

# (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  
2019年11月7日 (07.11.2019)



(10) 国际公布号  
**WO 2019/210725 A1**

- (51) 国际专利分类号:  
*H04L 12/801* (2013.01)
- (21) 国际申请号: PCT/CN2019/074385
- (22) 国际申请日: 2019年2月1日 (01.02.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201810421769.9 2018年5月4日 (04.05.2018) CN
- (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: 董建波 (DONG, Jianbo); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。付斌章 (FU, Binzhang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。谭焜 (TAN, Kun); 中

国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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: CONGESTION CONTROL METHOD, APPARATUS AND DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 拥塞控制方法、装置、设备及存储介质

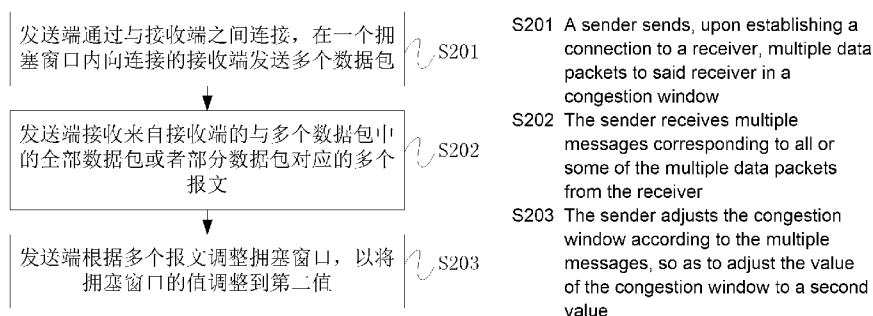


图 4

(57) Abstract: The present invention relates to the field of network communications, and provides a congestion control method, apparatus and device, and a storage medium. The congestion control method comprises: sending data packets to a receiver according to a first value of a congestion window, the number of the data packets being the first value; receiving multiple messages corresponding to all or some of multiple data packets, the multiple messages comprising one or more first messages, and one or more second messages; and adjusting, by a sender, the congestion window according to the multiple messages, so as to adjust the value of the congestion window to a second value. By means of the technical solution of the present invention, packet loss in a network can be reduced, and the stability of the network can be improved.

(57) 摘要: 本发明提供了一种拥塞控制方法、装置、设备及存储介质, 涉及网络通信领域。拥塞控制方法包括: 根据拥塞窗口的第一值, 向接收端发送第一值个数据包。接收多个数据包中的全部数据包或者部分数据包对应的多个报文, 该多个报文包括一个或多个第一报文, 以及一个或多个第二报文。发送端根据多个报文调整拥塞窗口, 以将拥塞窗口的值调整到第二值。利用本发明的技术方案能够减小网络中的丢包, 提高网络的稳定性。

WO 2019/210725 A1

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))。

## 拥塞控制方法、装置、设备及存储介质

### 技术领域

本申请涉及网络通信技术领域，尤其涉及一种拥塞控制方法、装置、设备及存储介质。

### 背景技术

随着网络通信技术的发展，云计算和大数据服务越来越多地被应用在各个方面。为了满足大量用户对云计算和大数据服务的要求，建立了大量的数据中心（Data Center, DC）。在每个数据中心内部都存在大量的通过数据中心网络连接的服务器。服务器可提供的很多应用服务均需要低延迟的网络，比如在线搜索、社交网络等应用服务。

现有的流控方法，连接的发送端根据接收端返回的报文，使用数据中心量化拥塞通知（Data Center Quantized Congestion Notification, DCQCN）作为拥塞控制策略，根据该策略调整发送端的发送速率，从而控制连接的数据传输以避免连接中拥塞的发生。然而这种拥塞控制策略，并不能保证数据传输的稳定性，连接的传输性能差，具体表现为，数据传输过程中，拥塞发生的概率高以及容易丢包，丢包率高。

### 发明内容

本申请提供了一种拥塞控制方法、装置、设备及存储介质，能够更准确地调整发送端的拥塞窗口，使用这样的拥塞窗口发送数据，可以降低连接发生拥塞的概率，在不对交换节点做控制的情况下，使交换节点具有稳定而较浅的待转发数据的队列深度，从而减小网络中的丢包，提高网络的稳定性。

第一方面，本申请提供了一种拥塞控制方法，拥塞控制方法包括：

发送端通过与接收端之间的连接（link），在一个拥塞窗口内向接收端发送多个数据包，多个数据包的数目为第一值，第一值为连接的拥塞窗口的值，其中，连接处于拥塞避免状态；发送端接收来自接收端的与多个数据包中的全部数据包或者部分数据包对应的多个报文，多个报文包括一个或多个第一报文，以及一个或多个第二报文，第一报文用于表示连接在传输多个数据包中的一个或多个数据包时拥塞，第二报文用于表示连接在传输多个数据包中的一个或多个数据包时不拥塞；发送端根据多个报文调整拥塞窗口，以将拥塞窗口的值调整到第二值，其中，多个报文中的第一报文，使拥塞窗口的值减小，多个报文中的第二报文，使拥塞窗口的值增加。

由于上述方法中，一个拥塞窗口对应的第一报文和第二报文都会对该拥塞窗口产生影响，则，第二值可以大于或者小于第一值，甚至在一些情况下，第一值和第二值可以相等。

以及，连接在传输多个数据包中的一个或多个数据包时拥塞。一种实现方式下，可以是，一个或多个数据包在传输的过程中，被该连接的某一个交换节点，或者接收端确定为发生拥塞；相应的，连接在传输多个数据包中的一个或多个数据包时不拥塞，一种实现方式下，可以是，一个或多个数据包在传输的过程中，被该连接任一交换节点以及接收端都未得到或生成用于指示拥塞的信息。

一个第一报文可以对应一个或多个数据包，一个第二报文也可以对应一个或多个数据包。第二报文可以是数据包的确认字符 ACK(Acknowledge)，或者其他。第一报文也可以是使用一个专门格式的报文以指示网络拥塞。例如可以是拥塞通知报文（Congestion Notification Packet, CNP）。

一种实现方式下，该多个报文可以都是反馈报文，每个报文还用于指示多个数据包中一定数目的数据包到达接收端，一定数目为一个或多个；第一报文携带拥塞反馈标识，第二报文不携带拥塞反馈标识。第一报文通过拥塞标识表示网络拥塞。例如该拥塞标识可以是基于显示拥塞通知反馈（Explicit Congestion Notification Echo）技术在ACK中打上的标签等。

第一报文对应和第二报文共同反映综合的网络拥塞程度的严重性，发送端可根据第一报文和第二报文共同反映综合的网络拥塞程度的严重性，有比例的减小和增加拥塞窗口的值。实现对拥塞窗口值的精细调整，从而使得网络能够快速对网络拥塞做出反应。在发送端对拥塞窗口的值的不断调整中，使得交换节点的队列深度得到控制，可达到一个较浅稳定的队列深度，减小网络中的丢包，避免网络性能出现跳变，提高了网络的稳定性。保证交换节点的队列缓存充足，降低了数据传输过程中的延时。而且，交换节点的队列深度始终保持在一个稳定的数值上，保证队列中一直存在数据包，也能够提高网络的带宽利用率，使网络保持高带宽的数据传输。一种实现方式下，多个报文中的一个第一报文，使拥塞窗口的值减小第三值，多个报文中的一个第二报文，使拥塞窗口的值增加第四值，第三值与一个第一报文指示的数据包的数目相关，第四值与一个第二报文指示的数据包的数目相关。

第一报文指示的数据包的数目和第二报文指示的数据包的数目，可进体现更加精确的网络拥塞的程度。发送端调整拥塞窗口的值减小与第一报文指示的数据包的数目相关的第三值，增加与第二报文指示的数据包的数目相关的第四值，能够进一步提高拥塞窗口调整的精细度，即使得交换节点的队列深度控制更加精准，进一步减少丢包，并提高网络的稳定性。

一种实现方式下，第二值为第一值、第一和的相反数与第二和的加和，第一和为多个报文中的每个第一报文对应的第三值相加的和，第二和为多个报文中的每个第二报文对应的第四值相加的和。

在一种可能的实现方式中，第四值还与第一值和增加系数相关。

一种实现方式下，发送端根据多个报文调整拥塞窗口，包括：发送端根据多个报文中的一个第一报文，将拥塞窗口的值减小第三值；以及发送端根据多个报文中的一个第二报文，将拥塞窗口的值增加第四值。实时对拥塞窗口的值进行调整，保证了拥塞窗口的值调整的及时性，能够进一步提高网络对网络拥塞做出反应的速度，进一步提高网络的稳定性。

一种实现方式下，发送端根据多个报文调整拥塞窗口，包括：发送端根据第一报文的数目，每个第一报文使拥塞窗口的值减小的量，第二报文的数目以及每个第二报文使拥塞窗口的值增加的量，计算出第二值，发送端将拥塞窗口的值从第一值调整到第二值。待基于该多个报文计算出第二值后，再调整，这样可以避免过于频繁地调整拥塞窗口，更加简便。

一种实现方式下，该方法还包括：向连接的接收端发送数目为第二值的多个数据包。

第二方面，本申请提供了一种拥塞控制装置，包括：发送模块，用于通过与接收端之间的连接，在一个拥塞窗口内向接收端发送多个数据包，多个数据包的数目为第一值，第一值为连接的拥塞窗口的值，其中，连接处于拥塞避免状态；接收模块，用于接收来自接收端的与多个数据包中的全部数据包或者部分数据包对应的多个报文，多个报文包括一个或多个第一报文，以及一个或多个第二报文，第一报文用于表示连接在传输多个数据包中的一个或多个数据包时拥塞，第二报文用于表示连接在传输多个数据包中的一个或多个数据包时不拥塞；控制模块，用于根据多个报文调整拥塞窗口，以将拥塞窗口的值调整到第二值，其中，多个报文中的第一报文，使拥塞窗口的值减小，多个报文中的第二报文，使拥塞窗口的值增加。

一种实现方式下，第一报文为拥塞通知报文CNP，第二报文为确认字符ACK。

一种实现方式下，多个报文为反馈报文，每个反馈报文还用于指示多个数据包中一定数

目的数据包到达接收端，一定数目为一个或多个；第一报文携带拥塞反馈标识，第二报文不携带拥塞反馈标识。

一种实现方式下，多个报文中的一个第一报文，使拥塞窗口的值减小第三值，多个报文中的一个第二报文，使拥塞窗口的值增加第四值，第三值与一个第一报文指示的数据包的数目相关，第四值与一个第二报文指示的数据包的数目相关。

一种实现方式下，第二值为第一值、第一和的相反数与第二和的加和，第一和为多个报文中的每个第一报文对应的第三值相加的和，第二和为多个报文中的每个第二报文对应的第四值相加的和。

一种实现方式下，在根据多个报文调整拥塞窗口的方面，控制模块用于根据多个报文中的一个第一报文，将拥塞窗口的值减小第三值；以及根据多个报文中的一个第二报文，将拥塞窗口的值增加第四值。

一种实现方式下，在根据多个报文调整拥塞窗口的方面，控制模块用于根据第一报文的数目，每个第一报文使拥塞窗口的值减小的量，第二报文的数目以及每个第二报文使拥塞窗口的值增加的量，计算出第二值，以及将拥塞窗口的值从第一值调整到第二值。

一种实现方式下，发送模块还用于向连接的接收端发送数目为第二值的多个数据包。

在一种可能的实现方式中，第四值还与第一值和增加系数相关。其中增加系数可以是预设的值。

第二方面以及第二方面的各个可能的实现方式中的拥塞控制装置也可达到与上述技术方案中的拥塞控制方法相同的技术效果。

第三方面，本申请提供了一种设备，包括存储介质、处理电路和通信接口；存储介质用于存储可执行程序；处理电路用于读取存储介质中存储的可执行程序，以在通信接口的配合下执行上述技术方案中的拥塞控制方法。其中，数据包的发送、报文的接收等均可由处理电路调用通信接口来实现，处理电路具体可实现调整拥塞窗口的步骤。本申请提供的设备可达到与上述技术方案中的拥塞控制方法相同的技术效果。

第四方面，本申请提供了一种存储介质，存储介质上存储有程序；程序被处理电路执行时实现如上述技术方案中的拥塞控制方法，可达到与上述技术方案中的拥塞控制方法相同的技术效果。

## 附图说明

图1为本申请实施例中一种网络应用场景图；

图2为本申请实施例中的一种拥塞控制方法的流程图；

图3为本申请实施例中一种拥塞标识设置概率与队列深度正相关的示意图；

图4为本申请实施例中一种拥塞控制方法的流程图；

图5为本申请实施例中一种设备的结构示意图；

图6为本发明实施例中一种拥塞控制装置的结构示意图。

## 具体实施方式

本申请提供了一种拥塞控制方法、装置、设备及存储介质，可应用于通信协议栈中的传输控制协议/用户数据报协议 (Transmission Control Protocol/User Datagram Protocol, TCP/UDP)层和网络协议(Internet Protocol, IP)层,以及以太网交换机(即 Ethernet Switch)和与主机通道适配器(即 Host Channel Adapter)之间的基于融合以太网的远程内存直接访问(Remote Direct Memory Access over Converged Ethernet, RoCE)。例如，一个典型的本申请的技术方案的应用场景，就是远程内存直接访问(Remote Direct Memory Access，

RDMA)网络。在一个示例中,本申请中的拥塞控制方法可应用于 RDMA 数据存取的场景中。比如,用户终端将数据直接写入远程服务器的内存,在数据写入的过程中无内核干预和内存拷贝发生。RDMA 可降低远程服务器侧数据处理的延迟。在以太网中,用户终端先将数据以数据包的形式发送至交换节点,再由交换节点将数据包发送至远程服务器。在数据存储至远程服务器的过程中,数据快速移动到了远程服务器的存储器中,数据交换的过程并没有经过远程服务器的中央处理器,从而减少了数据传输时远程服务器的中央处理器的开销。数据在用户终端和交换节点中的传输即可采用本申请中的拥塞控制方法。

图 1 为本申请实施例中一种网络应用场景图。如图 1 所示,应用缓存(即 Application Buffers)为应用程序中的用户缓存空间。一方面,数据可从应用缓存通过用户的套接字应用程序编程接口(Socket Application Programming Interface, Sockets API)进入操作系统内核(即 Kernel),在操作系统内核中通过套接字(即 Sockets)、传输控制协议(Transmission Control Protocol, TCP)、第四版网络协议/第六版网络协议(即 IPv4/IPv6)、网络设备(即 Network Device),到达操作系统内核中的设备驱动(Device Driver)。另一方面,数据也可从应用缓存通过远程内存直接访问应用于直接内存存取的应用程序编程接口(Remote Direct Memory Access Verbs Application Programming Interface, RDMA Verbs API)进入操作系统内核,利用操作系统内核旁路(即 Kernel Bypass)功能,直接到达操作系统内核中的设备驱动,实现了远程内存直接访问(Remote Direct Memory Access, RDMA)。操作系统内核中的缓存标头(即 Buffers Headers)可存储数据实际的内存地址、数据类型、缓存标头所在的链表、数据状态、数据标记等,在此并不限定。操作系统内核可通过服务器安装的主机通道适配器(Host Channel Adapter, HCA)利用无限带宽技术(即 InfiniBand)与无线带宽技术交换器(InfiniBand)进行数据交互。操作系统内核还可通过主机通道适配器利用互联网的广域远程内存直接访问协议(Internet Wide Area Remote Direct Memory Access Protocol, iWARP)、基于融合以太网的远程内存直接访问(Remote Direct Memory Access over Converged Ethernet, RoCE)等技术以太网交换机进行数据交互。

本申请实施例中拥塞控制方法的数据流动涉及图 1 所示的虚线框中的部分,即涉及网络协议和传输控制协议,以及交换节点和发送端之间的数据流通。应理解,本申请的技术方案涉及三端交互场景,即发送端,交换节点和接收端,但是一种实现方式下,本申请的技术改进在发送端,交换节点和接收端可以仍然使用现有技术中的方式给予配合。交换节点可以是路由器,交换机等,发送端可以是服务器,终端,云平台,甚至是某些网络设备,如智能网卡。比如,本申请实施例中拥塞控制方法可根据网络协议,比如根据第四版网络协议或第六版网络协议标记的拥塞标识,标记根据传输控制协议得到的拥塞反馈标识。根据报文,采用本申请实施例中的拥塞控制方法,调整发送端的拥塞窗口的值。又比如,在服务器(即发送端)与以太网交换机(即交换节点)之间的 RoCE 过程中采用本申请实施例中的拥塞控制方法,调整发送端的拥塞窗口的值。

本申请实施例中的拥塞控制方法主要应用于一个连接中的发送端。在一个连接中包括发送端、交换节点和接收端。发送端可将数据包发送至交换节点,数据包进入交换节点中的队列。交换节点可用于对数据包进行拥塞标识的设置(即打标)。交换节点按照队列中数据包的排列顺序,将数据包依次向接收端发送。接收端依据接收到的数据包向发送端发送报文。发送端根据报文调制拥塞窗口的值。

图 2 为本申请实施例中的一种拥塞控制方法的流程图。如图 2 所示,拥塞控制方法包括步骤 S101 至步骤 S106,为了便于理解,以三方交互的流程进行说明。

在步骤 S101 中，发送端通过与接收端之间的连接，在一个拥塞窗口内向交换节点发送多个数据包。

其中，发送端发送的这多个数据包的数目即为第一值，第一值为连接的拥塞窗口的值。该连接处于拥塞避免（即 Congestion Avoidance）状态。具体的，发送端通过交换节点向接收端发送多个数据包。也就是说，发送端将多个数据包向交换节点发送，交换节点再将这多个数据包向接收端发送。

在步骤 S102 中，交换节点接收多个数据包，按照拥塞标识设置概率使得数据包指示连接拥塞或指示连接不拥塞，并将数据包发送至接收端。

这里数据包指示连接拥塞即表示连接在传输多个数据包中的一个或多个数据包时拥塞。数据包指示连接不拥塞即表示连接在传输多个数据包中的一个或多个数据包时不拥塞。

在一个示例中，可为数据包设置拥塞标识，携带拥塞标识的数据包指示连接拥塞。将拥塞标识在网络第三层即 IP 层标记，即标记在数据包的 IP 包头上。拥塞标识表征网络发生拥塞。比如，拥塞标识可以为拥塞经历（Congestion Experienced, CE）。例如，CE 为 IP 数据包的包头上设置的 11。不携带拥塞标识的数据包指示连接不拥塞。需要说明的是，交换节点的队列深度位于队列正常范围内，拥塞标识设置概率与队列深度正相关。交换节点的队列深度为等待发送的数据包的数目。在一个示例中，队列正常范围与带宽延时积、队列正常范围的最大值、队列正常范围的最小值、增加系数、拥塞标识最大设置概率、第一报文对应的第三值、慢启动（即 Slow Start）状态中发送端接收到的第二报文对应的第五值等参数中的一个或多个相关。慢启动（即 Slow Start）状态为连接建立初始的一段时间内的状态。慢启动状态开始于连接建立，终止于发送端首次接收到第一报文之后。第一报文和第二报文的相关说明将会在下面具体介绍。

交换节点的队列深度位于队列正常范围内的情况下，拥塞标识设置概率与队列深度可线性正相关，也可为非线性正相关。比如，图 3 为本申请实施例中一种拥塞标识设置概率与队列深度正相关的示意图。如图 3 所示，横坐标为队列深度，纵坐标为拥塞标识设置概率。设队列正常范围的最小值为  $K_{min}$ ，队列正常范围的最大值为  $K_{max}$ 。队列深度大于或等于  $K_{min}$ ，且队列深度小于或等于  $K_{max}$ ，拥塞标识设置概率与队列深度线性正相关。队列深度小于  $K_{min}$ ，则拥塞标识设置概率为 0。队列深度大于  $K_{max}$ ，则拥塞标识设置概率为 100%。

在步骤 S103 中，接收端根据多个数据包中的全部数据包或者部分数据包，生成多个报文，并向发送端发送多个报文。

在一个示例中，接收端可将接收到的多个数据包中的全部数据包作为生成报文的依据。比如，发送端向接收端发送了 100 个数据包。在这 100 个数据包中，20 个数据包指示连接拥塞，80 个数据包指示连接不拥塞。则可根据这 100 个数据包，生成多个报文，并向发送端发送。

在另一个示例中，接收端也可将接收到的多个数据包中的部分数据包作为生成报文的依据。比如，发送端向接收端发送了 100 个数据包。在这 100 个数据包中，20 个数据包指示连接拥塞，80 个数据包指示连接不拥塞。接收端可随机或按照某种规律在 100 个数据包中选取 60 个数据包。根据选取的 60 个数据包，生成多个报文，并向发送端发送。

其中，多个报文包括一个或多个第一报文，以及一个或多个第二报文。第一报文用于表示连接在传输多个数据包中的一个或多个数据包时拥塞。第二报文用于表示连接在传输多个数据包中的一个或多个数据包时不拥塞。

在一个示例中，第一报文可以为用于指示网络拥塞而使用的一个专门的报文，例如专门

的拥塞通知报文 (Congestion Notification Packet, CNP) 等。第二报文可以是数据包的确认字符 (Acknowledge, ACK)。

在另一个示例中, 多个报文可以为反馈报文。每个反馈报文还可用于指示多个数据包中一定数目的数据包达到接收端。其中, 一定数目为一个或多个。也就是说, 每个反馈报文可对应一个或多个数据包。在多个报文为反馈报文的情况下, 第一报文携带拥塞反馈标识, 通过拥塞反馈标识表示连接在传输多个数据包中的一个或多个数据包时拥塞。例如该拥塞反馈标识可以是基于显示拥塞通知反馈 (Explicit Congestion Notification Echo) 技术在 ACK 中打上的标签等。具体的, 拥塞反馈标识可根据拥塞标识在报文中标记, 标记拥塞反馈标识的报文即为第一报文。具体的, 接收端接收到交换节点发送来的数据包, 若数据包的 IP 包头标记有拥塞标识, 则可将拥塞反馈标识在网络第四层即 TCP 层标记, 可标记在 TCP 包头上。第一报文的拥塞反馈标识可具体为显式拥塞通知反馈 (Explicit Congestion Notification Echo, ECE)。例如, ECE 为 TCP 数据包的包头上设置的 1。第二报文不携带拥塞反馈标识, 即接收端生成的第二报文中并不标记拥塞反馈报文。

在步骤 S104 中, 发送端接收多个数据包中的全部数据包或部分数据包对应的多个报文。

在步骤 S105 中, 发送端根据多个报文调整拥塞窗口, 以将拥塞窗口的值调整到第二值。

发送端接收多个报文, 可对多个报文进行解析, 从而确定多个报文中的第一报文和第二报文。比如, 报文为反馈报文, 则发送端接收到反馈报文后, 对反馈报文进行解析。若解析反馈报文得到拥塞反馈标识, 则确定该报文为第一报文。若解析反馈报文未得到拥塞反馈标识, 则确定该报文为第二报文。

第一报文使拥塞窗口的值减小, 第二报文使拥塞窗口的值增大。比如, 多个报文中的一个第一报文, 使拥塞窗口的值减小第三值。多个报文中的一个第二报文, 使拥塞窗口的值增加第四值。第三值与一个第一报文指示的数据包的数目相关。第四值与一个第二报文指示的数据包的数目相关。在一个示例中, 第三值与一个第一报文指示的数据包的数目正相关; 第四值与一个第二报文指示的数据包的数目正相关。

比如, 对应  $N$  个数据包的第一报文所对应的第三值, 为对应一个数据包的第一报文所对应的第三值的  $N$  倍。对应  $N$  个数据包的第二报文所对应的第四值, 为对应一个数据包的第二报文所对应的第四值的  $N$  倍。其中,  $N$  为正整数。

在一个示例中, 接收端设置每个第一报文和第二报文均各自对应固定数目个数据包, 固定数目为一个或多个。则每个第一报文对应的第三值均相等。每个第二报文对应的第四值相等。比如, 每个第一报文均对应两个指示连接拥塞的数据包, 一个指示连接拥塞的数据包对应的第一报文可调整拥塞窗口的值的减少量为  $\beta$ , 则每个第一报文对应的第三值为  $2\beta$ 。每个第二报文均对应两个指示连接不拥塞的数据包, 一个指示连接不拥塞的数据包对应的第二报文可调整拥塞窗口的值的增加量为  $\theta$ , 则每个第二报文对应的第四值为  $2\theta$ 。需要说明的是, 第一报文对应的数据包数目与第二报文对应的数据包数目可以相同, 也可以不同, 在此并不限定。

在另一个示例中, 接收端可不限定每个第一报文对应的数据包的数目和每个第二报文对应的数据包的数目。比如, 处于拥塞避免状态的连接中发送端接收到两个第一报文和两个第二报文。第一个第一报文对应一个数据包, 第二个第一报文对应两个数据包。一个指示连接拥塞的数据包对应的第一报文可调整拥塞窗口的值的减少量为  $\beta$ 。则第一个第一报文对应的第三值为  $\beta$ 。第二个第一报文对应的第三值为  $2\beta$ 。第一个第二报文对应两个数据包, 第二个第二报文对应一个数据包。一个指示连接不拥塞的数据包对应的第二报文可调整拥塞窗口

的值的增加量为  $\theta$ 。则第一条第二报文对应的第四值为  $2\theta$ 。第二条第二报文对应的第四值为  $\theta$ 。

每个报文可对应一个或多个数据包。每个报文对应的数据包数目可以相同。报文对应数据包的数目可在接收端与发送端之间预先设定，比如，预先设定每个报文对应一个数据包。又比如，预先设定每个报文对应两个数据包。每个报文对应的数据包数目也可以不同。可在报文中的空闲字节中设置数据包数目位，通过解析报文，读取数据包数目位上的信息，确定报文对应的数据包数目。需要说明的是，第一报文可以为特殊的专用于表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文，该特殊的报文可对应一个用于表示连接拥塞的数据包，或者对应多个专用于表示连接拥塞的数据包。该特殊的报文对应的专用于表示连接拥塞的数据包的数目的设置可参考与上述普通的报文对应的数据包的数目的设置。

针对每一个数据包各生成一个报文，可实现发送端的细粒度发送速率控制，从而提高拥塞控制的精度。但考虑到拥塞控制的资源占用等计算代价，可适当设置两个以上的数据包共享一条报文，从而达到计算代价和拥塞控制精度的平衡。

在一个示例中，在调整拥塞窗口的过程中，发送端每接收到一个报文，就可实时根据该报文进行拥塞窗口的调整。若发送端接收到的多个报文中的一个报文为第一报文，则根据接收到的这一个第一报文，将拥塞窗口的值减小与该第一报文对应的第三值。若发送端接收到的多个报文中一个报文为第二报文，则根据接收到的这一个第二报文，将拥塞窗口的值增大与该第二报文对应的第四值。

需要说明的是，上述说明内容中的一个指示连接拥塞的数据包对应的第一报文可调整拥塞窗口的值的减少量（即对应一个数据包的第一报文对应的第三值），以及一个指示连接不拥塞的数据包对应的第二报文可调整拥塞窗口的值的增加量（即对应一个数据包的第二报文对应的第四值）可根据具体工作场景和工作需求预先设定。在一个示例中，第四值还与第一值和增加系数相关。比如，对应一个数据包的第二报文对应的第四值为增加系数与第一值的商。设对应一个数据包的第二报文对应的第四值为  $\theta$ ；增加系数为  $\gamma$ ；第一值为  $Cwnd1$ 。增加系数也可根据具体工作场景和工作需求设定。则上述参数之间的关系如算式（1）所示：

$$\theta = \gamma / Cwnd1 \quad (1)$$

在具体操作中，发送端可根据第一报文的数目、每个第一报文使拥塞窗口的值减小的量、第二报文的数目以及每个第二报文使拥塞窗口的值增加的量，计算出第二值。发送端得到第二值，将拥塞窗口的值从第一值调整到第二值。根据第一报文的数目，以及每个第一报文使拥塞窗口的值减小的量，可以得到所有第一报文使拥塞窗口的值减小的总量。根据第二报文的数目，以及每个第二报文使拥塞窗口的值增加的量，可以得到所有第二报文使拥塞窗口的值增加的总量。根据所有第一报文使拥塞窗口的值减小的总量、所有第二报文使拥塞窗口的值增加的总量以及拥塞窗口的值，可以得到第二值。

也就是说，上述说明中的第二值为发送端接收到一个拥塞窗口所有报文多次调整后的拥塞窗口的值。比如，第二值为第一值、第一和的相反数与第二和的加和。其中，第一和为多个报文中的每个第一报文对应的第三值的加和。第二和为多个报文中的每个第二报文对应的第四值的加和。例如，处于拥塞避免状态的连接中一个拥塞窗口中，发送端接收到一个第一报文和两个第二报文。假设第一值为  $Cwnd1$ ；第二值为  $Cwnd2$ ；第一报文对应的第三值为  $\beta_1$ ；第一个第二报文对应的第四值为  $\theta_1$ ；第二个第二报文对应的第四值为  $\theta_2$ 。则可以得到算式（2）如下：

$$Cwnd2 = Cwnd1 - \beta_1 + \theta_1 + \theta_2 \quad (2)$$

本申请实施例中的拥塞控制方法中，接收端根据一个拥塞窗口的数据包中的拥塞标识，生成表示连接在传输多个数据包中的一个或多个数据包时拥塞的第一报文，和表示连接在传输多个数据包中的一个或多个数据包时不拥塞的第二报文。发送端根据多个报文中的第一报文和第二报文，有比例的减小和增加拥塞窗口的值，从而对拥塞窗口的值进行更精细的调整。在本申请实施例中，拥塞窗口的调整中存在李雅普诺夫平衡点，在对拥塞窗口的值的不断调整中，使得交换节点的队列深度得到控制，达到较浅的一个稳定队列深度。

比如，设对应一个数据包的第二报文对应的第四值如算式 (1) 所示；拥塞标识设置概率为  $p$ ；时刻  $t$  的拥塞窗口的值为  $Cwnd(t)$ ，则在时刻  $t$  经过一个往返时延 (Round-trip Time, RTT) 的拥塞窗口的值  $Cwnd(t+Rtt)$  计算见算式 (3)：

$$Cwnd(t+Rtt) = Cwnd(t) + \gamma \times (1-p) - \beta \times Cwnd(t) \times p \tag{3}$$

其中， $\gamma$  为上述说明内容中的增加系数。 $\beta$  为对应一个数据包的第二报文对应的第四值。根据算式 (3) 可以得到算式 (4)，算式 (4) 如下：

$$(dCwnd/dt) = (1/RTT) \times [(\gamma / Cwnd) \times (1-p) - \beta \times p] \tag{4}$$

令  $\Sigma (dCwnd/dt) = 0$ ，可得， $\Sigma Cwnd = (n \gamma / \beta) [(1/p) - 1]$ 。其中， $n$  为网络链路流的数目。当  $\Sigma Cwnd$  等于带宽延时积 (Bandwidth-Delay Product, BDP) 时达到李雅普诺夫平衡点，可计算得到较小的稳定队列深度  $Q$ ，具体计算如下：

$$Q = (K \times n \gamma) / [(BDP \times \beta) + n \gamma] \tag{5}$$

其中， $K$  为队列深度的最大值与队列深度的最小值的差值。

达到李雅普诺夫平衡点的情况下，交换节点中队列深度将稳定维持在较浅的稳定队列深度  $Q$ 。在现有技术中，交换节点中的队列深度上下波动非常大，且队列深度非常大，队列深度可达到 100 至 256。由于交换节点中的队列缓存有限，队列深度大导致交换节点中的队列缓存紧张，数据传输过程中的延时增大，使得丢包率增高。采用本申请实施例的拥塞控制方法，发送端在不断调整拥塞窗口的过程中，按照第一报文和第二报文，有比例地减小和增加拥塞窗口的值。使得拥塞窗口的值可稳定维持在一个较浅的稳定队列深度，比如，本申请实施例中的稳定队列深度可达到 10 至 20。稳定队列深度较浅，为交换节点留出了足够的队列缓存。从而降低了数据传输过程中的延时，使得丢包率更加可控，而更能抗击突发因素对丢包率等网络性能的破坏，使得数据传输更加稳定，质量更有保障。而且，采用本申请实施例中的拥塞控制方法，可以在不启用基于优先级的流控控制 (Priority-based Flow Control, PFC) 等复杂机制的情况下，更加精确地对拥塞窗口进行控制，提高网络的稳定性。可在避免死锁等稳定性问题的基础上，保证网络的稳定性，从而提升网络服务质量。

而且，在本申请实施例中，稳定队列深度不为 0 的情况下，即保证队列中一直存在数据包，也可保证较高的网络带宽利用率。从而达到了数据传输过程中延时低和网络带宽利用率高的协调平衡。

比如，在一个示例中，在相同网络条件下采用传输控制协议 (Transmission Control Protocol, TCP) 算法、数据中心量化拥塞通知 (Data Center Quantized Congestion Notification, DCQCN) 算法和本申请方案算法的效果如表一所示：

	利用的带宽	队列深度
TCP 算法	13.8Gbps	0 至 256
DCQCN 算法	25.2 Gbps	0 至 100
本申请方案算法	35.9 Gbps	0 至 20

表一

可以得知，本申请方案算法的带宽明显高于 TCP 算法和 DCQCN 算法，且本申请方案算法的队列深度明显低于 TCP 算法和 DCQCN 算法，且队列深度得到有效控制，变化较小，可保证极低丢包率。而且，队列深度始终保持在一个稳定的队列深度值上，也能够进一步提高网络的带宽的利用率。

在步骤 S106 中，发送端向连接的接收端发送数目为第二值的多个数据包。

调整拥塞窗口的值后，在下一个拥塞窗口发送数据包的过程中，发送端向交换节点发送调整后得到的第二值数目个数据包，以调整交换节点的队列深度。

在另一个实施例中，上述连接还可处于慢启动状态。在步骤 S101 之前，即在发送端的连接处于拥塞避免状态之前，发送端的连接处于慢启动（即 Slow Start）状态。慢启动状态开始于连接初始建立，终止于发送端首次接收到表示连接拥塞的第一报文之后。在慢启动状态中，发送端可根据处于慢启动状态的连接的拥塞窗口的第六值，向处于慢启动状态的连接的接收端发送多个数据包。处于慢启动状态的连接中的多个数据包的数目为上述第六值。

具体的，发送端通过交换节点向接收端发送数据包。也就是说，发送端先将数据包发送至交换节点，交换节点按照拥塞标识设置概率使得数据包指示连接拥塞或指示连接不拥塞，并将数据包发送至接收端。接收端根据数据包，生成慢启动报文。其中，慢启动报文为处于慢启动状态的连接中多个数据包对应的报文。慢启动报文包括一个第一报文和整数个第二报文。接收端将第一报文和第二报文发送至发送端。发送端根据处于慢启动状态的连接中的慢启动报文中的每一个报文调整拥塞窗口，以将拥塞窗口的值调整至第一值。

慢启动报文中的每个报文的作用与上述拥塞避免状态的连接中的报文基本相同。需要注意的是，在处于慢启动状态的连接中，每一个第一报文，使拥塞窗口减小第三值。每一个第二报文，使拥塞窗口增加第五值。

在慢启动状态终止后，可切换至拥塞避免状态。处于拥塞避免状态的连接开始于发送端首次接收到携带拥塞反馈标识的第一报文之后。在处于慢启动状态的连接中不断调整拥塞窗口的值，慢启动状态终止时拥塞窗口的值即为处于拥塞避免状态的连接开始时的拥塞窗口的值。

为了便于理解，下面举例说明从慢启动状态到拥塞避免状态的整个过程中拥塞控制方法中拥塞窗口的值的调整算法。

假设处于慢启动状态的连接中对应一个数据包的第一报文对应的第三值为  $\beta$ ；处于慢启动状态的连接中对应一个数据包的第二个报文对应的第五值为  $\alpha$ ；处于拥塞避免状态的连接中对应一个数据包的第一报文对应的第三值为  $\beta$ ；处于拥塞避免状态的连接中对应一个数据包的第二个报文对应的第四值为  $\theta$ 。设  $\theta$  如算式 (1) 所示。假设一个拥塞窗口中，发送端接收到了 10 条报文。每条报文均对应一个数据包。其中，10 条报文按顺序分别为第二报文、第二报文、第一报文、第二报文、第一报文、第一报文、第二报文、第二报文、第一报文和第二报文。

则前两条第二报文和第一条第一报文均为处于慢启动状态的连接中的报文，剩余部分的第一报文和第二报文均为处于拥塞避免状态的连接中的报文。设第六值为  $Cwnd3$ ，则第一值  $Cwnd1$  的计算可如算式 (6) 所示：

$$Cwnd1 = Cwnd3 + \alpha + \alpha - \beta = Cwnd3 - 2 \times \alpha - \beta \quad (6)$$

第二值  $Cwnd2$  的计算可如算式 (7) 所示：

$$\begin{aligned} Cwnd2 &= Cwnd1 + (\gamma / Cwnd1) - \beta - \beta + (\gamma / Cwnd1) + (\gamma / Cwnd1) - \beta + (\gamma / Cwnd1) \\ &= Cwnd3 - 2 \times \alpha - 4 \times \beta + 4 \times (\gamma / Cwnd1) \end{aligned} \quad (7)$$

则经过一个拥塞窗口不断的调整，下一拥塞窗口的值即第二值  $Cwnd2$  为  $Cwnd3 - 2 \times \alpha - 4 \times \beta + 4 \times (\gamma / Cwnd1)$ 。

需要理解的是，上述的参数  $\alpha$ ， $\beta$  和  $\gamma$ ，可以根据网络的拓扑和网络参数等信息，先在数据传输前，例如组网过程中，或者连接建立过程中使用上述举例的计算方式或者其他计算方式确定后，作为系统参数配置好，在连接传输数据的过程中（如上文描述的从 S101 开始的过程，或者慢启动状态中），直接使用即可。又例如，可预先设置期望的稳定队列深度  $Q$ ，设置可达到期望的稳定队列深度  $Q$  的  $\alpha$ ， $\beta$  和  $\gamma$ ，并将  $\alpha$ ， $\beta$  和  $\gamma$  作为系统参数配置。

下面将以一实例说明本申请实施例中拥塞控制方法的过程。设置发送端初始发送的拥塞窗口值为 100，即发送端初始发送的拥塞窗口为 100 个数据包。队列深度的最小值  $K_{min}=0$ ，队列深度的最大值  $K_{max}=40$ 。初始的队列深度为 20。初始的拥塞标识设置概率为 10%。对应一个数据包的第一报文的第三值  $\beta=0.5$ ，增加系数  $\gamma=1.0$ 。在队列正常范围内，拥塞标识设置概率的曲线公式为  $y=0.005x$ ，其中， $y$  为拥塞标识设置概率， $x$  为队列深度。假设，每个数据包均对应一条报文。

第一次拥塞窗口的值调整：发送端将根据初始的拥塞窗口值 100，将 100 个数据包发送至交换节点。交换节点按照初始的拥塞标识设置概率 10% 设置用于指示连接拥塞的数据包和指示连接不拥塞的数据包，则 100 个数据包包括 90 个指示连接不拥塞的数据包和 10 个指示连接拥塞的数据包。接收端接收 90 个指示连接不拥塞的数据包和 10 个指示连接拥塞的数据包，生成 90 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 10 个未表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文。发送端接收接收端发送的 90 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 10 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文，并进行第一次拥塞窗口值的调整。调整后的拥塞窗口值为  $100 + 1.0 \times (90/100) - 0.5 \times 10 \approx 96$ 。调整后的拥塞窗口值的计算方法具体可参见上述算式 (3)。

第二次拥塞窗口的值调整：发送端将根据第一次调整后的拥塞窗口值 96，将 96 个数据包发送至交换节点。由于拥塞窗口值减小，因此队列深度也对应减小。计算第二次拥塞窗口值调整所需的拥塞标识设置概率为  $[20 - (100 - 96)] \times 0.05 = 8\%$ 。交换节点按照拥塞标识设置概率 8% 设置用于指示连接拥塞的数据包和指示连接不拥塞的数据包，则 96 个数据包包括 88 个指示连接不拥塞的数据包和 8 个指示连接拥塞的数据包。接收端接收 88 个指示连接不拥塞的数据包和 8 个指示连接拥塞的数据包，生成 88 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 8 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文。发送端接收接收端发送的 88 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 8 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文，并进行第二次拥塞窗口值的调整。调整后的拥塞窗口值为  $96 + 1.0 \times (88/96) - 0.5 \times 8 \approx 93$ 。

第三次拥塞窗口的值调整：发送端将根据第二次调整后的拥塞窗口值 93，将 93 个数据包发送至交换节点。由于拥塞窗口值减小，因此队列深度也对应减小。计算第二次拥塞窗口值调整所需的拥塞标识设置概率为  $[20 - (100 - 93)] \times 0.05 = 6.5\%$ 。交换节点按照拥塞标识设置概率 6.5% 设置用于指示连接拥塞的数据包和指示连接不拥塞的数据包，则 93 个数据包包括 87 个指示连接不拥塞的数据包和 6 个指示连接拥塞的数据包。接收端接收 87 个指示连接不拥塞的数据包和 6 个指示连接拥塞的数据包，生成 87 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 6 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文。发送端接收接收端发送的 87 条表示连接在传输多个数据包中的一个或多个数据包

时不拥塞的报文和 6 条表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文，并进行第三次拥塞窗口值的调整。调整后的拥塞窗口值为  $93+1.0 \times (87/93)-0.5 \times 6 \approx 91$ 。

依次类推，直至第八次拥塞窗口的值调整。

第八次拥塞窗口的值调整：发送端将根据第七次调整后的拥塞窗口值 86，将 86 个数据包发送至交换节点。由于拥塞窗口值减小，因此队列深度也对应减小。计算第二次拥塞窗口值调整所需的拥塞标识设置概率为  $[20-(100-86)] \times 0.05=3\%$ 。交换节点按照拥塞标识设置概率 3% 设置用于指示连接拥塞的数据包和指示连接不拥塞的数据包，则 86 个数据包包括 83 个指示连接不拥塞的数据包和 3 个指示连接拥塞的数据包。接收端接收 83 个指示连接不拥塞的数据包和 3 个指示连接拥塞的数据包，生成 83 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 3 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文。发送端接收接收端发送的 83 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 3 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文，并进行第八次拥塞窗口值的调整。调整后的拥塞窗口值为  $86+1.0 \times (83/86)-0.5 \times 3 \approx 85$ 。

第九次拥塞窗口的值调整：发送端将根据第八次调整后的拥塞窗口值 85，将 85 个数据包发送至交换节点。由于拥塞窗口值减小，因此队列深度也对应减小。计算第二次拥塞窗口值调整所需的拥塞标识设置概率为  $[20-(100-85)] \times 0.05=2.5\%$ 。交换节点按照拥塞标识设置概率 2.5% 设置用于指示连接拥塞的数据包和指示连接不拥塞的数据包，则 85 个数据包包括 83 个指示连接不拥塞的数据包和 2 个指示连接拥塞的数据包。接收端接收 83 个指示连接不拥塞的数据包和 2 个指示连接拥塞的数据包，生成 83 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 2 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文。发送端接收接收端发送的 83 个表示连接在传输多个数据包中的一个或多个数据包时不拥塞的报文和 2 个表示连接在传输多个数据包中的一个或多个数据包时拥塞的报文，并进行第九次拥塞窗口值的调整。调整后的拥塞窗口值为  $85+1.0 \times (83/85)-0.5 \times 2 \approx 85$ 。

依次类推，在之后的拥塞窗口的值调整中，队列深度稳定为 5，不再发生较大波动。从而保证了较小的稳定队列深度，为交换节点留出了足够的队列缓存，降低了数据传输过程中的延时，从而降低了丢包率。

下面将以发送端为执行主体说明本申请实施例中拥塞控制方法的实施步骤。图 4 为本申请实施例中一种拥塞控制方法的流程图。图 4 所示的拥塞控制方法应用于一个连接的发送端。如图 4 所示，拥塞控制方法包括步骤 S201 至步骤 S203。

在步骤 S201 中，发送端通过与接收端之间的连接，在一个拥塞窗口内向连接的接收端发送多个数据包。

其中，多个数据包的数目为第一值，第一值为连接的拥塞窗口的值。其中，连接处于拥塞避免状态。

在步骤 S202 中，发送端接收来自接收端的与多个数据包中的全部数据包或者部分数据包对应的多个报文。

其中，多个报文包括一个或多个第一报文，以及一个或多个第二报文，第一报文用于表示连接在传输多个数据包中的一个或多个数据包时拥塞，第二报文用于表示连接在传输多个数据包中的一个或多个数据包时不拥塞。

在步骤 S203 中，发送端根据多个报文调整拥塞窗口，以将拥塞窗口的值调整到第二值。

其中，多个报文中的一个第一报文，使拥塞窗口减小。多个报文中的一个第二报文，使拥塞窗口增加。

上述步骤 S201 至步骤 S203 的相关说明请参见上述实施例中步骤 S101 至步骤 S106 的相关说明，在此不再赘述。

下面介绍本申请实施例中的一种网卡以及一种设备，例如该设备可以是终端或者服务器，甚至是某些网络设备，例如云平台等。图 5 为本申请实施例中一种设备 30 的结构示意图。如图 5 所示，设备包括处理器 31、存储器 32 和网卡 33。其中网卡 33 可包括处理电路（例如 ASIC 电路，FPGA，DSP 等，图 5 中将处理电路示意为处理芯片 331）、和存储介质 332 和通信接口 333。在一个示例中，网卡 33 还可包括连接接口（图 5 中将连接接口示意为外部组件连接快速接口 334）。在一个示例中，上述实施例中的拥塞控制方法可应用于网卡 33。也就是说，网卡 33 中的存储介质 332 用于存储程序。处理芯片 331 用于执行程序，程序被处理芯片 332 执行时可实现上述实施例中的拥塞控制方法。外部组件连接快速接口 334 用于将网卡 33 连接在设备 30 上。通信接口 333 被处理芯片 331 调用，使得网卡向其他设备发送数据包，以及接收上述的报文。存储介质 332 还可用于存储拥塞控制方法涉及到的参数，比如处于慢启动状态的连接中对应一个数据包的第一报文对应的第三值  $\beta$ 、处于慢启动状态的连接中对应一个数据包的第二报文对应的第四值  $\theta$ 、处于拥塞避免状态的连接中对应一个数据包的第一报文对应的第三值  $\beta$ 、处于拥塞避免状态的连接中对应一个数据包的第二报文对应的第五值  $\alpha$ ，以及拥塞窗口的值等参数。需理解，图 5 中描述的网卡可以安装在各种设备上，本申请不限定设备的结构，图 5 只做示意性说明。

应理解的是，类似于网卡，本申请所描述的方法还可以应用于芯片或者其他装置，这些装置的构成类似网卡，包括存储介质，处理电路和通信接口，且三者之间电连接，以实现本申请上述的方法，本申请不再赘述。

在另一个示例中，上述实施例中的拥塞控制方法可应用于设备 30。也就是说，设备中的存储器 32 相当于用于存储程序的存储介质。处理器 31 相当于用于执行存储介质中的程序的处理电路，程序被处理器 31 执行时可实现本申请所记载的拥塞控制方法。设备中的存储器 32 属于存储介质中的一种，处理器 31 属于处理电路的一种。而网卡 33 则可认为是设备 30 的通信接口一种实现。一种实现方式下，设备 30 的通信接口还可以包括天线等硬件器件。通信接口主要用于实现本申请实施例中各模块、装置、单元和/或设备之间的通信。也可通过通信接口接入输入设备和/或输出设备。

上述处理器 31 可以包括中央处理器(CPU)，或者特定集成电路(ASIC)，或者可以被配置成实施本申请实施例的一个或多个集成电路。

存储器 32 可以包括用于数据或指令的大容量存储器。举例来说而非限制，存储器 32 可包括 HDD、软盘驱动器、闪存、光盘、磁光盘、磁带或通用串行总线(USB)驱动器或者两个或更多个以上这些的组合。在合适的情况下，存储器 32 可包括可移除或不可移除(或固定)的介质。在合适的情况下，存储器 32 可在设备 30 的内部或外部。在特定实施例中，存储器 32 是非易失性固态存储器。在特定实施例中，存储器 32 包括只读存储器(ROM)。在合适的情况下，该 ROM 可以是掩模编程的 ROM、可编程 ROM(PROM)、可擦除 PROM(EPROM)、电可擦除 PROM(EEPROM)、电可改写 ROM(EAROM)或闪存或者两个或更多个以上这些的组合。

处理器 31 通过读取存储器 32 中存储的可执行程序代码来运行与可执行程序代码对应的程序，以用于执行上述各个实施例中的发送端的拥塞控制方法。

在一个示例中，设备 30 还包括总线 35 其中，如图 5 所示，处理器 31、存储器 32、网卡 33 通过总线 34 连接并完成相互间的通信。

总线 34 包括硬件、软件或两者，将设备的部件彼此耦接在一起。举例来说而非限制，总

线 34 可包括加速图形端口 (AGP) 或其他图形总线、增强工业标准架构 (EISA) 总线、前端总线 (FSB)、超传输 (HT) 互连、工业标准架构 (ISA) 总线、无限带宽互连、低引脚数 (LPC) 总线、存储器总线、微信道架构 (MCA) 总线、外围组件互连 (PCI) 总线、PCI-Express (PCI-X) 总线、串行高级技术附件 (SATA) 总线、视频电子标准协会局部 (VLB) 总线或其他合适的总线或者两个或更多个以上这些的组合。在合适的情况下，总线 34 可包括一个或多个总线。尽管本申请实施例描述和示出了特定的总线，但本申请考虑任何合适的总线或互连。

本申请实施例还提供了一种存储介质，该存储介质上存储有程序。程序被处理电路执行时实现如本申请所记载的拥塞控制方法。存储介质可以包括能够存储或传输信息的任何介质。存储介质的例子包括电子电路、半导体存储器设备、ROM、闪存、可擦除 ROM (EROM)、软盘、CD-ROM、光盘、硬盘、光纤介质、射频 (RF) 链路，等等。程序、代码段可以经由诸如因特网、内联网等的计算机网络被下载。

图 6 为本发明实施例中一种拥塞控制装置的结构示意图。如图 6 所示，拥塞控制装置 400 包括发送模块 401、接收模块 402 和控制模块 403。其中，发送模块 401 和接收模块 402 可以通过处理电路执行存储介质中的可执行代码，调用通信接口实现，控制模块 403 可以通过处理电路执行存储介质中的可执行程序实现，关于处理电路、存储介质和通信接口的具体形态，可以参考图 5 对应的描述。关于图 6 所示装置的其他实现细节和有益效果的描述，请参考本申请的记载。

发送模块 401，用于通过与接收端之间的连接，在一个拥塞窗口内向接收端发送多个数据包。

其中，多个数据包的数目为第一值。第一值为连接的拥塞窗口的值。其中，连接处于拥塞避免状态。

在一个示例中，发送模块 401 还用于向连接的接收端发送数目为第二值的多个数据包。

接收模块 402，用于接收来自接收端的与多个数据包中的全部数据包或者部分数据包对应的多个报文。

多个报文包括一个或多个第一报文，以及一个或多个第二报文，第一报文用于表示连接在传输多个数据包中的一个或多个数据包时拥塞，第二报文用于表示连接在传输多个数据包中的一个或多个数据包时不拥塞。

在一个示例中，第一报文为拥塞通知报文，第二报文为确认字符。

在另一个示例中，多个报文为反馈报文。每个反馈报文还用于指示多个数据包中一定数目的数据包到达接收端。一定数目为一个或多个。第一报文携带拥塞反馈标识，第二报文不携带拥塞反馈标识。

控制模块 403，用于根据多个报文调整拥塞窗口，以将拥塞窗口的值调整到第二值。

其中，多个报文中的第一报文，使拥塞窗口的值减小，多个报文中的第二报文，使拥塞窗口的值增加。

在一个示例中，多个报文中的一个第一报文，使拥塞窗口的值减小第三值，多个报文中的一个第二报文，使拥塞窗口的值增加第四值，第三值与一个第一报文指示的数据包的数目相关，第四值与一个第二报文指示的数据包的数目相关。

在一个示例中，在根据多个报文调整拥塞窗口的方面，控制模块 403 用于根据多个报文中的一个第一报文，将拥塞窗口的值减小第三值；以及根据多个报文中的一个第二报文，将拥塞窗口的值增加第四值。

在一个示例中，在根据多个报文调整拥塞窗口的方面，控制模块 403 用于根据第一报文

的数目，每个第一报文使拥塞窗口的值减小的量，第二报文的数目以及每个第二报文使拥塞窗口的值增加的量，计算出第二值，以及将拥塞窗口的值从第一值调整到第二值。

上述第二值为第一值、第一和的相反数与第二和的加和，第一和为多个报文中的每个第一报文对应的第三值相加的和，第二和为多个报文中的每个第二报文对应的第四值相加的和。

在一个示例中，第四值还与第一值和增加系数相关。其中增加系数可以是预设的值。

本说明书的各个部分均采用递进的方式进行描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点介绍的都是与其他实施例不同之处。尤其，对于装置实施例、设备实施例和存储介质实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例部分的说明即可。

以上，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以权利要求的保护范围为准。

## 权 利 要 求 书

1、一种拥塞控制方法，其特征在于，所述拥塞控制方法包括：

所述发送端通过与接收端之间的连接，在一个拥塞窗口内向所述接收端发送多个数据包，所述多个数据包的数目为第一值，所述第一值为所述连接的拥塞窗口的值，其中，所述连接处于拥塞避免状态；

所述发送端接收来自接收端的与所述多个数据包中的全部数据包或者部分数据包对应的多个报文，所述多个报文包括一个或多个第一报文，以及一个或多个第二报文，所述第一报文用于表示所述连接在传输所述多个数据包中的一个或多个数据包时拥塞，所述第二报文用于表示所述连接在传输所述多个数据包中的一个或多个数据包时不拥塞；

所述发送端根据所述多个报文调整所述拥塞窗口，以将所述拥塞窗口的值调整到第二值，其中，所述多个报文中的第一报文，使所述拥塞窗口的值减小，所述多个报文中的第二报文，使所述拥塞窗口的值增加。

2. 根据权利要求 1 所述的方法，其特征在于，所述第一报文为拥塞通知报文 CNP，所述第二报文为确认字符 ACK。

3. 根据权利要求 1 所述的方法，其特征在于，所述多个报文为反馈报文，每个所述反馈报文还用于指示所述多个数据包中一定数目的数据包到达所述接收端，所述一定数目为一个或多个；

所述第一报文携带拥塞反馈标识，所述第二报文不携带拥塞反馈标识。

4. 根据权利要求 3 所述的方法，其特征在于，所述多个报文中的一个第一报文，使所述拥塞窗口的值减小第三值，所述多个报文中的一个第二报文，使所述拥塞窗口的值增加第四值，所述第三值与一个所述第一报文指示的数据包的数目相关，所述第四值与一个所述第二报文指示的数据包的数目相关。

5、根据权利要求 3 或 4 所述的方法，其特征在于，所述第二值为所述第一值、第一和的相反数与第二和的加和，所述第一和为所述多个报文中的每个第一报文对应的第三值相加的和，所述第二和为所述多个报文中的每个第二报文对应的第四值相加的和。

6、根据权利要求 3 所述的方法，其特征在于，所述发送端根据所述多个报文调整所述拥塞窗口，包括：所述发送端根据所述多个报文中的一个所述第一报文，将所述拥塞窗口的值减小所述第三值；以及

所述发送端根据所述多个报文中的一个所述第二报文，将所述拥塞窗口的值增加所述第四值。

7. 根据权利要求 3 所述的方法，其特征在于，所述发送端根据所述多个报文调整所述拥塞窗口，包括：

所述发送端根据所述第一报文的数目，每个所述第一报文使所述拥塞窗口的值减小的量，所述第二报文的数目以及每个所述第二报文使所述拥塞窗口的值增加的量，计算出所述第二值，

所述发送端将所述拥塞窗口的值从所述第一值调整到所述第二值。

8、根据权利要求 1 至 7 中任意一项所述的方法，其特征在于，所述方法还包括：

所述发送端向所述连接的接收端发送数目为所述第二值的多个数据包。

9、一种拥塞控制装置，其特征在于，包括：

发送模块，用于通过与接收端之间的连接，在一个拥塞窗口内向所述接收端发送多个数

据包，所述多个数据包的数目为第一值，所述第一值为所述连接的拥塞窗口的值，其中，所述连接处于拥塞避免状态；

接收模块，用于接收来自接收端的与所述多个数据包中的全部数据包或者部分数据包对应的多个报文，所述多个报文包括一个或多个第一报文，以及一个或多个第二报文，所述第一报文用于表示所述连接在传输所述多个数据包中的一个或多个数据包时拥塞，所述第二报文用于表示所述连接在传输所述多个数据包中的一个或多个数据包时不拥塞；

控制模块，用于根据所述多个报文调整所述拥塞窗口，以将所述拥塞窗口的值调整到第二值，其中，所述多个报文中的第一报文，使所述拥塞窗口的值减小，所述多个报文中的第二报文，使所述拥塞窗口的值增加。

10. 根据权利要求9所述的装置，其特征在于，所述第一报文为拥塞通知报文CNP，所述第二报文为确认字符ACK。

11. 根据权利要求9所述的装置，其特征在于，所述多个报文为反馈报文，每个所述反馈报文还用于指示所述多个数据包中一定数目的数据包到达所述接收端，所述一定数目为一个或多个；

所述第一报文携带拥塞反馈标识，所述第二报文不携带拥塞反馈标识。

12. 根据权利要求11所述的装置，其特征在于，所述多个报文中的一个第一报文，使所述拥塞窗口的值减小第三值，所述多个报文中的一个第二报文，使所述拥塞窗口的值增加第四值，所述第三值与一个所述第一报文指示的数据包的数目相关，所述第四值与一个所述第二报文指示的数据包的数目相关。

13. 根据权利要求12所述的装置，其特征在于，所述第二值为所述第一值、第一和的相反数与第二和的加和，所述第一和为所述多个报文中的每个第一报文对应的第三值相加的和，所述第二和为所述多个报文中的每个第二报文对应的第四值相加的和。

14. 根据权利要求12所述的装置，其特征在于，在根据所述多个报文调整所述拥塞窗口的方面，所述控制模块用于根据所述多个报文中的一个所述第一报文，将所述拥塞窗口的值减小所述第三值；以及根据所述多个报文中的一个所述第二报文，将所述拥塞窗口的值增加所述第四值。

15. 根据权利要求12所述的装置，其特征在于，在根据所述多个报文调整所述拥塞窗口的方面，所述控制模块用于根据所述第一报文的数目，每个所述第一报文使所述拥塞窗口的值减小的量，所述第二报文的数目以及每个所述第二报文使所述拥塞窗口的值增加的量，计算出所述第二值，以及将所述拥塞窗口的值从所述第一值调整到所述第二值。

16. 根据权利要求9至15中任意一项所述的装置，其特征在于，所述发送模块还用于向所述连接的接收端发送数目为所述第二值的多个数据包。

17. 一种设备，其特征在于，包括存储介质、处理电路和通信接口；

所述存储介质用于存储可执行程序；

所述处理电路用于读取所述存储介质中存储的可执行程序，以在所述通信接口的配合下，执行权利要求1至8任意一项所述的拥塞控制方法。

18. 一种存储介质，其特征在于，所述存储介质上存储有程序；所述程序被处理电路执行时实现如权利要求1至8任意一项所述的拥塞控制方法。

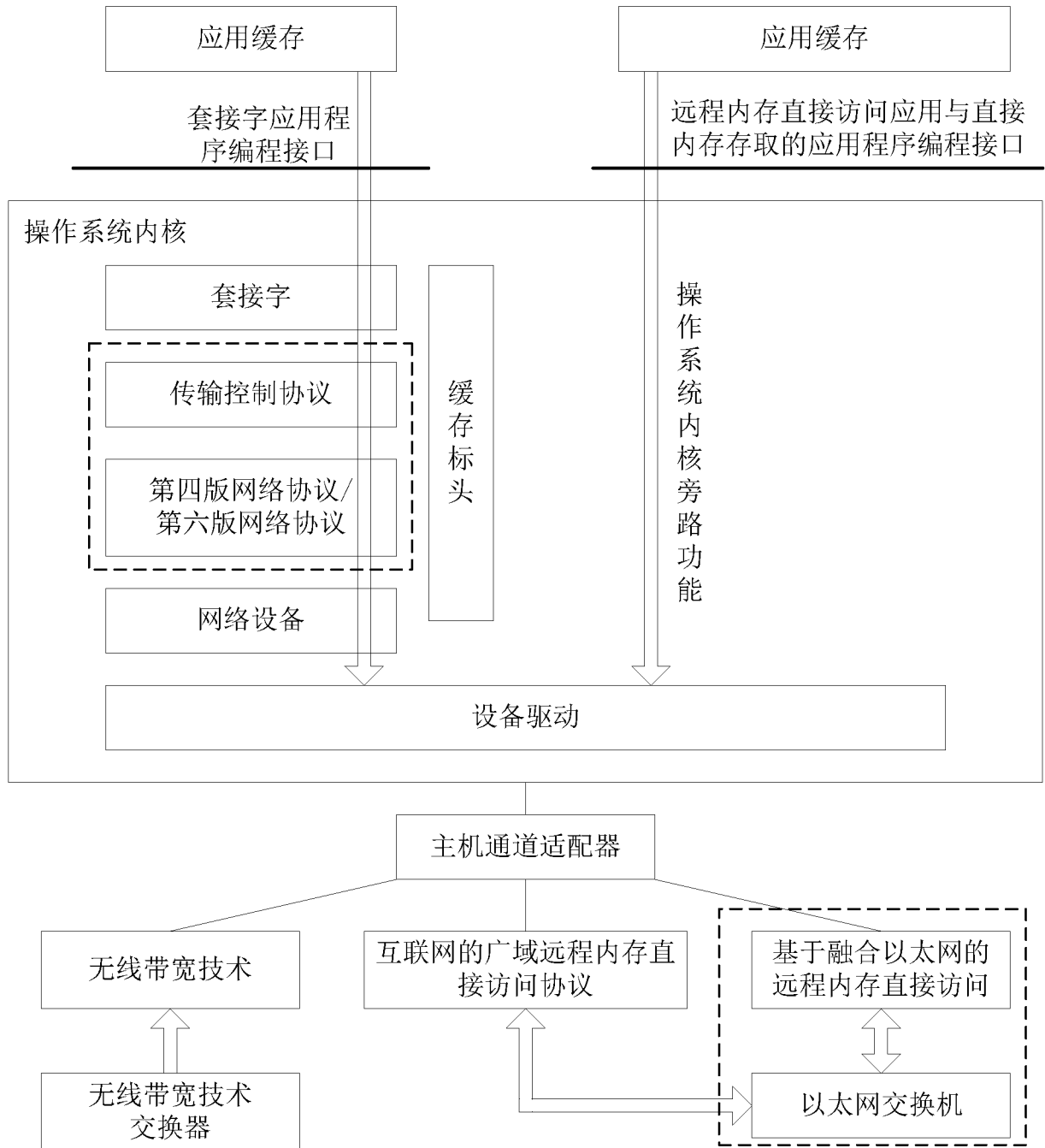


图 1

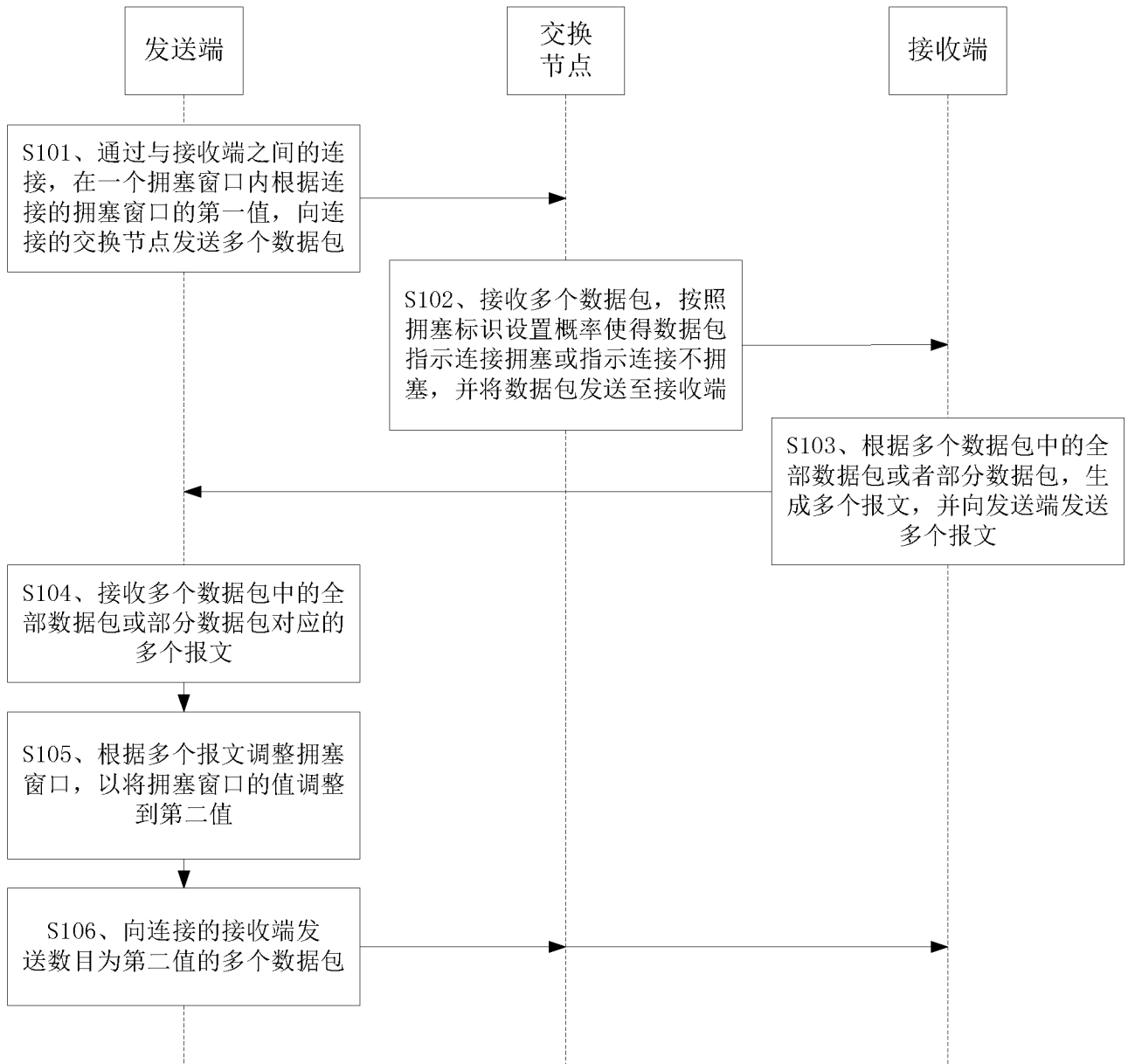


图 2

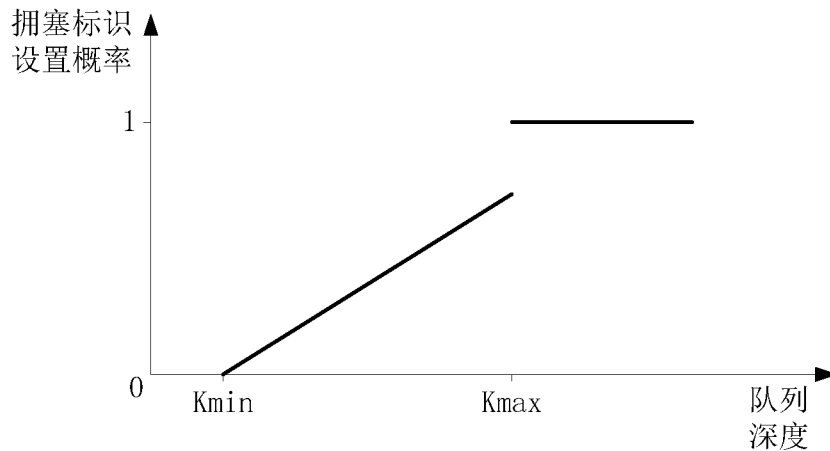


图 3

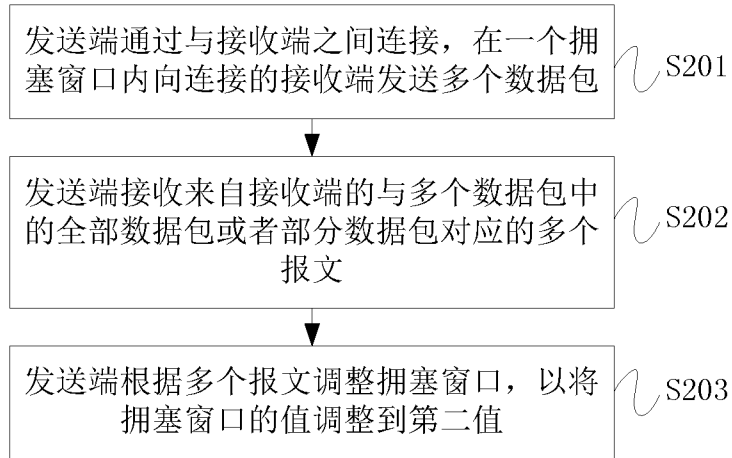


图 4

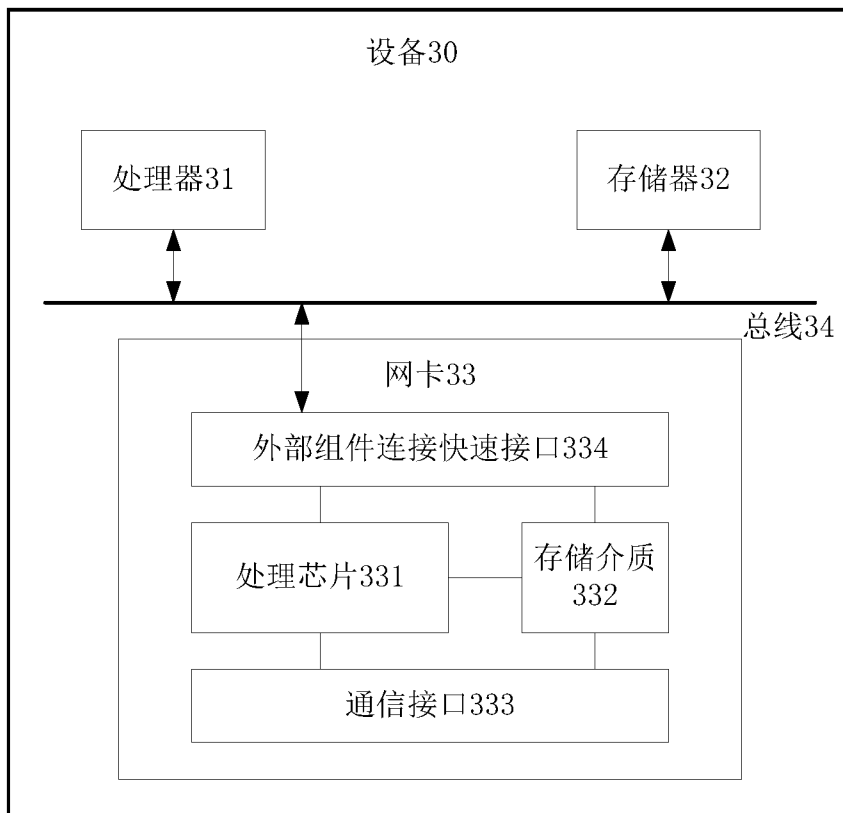


图 5

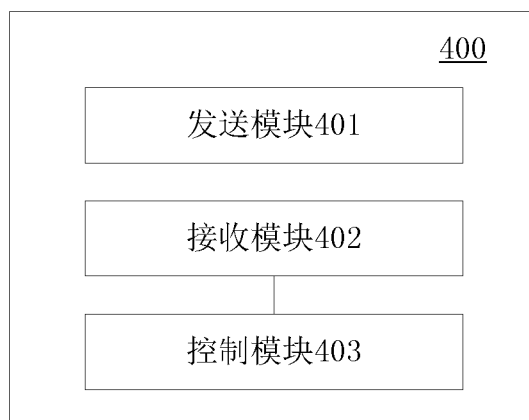


图 6

## INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2019/074385****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)

CNABS; CNTXT; CNKI; VEN; USTXT; EPTXT; WOTXT: 拥塞, 控制, 窗口, 增, 加, 减小, 调整, 通知, 反馈, 响应, congestion, control, ACK, CNP, window, modify, increase, decrease, notification, ECN, echo

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 103051555 A (BEIHANG UNIVERSITY) 17 April 2013 (2013-04-17) description, paragraphs [0019]-[0032]	1-18
Y	CN 103905327 A (INSTITUTE OF ACOUSTICS, CHINESE ACADEMY OF SCIENCES) 02 July 2014 (2014-07-02) description, paragraphs [0018]-[0045]	1-18
A	CN 103297346 A (BEIHANG UNIVERSITY) 11 September 2013 (2013-09-11) entire document	1-18
A	US 2014126357 A1 (LOCKHEED MARTIN CORP.) 08 May 2014 (2014-05-08) entire document	1-18

 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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

**08 March 2019**

Date of mailing of the international search report

**17 April 2019**

Name and mailing address of the ISA/CN

**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

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2019/074385**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	103051555	A	17 April 2013	None			
CN	103905327	A	02 July 2014	CN	103905327	B	17 May 2017
CN	103297346	A	11 September 2013	CN	103297346	B	10 May 2017
US	2014126357	A1	08 May 2014	US	8923123	B2	30 December 2014

国际检索报告

国际申请号

PCT/CN2019/074385

<p><b>A. 主题的分类</b></p> <p>H04L 12/801(2013.01)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>B. 检索领域</b></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;USTXT;EPTXT;WOTXT: 拥塞, 控制, 窗口, 增, 加, 减小, 调整, 通知, 反馈, 响应, congestion, control, ACK, CNP, window, modify, increase, decrease, notification, ECN, echo</p>																	
<p><b>C. 相关文件</b></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>Y</td> <td>CN 103051555 A (北京航空航天大学) 2013年 4月 17日 (2013 - 04 - 17) 说明书第[0019]-[0032]段</td> <td>1-18</td> </tr> <tr> <td>Y</td> <td>CN 103905327 A (中国科学院声学研究所) 2014年 7月 2日 (2014 - 07 - 02) 说明书第[0018]-[0045]段</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 103297346 A (北京航空航天大学) 2013年 9月 11日 (2013 - 09 - 11) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>US 2014126357 A1 (LOCKHEED MARTIN CORP) 2014年 5月 8日 (2014 - 05 - 08) 全文</td> <td>1-18</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	Y	CN 103051555 A (北京航空航天大学) 2013年 4月 17日 (2013 - 04 - 17) 说明书第[0019]-[0032]段	1-18	Y	CN 103905327 A (中国科学院声学研究所) 2014年 7月 2日 (2014 - 07 - 02) 说明书第[0018]-[0045]段	1-18	A	CN 103297346 A (北京航空航天大学) 2013年 9月 11日 (2013 - 09 - 11) 全文	1-18	A	US 2014126357 A1 (LOCKHEED MARTIN CORP) 2014年 5月 8日 (2014 - 05 - 08) 全文	1-18
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
Y	CN 103051555 A (北京航空航天大学) 2013年 4月 17日 (2013 - 04 - 17) 说明书第[0019]-[0032]段	1-18															
Y	CN 103905327 A (中国科学院声学研究所) 2014年 7月 2日 (2014 - 07 - 02) 说明书第[0018]-[0045]段	1-18															
A	CN 103297346 A (北京航空航天大学) 2013年 9月 11日 (2013 - 09 - 11) 全文	1-18															
A	US 2014126357 A1 (LOCKHEED MARTIN CORP) 2014年 5月 8日 (2014 - 05 - 08) 全文	1-18															
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																	
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” 同族专利的文件</p>																	
<p>国际检索实际完成的日期</p> <p>2019年 3月 8日</p>		<p>国际检索报告邮寄日期</p> <p>2019年 4月 17日</p>															
<p>ISA/CN的名称和邮寄地址</p> <p>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</p> <p>传真号 (86-10)62019451</p>		<p>受权官员</p> <p>马小瑜</p> <p>电话号码 86-(512)-88996083</p>															

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2019/074385

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	103051555	A	2013年 4月 17日	无			
CN	103905327	A	2014年 7月 2日	CN	103905327	B	2017年 5月 17日
CN	103297346	A	2013年 9月 11日	CN	103297346	B	2017年 5月 10日
US	2014126357	A1	2014年 5月 8日	US	8923123	B2	2014年 12月 30日

表 PCT/ISA/210 (同族专利附件) (2015年1月)