THE NETWORK FLOW CLASSIFICATION RULE

(57) **Abstract:** A network communication method and apparatus, an electronic device and a storage device. The method comprises: classifying, before data packets to be sent enter sending queues, the data packets by means of a filter according to a network flow classification rule and setting corresponding network flow class identifiers, the network flow classification rule being used for matching the features of the data packets (S201); and selecting, according to the network flow class identifiers, sending queues to forward the data packets, wherein the filter is used for setting corresponding network flow class identifiers according to a matching relationship between the data packets and the network flow classification rule (S202). The method uses network flow class identifiers for multi-queue sending to implement flow control, thereby increasing the number of network flow classes sent by means of multiple queues.

(57) **摘要**: 一种网络通信的方法、装置、电子设备及存储设备，所述方法包括：在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征（S201）；根据所述网络流类别标识，选择发送队列转发所述数据包；其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识（S202）。该方法使用网络流类别标识进行多队列发送实现流量控制，提高了通过多队列发送的网络流分类的数目。

WO 2020/087523 A1

[见续页]

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))。

网络通信的方法、装置及电子设备

技术领域

本申请涉及网络通信领域，具体涉及一种网络通信的方法、装置、电子设备及存储设备。本申请还涉及第二种网络通信的方法及装置。

背景技术

Linux 作为开源软件，因其具有灵活性和可扩展性，符合作为集中式服务平台提供可伸缩的共享计算能力的云计算的发展趋势，是云计算业界广泛采用的基础软件。云计算提供的计算能力，包括处理器、内存、存储和网络接口，需要处理海量数据，因此网络数据包的流控方案和吞吐量显得尤其重要。

Linux 平台的网络流控方案一般是基于 TC（Traffic Control，或流量控制器）进行内核流量控制，TC 模块支持按照 socket 类别标识来分类网络包，并通过令牌桶技术来限定特定网络流可用带宽的最大和最小值从而达到进行流控的目的。TC 模块由于使用了全局自旋锁（spin_lock），在任意时刻仅允许一个处理器核处理一个网络包，为提高数据包处理的并发性和吞吐量，现有的 TC 技术支持多队列网络流控方案，具体的，通过 Linux 的 mq_prio 模块实现基于网络包的 socket 优先级来分发网络包到多个网卡发送和接收队列上，从而允许多个处理器核分别处理位于多个队列上的网络包，其中，socket 优先级是指 IP 协议中的定义的 TOS（Type of Service）优先级，共 8 个级别，这样提高了数据包处理的并发性以及提高数据包吞吐量，数据包吞吐量以 PPS（Packet per Second，或每秒钟处理的网络数据包数量）表示。

现有的 Linux TC 网络流控技术，虽然支持多队列，但是仅能使用 socket 优先级对网络流分类，由于协议限定仅支持 8 个级别，这极大地限制了网络流分类的能力。由于无法根据网络流的类别标识对网络流进行灵活分类，从而不能按照网络流分发网络包到多个发送和接收队列上，导致实际应用中存在网络数据包转发性能瓶颈的问题。例如，现有的 Linux TC 技术应用于多种网络应用带宽隔离的场景时，网络处理吞吐量存在性能瓶颈，在商用多核处理器（如 Intel Xeon CPU E5-2630 2.30GHz）上，吞吐量无法超过 600K PPS，即每秒处理的网络数据包数不超过 60 万个包，因而，在 10Gbps（或者更高传输速率的）网卡以及网络包大小小于 2K 字节的情况下仅能利用部分可用带宽，例如，包长为 512

字节大小的网络包仅能使用 2G 带宽，利用率为 20%，并且扩充处理器核数也不能解决此性能瓶颈问题。

发明内容

本申请提供一种用于多队列网络流控的方法，以解决现有的 Linux TC 网络流控技术中存在的由于无法根据网络流的类别标识对网络流进行灵活分类，从而导致的实际应用中存在网络数据包转发性能瓶颈的问题。

本申请提供的一种网络通信的方法，包括:

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

可选的，所述发送队列，是关联了排队规则的序号连续的类别队列；其中，所述类别队列，使用主序列号和从序列号进行标识，所述类别队列的主序列号用于关联网络设备；所述排队规则，关联了所述过滤器，使用主序列号和从序列号进行标识，所述排队规则的主序列号取值于所述类别队列的从序列号。

可选的，所述数据包的特征包括如下至少之一：IP 五元组信息、socket 信息。

可选的，所述过滤器，是基于 Linux 流量控制的过滤器;

相应的，所述通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，包括:

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征，按照所述网络流分类规则对所述数据包分类;

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征与网络流分类规则的匹配关系，设置所述数据包的网络类别标识;

所述数据包的 IP 头部特征至少包括下述任一信息: 源 IP、目的 IP、源 socket 端口、目的 socket 端口、协议类型。

可选的，所述根据所述网络流类别标识，选择发送队列转发所述数据包，包括:

通过在 Linux 内核引入的发送队列调度模块，根据所述排队规则关联的过滤器的网络流类别标识选择发送队列。

可选的，包括:

提供 Linux 控制流量命令的扩展命令，用于将所述网络流类别标识关联到所述排队规则。

可选的，所述根据所述排队规则关联的过滤器的网络流类别标识选择发送队列，包括:

根据数据包的网络流类别标识，确定所述网络流类别标识对应的排队规则;

根据所述排队规则的主序列号，查找所述排队规则的主序列号到所述发送队列号的哈希表，确定发送队列号。

可选的，还包括:

根据选择的发送队列关联的排队规则对应的网络流的优先级，确定所述数据包离队次序，所述排队规则，用于对所述发送队列的数据包排序。

可选的，还包括:

各发送队列根据进入队列的数据包的离队次序，依次发送所述数据包。

本申请还提供一种网络通信的方法，包括:

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器;

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识;

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

可选的，所述网络流分类规则，包含用于匹配数据包头部特征的信息；所述数据包头部特征，至少包括下述任一信息：源 IP 地址、目的 IP 地址、数据包接收端对应的 socket 和协议。

可选的，所述根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包，包括:

通过所述接收端的排队规则关联的过滤器，将所述数据包重定向到中间功能模块;

根据所述网络流类别标识，选择所述中间功能模块的不同发送队列，转发

所述数据包。

可选的，还包括:

使用接收线程重用技术重用接收端线程上下文。

可选的，所述接收线程，是用于处理接收端的数据包的接收的线程；相应的，所述使用接收线程重用技术重用接收端线程上下文，包括:

直接使用所述接收线程，按照所述接收端的排队规则将所述数据包转发到中间功能模块；所述中间功能模块，用于接收重定向来的数据包，按照所述排队规则对应的网络流的优先级将所述重定向来的数据包发送回被重定向前的网络设备。

此外，本申请还提供一种网络通信的装置，包括:

标识单元，用于在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

转发单元，用于根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

此外，本申请还提供一种网络通信的装置，包括:

网络流分类单元，用于按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器;

标识单元，用于基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识;

转发单元，用于根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

此外，本申请还提供一种电子设备，包括:

存储器，以及处理器;

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令:

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

根据所述网络流类别标识，选择发送队列转发所述数据包；

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

此外，本申请还提供一种电子设备，包括：

存储器，以及处理器；

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令：

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器；

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识；

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

此外，本申请还提供一种存储设备，存储有指令，所述指令能够被处理器加载并执行以下步骤：

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征；

根据所述网络流类别标识，选择发送队列转发所述数据包；

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

此外，本申请还提供一种存储设备，存储有指令，所述指令能够被处理器加载并执行以下步骤：

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器；

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识；

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

与现有技术相比，本申请具有以下优点:

本申请提供的网络通信的方法、装置、电子设备及存储设备，通过使用网络流类别标识来确定发送队列由于所述网络流类别标识的可用空间远大于socket优先级的数目，因此大大提高了通过多队列发送的网络流分类的数目，从而解决了现有的网络流量控制技术中由于不支持网络流灵活分类而导致的数据包处理的性能瓶颈问题。进一步的，将网络流控模块关联到多个队列上，实现了对基于网络流类别标识进行网络流量控制的支持。

附图说明

图 1 是本申请其一实施例提供的一种网络通信的方法的实际部署环境的示意图;

图 2 是本申请其一实施例提供的一种网络通信的方法的处理流程图;

图 3 是本申请其一实施例提供的一种网络通信的方法的发送队列分组示意图;

图 4 是本申请其一实施例提供的一种网络通信的方法的数据包流向示意图;

图 5 是本申请提供的第二种网络通信的方法实际部署的网络环境示意图;

图 6 是本申请提供的第二种网络通信的方法实施例的处理流程图;

图 7 是本申请提供的第二种网络通信的方法实施例的数据包流向示意图;

图 8 是本申请提供的一种网络通信的装置实施例的示意图;

图 9 是本申请提供的第二种网络通信的装置实施例的示意图;

图 10 是本申请提供的一种电子设备实施例的示意图。

具体实施方式

在下面的描述中阐述了很多具体细节以便于充分理解本申请。但是本申请能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本申请内涵的情况下做类似推广，因此本申请不受下面公开的具体实施的限制。

本申请提供一种网络通信的方法、装置、电子设备及存储设备。本申请还提供第二种网络通信的方法、装置、电子设备及存储设备。在下面的实施例中逐一进行详细说明。

本申请第一实施例提供一种网络通信的方法。以下结合图 1 至图 4 对本申请提供的一种网络通信的方法的实施例进行说明。

网络通信中，以类比水流来描述数据包在网络中的传输，从一个特定源发

送到一个特定目的的一系列数据包称为一个网络流（或称为网络数据流，或简称为流）。例如，在一个源 IP 地址和一个目的 IP 地址之间单向传输的具有相同 IP 五元组信息的数据包定义为一个网络流。所谓 IP 五元组信息，包括：源 IP 地址、目的 IP 地址、源端口、目的端口、协议号。可以使用不同的标识来标记不同的网络流，用于标记网络流的标识称为网络流类别标识（也称为类别标识，或 classid）。

针对不同的网络流采取不同的转发策略，从而对不同业务提供差异化服务，提高带宽利用率，是云计算和互联网行业节约网络资源的重要手段，因此，网络流的流控方案和吞吐量显得尤为重要。云计算和互联网行业中常用的 Linux 平台，基于 Linux TC（Traffic Control，或流量控制器）进行内核流量控制。网络数据包在 Linux 内的转发流程，从入口（输入网卡）接收进来，查找路由，确定是发给本机的，转发给上层协议处理，例如 TCP，如果确定需要转发的，则从出口（输出网卡）转发出去，Linux 原生的 TC 功能的网络流控是在出口进行控制，TC 模块由于使用了全局自旋锁（spin_lock），在任意时刻仅允许一个处理器核处理一个网络包，为了提高数据包处理的并发性和吞吐量，现有的 TC 技术提供多队列网络流控方案，支持按照 socket 类别标识来分类网络包，并通过令牌桶技术来限定特定网络流可用带宽的最大和最小值从而达到进行流控的目的。

本申请实施例提供的网络通信的方法，一般部署于数据包的发送端，所述发送端，可以是网络设备，例如，可以是网络接口板（Network Interface Card，或 NIC），网络接口板也称为网卡；所述发送端也可以是类似于所述网络设备的虚拟网络设备，通过虚拟网络设备发送数据包类似于通过网络设备的物理口发送数据包。例如，中间功能模块 IFB（Intermediate Functoinal Block）。所述网络通信的方法，支持基于网络流类别标识对所述网络流类别标识标记的网络流进行流量控制。具体为根据网络流类别标识来分发数据包到多个发送队列上，从而对网络流进行灵活的分类，由于网络流类别标识的可用空间大，因此类别标识的数目远大于 socket 优先级的数目，大大提高了网络流分类的能力，使得网络 PPS 可以随处理器核数线性增长，例如，在配置的 24 核处理器上可以达到 14.4M/s，从而提高了网络带宽利用率，以及数据中心服务器的资源利用率。

下述以基于 Linux TC 实现的所述网络通信的方法的实施例为例，说明本申请提供的网络通信的方法。

在阐述本实施例前 Linux TC 进行流控的基本概念：排队规则、类别、过滤

器。每个网卡都与一个排队规则相联系，当内核将数据包从网卡转发出去，都会将数据包根据网卡对应的排队规则加入到队列中。要实现强大的流量控制，需要功能复杂的队列，即有类别排队规则的队列（Classful），需要使用过滤器（filter）将数据包分成不同的类别（class），再根据不同的排队规则，按照不同顺序发送可分类队列中的数据包，从而实现流量控制。

图 1 示出了所述网络通信的方法实际部署的网络环境，包括：网络应用 101 构造网络数据包，通过 socket 机制发送到 TCP/IP 层 102；TCP/IP 层 102 对所述网络数据包增加 TCP/IP 协议头信息后，发送到 net_cls cGroup 数据包标记模块 103；net_cls cGroup 数据包标记模块 103 对添加了 TCP/IP 协议头信息的网络数据包增加类别标识后，发送到 mq_cls 发送队列调度模块 104；mq_cls 发送队列调度模块 104 选择发送队列将所述增加了 TCP/IP 协议头信息的网络数据包，通过网卡驱动 105 转发到网卡的出接口，例如 ethernet 接口。其中，可以在 mq_cls 发送队列调度模块 104 选择发送队列，控制包含 TCP/IP 协议头信息的网络数据包的入发送队列和出发送队列，来实现网络流量控制。对于网络流的流量控制可以在出口进行控制，包括整形（shaping）、调度（scheduling）、丢包（dropping）等控制，Linux 内核将数据包从网卡转发出去在出口进行流量控制。其中，所谓 shaping，当流量被限制，它的传输速率就被控制在某个值以下，限制值可以大大小于有效带宽，这样可以平滑突发数据流量，使网络更为稳定；所谓 scheduling，通过调度数据包的传输，可以在带宽范围内，按照优先级分配带宽；所谓 dropping 如果流量超过某个设定的带宽，就丢弃数据包。

所述 net_cls cGroup 是对数据包进行分类并标记的子系统，net_cls cgroup 对数据包的标记可以用作所述网络流类别标识，并且 Linux 流量控制 TC（traffic control）能够识别 net_cls cGroup 标记后的数据包。

图 2 是本申请其一实施例提供的网络通信的方法处理流程图。

图 2 所示的网络通信的方法，包括：步骤 S201 至 S202。

步骤 S201，在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征。

所述网络流分类规则，用于匹配所述数据包的特征；其中，所述数据包的特征包括如下至少之一：IP 五元组信息、socket 信息。例如，设定特定的 IP 五元组信息作为网络流分类规则，则匹配到所述特定的 IP 五元组信息的数据包归

类到一个特定的网络流。

本步骤是对匹配上所述网络流分类规则的数据包设置相应的网络流类别标识。

本申请实施例中，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识。例如，过滤器的网络流分类规则包含的用于匹配数据包的特征的信息为目的 IP 地址 192.168.0.1 和目的端口号 80，目的 IP 地址 192.168.0.1 和目的端口号 80 的网络流的网络流类别标识为 10:1；则过滤器将匹配到目的 IP 地址 192.168.0.1 和目的端口号 80 的数据包过滤出来，并针对匹配上所述网络流分类规则的数据包设置上网络流类别标识 10:1。

本申请实施例中，所述过滤器，可以是基于 Linux 流量控制的过滤器；

相应的，所述通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，包括：

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征，按照所述网络流分类规则对所述数据包分类；

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征与网络流分类规则的匹配关系，设置所述数据包的网络类别标识；

所述数据包的 IP 头部特征至少包括下述任一信息：源 IP、目的 IP、源 socket 端口、目的 socket 端口、协议类型。

其中，所谓协议类型，包括四层协议和/或七层协议，例如 UDP、HTTP 等。

本申请实施例中，所述过滤器，可以是 Linux cgroup 过滤器。所谓 cgroup 过滤器，是 Linux cgroup(Linux control group，或 Linux 控制组)中的子系统 net_cls cgroup 提供的过滤器。所谓 Linux cgroup，是指 Linux 内核提供资源管理的各子系统，其中 net_cls cgroup 是分类并标记数据包的子系统，net_cls cgroup 对数据包的标记可以用作所述网络流类别标识，Linux TC 能够识别 net_cls cgroup 子系统中配置的每个具体 cgroup 过滤器标记后的数据包。例如，建立并配置具体的 net_cls cgroup，再通过 TC filter 命令将具体 net_cls cgroup 标记的数据包归类到网络流类别标识对应的类别。

步骤 S102，根据所述网络流类别标识，选择发送队列转发所述数据包。

所述发送队列，是关联了排队规则的序号连续的类别队列；其中，所述类别队列，使用主序列号和从序列号进行标识，所述类别队列的主序列号用于关联网络设备；所述排队规则，关联了所述过滤器，使用主序列号和从序列号进

行标识，所述排队规则的主序列号取值于所述类别队列的从序列号。具体的，本申请实施例中，通过 Linunx TC 中的类（class）利用排队规则管理数据包，向发送队列插入数据包，由于将发送队列分组标识，因此称为类别队列。

本实施例中的排队规则是有类别的排队规则(即 classful queueing discipline)，无类别的排队规则是对于数据包采用先入先出的方式通过队列，则除了缓存网络接口一时无法处理的数据包外不对进入队列的数据包进行任何其他处理。所谓有类别的排队规则，包括 CBQ、HTB 和 Prior Qdisc，其中 CBQ（class based queueing，或基于类别排队），实现了丰富的连接共享类别结构，既有限制带宽（shaping）能力，也具有带宽优先级管理能力，带宽限制是通过计算连接的空闲时间完成的，空闲时间的计算标准是数据包离队事件的频率和下层连接（数据链路层）的带宽；HTB（hierarchy token bucket，或分层令牌桶），通过 TBF（token bucket filter，或令牌桶过滤器）实现带宽限制，允许特定的类别可以突破带宽上限，占用别的类的带宽；Prio Qdisc 这类规则，是将属于高优先级的数据包发送完毕后才发送低优先级数据包，即不同优先级的数据包顺序离队。排队规则可以包含一些类别，不同的类别中可以包含更深入的排队规则，通过这些细分的排队规则可以为进入的队列的数据包排队，通过设置各种类别数据包的离队次序，设置网络数据流量的优先级。所谓过滤器（filter），是用于对数据包分类，决定数据包按照何种排队规则进入队列。无论何时数据包进入一个划分子类的类别中，都需要进行分类。使用过滤器分类时，内核调用附属于这个类（class 的所有过滤器，直到返回一个判决。如果没有判决返回，就作进一步的处理，而处理方式和排队规则有关。

具体到本实施例，所述排队规则、类和过滤器都有 ID 来标识，ID 由一个主序列号和一个从序列号组成，两个数字用一个冒号分开。其中，排队规则的主序列号，叫做句柄(handle)，采用象“1:”这样的表达方式。从序列号作为类的命名空间。在同一个排队规则里面的类分享该排队规则的主序列号，但是每个类都有自己的从序列号，叫做类识别符(classid)。类识别符只与父排队规则有关，和父类无关。

本步骤是根据所述步骤 S201 设置的网络流类别标识，选择发送队列转发所述网络流类别标识标记的数据包。

本申请实施例中，对 Linux TC 的功能模块进行了扩展，包括下述两个方面：

一方面，Linux 内核引入发送队列调度模块。通过 Linux 内核引入的发送队

列调度模块，根据所述排队规则关联的过滤器的网络流类别标识选择发送队列。

另一方面，提供 Linux 控制流量命令的扩展命令。所述扩展命令，用于将所述网络流类别标识关联到所述排队规则。

本申请实施例中，使用过滤器对应的网络流分类规则对数据包进行匹配处理，过滤器附属于类。转发数据包的网络设备（例如网卡）需要创建根队列排队规则，在根队列排队规则下建立类以及类的子类，每个类只有一个父类，每个类可以有多个子类，这些类组成一个树，称为 TC 树。数据包与网络流分类规则进行匹配时，根据 TC 树分层进行匹配，按照每一层的处理逻辑从根直到叶子节点，真正排入真实队列通过网络设备转发数据包。例如，图 3 所示的发送队列分组示意图中包含的 TC 树，所述 TC 树包括:

根队列排队规则（Root QDISC）301，所述根队列排队规则的主序列号为 8001:;

所述根队列排队规则 301 包含两个类 class3202 和 class 303，class 302 的类 ID 是 8001:ab51，class 303 的类 ID 是 8001:ab52;

所述 class 302 的叶子排队规则 HTB 304;

所述 class 303 的叶子排队规则 HTB 305。

本申请实施例的一个具体方式包括: 通过 Linux 控制流量命令的扩展命令将网络流类别标识对应的过滤器关联到排队规则；再通过发送队列调度模块根据排队规则关联的过滤器的网络流类别标识选择发送队列。以下对所述发送队列调度模块的引入以及命令扩展进行说明。

基于网络流类别标识的队列分配方法需要定义灵活且可重用的接口，通过引入 mq_cls 发送队列调度模块实现 socket 类别标识设置，给分类后的数据包使用 SO_PRIORITY 选项在 socket->sk_classid 域设置类 ID，所述类 ID 即为网络流类别标识，并支持根据所述类 ID 实现发送端队列选择。具体的 mq_cls 模块包括如下内容:

mq_cls 支持将某个 TC qdisc（queueing discipline）关联到一组连续的发送队列上，所述排队规则的类 ID 的主序列号（major number）被用来作为 mq_cls 的 class ID（类 ID）选择某一组发送队列。扩展 Linux 现有的 TC qdisc 命令，增加对于 mq_cls 发送队列调度模块的配置命令，扩展后的新命令格式为:

tc qdisc (add | change | del) dev (dev) root [handle major:] mqcls [numtc tcs] [classes class0 class1 …] [queues count1@offset1 count2@offset2 …]

其中，各命令参数含义如下：

qdisc 表示队列规则相关命令；

(add|change|del) 增加、修改、删除队列规则的相关命令；

dev 指定设备的关键字；

(dev) 设备名；

root 操作根队列规则；

[handle major:] [队列规则标识 队列规则的主序列号:]；

mqcls 队列规则标识名；

[numtc tcs] 指定 mqcls 的子类别个数；

[classes class0 class1 ...] 指定 mqcls 的子类别序号（从序列号）。Class0 指的是第一个子类别的序号；

[queues count1@offset1 count2@offset2 ...] 指定 mqcls 的子类别所对应的队列分组， count1 指定第一个子类别所对应的队列分组中队列的个数。Offset1 指定第一个子类别所对应的队列分组中第一个队列的以 0 为基准的偏移。

举例如下：沿用图 3 所示的发送队列分组示意图，说明网络设备 eth0 的发送队列分组。将 TXQ0 和 TXQ1 作为一组发送队列 306，将 TXQ2 和 TXQ23 作为一组发送队列 307；

class 302 将网络流类别标识为 ab51 的数据包根据 HTB 排队规则，插入到发送队列 306；

class 303 将网络流类别标识为 ab52 的数据包根据 HTB 排队规则，插入到发送队列 307。

将 HTB (Hiearchical Token Bucket) qdisc 挂载到发送队列 TXQ0 和 TXQ1 上配置命令如下：

命令 1：tc qdisc add dev eth0 root mqcls numtc 1 classes ab51 queues 2@0

命令 2：tc qdisc add dev eth0 parent 8001:ab51 handle ab51:0 htb

通过命令 1 将网络设备 eth0 的发送队列 TXQ0 和 TXQ1 划分为一个分组，通过命令 2 将 HTB 排队规则挂载到这个分组上。

将 HTB qdisc 挂载到发送队列 TXQ2 至 TXQ23 的配置命令如下：

命令 3：tc qdisc change dev eth0 root mqcls numtc 2 classes ab51 ab52 queues 2@0 22@2

命令 4：tc qdisc add dev eth0 parent 8001:ab52 handle ab52:0 htb

通过命令 3 将网络设备 eth0 的发送队列 TXQ2 和 TXQ23 划分为一个分组，通过命令 4 将 HTB 排队规则挂载到这个分组上。

需要注意的是，本申请实施例中沿用了 Linux TC 对排队规则句柄（qdisc handle）以及类 ID（class ID）取值的规则，关联在发送队列分组上的排队规则必须使用分组对应的类的类句柄的从序列号（minor number）作为其根句柄（root handle）的主序列号（major number）。例如，命令 2 中的 handle 参数的值为“ab51:0”，这取决于命令 2 中的 classes 参数的值为 “ab51”。这样关联到 Linux TC 的类的 cgroup 过滤器（根据 net_cls cgroup 中的 class ID 来进行分类的过滤器）能够将所述类的类 ID 作为网络流类别标识对数据包进行分类，并对匹配到所述类的数据包设置标记。

本申请实施例中，具体通过下述处理根据所述排队规则关联的过滤器的网络流类别标识选择发送队列，包括:

确定所述网络流类别标识对应的排队规则;

根据所述排队规则的主序列号，查找所述排队规则的主序列号到所述发送队列号的哈希表，确定发送队列号。

实际实施中，为了实现发送队列的快速查找，在 Linux 内核中增加的所述排队规则的句柄的主序列号到所述发送队列号的哈希表，通过哈希表快速查找发送队列号。具体的，对 netdev_pick_tx()函数进行修改，使得 skb 的 net_cls 中的 major number 被用来选择 txq，Linux 内核中增加一个从 net_cls major number 到 txq number 的哈希表，可以做 txq 快速查找，从而通过 major number 选择 txq 进而选择其对应 qdisc。

另外，本申请实施例中，在选择发送队列转发所述数据包的处理过程中，还包括下述处理:

根据选择的发送队列所关联的排队规则对应的网络流的优先级，确定所述数据包离队次序，所述排队规则，用于对所述发送队列的数据包排序。各发送队列根据进入队列的数据包的离队次序，依次发送所述数据包。实现了使用网络流的类别标识来确定发送队列，并将网络流控模块关联到多个队列上从而实现了对基于网络类别标识(network class ID)进行流控的支持。

图 4 示出了上述网络通信的方法中数据包在发送端的转发处理流向。包括:

网络应用 401 构造数据包;

通过网络接口层 socket 402 将所述数据包发送到传输层/网络层 TCP/IP 403;

TCP/IP 403 对所述数据包添加协议头，例如端口号、IP 地址、检验和，将添加协议头的数据包发送到 TXQ Selection 404，所述 TXQ Selection 是发送队列选择模块；

TXQ Selection 404 实际部署了 mq_cls，用于选择网络设备（例如网卡，或称 network interface card，缩写为 NIC）的发送队列，将所述数据包发送到 Traffic Control 405 进行流量控制，图示进行流量控制的排队规则是 HTB；

经过 Traffic Control 405 进行流量控制处理的数据包通过网卡驱动（NIC Driver）406 入网卡，并到达实际发送队列出网卡（NIC） 407。

以上述网络通信的方法的实施例为基础，本申请第二实施例提供第二种网络通信的方法。

以下结合图 5 至图 7 对本申请提供的第二种网络通信的方法的实施例进行说明。其中，图 5 是本申请提供的所述网络通信的方法实际部署的网络环境示意图；图 6 是本申请提供的第二种网络通信的方法处理流程图；图 7 是本申请提供的第二种网络通信的方法中接收端的数据包的转发流向示意图。

由于本实施例是以本申请第一实施例提供的网络通信的方法的实施例为基础，其描述是简要的，相关部分参见本申请第一实施例的描述即可。

图 6 所示的网络通信的方法，包括：步骤 S601 至步骤 S603。

步骤 S601，按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器。

本申请第二实施例提供的网络通信的方法一般部署于网络上传输的数据包的接收端，实现接收端的流量控制。所述接收端，是网络设备，例如，可以是网络接口板（Network Interface Card，或 NIC），网络接口板也称为网卡。具体的，接收端从网络侧接收的数据包经过网卡的接收队列之后，被重定向到 IFB（Intermediate Functional Block，或中间功能模块）的发送路径，当数据包经过 IFB 转发时进行流量控制。

所谓 IFB，模拟网卡转发数据包，并且提供流量控制功能。由于流量控制模块 TC 关联到了 IFB，因此 IFB 能够使用 TC 提供流量控制功能。IFB 不改变数据包的传输方向，也就是说，网卡接收队列的数据包重定向到 IFB 后，经过 IFB

的流量控制功能的处理后，又转发回被重定向之前的网卡继续进行接收处理。例如，从一块网卡发送数据包还是从一块网卡接收数据包，重定向到 IFB 经过流量控制功能的处理之后，通过在 IFB 的 dev_queue_xmit 函数中处理转发回被重定向之前的网卡。

本实施例提供的网络通信的方法，目的是通过网络流类别标识来完成接收端的流量控制。为了能够确定到达 IFB 的数据包的网络流类别标识，因此引入了 cls_sk 过滤器模块（也叫 cls_sk filter）。所述 cls_sk 过滤器模块，关联到接收端的排队规则，提供 cls_sk 过滤器，确定接收端接收到的每个数据包的 socket 端口号，并根据所述数据包的特征设置所述数据包的网络流类别标识。

图 5 示出了本实施例提供的所述网络通信的方法实际部署的网络环境，包括：数据包入网卡接收队列；

所述数据包通过网卡驱动 501 到达接收端排队规则 502，所述接收端排队规则也称为 ingress QDISC；

通过接收端排队规则 502 关联的 cls_sk filter 503 查询所述数据包的 socket 端口号，根据所述数据包的特征确定所述数据包对应的网络流类别，并给所述数据包设置对应的网络流类别标识；

所述 cls_sk filter 503 将使用网络流类别标识标记的数据包重定向到 IFB 504 的发送路径，所述 IFB 504 的发送路径，关联了用于选择发送队列的 mq_cls 发送队列调度模块；

通过 IFB 504 的发送路径的排队规则，根据所述数据包的网络流类别标识控制所述数据包转发回 TCP/IP 505 进行网络协议栈的处理，从而达到流量控制的目的，所述 TCP/IP 是指网络协议栈模块；

TCP/IP 505 将处理后的数据包发送到网络应用 506 进一步处理。

本步骤是对接收端接收到的网络数据包进行分类。

本申请实施例中，具体按照过滤器的网络流分类规则对从网络侧接收到的数据包进行分类。具体的，使用 cls_sk 过滤器查询所述数据包的 socket 端口号，根据所述数据包的特征确定所述数据包对应的网络流类别。所述网络流分类规则用于匹配所述数据包的特征，具体包含用于匹配数据包头部特征的信息；所述数据包头部特征，至少包括下述任一信息：源 IP 地址、目的 IP 地址、数据包接收端对应的 socket 和协议。所述过滤器是关联到所述接收端的排队规则的过滤器。

步骤 S602，基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识。

本步骤是对数据包设备网络流类别标识。

本申请实施例中，通过使用 cls_sk 过滤器对所述数据包进行分类后，将匹配上 cls_sk 过滤器的网络流分类规则的数据包设置网络流类别标识。

步骤 S603，根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

本步骤是按照接收端的排队规则将数据包转发到内核的网络协议栈处理，实现接收端的流量控制。

本申请实施例中，具体进行下述操作，实现根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包，包括:

通过所述接收端的排队规则关联的过滤器，将所述数据包重定向到中间功能模块;

根据所述网络流类别标识，选择所述中间功能模块的发送队列，转发所述数据包。

具体的，所述中间功能模块是指 IFB。例如，通过 IFB 关联的 mq_cls 发送队列调度模块，选择 IFB 的发送队列，并根据 IFB 的排队规则，例如 HTB 进行流量控制。

本申请实施例中，还包括: 使用接收线程重用技术重用接收端线程上下文。优选的，支持通过开关控制是否使能接收线程重用功能。所述接收线程，是用于处理接收端的数据包的接收的线程。

一个特定线程的上下文中的内容被其他模块所使用，可以通过内存共享、消息队列等机制实现，也可以直接重用所述特定线程完成其他模块要求的至少部分工作。本申请实施例中，所谓接收线程重用技术，是指将接收线程用于处理除了数据包的接收之外的其他数据包处理工作，从而可以使得接收端的上下文中保存的内容可以直接用于所述其他数据包处理工作。具体的，中间功能模块 IFB 中重用所述接收线程。例如，直接使用所述接收线程按照所述接收端的排队规则将所述数据包转发到中间功能模块(IFB)，而不使用 tasklet 机制处理，可以避开 tasklet 可能带来的处理器核竞争。所述中间功能模块，用于接收重定向来的数据包，按照所述排队规则对应的网络流的优先级将所述重定向来的数据包发送回被重定向前的网络设备。当所述重定向来的数据包，是由接收端重

定向来的，中间功能模块对所述重定向来的数据包进行处理后发送回被重定向前的接收端。

另外，接收线程重用，也可以是使用接收线程直接执行网络协议栈，这样可以提高接收端流量控制处理的并发度。

图 7 示出了接收端的从网络侧接收到数据包后的流向。数据包通过网络设备（网卡）被接收，接收后的数据包经过所述网络通信的方法进行处理，以及处理后的数据包被发送到网络应用进一步使用，这些过程中的数据包流向参见图 7，包括:

NIC 701 是网卡，所述网卡有 n+1 个接收队列 RXQ0，RXQ1，RXQ2，...，RXQn；数据包进入 NIC 701 的接收队列;

数据包通过 NIC 701 的接收队列到达 NIC Driver 702，NIC Driver 是网卡驱动;

数据包由 NIC Driver 702 到达 Ingress QDISC 703，Ingress QDISC 是接收端排队规则，也称为接收端流量控制调度器;

数据包由 Ingress QDISC 703 到达接收端排队规则关联的 cls_sk 过滤器，即图中的 cls_sk/mirred action 704，cls_sk 确定所述数据包的网络流类别并给所述数据包设置网络流类别标识; 图示的 mirred action 是指 cls_sk 过滤器执行重定向动作，所述数据包将被重定向到中间功能模块 IFB 706;

Traffic Control 705 是 IFB 706 关联的流量控制模块，按照 IFB 706 关联的流量控制模块选择 IFB 706 发送路径上的发送队列（见图示的 TXQ Selection，发送队列是重用的接收队列，有 n 个发送队列），并按照 IFB 706 关联的排队规则（例如图示的 HTB）执行流量控制后，由 IFB 706 发送回重定向前的网卡的接收队列继续进行接收处理，继续进行接收处理包括发送到 TCP/IP 707 进行网络协议栈处理，图示的 TCP/IP 表示所述网络协议栈是 IP 协议栈;

TCP/IP 707 将处理后的数据包通过 socket 层发送给网络应用 708 进一步处理或使用。

与本申请第一实施例提供的一种网络通信的方法的实施例相对应，本申请还提供了第一种网络通信的装置。

参照图 8，其示出了根据本申请提供的第一种网络通信的装置示意图。由于装置实施例基本相似于方法实施例，所以描述得比较简单，相关的部分请参见

对应的方法实施例的对应说明即可。

本申请提供第一种网络通信的装置，包括:

标识单元 801，用于在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

转发单元 802，用于根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

可选的，所述发送队列，是关联了排队规则的序号连续的类别队列；其中，所述类别队列，使用主序列号和从序列号进行标识，所述类别队列的主序列号用于关联网络设备；所述排队规则，关联了所述过滤器，使用主序列号和从序列号进行标识，所述排队规则的主序列号取值于所述类别队列的从序列号。

可选的，所述数据包的特征包括如下至少之一：IP 五元组信息、socket 信息。

可选的，所述过滤器，是基于 Linux 流量控制的过滤器;

相应的，所述标识单元 801，具体用于:

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征，按照所述网络流分类规则对所述数据包分类;

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征与网络流分类规则的匹配关系，设置所述数据包的网络类别标识;

所述数据包的 IP 头部特征至少包括下述任一信息: 源 IP、目的 IP、源 socket 端口、目的 socket 端口、协议类型。

可选的，所述转发单元 802，具体用于:

通过在 Linux 内核引入的发送队列调度模块，根据所述排队规则关联的过滤器的网络流类别标识选择发送队列。

可选的，所述网络通信的装置，还包括命令单元，用于:

提供 Linux 控制流量命令的扩展命令，所述扩展命令用于将所述网络流类别标识关联到所述排队规则。

可选的，所述转发单元 802，具体用于:

根据数据包的网络流类别标识，确定所述网络流类别标识对应的排队规则;

根据所述排队规则的主序列号，查找所述排队规则的主序列号到所述发送队列号的哈希表，确定发送队列号。

可选的，所述网络通信的装置，还包括离队次序确定单元，用于：

根据选择的发送队列关联的排队规则对应的网络流的优先级，确定所述数据包离队次序，所述排队规则，用于对所述发送队列的数据包排序。

与本申请提供的第二种网络通信的方法的实施例相对应，本申请还提供了第二种网络通信的装置。

参照图 9，其示出了根据本申请提供的第二种网络通信的装置示意图。由于装置实施例基本相似于方法实施例，所以描述得比较简单，相关的部分请参见对应的方法的实施例的对应说明即可。

本申请提供第二种用网络通信的装置，包括：

网络流分类单元 901，用于按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器；

标识单元 902，用于基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识；

转发单元 903，用于根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

可选的，所述网络流分类规则，包含用于匹配数据包头部特征的信息；所述数据包头部特征，至少包括下述任一信息：源 IP 地址、目的 IP 地址、数据包接收端对应的 socket 和协议。

可选的，所述转发单元 903，具体用于：

通过所述接收端的排队规则关联的过滤器，将所述数据包重定向到中间功能模块；

根据所述网络流类别标识，选择所述中间功能模块的发送队列，转发所述数据包。

可选的，所述网络通信的装置，还包括接收线程重用单元，用于：

使用接收线程重用技术重用接收端线程上下文。

可选的，所述接收线程，是用于处理接收端的数据包的接收的线程；相应的，所述接收线程重用单元，具体用于：

直接使用所述接收线程，按照所述接收端的排队规则将所述数据包转发到中间功能模块；所述中间功能模块，用于接收重定向来的数据包，按照所述排队规则对应的网络流的优先级将所述重定向来的数据包发送回被重定向前的网络设备。

本申请还提供了一种电子设备的实施例，所述电子设备用于实现本申请第一实施例提供的网络通信的方法，参照图 10，其示出了本实施例提供的一种电子设备的示意图。

本申请提供的所述电子设备实施例描述得比较简单，相关的部分请参见本申请第一实施例的对应说明即可。

本申请提供一种电子设备，包括:

存储器 1001，以及处理器 1002;

所述存储器 1001 用于存储计算机可执行指令，所述处理器 1002 用于执行所述计算机可执行指令:

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

可选的，所述发送队列，是关联了排队规则的序号连续的类别队列；其中，所述类别队列，使用主序列号和从序列号进行标识，所述类别队列的主序列号用于关联网络设备；所述排队规则，关联了所述过滤器，使用主序列号和从序列号进行标识，所述排队规则的主序列号取值于所述类别队列的从序列号。

可选的，所述数据包的特征包括如下至少之一：IP 五元组信息、socket 信息。

可选的，所述过滤器，是基于 Linux 流量控制的过滤器;

相应的，所述处理器 1002 还用于执行下述计算机可执行指令:

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征，按照所述网络流分类规则对所述数据包分类;

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征与网络

流分类规则的匹配关系，设置所述数据包的网络类别标识;

所述数据包的IP头部特征至少包括下述任一信息: 源IP、目的IP、源socket端口、目的socket端口、协议类型。

可选的，所述处理器1002还用于执行下述计算机可执行指令:

通过在Linux内核引入的发送队列调度模块,根据所述排队规则关联的过滤器的网络流类别标识选择发送队列。

可选的，所述处理器1002还用于执行下述计算机可执行指令:

提供Linux控制流量命令的扩展命令，所述扩展命令，用于将所述网络流类别标识关联到所述排队规则。

可选的，所述处理器1002还用于执行下述计算机可执行指令:

根据数据包的网络流类别标识，确定所述网络流类别标识对应的排队规则;

根据所述排队规则的主序列号，查找所述排队规则的主序列号到所述发送队列号的哈希表，确定发送队列号。

可选的，所述处理器1002还用于执行下述计算机可执行指令:

根据选择的发送队列关联的排队规则对应的网络流的优先级，确定所述数据包离队次序，所述排队规则，用于对所述发送队列的数据包排序。

本申请还提供第二种电子设备的实施例，所述电子设备用于实现本申请第二实施例提供的网络通信的方法，所述电子设备的示意图类似于图10。

本申请提供的所述电子设备实施例描述得比较简单，相关的部分请参见本申请第二实施例的对应说明即可。

本申请提供第二种电子设备，包括:

存储器，以及处理器;

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令:

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器;

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识;

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述

数据包。

可选的，所述网络流分类规则，包含用于匹配数据包头部特征的信息；所述数据包头部特征，至少包括下述任一信息：源 IP 地址、目的 IP 地址、数据包接收端对应的 socket 和协议。

可选的，所述处理器还用于执行下述计算机可执行指令：

通过所述接收端的排队规则关联的过滤器，将所述数据包重定向到中间功能模块；

根据所述网络流类别标识，选择所述中间功能模块的发送队列，转发所述数据包。

可选的，所述处理器还用于执行下述计算机可执行指令：

使用接收线程重用技术重用接收端线程上下文。

可选的，所述接收线程，是用于处理接收端的数据包的接收的线程；相应的，所述处理器还用于执行下述计算机可执行指令：

直接使用所述接收线程，按照所述接收端的排队规则将所述数据包转发到中间功能模块；所述中间功能模块，用于接收重定向来的数据包，按照所述排队规则对应的网络流的优先级将所述重定向来的数据包发送回被重定向前的网络设备。

本申请还提供一种存储设备的实施例。

本申请提供一种存储设备，包括：存储器和处理器，其中，存储器存储有指令，所述指令能够被处理器加载并执行以下步骤：

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征；

根据所述网络流类别标识，选择发送队列转发所述数据包；

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

本申请还提供第二种存储设备的实施例。

本申请提供的第二种存储设备，包括：存储器和处理器，其中，存储器存储有指令，所述指令能够被处理器加载并执行以下步骤：

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器；

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识；

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

在一个典型的配置中，计算设备包括一个或多个处理器 (CPU)、输入/输出接口、网络接口和内存。

内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器 (RAM) 和/或非易失性内存等形式，如只读存储器 (ROM) 或闪存(flash RAM)。内存是计算机可读介质的示例。

1、计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非暂存电脑可读媒体 (transitory media)，如调制的数据信号和载波。

2、本领域技术人员应明白，本申请的实施例可提供为方法、系统或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本申请虽然以较佳实施例公开如上，但其并不是用来限定本申请，任何本领域技术人员在不脱离本申请的精神和范围内，都可以做出可能的变动和修改，因此本申请的保护范围应当以本申请权利要求所界定的范围为准。

权 利 要 求 书

1、一种网络通信的方法，其特征在于，包括:

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

2、根据权利要求 1 所述的方法，其特征在于，所述发送队列，是关联了排队规则的序号连续的类别队列；其中，所述类别队列，使用主序列号和从序列号进行标识，所述类别队列的主序列号用于关联网络设备；所述排队规则，关联了所述过滤器，使用主序列号和从序列号进行标识，所述排队规则的主序列号取值于所述类别队列的从序列号。

3、根据权利要求 1 所述的方法，其特征在于，所述数据包的特征包括如下至少之一：IP 五元组信息、socket 信息。

4、根据权利要求 3 所述的方法，其特征在于，所述过滤器，是基于 Linux 流量控制的过滤器;

相应的，所述通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，包括:

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征，按照所述网络流分类规则对所述数据包分类;

通过基于 Linux 流量控制的过滤器，根据所述数据包的 IP 头部特征与网络流分类规则的匹配关系，设置所述数据包的网络类别标识;

所述数据包的 IP 头部特征至少包括下述任一信息: 源 IP、目的 IP、源 socket 端口、目的 socket 端口、协议类型。

5、根据权利要求 2 所述的方法，其特征在于，所述根据所述网络流类别标识，选择发送队列转发所述数据包，包括:

通过在 Linux 内核引入的发送队列调度模块，根据所述排队规则关联的过滤器的网络流类别标识选择发送队列。

6、根据权利要求 2 所述的方法，其特征在于，还包括:

提供 Linux 控制流量命令的扩展命令，用于将所述网络流类别标识关联到所

述排队规则。

7、根据权利要求 5 所述的方法，其特征在于，所述根据所述排队规则关联的过滤器的网络流类别标识选择发送队列，包括:

根据数据包的网络流类别标识，确定所述网络流类别标识对应的排队规则;

根据所述排队规则的主序列号，查找所述排队规则的主序列号到所述发送队列号的哈希表，确定发送队列号。

8、根据权利要求 1 所述的方法，其特征在于，还包括:

根据选择的发送队列关联的排队规则对应的网络流的优先级，确定所述数据包离队次序，所述排队规则，用于对所述发送队列的数据包排序。

9、根据权利要求 8 所述的方法，其特征在于，所述根据所述网络流类别标识，选择发送队列转发所述数据包，包括:

各发送队列根据进入队列的数据包的离队次序，依次发送所述数据包。

10、一种网络通信的方法，其特征在于，包括:

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器;

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识;

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

11、根据权利要求 10 所述的方法，其特征在于，所述网络流分类规则，包含用于匹配数据包头部特征的信息; 所述数据包头部特征，至少包括下述任一信息: 源 IP 地址、目的 IP 地址、数据包接收端对应的 socket 和协议。

12、根据权利要求 10 所述的方法，其特征在于，所述根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包，包括:

通过所述接收端的排队规则关联的过滤器，将所述数据包重定向到中间功能模块;

根据所述网络流类别标识，选择所述中间功能模块的不同发送队列，转发所述数据包。

13、根据权利要求 10 所述的方法，其特征在于，还包括:

使用接收线程重用技术重用接收端线程上下文。

14、根据权利要求 13 所述的方法，其特征在于，所述接收线程，是用于处理接收端的数据包的接收的线程；相应的，所述使用接收线程重用技术重用接收端线程上下文，包括:

直接使用所述接收线程，按照所述接收端的排队规则将所述数据包转发到中间功能模块；所述中间功能模块，用于接收重定向来的数据包，按照所述排队规则对应的网络流的优先级将所述重定向来的数据包发送回被重定向前的网络设备。

15、一种网络通信的装置，其特征在于，包括:

标识单元，用于在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

转发单元，用于根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

16、一种网络通信的装置，其特征在于，包括:

网络流分类单元，用于按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器;

标识单元，用于基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识;

转发单元，用于根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

17、一种电子设备，其特征在于，包括:

存储器，以及处理器;

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令:

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系

设置对应的网络流类别标识。

18、一种电子设备，其特征在于，包括:

存储器，以及处理器;

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令:

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器;

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识;

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

19、一种存储设备，其特征在于，存储有指令，所述指令能够被处理器加载并执行以下步骤:

在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征;

根据所述网络流类别标识，选择发送队列转发所述数据包;

其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识。

20、一种存储设备，其特征在于，存储有指令，所述指令能够被处理器加载并执行以下步骤:

按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器;

基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识;

根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包。

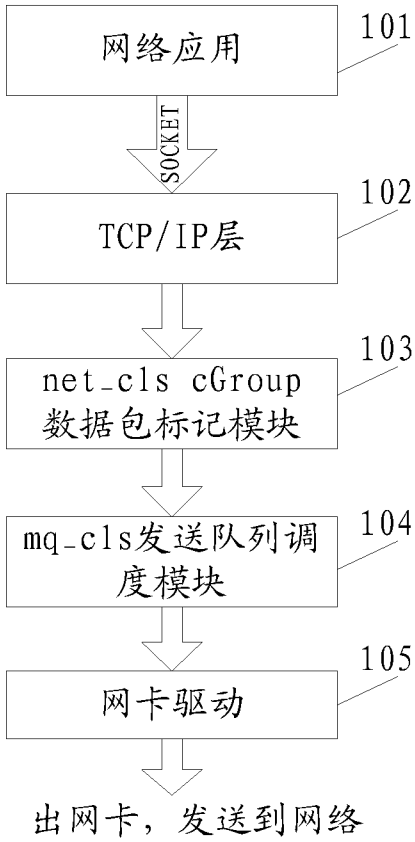
网络应用
101
SOCKET
TCP/IP层
102
net_cls cGroup
数据包标记模块
103
mq_cls发送队列调
度模块
104
网卡驱动
105
出网卡，发送到网络

图 1

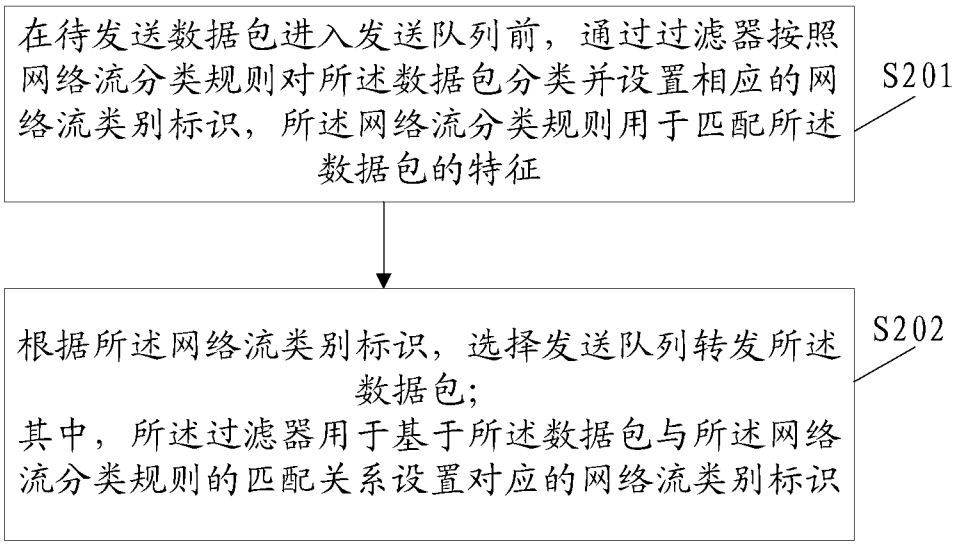
在待发送数据包进入发送队列前，通过过滤器按照网络流分类规则对所述数据包分类并设置相应的网络流类别标识，所述网络流分类规则用于匹配所述数据包的特征
S201
根据所述网络流类别标识，选择发送队列转发所述数据包;
其中，所述过滤器用于基于所述数据包与所述网络流分类规则的匹配关系设置对应的网络流类别标识
S202

图 2

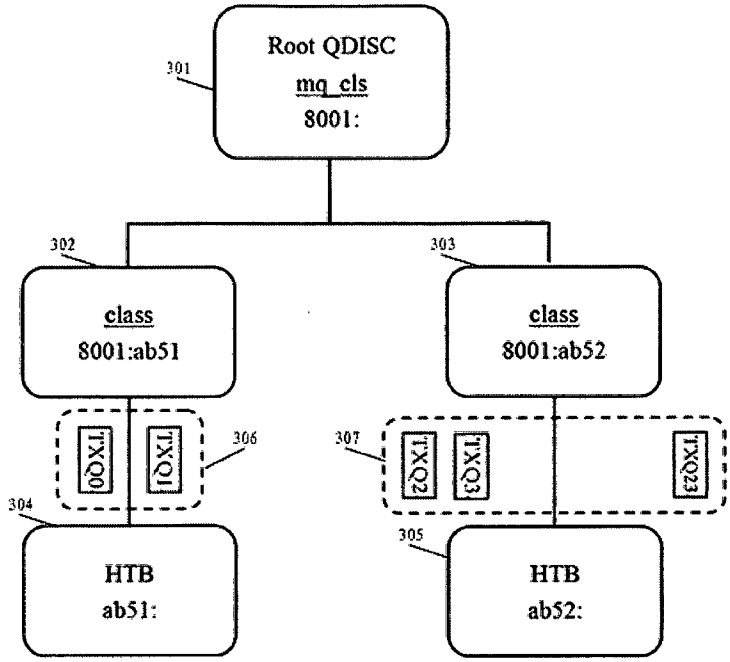
Root QDISC
mq cls
8001:
301
302
303
class
8001:ab51
class
8001:ab52
306
307
TXQ0
TXQ1
TXQ2
TXQ3
TXQ23
304
305
HTB
ab51:
HTB
ab52:
举例所对应的TC树和发送队列分组

图 3

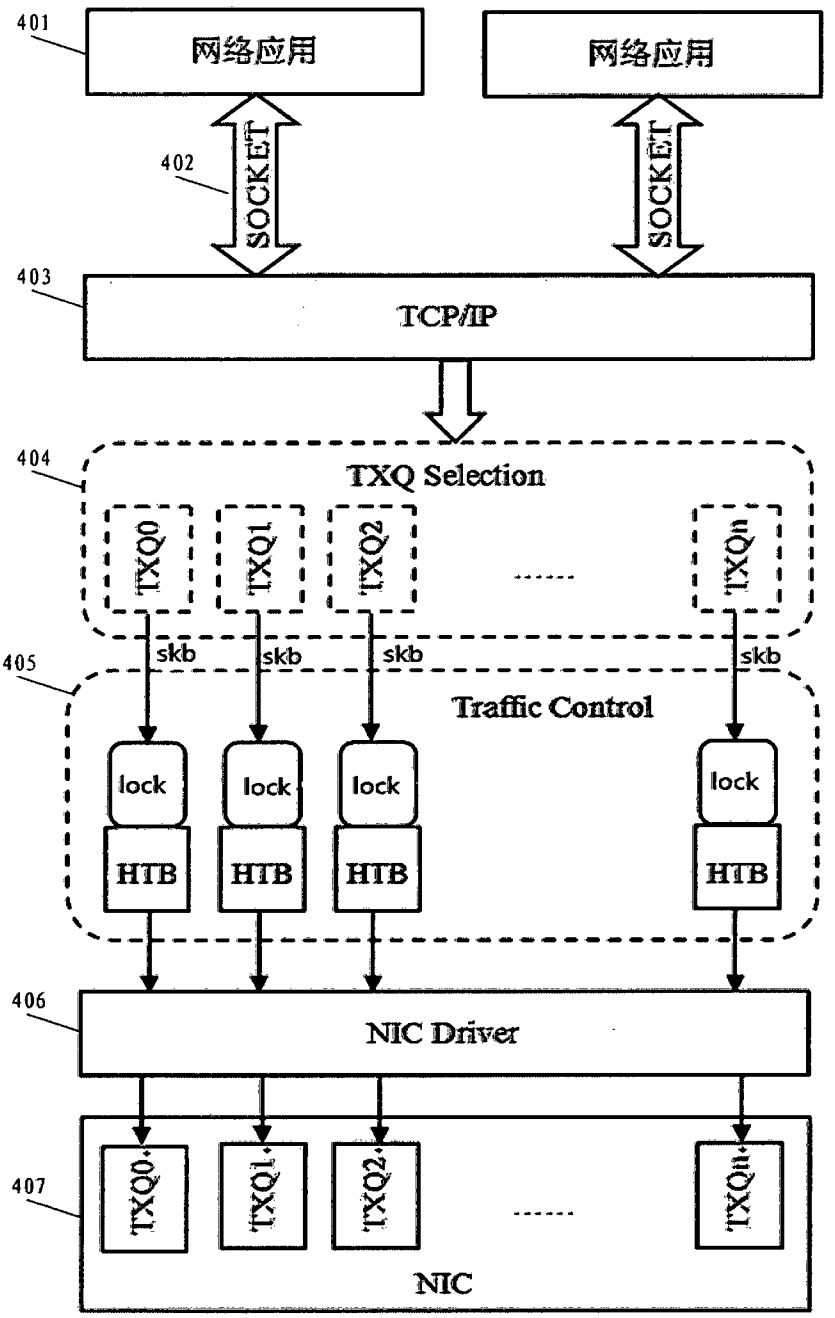
401
网络应用
网络应用
402
SOCKET
SOCKET
403
TCP/IP
404
TXQ Selection
TXQ0
TXQ1
TXQ2
.....
TXQn
skb
skb
skb
skb
405
Traffic Control
lock
lock
lock
lock
HTB
HTB
HTB
HTB
406
NIC Driver
407
TXQ0'
TXQ1'
TXQ2'
.....
TXQn'
NIC

图 4

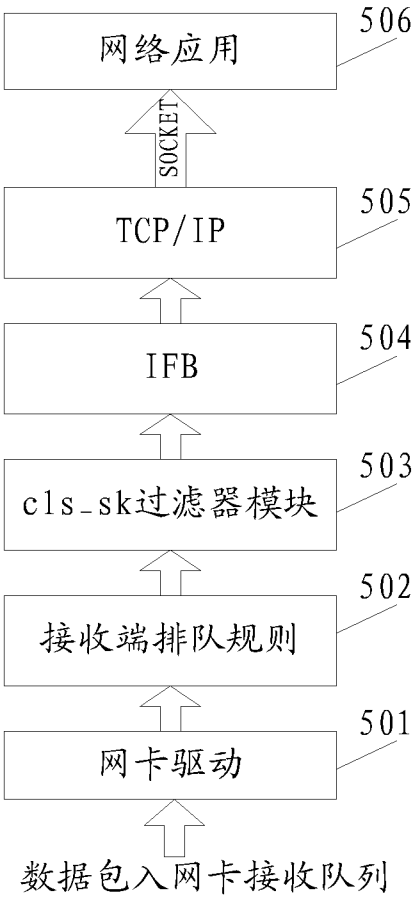
网络应用
506
SOCKET
TCP/IP
505
IFB
504
cls_sk过滤器模块
503
接收端排队规则
502
网卡驱动
501
数据包入网卡接收队列

图 5

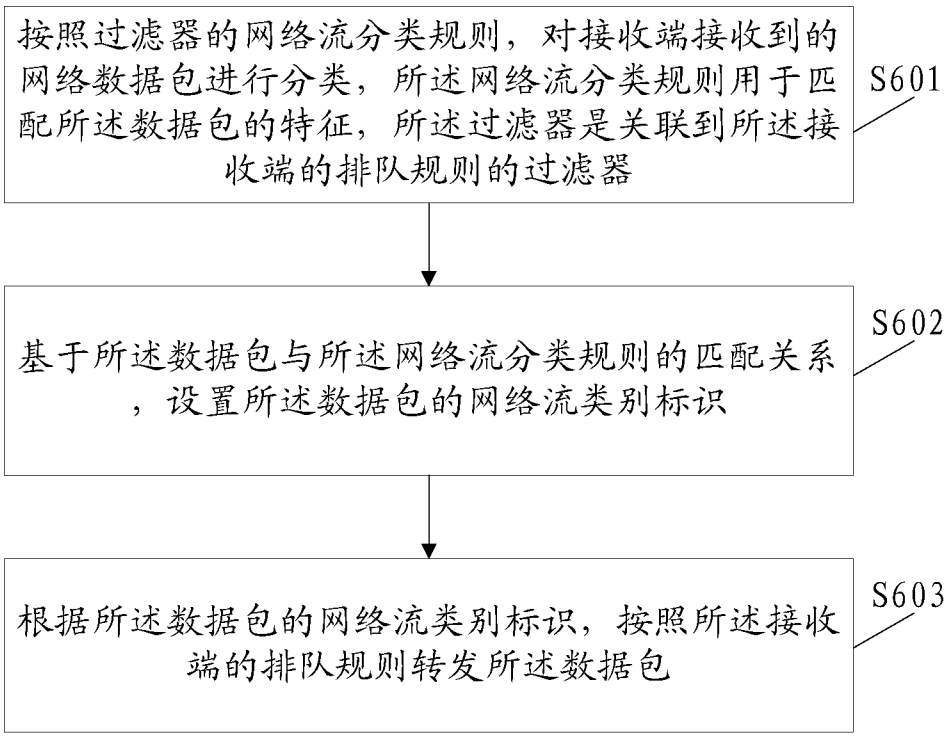
按照过滤器的网络流分类规则，对接收端接收到的网络数据包进行分类，所述网络流分类规则用于匹配所述数据包的特征，所述过滤器是关联到所述接收端的排队规则的过滤器
S601
基于所述数据包与所述网络流分类规则的匹配关系，设置所述数据包的网络流类别标识
S602
根据所述数据包的网络流类别标识，按照所述接收端的排队规则转发所述数据包
S603

图 6

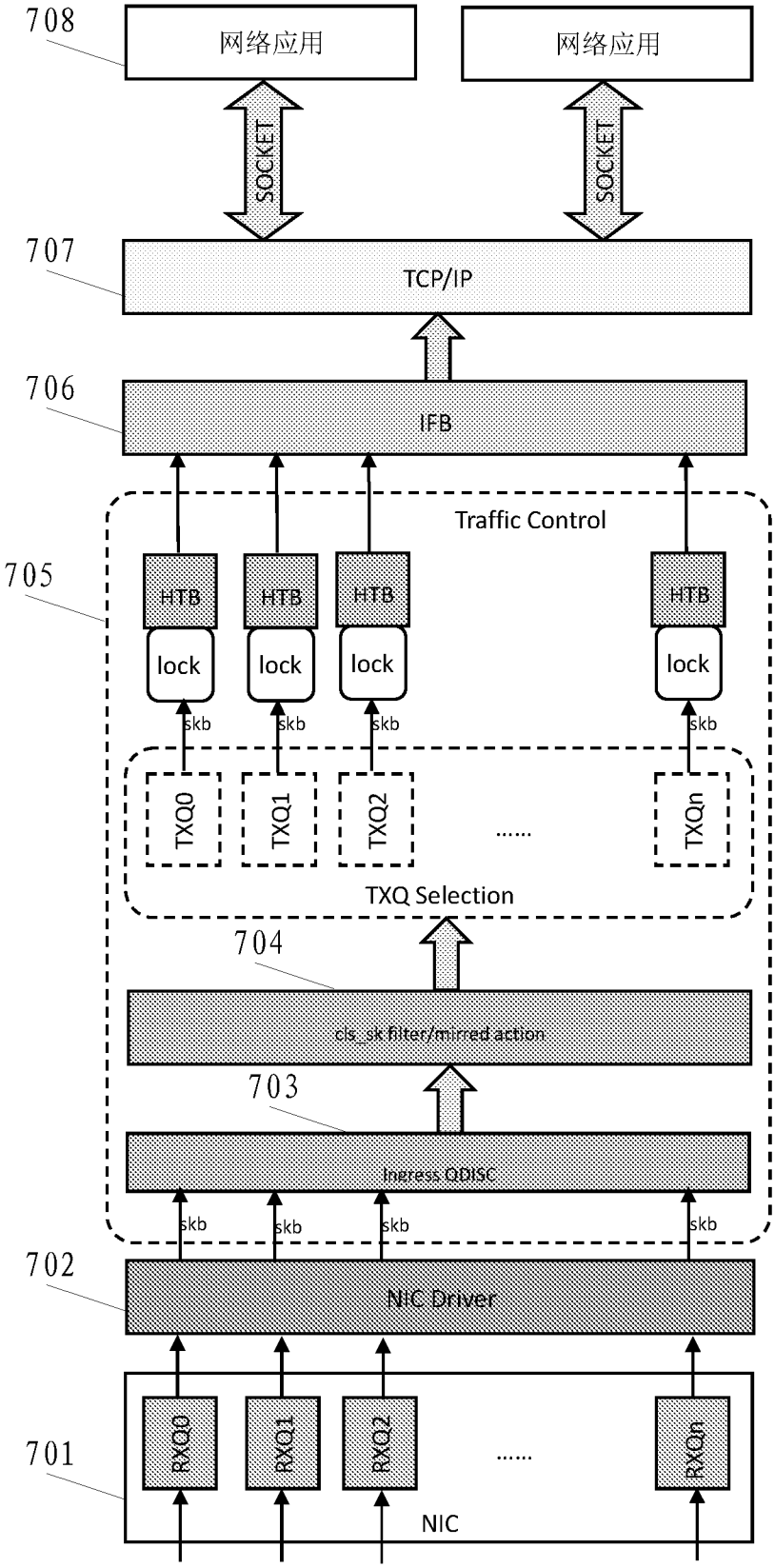
708
网络应用
网络应用
SOCKET
SOCKET
707
TCP/IP
706
IFB
Traffic Control
705
HTB
HTB
HTB
HTB
lock
lock
lock
lock
skb
skb
skb
skb
TXQ0
TXQ1
TXQ2
.....
TXQn
TXQ Selection
704
cls_sk filter/mirred action
703
Ingress QDISC
skb
skb
skb
skb
702
NIC Driver
RXQ0
RXQ1
RXQ2
.....
RXQn
701
NIC

图 7

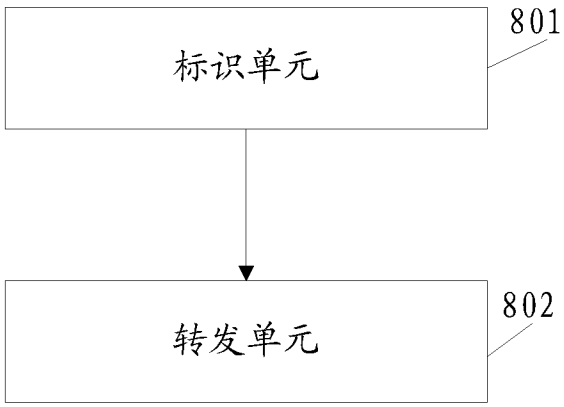
标识单元
801
转发单元
802

图 8

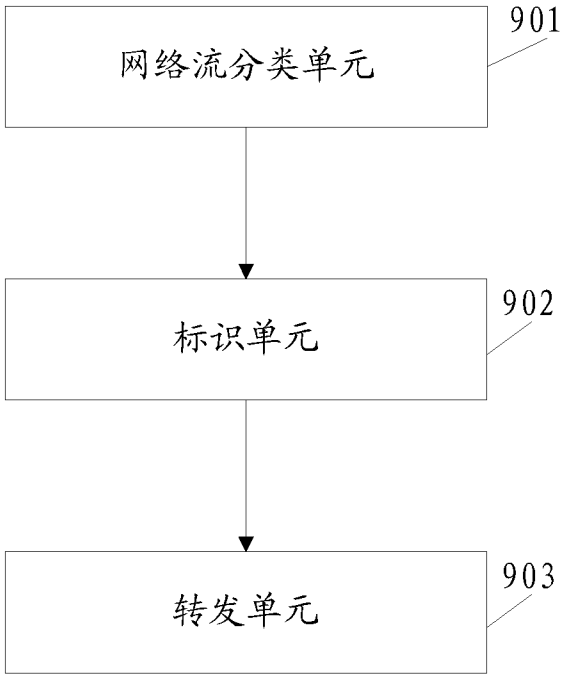
网络流分类单元
901
标识单元
902
转发单元
903

图 9

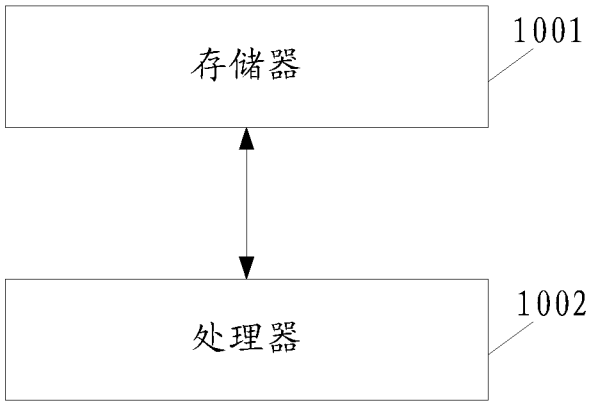
存储器
1001
处理器
1002

图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/113789

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/08(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)

WPI, EPODOC, CNPAT, CNKI, IEEE: 数据包, 网络流, 标识, 类别, 流量控制, 过滤器, data, packet, network, stream, ID, type, TC, filter

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	宗希鹏 (ZONG, Xipeng). "中国优秀硕士学位论文全文数据库 信息科技辑 (Chinese Master's Theses Full-text Database, Information Science and Technology)" *自适应业务分级转发控制技术研究及实现 (The Research and Implementation of Self Adaptive Traffic Classification Control Technology)*, No. 9, 15 September 2011 (2011-09-15), sections 2.3.1.3 and 3.2	1-20
A	CN 102035748 A (SHENZHEN SINFOR ELECTRONIC TECHNOLOGY CO., LTD.) 27 April 2011 (2011-04-27) entire document	1-20
A	CN 102238743 A (LEADCORE TECHNOLOGY CO., LTD.) 09 November 2011 (2011-11-09) entire document	1-20
A	CN 106385386 A (CHENGDU FEIYUXING TECHNOLOGY CO., LTD.) 08 February 2017 (2017-02-08) entire document	1-20
A	US 2017070554 A1 (VANTRIX CORPORATION) 09 March 2017 (2017-03-09) entire document	1-20

☐ 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
"D" document cited by the applicant in the international application
"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	Date of mailing of the international search report
08 July 2019	**26 July 2019**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/113789

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	102035748	A	27 April 2011	None			
CN	102238743	A	09 November 2011	None			
CN	106385386	A	08 February 2017	None			
US	2017070554	A1	09 March 2017	EP	3371980	A1	12 September 2018
				CN	108476345	A	31 August 2018
				CA	3042326	A1	11 May 2017
				WO	2017075692	A1	11 May 2017
				KR	20180079320	A	10 July 2018
				JP	2018537885	A	20 December 2018

国际检索报告

国际申请号
PCT/CN2018/113789

A. 主题的分类

H04L 29/08(2006.01)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）)

WPI，EPODOC，CNPAT，CNKI，IEEE：数据包，网络流，标识，类别，流量控制，过滤器，data，packet，network，stream，ID，type，TC，filter

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
X	宗希鹏. ″《中国优秀硕士学位论文全文数据库 信息科技辑》″ 自适应业务分级转发控制技术研究及实现，第9期，2011年 9月 15日 (2011-09-15)， 第2.3.1.3、3.2部分	1-20
A	CN 102035748 A (深圳市深信服电子科技有限公司) 2011年 4月 27日 (2011-04-27) 全文	1-20
A	CN 102238743 A (联芯科技有限公司) 2011年 11月 9日 (2011-11-09) 全文	1-20
A	CN 106385386 A (成都飞鱼星科技股份有限公司) 2017年 2月 8日 (2017-02-08) 全文	1-20
A	US 2017070554 A1 (VANTRIX CORPORATION) 2017年 3月 9日 (2017-03-09) 全文	1-20

☐ 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	国际检索报告邮寄日期
2019年 7月 8日	2019年 7月 26日

ISA/CN的名称和邮寄地址	受权官员
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	朱殿尧 电话号码 86-(10)-53961412

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/113789

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	102035748	A	2011年 4月 27日	无			
CN	102238743	A	2011年 11月 9日	无			
CN	106385386	A	2017年 2月 8日	无			
US	2017070554	A1	2017年 3月 9日	EP	3371980	A1	2018年 9月 12日
				CN	108476345	A	2018年 8月 31日
				CA	3042326	A1	2017年 5月 11日
				WO	2017075692	A1	2017年 5月 11日
				KR	20180079320	A	2018年 7月 10日
				JP	2018537885	A	2018年 12月 20日