

说明书
发明名称：报文处理方法、装置、芯片、设备、存储介质及程序产品

相关申请的交叉引用

本申请要求于 2024 年 1 月 29 日提交中国专利局、申请号为 2024101230023、发明名称为“报文处理方法、装置、芯片、设备、存储介质及程序产品”的中国专利申请的优先权，所述专利申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及芯片技术领域，特别是涉及一种报文处理方法、装置、芯片、设备、存储介质及程序产品。

背景技术

可编程拥塞控制（PCC, programmable congestion control）框架允许用户结合实际应用场景，使能 PCC 框架所包含的各种事件信号，自行设计和实现拥塞控制算法，从而给予用户巨大的灵活性以实现最优拥塞控制解决方案，来处理服务器集群的拥塞问题。

相关技术中，支持 PCC 框架的设备基于事件响应模型工作，然而，当前 PCC 事件上报全部集中在源端，接收端 PCC 框架不运行，报文处理事件对应的处理算法的可扩展性较低，用户难以灵活地开发接收端驱动的拥塞控制算法。

发明内容

根据本申请的各种实施例，提供一种报文处理方法、装置、芯片、设备、存储介质及程序产品。

第一方面，本申请提供了一种报文处理方法。应用于报文处理设备，所述报文处理设备包括接收模块、发送模块、过滤模块和算法执行模块，所述方法包括：

所述接收模块接收待处理报文，并根据所述待处理报文的类型，触发所述待处理报文对应的报文处理事件；

所述过滤模块根据预置的事件使能信息，过滤所述报文处理事件中的未使能事件，并向所述算法执行模块发送事件请求，所述事件请求中包括有过滤后的报文处理事件的事件信息；

所述算法执行模块根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息，并通过所述过滤模块将所述事件响应信息发送给所述发送模块；

所述发送模块根据所述事件响应信息调节报文发送行为。

在其中一个实施例中，所述向所述算法执行模块发送事件请求之前，所述方法还包括：

所述过滤模块通过对所述过滤后的报文处理事件的初始信息进行预处理，以生成所述过滤后的报文处理事件的事件信息；

其中，所述预处理包括以下至少一项：计算处理、裁剪处理、合并处理、丢弃处理。

在其中一个实施例中，所述待处理报文中包括网络转发设备添加的拥塞标记信息和/或网络遥测信息，所述网络转发设备用于在两个报文处理设备之间转发报文。

在其中一个实施例中，所述报文处理设备中还包括管控模块，在所述算法执行模块根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息之前，所述方法还包括：

所述管控模块创建所述过滤后的报文处理事件的处理资源并通过热升级启动所述算法执行模块中的算法进程；

所述算法执行模块根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息，包括：

所述算法执行模块根据算法参数和所述事件信息，在所述算法进程中运行所述过滤后的报文处理事件的处理算法，得到所述事件响应信息。

在其中一个实施例中，在所述管控模块创建所述过滤后的报文处理事件的处理资源并通过热升级启动所述算法执行模块中的算法进程之后，所述方法还包括：

所述管控模块通过拥塞控制队列，获取所述算法执行模块在运行所述处理算法时的拥塞控制条目，所述拥塞控制条目中包括所述处理算法的算法状态信息；

所述管控模块根据所述算法状态信息，调整所述算法参数。

在其中一个实施例中，所述管控模块通过拥塞控制队列，获取所述算法执行模块在运行所述处理算法时的拥塞控制条目，包括：

所述算法执行模块创建所述拥塞控制条目；

所述算法执行模块根据所述拥塞控制队列的头部指针信息和尾部指针信息，确定所述拥塞控制队列是否处于满队列状态；

若所述拥塞控制队列未处于满队列状态，则所述算法执行模块在所述拥塞控制队列队尾添加所述拥塞控制条目并

翻转所述拥塞控制队列的标记信息；

所述管控模块根据所述标记信息获取所述拥塞控制队列中的拥塞控制条目，并跟新所述拥塞控制队列的头部指针信息。

在其中一个实施例中，在所述管控模块通过拥塞控制队列，获取所述算法进程在运行所述处理算法时的算法状态信息之后，所述方法还包括：

若所述算法状态信息异常，则所述管控模块销毁所述过滤后的报文处理事件的处理资源并退出所述算法进程。

在其中一个实施例中，所述算法参数包括粒度配置参数、事件使能参数、反馈使能参数、算法执行参数和硬件配置参数。

在其中一个实施例中，所述事件响应信息中包括每条报文传输的网络路径的优先级和报文发送模式；所述发送模块根据所述事件响应信息调节报文发送行为，包括：

所述发送模块根据所述每条报文传输的网络路径的优先级和所述报文发送模式，调节报文发送行为；

其中，所述报文发送模式包括基于速率的发送模式、基于窗口的发送模式、基于令牌的发送模式或基于信用的发送模式。

在其中一个实施例中，所述事件响应信息用于调节该事件响应信息对应的网络路径的报文发送行为，或者，所述事件响应信息用于调节所述报文处理设备对应的全部网络路径的报文发送行为。

在其中一个实施例中，所述事件请求通过事件请求队列在所述算法执行模块和所述过滤模块之间交互，所述事件响应信息通过事件响应队列在所述算法执行模块和所述过滤模块之间交互。

在其中一个实施例中，所述事件请求队列包括活动子队列和提交子队列；

所述向所述算法执行模块发送事件请求，包括：

所述过滤模块将所述事件请求发送给所述活动子队列中；

所述过滤模块通过所述活动子队列对所述事件请求进行聚合；

若所述提交子队列中的事件数量大于等于第一数量阈值时，所述过滤模块控制所述活动子队列中的事件请求阻塞等待，直至满足阻塞结束条件；所述阻塞结束条件包括：所述提交子队列中的事件数量小于所述第一数量阈值，或者，所述提交子队列中的事件数量未满足且所述活动子队列中的事件请求大于等于第二数量阈值；

若所述提交子队列中的事件数量小于第一数量阈值时，所述过滤模块将所述活动子队列中的事件请求发送至所述提交子队列，以使所述算法执行模块读取所述提交子队列中的事件请求。

在其中一个实施例中，在所述过滤模块通过所述活动子队列对所述事件请求进行聚合之后，所述方法还包括：

所述过滤模块根据所述事件聚合的类型，生成聚合后的事件请求的事件信息；其中，所述事件聚合的类型包括累加型和覆盖型。

在其中一个实施例中，所述方法还包括：

若所述事件请求队列中的事件数量已满，所述过滤模块根据事件丢弃配置信息，丢弃所述事件请求队列中的事件请求，或者，所述过滤模块向所述接收模块发送指示信息，所述指示信息用于指示所述接收模块停止向所述过滤模块发送所述报文处理事件。

在其中一个实施例中，所述通过热升级启动所述算法执行模块中的算法进程，包括：

管控模块将所述处理算法对应的编译文件存放在目标存储位置，以使所述报文处理设备的处理器通过在所述目标存储位置中获取并加载所述编译文件，启动所述算法进程。

第二方面，本申请提供了一种报文处理装置。所述报文处理装置包括接收模块、发送模块、过滤模块和算法执行模块；

所述接收模块，用于接收待处理报文，并根据所述待处理报文的类型，触发所述待处理报文对应的报文处理事件；

所述过滤模块，用于根据预置的事件使能信息，过滤所述报文处理事件中的未使能事件，并向所述算法执行模块发送事件请求，所述事件请求中包括有过滤后的报文处理事件的事件信息；

所述算法执行模块，用于根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息，并通过所述过滤模块将所述事件响应信息发送给所述发送模块；

所述发送模块，用于根据所述事件响应信息调节报文发送行为。

第三方面，本申请还提供了一种芯片，所述芯片包括如第二方面所述的报文处理装置。

第四方面，本申请还提供了一种计算机设备。所述计算机设备包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现上述第一方面的报文处理方法。

第五方面，本申请还提供了一种计算机可读存储介质。所述计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述第一方面的报文处理方法。

第六方面，本申请还提供了一种计算机程序产品。所述计算机程序产品，包括计算机程序，该计算机程序被处理器执行时实现上述第一方面的报文处理方法。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，根据这些附图获得其他的附图仍属于本申请的范畴。

- 图 1 为本申请实施例提供的一种报文处理方法的应用环境示意图；
- 图 2 为本申请实施例提供的一种报文处理设备的结构示意图；
- 图 3 为本申请实施例提供的一种 PCC 框架端点运行流程示意图；
- 图 4 为本申请实施例提供的一种算法热升级的流程示意图；
- 图 5 为本申请实施例提的一种管控模块和算法执行模块的交互示意图；
- 图 6 为本申请实施例提的另一种管控模块和算法执行模块的交互示意图；
- 图 7 为本申请实施例提供的一种算法执行模块的编译示意图；
- 图 8 为本申请实施例提供的一种管控模块的编译示意图；
- 图 9 为本申请实施例提供的一种报文处理方法的流程示意图；
- 图 10 为本申请实施例提供的一种报文处理装置的结构框图；
- 图 11 本申请实施例提供的一种计算机设备的内部结构图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

下面首先对于相关技术进行说明。

相关技术中，支持 PCC 框架的设备基于事件-响应模型工作。支持 PCC 框架的设备支持用户使能的事件信号有 6 种，包括固件信号（DOCA_PCC_DEV_EVT_FW）、拥塞通知报文（Congestion Notification Packet，CNP）、报文发送信号（DOCA_PCC_DEV_EVT_ROCE_CNP）、报文应答信号（DOCA_PCC_DEV_EVT_ROCE_ACK）、非预期报文应答信号（DOCA_PCC_DEV_EVT_ROCE_NACK）、网络往返时间（Round-Trip Time，RTT）信号。示例性的，如表 1 所示，不同的事件信号对应的事件含义、数据内容及数据粒度并不相同。

表 1

事件类型	事件含义	数据字段	数据粒度
固件信号	无描述	无表意字段	无
拥塞通知报文	交换机在触发ECN水线后，对报文标记ECN CE，数据接收端经过基于时间阈值过滤后生成响应报文发回给数据发送端。	num_coalesced（CNP聚合数量）	无
报文发送信号	数据发送端已发送字节数或报文数的统计。	sent_32bytes（发送字节数）；sent_pkts（发送报文数）	32字节；个
报文应答信号	数据发送端判断接收端已应答报文数的统计。	num_coalesced（ACK聚合报文数）	个
非预期报文应答信号	数据发送端判断未按序发送报文数的统计。	num_coalesced（NACK聚合报文数）	个
网络往返时间信号	RTT探测报文在发送端与接收端所标记的各个时间戳。	req_send_timestamp（请求发送时间戳）；req_recv_timestamp（请求接收时间戳）；resp_send_timestamp（响应发送时间戳）	纳秒（ns）

此外，除表 1 中所示的数据字段外，各事件上报时均包含公共数据字段，包括端口号（port_num）、流标识符（flow_tag）、报文序号（sn）、事件上报时间戳（timestamp）。相关技术中，事件信号的响应信息包括速率（rate，单位 10KB/s）和 RTT 事件探测指示（rtt_req，1 代表发送 RTT 探测报文，0 代表不发送 RTT 探测报文）。

相关技术中，还支持报文处理事件对应的处理算法的热升级。若执行运行算法命令，则算法可能处于活跃状态、

待命状态、停止状态和出错状态。

其中，在活跃状态下，该算法为唯一活跃算法，接收上报的各个事件请求并执行相应事件的处理过程。在待命状态下，当前算法没有实际运转，已有处于活跃状态的算法在执行事件处理，仅当处于活跃状态的算法退出、停用或出错后，当前待命状态算法切换为活跃状态。在停止状态下，算法进程被网卡固件停用，则销毁相关资源。在出错状态下，当前算法进程出错，则销毁相关资源。

然而，当前 PCC 事件上报全部集中在源端（发送端），目的端（接收端）PCC 框架不运行，报文处理事件对应的处理算法的可扩展性较低，用户难以灵活地开发接收端驱动（报文接收端驱动是指由报文接收端告知报文发送端下一轮的拥塞调节方式）的拥塞控制算法。

为解决上述问题，本申请实施例提供一种报文处理方法、装置、设备、存储介质及程序产品，该报文处理方法使得在报文接收端可以进行事件的处理，从而使得报文接收端支持可编程拥塞控制框架，具备事件上报与响应能力，从而有效提高了报文处理事件对应的处理算法的可扩展性。

本申请实施例提供的报文处理方法，可以应用于如图 1 所示的应用环境中。如图 1 所示，两个报文处理设备之间传输待处理报文。其中，一个报文处理设备为报文发送端 101（或者，称为源端），一个报文处理设备为报文接收端 102（或者，称为目的端）。报文发送端 101 发送待处理报文，经过网络转发设备 103 的转发，发送给报文接收端 102。随后，报文接收端 102 对待处理报文进行处理，并生成事件响应信息，并仅通过网络转发设备 103，反馈给报文发送端 101。

其中，报文发送端 101 发送的待处理报文可以包括数据报文和探测请求报文。其中，探测请求报文包括但不限于 RTT 探测请求、网络遥测信息（Telemetry，如 INT 或 IFA）探测请求、带宽探测请求、窗口探测请求、信用探测请求、令牌探测请求等。

报文接收端 102 发送的事件响应信息，可以为确认（Acknowledgement，ACK）报文、否认（Negative Acknowledgement，NAK）报文、CNP 报文、（Explicit Congestion Notification Echo，ECN-Echo）报文和探测响应报文。探测响应报文包括但不限于 RTT 探测响应、网络遥测信息探测响应、带宽探测响应、窗口探测响应、信用探测响应、令牌探测响应。

在一些实施例中，探测请求报文与探测响应报文可实现为带内（in-band）或带外（out-of-band）的形式。带内形式可以为探测请求报文头携带在数据报文中，带外形式可以为探测请求报文为独立的报文。

应理解，在本申请中，任意报文处理设备既可以为报文发送端，也可以为报文接收端。示例性的，同一个报文处理设备，在发送待处理报文时可以作为报文发送端，在接收待处理报文时可以作为报文接收端。

图 2 为本申请实施例提供的一种报文处理设备的结构示意图，如图 2 所示，该报文处理设备，包括接收模块、发送模块、过滤模块和算法执行模块。过滤模块分别与接收模块、发送模块和算法执行模块连接。

其中，接收模块在收到待处理报文后可以根据报文类型进行相应处理，触发报文处理事件。接收模块可触发的报文处理事件可以为报文接收（Receive，RX）事件。相似的，发送模块也可以触发报文处理事件，发送模块可触发的报文处理事件可以为报文发送（Transmit，TX）事件。

示例性的，表 1 为本申请实施例提供的一种报文处理事件的分类表。如表 2 所示，报文处理设备作为报文发送端和报文接收端时，可以触发不同的报文处理事件。

表 2

报文处理事件类型	报文处理事件
发送端 + TX	新流事件
	报文已发送事件
	RTT探测报文已发送事件
	网络遥测探测报文已发送事件
	带宽探测报文已发送事件
	窗口探测报文已发送事件
	源端令牌请求事件
	令牌探测报文已发送事件
	信用探测报文已发送事件
	超时事件
发送端 + RX	报文接收确认事件
	非预期报文确认事件
	拥塞通知报文事件

	ECN回应报文事件
	RTT探测响应报文接收事件
	网络遥测信息探测响应报文接收事件
	带宽探测响应报文接收事件
	窗口探测响应报文接收事件
	令牌探测响应报文接收事件
	信用探测响应报文接收事件
接收端 + TX	带宽探测响应报文发送事件
	窗口探测响应报文发送事件
	令牌探测响应报文发送事件
	信用探测响应报文发送事件
接收端 + RX	数据报文接收事件
	带宽探测请求报文接收事件
	窗口探测请求报文接收事件
	令牌探测请求报文接收事件
	信用探测请求报文接收事件

如表 2 所示，报文处理设备作为报文发送端时，发送模块可以触发新流事件、报文已发送事件、RTT 探测报文已发送事件、网络遥测探测报文已发送事件、带宽探测报文已发送事件、窗口探测报文已发送事件、源端令牌请求事件、令牌探测报文已发送事件、信用探测报文已发送事件、超时事件。报文处理设备作为报文发送端时，接收模块可以触发报文接收确认事件、非预期报文确认事件、拥塞通知报文事件、ECN 回应报文事件、RTT 探测响应报文接收事件、网络遥测信息探测响应报文接收事件、带宽探测响应报文接收事件、窗口探测响应报文接收事件、令牌探测响应报文接收事件、信用探测响应报文接收事件。

继续参考表 2，报文处理设备作为报文接收端时，发送模块可以触发带宽探测响应报文发送事件、窗口探测响应报文发送事件、令牌探测响应报文发送事件和信用探测响应报文发送事件。报文处理设备作为报文接收端时，接收模块可以触发数据报文接收事件、带宽探测请求报文接收事件、窗口探测请求报文接收事件、令牌探测请求报文接收事件和信用探测请求报文接收事件。

下面对于本申请实施例涉及的各种待处理报文进行说明。

其中，新流事件，是初始化算法参数及发送模块的首次配置。该事件所包含的主要数据字段为事件时间戳（timestamp），粒度受算法参数约束。示例性的，事件时间戳可以为 1 纳秒（ns）。

在一些实施例中，新流事件当流或队列组（Queue Pair，QP）建链成功后立即上报。在另一些实施例中，当已创建流在间隔时间超过一定阈值没有报文发送后也可以上报新流事件。

其中，报文已发送事件，用于更新发送窗口和飞行报文数量。报文已发送事件所包含的主要数据字段为事件时间戳、发送报文数，粒度受算法参数约束。示例性的，事件时间戳粒度可以为 1ns，发送报文数粒度为 1B 或 1 个报文。

在一些实施例中，当发送模块发送一定量的报文且受算法响应值或配置约束不能再发送新的报文后，上报报文已发送事件。

其中，RTT 探测报文已发送事件，用于通知算法执行模块 RTT 发送探测已执行，避免不必要的重复探测。RTT 探测报文已发送事件包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度可以为 1ns。

在一些实施例中，当发送模块完成 RTT 探测报文的发送后，上报 RTT 探测报文已发送事件。

其中，网络遥测探测报文已发送事件，用于通知算法执行模块网络遥测发送探测已执行，避免不必要的重复探测。网络遥测探测报文已发送事件包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns。

在一些实施例中，当发送模块完成网络遥测探测报文的发送后，上报网络遥测探测报文已发送事件。其中，网络遥测可以采用带内网络遥测（In-Band Network Telemetry，INT）或带内流量分析仪（In-Band Flow Analyzer，IFA）技术。

其中，带宽探测报文已发送事件，用于通知算法模块带宽发送探测已执行，避免不必要的重复探测。该事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns。

在一些实施例中，当发送模块完成带宽探测报文的发送后，上报带宽探测报文已发送事件。

其中，窗口探测报文已发送事件，用于通知算法模块窗口发送探测已执行，可避免不必要的重复探测。窗口探测报文已发送事件，所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns。

在一些实施例中，当发送模块完成窗口探测报文的发送后，上报窗口探测报文已发送事件。

其中，源端令牌请求事件，用于请求算法执行模块发放新的令牌，以支持发送模块继续执行发送任务。源端令牌请求事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns。

在一些实施例中，当发送模块基于令牌模式发送报文，调度当前流执行发送任务但无可用令牌时，上报源端令牌请求事件。

- 5 其中，令牌探测报文已发送事件，用于通知算法模块令牌发送探测已执行，可避免不必要的重复探测。该事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns。

在一些实施例中，当发送模块完成令牌探测报文的发送后，上报令牌探测报文已发送事件。

其中，信用探测报文已发送事件，用于通知算法执行模块令牌发送探测已执行，可避免不必要的重复探测。该事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns。

- 10 在一些实施例中，当发送模块完成信用探测报文的发送后，上报信用探测报文已发送事件。在另一些实施例中，算法层面可以考虑直接使用令牌探测报文已发送事件指代信用探测报文已发送事件。

其中，超时事件，用于通知算法执行模块潜在的报文丢失信息，使其通过生成相应的事件响应约束后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns。

- 15 在一些实施例中，当发送模块检测到发送的数据报文在一定时间内未收到确认，发送模块可以重新传输该报文，并上报超时事件。其中，上述一定时间可以为设置时间（retransmission timeout，RTO）

其中，报文接收确认事件，用于更新发送窗口和飞行报文数量。该事件所包含的主要数据字段为事件时间戳，确认接收报文数（pkt_ack_num），粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns，确认接收报文数粒度为 1B 或 1 个报文。

在一些实施例中，当源端的接收模块收到目的端报文发送模块发出的 ACK 报文后，上报报文接收确认事件。

- 20 其中，非预期报文确认事件，用于通知算法执行模块潜在的报文丢失信息，使其通过生成相应的事件响应约束后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，非预期报文确认数（pkt_nak_num），粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns，非预期报文确认数粒度为 1B 或 1 个报文。

- 25 在一些实施例中，目的端的接收模块收到一个数据包序列号（Packet Sequence Number，PSN）不按序到达的报文会触发送模块发送 NAK 报文。源端的接收模块收到 NAK 报文后上报非预期报文确认事件。其中，若支持选择性重传，NAK 信号可包含选择性确认（selective ACK，SACK）子信号。

其中，拥塞通知报文事件，用于通知算法执行模块潜在的拥塞信息，使其通过生成相应的事件响应约束后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，聚合 CNP 报文数（cnp_num），粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns，聚合 CNP 报文数粒度为 1 个。

- 30 在一些实施例中，目的端接收模块收到网络转发设备标记的显示拥塞通知（Explicit Congestion Notification，ECN）遇到拥塞（Congestion Experienced，CE）的报文后，可以根据 ECN 过滤配置参数进行过滤，如满足过滤条件，则触发送模块发送 CNP 报文。源端报文接收模块收到 CNP 报文后上报拥塞通知报文事件。其中，ECN CE，用 11 比特表示。

- 35 其中，ECN 回应报文事件，用于通知算法执行模块潜在的拥塞信息，使其通过生成相应的事件响应约束后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，聚合 ECN-Echo 报文数（ece_num），粒度受算法参数约束。示例性的，事件时间戳粒度为 1ns，ece_num 粒度为 1 个。

在一些实施例中，目的端的接收模块收到网络转发设备标记 ECN CE 的报文后，触发报文发送模块发送 ECN-Echo 报文。源端的接收模块收到 ECN-Echo 报文后上报 ECN 回应报文事件。

- 40 其中，RTT 探测响应报文接收事件，用于通知算法执行模块当前探测到的 RTT 大小，用于更新算法运行时状态信息，使其通过生成相应的事件响应调节后续报文发送行为。该事件所包含的主要数据字段为探测请求发送时间戳（ts_rtt_req_sent）、请求接收时间戳（ts_rtt_req_recv）、响应发送时间戳（ts_rtt_rsp_sent）、响应接收时间戳（ts_rtt_rsp_recv），粒度受算法参数约束。

在一些实施例中，过滤模块可以进一步将上述四个数据字段减少为两个，包括：rtt_fabric 与 rtt_host，计算方式为 $rtt_host = ts_rtt_rsp_sent - ts_rtt_req_recv$ ， $rtt_fabric = ts_rtt_rsp_recv - ts_rtt_req_sent - rtt_host$ 。示例性的，以上数据字段粒度均为 1ns。

- 45 在一些实施例中，源端的接收模块在收到目的端的发送模块发送的 RTT 探测响应报文后上报 RTT 探测响应报文接收事件。

其中，网络遥测信息探测响应报文接收事件，用于通知算法执行模块当前探测到的网络遥测信息，用于更新算法运行时状态信息，使其通过生成相应的事件响应调节后续报文发送行为。该事件所包含的主要数据字段为有效遥测数量（`valid_telemetry_num`）、遥测数据信息数组（`telemetry_data`）。每个遥测数据信息包括但不限于：TX 转发字节数（`tx_bytes`）、TX 时间戳（`tx_timestamp`）、缓冲区队列长度（`queue_length`）。上述数据字段粒度受网络转发设备配置约束。示例性的，`tx_bytes` 粒度为 1B，`tx_timestamp` 粒度为 1ns，`queue_length` 粒度为 256B。

在一些实施例中，源端的接收模块在收到目的端的发送模块发送的网络遥测信息探测响应报文后上报网络遥测信息探测响应报文接收事件。

其中，带宽探测响应报文接收事件，用于通知算法执行模块当前探测到的可用带宽大小，用于更新算法速率，使其通过生成相应的事件响应调节后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，探测带宽大小（`probe_bw`），粒度受算法参数约束。示例性的，`timestamp` 粒度为 1ns，`probe_bw` 粒度为 1MB/s。

在一些实施例中，源端的接收模块在收到目的端的发送模块发送的带宽探测响应报文后上报该事件。

其中，窗口探测响应报文接收事件，用于通知算法执行模块当前探测到的可用窗口大小，用于更新算法拥塞窗口，使其通过生成相应的事件响应调节后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，探测窗口大小（`probe_wnd`），粒度受算法参数约束。示例性的，`timestamp` 粒度为 1ns，`probe_wnd` 粒度为 1B 或 1 个报文。

在一些实施例中，源端的接收模块在收到目的端的发送模块发送的窗口探测响应报文后上报窗口探测响应报文接收事件。

其中，令牌探测响应报文接收事件，用于通知算法执行模块当前探测到的可用令牌大小，用于更新算法令牌数量，使其通过生成相应的事件响应调节后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，探测令牌大小（`probe_token`），粒度受算法参数约束。典型地，`timestamp` 粒度为 1ns，`probe_token` 粒度为 1B 或 1 个报文。

在一些实施例中，源端的接收模块在收到目的端的发送模块发送的令牌探测响应报文后上报令牌探测响应报文接收事件。

其中，信用探测响应报文接收事件，用于通知算法执行模块当前探测到的可用令牌大小，用于更新算法令牌数量，使其通过生成相应的事件响应调节后续报文发送行为。该事件所包含的主要数据字段为事件时间戳，探测信用大小（`probe_credit`），粒度受算法参数约束。典型地，`timestamp` 粒度为 1ns，`probe_credit` 粒度为 1B 或 1 个报文。

在一些实施例中，算法层面可以考虑直接使用令牌探测响应报文接收事件指代信用探测响应报文接收事件。

在一些实施例中，源端的接收模块在收到目的端的发送模块发送的信用探测响应报文后上报信用探测响应报文接收事件。

其中，带宽探测响应报文发送事件，用于通知算法执行模块带宽探测响应已执行，可避免一段时间内的不必要的重复响应。该事件所包含的主要数据字段为事件时间戳 `timestamp`，粒度受算法参数约束。示例性的，`timestamp` 粒度为 1ns。

在一些实施例中，目的端的发送模块发送带宽探测响应报文后上报带宽探测响应报文发送事件。

其中，窗口探测响应报文发送事件，用于通知算法执行模块窗口探测响应已执行，可避免一段时间内的不必要的重复响应。该事件所包含的主要数据字段为事件时间戳 `timestamp`，粒度受算法参数约束。示例性的，`timestamp` 粒度为 1ns。

在一些实施例中，目的端的发送模块发送窗口探测响应报文后上报该窗口探测响应报文发送事件。

其中，令牌探测响应报文发送事件，用于通知算法执行模块令牌探测响应已执行，可避免一段时间内的不必要的重复响应。该事件所包含的主要数据字段为事件时间戳 `timestamp`，粒度受算法参数约束。示例性的，`timestamp` 粒度为 1ns。

在一些实施例中，目的端的发送模块发送令牌探测响应报文后上报该令牌探测响应报文发送事件。

其中，信用探测响应报文发送事件，用于通知算法执行模块信用探测响应已执行，可避免一段时间内的不必要的重复响应。该事件所包含的主要数据字段为事件时间戳 `timestamp`，粒度受算法参数约束。示例性的，`timestamp` 粒度为 1ns。可选地，算法层面可以考虑直接使用令牌探测响应报文发送事件指代信用探测响应报文发送事件。

其中，数据报文接收事件，用于更新目的端的算法运行状态信息，以支持响应源端发送的各种探测事件。该事件所包含的主要数据字段为事件时间戳，接收报文数（`pkt_rcv_num`），粒度受算法参数约束。示例性的，`timestamp` 粒度为 1ns，`pkt_rcv_num` 粒度为 1B 或 1 个报文。

在一些实施例中，当目的端的接收模块收到源端的发送模块发出的数据报文后，上报数据报文接收事件。

其中，带宽探测请求报文接收事件，用于触发目的端的算法计算可用带宽大小，以生成带宽探测响应。该事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，timestamp 粒度为 1ns。

在一些实施例中，当目的端的接收模块收到源端的发送模块发出的带宽探测请求报文后，上报带宽探测请求报文接收事件。

- 5 其中，窗口探测请求报文接收事件，用于触发目的端算法计算可用窗口大小，以生成窗口探测响应。该事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，timestamp 粒度为 1ns。

在一些实施例中，当目的端的接收模块收到源端的发送模块发出的窗口探测请求报文后，上报窗口探测请求报文接收事件。

- 10 其中，令牌探测请求报文接收事件，用于触发目的端的算法计算可用令牌大小，以生成令牌探测响应。该事件所包含的主要数据字段为事件时间戳 timestamp，粒度受算法参数约束。示例性的，timestamp 粒度为 1ns。

在一些实施例中，当目的端的接收模块收到源端的发送模块发出的令牌探测请求报文后，上报该令牌探测请求报文接收事件。

- 15 其中，信用探测请求报文接收事件，用于触发目的端的算法启用信用分发机制，以针对当前或后续事件生成信用探测响应。该事件所包含的主要数据字段为事件时间戳，粒度受算法参数约束。示例性的，timestamp 粒度为 1ns。可选地，算法层面可以考虑直接使用令牌探测请求报文接收事件指代信用探测请求报文接收事件。

在一些实施例中，当目的端的接收模块收到源端的发送模块发出的信用探测请求报文后，上报信用探测请求报文接收事件。

继续参考图 2，过滤模块，用于收集发送模块与接收模块的初始信息，根据用户预置的事件使能信息，过滤未使能的事件，并对使能的事件的初始信息进行预处理，以生成过滤后的报文处理事件的事件信息。

- 20 其中，预处理包括以下至少一项：计算处理、裁剪处理、合并处理、丢弃处理。

算法执行模块，用于根据过滤模块上报的事件信息，执行过滤后的报文处理事件的处理算法，得到事件响应信息，并通过过滤模块将事件响应信息发送给发送模块。

在一些实施例中，过滤模块将算法执行模块返回的事件响应信息发送给发送模块前，也可以类似于上述预处理过程做适当转化处理。

- 25 在一些实施例中，发送模块，可以根据事件响应信息调节报文发送行为。应理解，本申请实施例对于发送模块如何调节报文发送行为不做限制，在一些实施例中，调节报文发送行为可以包括改变发送速率或发送窗口大小、改变发送路径优先级、发送探测请求或响应报文等。

继续参考图 2，两个报文处理设备传递待处理报文时，可以通过网络转发设备转发。在一些实施例中，网络转发设备在转发待处理报文时，可以在待处理报文中添加拥塞标记信息和/或网络遥测信息。

- 30 示例性的，拥塞标记信息可以为 ECN 标记，当 ECN 标记为 1 时表示拥塞，当 ECN 标记为 0 时表示未拥塞。网络遥测信息可以为网络遥测信息元数据（telemetry metadata）。示例的，网络转发设备可以为交换机和路由器等。

下面对于如何根据事件信息执行过滤后的报文处理事件的处理算法得到事件响应信息进行说明。

- 35 在一些实施例中，报文处理设备中还包括管控模块。管控模块可以先创建过滤后的报文处理事件的处理资源并通过热升级启动算法执行模块中的算法进程。随后，算法执行模块根据算法参数和事件信息，在算法进程中运行过滤后的报文处理事件的处理算法，得到事件响应信息。

其中，上述处理资源为算法进程在运行时使用的资源，可以包括处理算法对应的二进制算法文件的 Host 本地存储空间、拥塞控制队列（Congestion Control Queue，CCQ）等。

- 40 图 3 为本申请实施例提供的一种 PCC 框架端点运行流程示意图。如图 3 所示，上述管控模块负责控制面任务，算法处理模块负责数据面任务。管控模块运行主机（Host）侧，也可以运行在 DPU 侧。若管控模块运行在主机侧，则报文处理设备可以为带有 DPU 的服务器，若管控模块运行在 DPU 侧，则报文处理设备可以为 DPU 本身。

管控模块可以先创建资源，与算法执行模块交互，从而通过热升级启动算法进程。随后，算法执行模块根据算法参数和事件信息，在算法进程中运行过滤后的报文处理事件的处理算法，得到事件响应信息。

- 45 应理解，本申请实施例对于如何通过热升级启动算法进程不做限制，在一些实施例中，管控模块可以将处理算法对应的编译文件存放在目标存储位置，以使报文处理设备的处理器通过在目标存储位置中获取并加载编译文件，启动算法进程。

其中，上述报文处理设备的处理器，或报文处理设备本身，可以为 DPU，DPU 上可以包括片上系统（System on a

Chip, SoC)。上述编译文件可以为编译好的二进制的算法代码文件，上述目标存储位置可以为文件系统路径下的指定位置。

示例性的，图 4 为本申请实施例提供的一种算法热升级的流程示意图，如图 4 所示，PCC 热升级在管控模块执行启动算法进程的時刻触发。Host 侧的管控模块执行启动算法进程时，将编译好的二进制的算法代码文件，从函数参数所指定的文件系统路径加载到 Host 内存特定位置，并通知 SoC 该内存地址及大小。SoC 收到算法文件存储的内存地址及大小后，通过直接内存访问操作读取该二进制的算法代码文件内容，获取到 SoC 本地内存。

随后，SoC 将二进制的算法代码文件录到固件中，以持久化算法版本。SoC 判断当前是否已有算法进程处于运行状态，如有，则从固件中加载新算法，切换指令缓存中的算法指令数据为新的算法指令数据，然后执行后续流程，以启动算法进行。

在一些实施例中，上述切换方案使用双指令缓存（dual instruction cache）机制，新的算法指令数据不覆盖正处于运行的算法指令数据，而是放在备用指令缓存中，仅当新算法指令数据完全加载完毕，SoC 切换到备用指令缓存，使其成为默认指令缓存，原指令缓存降级为备用指令缓存。若不存在运行中的算法进程，则直接加载二进制的算法代码文件到默认指令缓存中执行算法进程。后续事件均由新的算法进程处理并生成响应返回给报文处理模块。最后，SoC 还需将算法二进制文件烧录到固件中，以持久化算法版本。

在一些实施例中，若管控模块启动算法时指定算法二进制文件为空值（NULL），可以不执行后续热升级动作，直接使用当前算法进程，或者，可以使用默认路径下的二进制的算法代码文件（如有）执行后续热升级动作。

在本申请中，基于热升级机制，支持无感知的算法切换，最小化对于业务流量的影响。并且，在管控模块执行加载算法时即执行算法切换，无需其他繁琐的配置步骤，最小化用户的学习与管理成本。

应理解，本申请实施例对于上述算法参数不做限制，在一些实施例中，启动算法进程时，可优先使用默认的算法参数配置，在算法进行运行以后也可以实时调节相关的算法参数的配置。

其中，修改算法参数的配置后，该配置对网卡硬件生效，并影响事件上报行为。

下面对于管控模块如何调整算法参数进行说明。

在一些实施例中，管控模块通过拥塞控制队列，获取算法执行模块在运行处理算法时的拥塞控制条目，拥塞控制条目中包括处理算法的算法状态信息。随后，管控模块根据算法状态信息，调整算法参数。

其中，上述拥塞控制条目（Congestion Control Entry, CCE）为用户在实现算法代码时自定义的数据结构。本申请实施例对于 CCE 数据的大小和数据结构不做限制。示例性的，上述 CCE 数据结构大小为 32B，CCE 数据结构可以如表 3 所示。

表 3

字段名称	字段含义
flow_id	流标识符，标识端点全局唯一的流 ID
max_rtt	算法进程探测到的最大 RTT（仅当使能 RTT 事件该值有效）
min_rtt	算法进程探测到的最小 RTT（仅当使能 RTT 事件该值有效）
owner	翻转标记位，算法每次写 CCE 时对该比特位执行翻转
data	其他用户自定义数据字段

示例性的，图 5 为本申请实施例提供的一种管控模块和算法执行模块的交互示意图。如图 5 所示，拥塞控制队列（Congestion Control Queue, CCQ）可以为一个基于 CCE 结构体数组实现的先入先出队列，逻辑上可抽象为一个环形缓冲区，在主机内存上。头部指针（Head）与尾部指针（Tail）指针均向前（图中右侧方向）移动，如当前 Head 或 Tail 指针已在到队尾，则下一次移动到队头。初始状态下 Head 与 Tail 指针均指向队头第一个 CCE 元素。record 是一个预分配的固定大小的缓冲区，用于存放 Head 指针信息，也在主机内存上。

在一些实施例中，上述算法执行模块创建拥塞控制条目。随后，算法执行模块根据拥塞控制队列的头部指针信息和尾部指针信息，确定拥塞控制队列是否处于满队列状态。若拥塞控制队列未处于满队列状态，则算法执行模块在拥塞控制队列队尾添加拥塞控制条目并翻转拥塞控制队列的标记信息。最后，管控模块根据标记信息获取拥塞控制队列中的拥塞控制条目，并跟新拥塞控制队列的头部指针信息。

示例性的，每当算法进程完成一次事件处理，就创建一个新的 CCE 结构体实例并填充字段，尝试向 CCQ 的队尾添加该 CCE，并翻转标记信息（例如，CCE owner 标记位）。仅当 CCQ 队列为未满状态下，算法执行模块可以成功向 CCQ 中添加 CCE，并更新 Tail 指针，向前移动一位。算法执行模块可以从 record 缓冲区中获取最新 Head 指针信息并通过与 Tail 指针进行比较，从而判定队列是否为已满状态。

- 示例性的，管控进程可以周期性地或轮询 CCQ 中 Head 指针指向的 CCE，若 CCE 中的 owner 标记位发生了翻转，则确定该 CCE 为有效数据，管控模块获取该数据进行分析，并将 Head 指针向前移动一位。最后，管控进程还需将 Head 指针信息根据如偏移大小更新到 record，以便算法进程可计算 CCQ 队列是否可放入新的 CCE 元素。
- 其中，上述拥塞控制队列的标记信息的翻转可以为由 0 变为 1，或者由 1 变为 0。
- 5 应理解，本申请实施例对于如何根据算法状态信息调整算法参数不做限制，在一些实施例中，可以采用机器学习或深度学习等人工智能模型自适应地根据 CCE 所包含的算法状态信息调节算法参数，从而达到更好的拥塞控制效果。
- 在一些实施例中，若算法状态信息异常，则管控模块销毁过滤后的报文处理事件的处理资源并退出算法进程。
- 在本申请中，管控模块可以通过拥塞控制队列，获取算法执行模块在运行处理算法时的拥塞控制条目，进而获取算法状态信息，并根据算法状态信息调节算法参数，配置到网卡硬件，进而通过算法参数约束硬件事件上报行为和更新算法行为。通过该方式，可以支撑用户设计更加智能的算法。
- 10 下面对于本申请实施例涉及的算法参数进行说明。
- 在一些实施例中，算法参数包括粒度配置参数、事件使能参数、反馈使能参数、算法执行参数和硬件配置参数。通过上述算法参数，可以影响硬件的事件上报行为、报文发送行为及算法执行行为。
- 其中，粒度配置参数，用于确定事件上报与响应各数据字段的基本单位。示例性的，表 4 为本申请实施例提供的一种粒度配置参数的示意图，该粒度配置参数的示意图中包括粒度配置参数的各个参数名称等信息。
- 15 种粒度配置参数的示意图，该粒度配置参数的示意图中包括粒度配置参数的各个参数名称等信息。

表 4

参数名称	参数含义	典型配置值
ts_unit	时间戳基本单位	1ns
pkt_mode	是否使用报文模式	1（使用报文模式）
pkt_unit	报文大小基本单位	1B（未启用报文模式）1个报文（启用报文模式）
bw_unit	速率基本单位	1MB/s
event_unit	CNP/ECN事件数量单位	1个事件

- 需要说明的是，若报文处理事件为网络遥测探测报文已发送事件，则字段的粒度依赖于网络转发设备配置，不受上述粒度配置参数约束。
- 其中，事件使能参数，用于配置事件信号的情况。仅当某一事件信号使能时，算法模块才能收到来自硬件上报的对应事件信息，发送模块或接收模块才能处理对应事件的探测请求或响应。
- 20 在一些实施例中，事件使能参数可以设置为一个 32 比特的数据字段（enable_event），每个比特对应一个事件的发生情况，该比特为 1 则使能，为 0 则不使能，从而使得多个事件可以同时使能。
- 在一些实施例中，可以设置一个 32 比特的事件丢弃使能数据字段（drop_event），每个比特对应一个事件的可丢弃配置，该比特为 1 则可丢弃，为 0 为则不可丢弃。仅当某一事件使能时，才需要关注该事件对应的丢弃使能类型。
- 25 其中，反馈使能参数即算法反馈配置参数，用于要求硬件以何种形式执行报文发送动作。反馈使能参数可包括多路径使能参数（multi_path_enable）和报文发送模式参数（packet_send_mode）。多路径使能参数，用于表示报文发送是否使用多条可用网络路径，若使能则同一条流的不同报文可以使用不同的网络路径，否则同一条流的所有报文仅使用一条固定（默认）网络路径。报文发送模式参数，用于表示硬件在单一网络路径上如何控制发送报文的快慢或数量。示例性的，报文发送模式包括基于速率的发送模式、基于窗口的发送模式、基于令牌的发送模式或基于信用的发送模式。
- 30 在一些实施例中，通过上述多路径使能参数，可以支持硬件的多路径报文发送。发送模块根据每条报文传输的网络路径的优先级和报文发送模式，调节报文发送行为。
- 示例性的，通过多路径使能参数，使能多路径支持时，算法执行模块针对任意一个事件的响应，可以约束多个路径的报文发送快慢或数量。算法模块反馈包括每条网络路径的优先级和报文发送模式。根据每条路径的对应的优先级和报文发送模式执行发包任务。
- 35 示例性的，发送模块先使用优先级最高的网络路径发送报文，直到该路径不满足报文发送条件，然后使用次优网络路径发送报文，以此类推。当所有网络路径均不能再发送报文后，需挂起当前流等待下一次调度。
- 在一些实施例中，基于速率的发送模式中，发送模块按照算法执行模块反馈的速率值约束报文发送。
- 示例性的，基于速率的发送模式中，发送模块可以将速率转化为报文发送间隔，发送完上一报文（或报文段）后满足计算出的时间间隔才可以发送下一报文（或报文段）。
- 40 示例性的，基于速率的发送模式中，发送模块还可以将速率转化为令牌，通过公式 $\text{token} = \text{rate} * \text{interval}$ ，然后使

- 用令牌模式约束报文发送，其中，rate 为发送速率，interval 为两次发送调度的间隔时间，token 为可用令牌数量。
- 示例性的，基于速率的发送模式中，报文发送端可以基于报文接收端的驱动确定发送速率大小（是指由报文接收端告知报文发送端下一次调节的发送速率大小）。（需使能以下事件：带宽探测报文已发送事件、带宽探测响应报文接收事件、带宽探测响应报文发送事件、带宽探测请求报文接收事件），即报文发送端根据算法执行模块指示定期执行带宽探测，等待报文接收端发回带宽探测响应报文，并根据该响应报文包含的值返回给发送模块，设定发送速率大小。
- 5 在一些实施例中，基于窗口的发送模式中，发送模块按照算法执行模块反馈的拥塞窗口值约束飞行报文数量。仅当飞行报文数小于当前拥塞窗口大小时，可以发送新的报文。否则，需将当前流挂起等待下一次调度，当 ACK 报文更新飞行报文大小或算法模块反馈更大的拥塞窗口值后可以继续发送报文。
- 示例性的，窗口模式可以采用报文接收端驱动的实现（是指由报文接收端告知报文发送端下一次调节的窗口大小）（需使能以下事件：窗口探测报文已发送事件、窗口探测响应报文接收事件、窗口探测响应报文发送事件、窗口探测请求报文接收事件），即发送方根据算法执行模块指示定期执行窗口探测，等待接收方发回窗口探测响应报文，并根据该响应报文包含的值返回给发送模块，设定拥塞窗口大小。
- 10 在一些实施例中，基于令牌的发送模式中，发送模块按照算法执行模块反馈的令牌值约束报文发送。发送模块仅在存在可用令牌且令牌大小大于待发送报文大小时可以执行报文发送。若令牌大小不足，则需上报源端令牌请求事件请求算法执行模块发放新的可用令牌，若使用新的发放令牌仍然不满足报文发送条件，需将当前流挂起等待下一次调度。算法执行模块仅在收到源端令牌请求事件后才响应令牌发放，其余事件请求不发放令牌。
- 15 示例性的，令牌模式可以采用接收方驱动的实现（是指由报文接收端告知报文发送端下一次调节的令牌大小）（需使能以下事件：令牌探测报文已发送事件、令牌探测响应报文接收事件、令牌探测响应报文发送事件、令牌探测请求报文接收事件），即发送方根据算法执行模块指示定期执行令牌探测，等待接收方发回令牌探测响应报文，并根据该响应报文包含的值返回给发送模块，添加可用令牌大小。
- 20 在一些实施例中，基于信用的发送模式中，发送模块按照算法执行模块反馈的信用值约束报文发送。与令牌模式相似，发送模块仅在存在可用信用值且信用大于待发送报文大小时可以执行报文发送。当信用不足时，需将当前流挂起等待下一次调度。
- 示例性的，基于信用的发送模式与基于令牌的发送模式的主要区别有两点：第一，基于信用的发送模式下源端发送报文需要等待目的端授权信用值，而令牌模式无此约束，既可以源端计算也可以目的端计算；第二，基于信用的发送模式下算法执行模块在收到任意事件后均可响应信用授权（即信用发放），发送模块需将有效的响应信用值进行累加，限制报文发送数量，而令牌模式只有在向算法模块发起令牌请求后才会获得新的可用令牌。
- 25 示例性的，信用模式可以采用接收方驱动的实现（是指由报文接收端告知报文发送端下一次调节的信用值大小）（需使能以下事件：信用探测报文已发送事件、信用探测响应报文接收事件、信用探测响应报文发送事件，信用探测请求报文接收事件），即发送方首次发送信用探测报文请求后等待接收方后续发回的多次信用探测响应报文，并根据该响应报文包含的值返回给发送模块，添加可用信用大小。
- 30 示例性的，算法层面可以使用令牌模式相关的事件信号指代信用模式相关的信号，并实现信用模式。示例性的，上述工作模式仅约束数据报文的发送行为，不约束探测请求/响应报文与 ACK、NAK、CNP、ECN-Echo 等报文。
- 其中，算法执行参数为算法运行所需的可配调节参数，又可称为模板参数。不同的算法对应的算法执行参数不同。示例性的，表 5 为本申请实施例提供的一种数据中心量化拥塞通知（Data Center Quantized Congestion Notification, DCQCN）算法的算法执行参数的示意图，如表 5 所示，用户可以按需将自定义算法的模板参数配置到硬件寄存器域段中，以支持实时动态地算法调整。
- 35

表 5

用途	参数	描述	推荐值
RP 升速	rp_timer_update_interval	将硬件带过来的时间戳，与 ctx 记录的时间戳相减，差值每超过该间隔 1 倍，T 值加 1，触发一次升速事件。	55us
	rp_byte_reset	硬件距离上一次重置的时间段内所发送的字节数，每超过该阈值 1 倍，BC 值加 1，触发一次升速事件。	1MB
	rp_f	BC 和 T 的检测门限，如果 $\text{Max}(T, BC) < F$ ，则处于快速恢复阶段；如果 $\text{Min}(T, BC) > F$ ，则进入 Hyper Increase 阶段；否则，进入 Additive	5

		Increase 阶段。	
	rp_ai_rate	Additive Increase 阶段每次 Rt 的增量。	25MB/s
	rp_hai_rate	Hyper Increase 阶段每次 Rt 的增量。	100MB/s
	rp_max_rate	Rc 和 Rt 的初始值，也是 Rc 能增长到的最大值。	25GB/s
RP 更新 α	rp_g	算法中 g 的值。	1/256
	rp_init_alpha	α 的初始值。	1
RP 降速	rp_rate_reduce_monitor_period	两次降速事件的最小时间间隔，在该时间内再次收到 CNP，会被忽略，不会触发降速事件，用于过滤频繁 CNP 消息。	50us
	rp_first_cnp_rate	第一次收到 CNP 的降速速率	12.5GB/s
	rp_max_dec_fac	一次降速所允许的最大降速幅度	1/2
	rp_min_rate	降速所允许的最小值，以防止饥饿	1MB/s
	rp_clamp_target_rate	是否每次降速都改变 target rate：1 代表是，0 代表否（仅第一次降速改变 target rate，后面的连续降速跳过 target rate 变更，直到升速后第一次降速）	1

其中，硬件配置参数为除了算法响应值以外约束硬件报文发送行为的其他参数，下表 6 列出了典型的硬件配置参数及功能，也可根据需要引入更多参数。

表 6

参数名称	参数含义	典型配置值
max_burst_size	最大单次报文突发发送大小，报文突发达到该数量后无视报文发送模式，需要等待下一次调度。	250KB
max_in_flight	硬件允许的最大飞行报文大小，飞行报文达到该数量后无视报文发送模式，需要停止继续发包。	500KB

5 本申请中，基于大量用户可配的算法配置参数，如粒度配置参数、事件使能参数、反馈使能参数、算法执行参数和硬件配置参数等，用户可以根据实际算法应用场景，灵活调节参数配置，以支撑高效开发场景适配的拥塞控制算法并提升算法的有效性。

下面对于报文处理设备中的事件交互模型进行说明。

10 在本申请中，报文处理设备中的事件交互模型可以为请求响应模型，在请求响应模型中，针对过滤模块上报的每一个请求（request），算法执行模块必定会生成一个对应的事件响应信息（response）。

示例性的，表 7 为本申请实施例提供的一种事件请求的结构示意，表 8 为本申请实施例提供的一种事件响应信息的结构示意。

表 7

序号	数据字段	含义
1	flow_id	流标识符，用于标识端点唯一的流
2	path_num	使用的路径数量，多路径模式可用，默认为 1
3	path_id	该事件请求的路径标识符，多路径模式可用，仅当path_num大于1时为有效值
4	event_type	事件请求类型，每比特对应一种事件信号：1为包含该事件，0为不包含
5	reflection_type	算法反馈模式（硬件报文发送模式）：基于速率、基于窗口、基于令牌、基于信用。
6	ts_unit	时间戳单位
7	pkt_mode	是否使用报文模式
8	pkt_unit	报文大小基本单位

9	bw_unit	速率基本单位
10	event_unit	CNP/ECN事件数量单位
11	timestamp	事件上报时间戳
12	pkt_sent_num	已发送报文数；仅当event_type包含报文已发送事件时有效
13	pkt_ack_num	确认接收报文数；仅当event_type包含报文接收确认事件时有效
14	pkt_nak_num	非预期报文确认数；仅当event_type包含非预期报文确认事件时有效
15	cnp_num	CNP报文数；仅当event_type包含拥塞通知报文事件时有效
16	ece_num	ECN-Echo报文数；仅当event_type包含ECN回应报文事件时有效
17	rtt_fabric	网络传播时延；仅当event_type包含RTT探测响应报文接收事件时有效
18	rtt_host	接收端处理时延；仅当event_type包含RTT探测响应报文接收事件时有效
19	probe_bw	探测带宽大小
20	probe_wnd	探测窗口大小
21	probe_token	探测令牌大小
22	probe_credit	探测信用大小
23	valid_telemetry_num	有效遥测数量；仅当event_type包含网络遥测信息探测响应报文接收事件时有效
24	telemetry_data	遥测数据信息数组，每个遥测数据信息包括但不限于：TX转发字节数tx_bytes、TX时间戳 tx_timestamp、缓冲区队列长度 queue_length；仅当event_type包含网络遥测信息探测响应报文接收事件时有效

表 8

序号	数据字段	含义
1	flow_id	流标识符，用于标识端点唯一的流
2	status	事件响应状态：正常或出错
3	path_num	使用的路径数量，多路径模式可用，默认为1
4	path_id	该事件响应的路径标识符，多路径模式可用，仅当path_num大于1时为有效值
5	priority	路径优先级，仅当path_num大于1时为有效值
6	value_update	指示value字段是否有效
7	value	更新后的速率/窗口/令牌/信用值，仅当value_update为1时为有效值
8	rtt_probe	是否发送RTT探测报文
9	telemetry_probe	是否发送网络遥测探测报文
10	bw_probe	是否发送带宽探测报文
11	wnd_probe	是否发送窗口探测报文
12	token_probe	是否发送令牌探测报文
13	credit_probe	是否发送信用探测报文

在一些实施例中，事件响应信息用于调节该事件响应信息对应的网络路径的报文发送行为，或者，事件响应信息用于调节报文处理设备对应的全部网络路径的报文发送行为。

5 示例性的，可以针对每条网络路径，在事件响应信息的数据结构中添加序号 5 至序号 13 的数据字段。即一条流任意网络路径的事件请求可以响应全部网络路径，约束不同网络路径的优先级及报文发送行为。

在一些实施例中，事件请求通过事件请求队列在算法执行模块和过滤模块之间交互，事件响应信息通过事件响应队列在算法执行模块和过滤模块之间交互。

其中，事件请求队列包括活动子队列和提交子队列。相应的，过滤模块将事件请求发送给活动子队列中。过滤模块通过活动子队列对事件请求进行聚合。若提交子队列中的事件数量大于等于第一数量阈值时，过滤模块控制活动子

队列中的事件请求阻塞等待，直至满足阻塞结束条件。若提交子队列中的事件数量小于第一数量阈值时，过滤模块将活动子队列中的事件请求发送至提交子队列，以使算法执行模块读取提交子队列中的事件请求。

示例性的，图 6 为本申请实施例提的另一种管控模块和算法执行模块的交互示意图。如图 6 所示，事件请求先进入活动子队列，再进入提交子队列。活动队列可以对事件请求进行聚合，而提交队列中的事件请求不能再进行更改。算法执行模块仅从提交队列中读取事件请求，进行处理并生成事件响应信息，放入事件响应队列。如图 6 所示，新事件为流 1 的事件信号，可以聚合到活动子队列中流 1 的已有事件中。当提交子队列深度小于第一数量阈值 M 时，活动子队列中的每一个新事件可以无阻塞地放入提交子队列的队尾。当提交子队列的排队数量大于等于 M 时，活动子队列中的新事件阻塞等待，允许合并该流的新事件，直到以下条件可以结束阻塞进入提交子队列。

其中，阻塞结束条件包括：提交子队列中的事件数量小于第一数量阈值，或者，提交子队列中的事件数量未满足且活动子队列中的事件请求大于等于第二数量阈值。

应理解，本申请实施例对于上述第一数量阈值和第二数量阈值不做限制，在一些实施例中，上述第一数量阈值为 M，上述第二数量阈值可以为 N，N 和 M 均为大于 0 的整数。示例性的，提交子队列与活动子队列的深度均可以设置为 32，M 与 N 可以分别设置为 16 和 24。

在一些可选的实施方式中，当 DPU 上的 SoC 为多核或多线程架构时，可以实现多个事件请求队列与事件响应队列，以提升事件处理性能。不同算法核可并行处理多个事件队列上的事件请求。

下面对于如何聚合事件请求进行说明。

在一些实施例中，在过滤模块通过活动子队列对事件请求进行聚合之后，过滤模块可以根据事件聚合的类型，生成聚合后的事件请求的事件信息。

应理解，事件聚合可以减少对于算法核的算力依赖，减少算力损耗，节约资源和成本。本申请中既支持同一条流的单一事件信号聚合，又支持同一条流的多个不同事件信号的聚合。

其中，事件聚合的类型包括累加型和覆盖型。累加型为在事件数据字段上进行累加，从而生成聚合后的事件请求的事件信息，覆盖型为直接覆盖事件数据字段的旧值，使用新值，从而生成聚合后的事件请求的事件信息。示例性的，表 9 为本申请实施例提供的一种单一事件聚合的示意表。

表 9

聚合类型	报文处理事件	聚合方法
累加型	报文已发送事件	将新事件中的pkt_sent_num累加到旧事件中的pkt_sent_num取值上，并替换事件时间戳为新事件时间戳
	报文接收确认事件	将新事件中的pkt_ack_num累加到旧事件中的pkt_ack_num取值上，并替换事件时间戳为新事件时间戳
	非预期报文确认事件	将新事件中的pkt_nak_num累加到旧事件中的pkt_nak_num取值上，并替换事件时间戳为新事件时间戳
	报文已接收事件	将新事件中的pkt_recv_num累加到旧事件中的pkt_recv_num取值上，并替换事件时间戳为新事件时间戳
	拥塞通知报文事件	将新事件中的cnp_num累加到旧事件中的cnp_num取值上，并替换事件时间戳为新事件时间戳
	ECN回报文事件	将新事件中的ece_num累加到旧事件中的ece_num取值上，并替换事件时间戳为新事件时间戳
	令牌探测响应报文接收事件	将新事件中的probe_token累加到旧事件中的

		probe_token取值上，并替换事件时间戳为新事件时间戳
	信用探测响应报文接收事件	将新事件中的probe_credit累加到旧事件中的probe_credit取值上，并替换事件时间戳为新事件时间戳
覆盖型	新流事件 超时事件 RTT探测请求报文发送事件 网络遥测信息发送事件 窗口探测请求报文发送事件 带宽探测请求报文发送事件 本地令牌事件 令牌探测请求报文接收事件 信用探测请求报文接收事件 RTT探测请求报文接收事件 网络遥测信息探测信息接收事件 带宽探测请求报文接收事件 窗口探测请求报文接收事件 令牌探测请求报文接收事件 信用探测请求报文接收事件 RTT探测报文已发送事件 网络遥测信息探测响应报文事件 带宽探测报文已发送事件 窗口探测报文已发送事件 令牌探测报文已发送事件 信用探测报文已发送事件 RTT探测响应报文接收事件 网络遥测信息探测响应报文接收事件	使用新的数据字段覆盖全部旧的数据字段

需要说明的是，在事件聚合的过程中多个事件请求的聚合使用表 9 所示的事件请求结构的不同数据字段，仅事件时间戳使用最新的事件时间戳进行覆盖。

下面对于事件丢弃机制进行说明。

5 在一些实施例中，若事件请求队列中的事件数量已满，过滤模块根据事件丢弃配置信息，丢弃事件请求队列中的事件请求，或者，过滤模块向接收模块发送指示信息，指示信息用于指示接收模块停止向过滤模块发送报文处理事件。

应理解，由于事件请求队列深度有限，当事件请求队列已满时，若过滤模块需要上报一个新事件时，则无法直接添加到事件请求队列中，只能选择丢弃该事件或反压上游停止发送新事件。通过上述丢弃配置信息可以进行事件请求队列中的事件丢弃，通过上述指示信息，可以实现反压上游停止发送新事件。

10 示例性的，选择反压上游，则可能会导致反压的逐级放大导致硬件的事件处理失能，造成不可预估的风险。因此，可以采用丢弃配置信息来丢弃事件，丢弃配置信息可以如表 10 所示。

表 10

事件丢弃选项	事件信号类型	原因分析
不可丢弃	新流事件 超时事件	稀有且非算法可触发
	报文已发送事件 报文接收确认事件 非预期报文确认事件 报文已接收事件	保证窗口计算正确性
	令牌本地事件 令牌探测请求报文接收事件 令牌探测响应报文接收事件	保证令牌（探测）可实行且返回（探测）结果可用
	信用探测请求报文接收事件 信用探测响应报文发送事件	保证信用探测可实行且返回探测结果可用

	拥塞通知报文事件 ECN回应报文事件	非算法可触发
可丢弃	RTT探测报文已发送事件 网络遥测探测报文已发送事件 带宽探测报文已发送事件 窗口探测报文已发送事件 令牌探测报文已发送事件 信用探测报文已发送事件 带宽探测响应报文发送事件 窗口探测响应报文发送事件 令牌探测响应报文发送事件 信用探测响应报文发送事件	不影响实际探测执行且算法可持续触发
	RTT探测响应报文接收事件 网络遥测信息探测响应报文接收事件 带宽探测响应报文接收事件 窗口探测响应报文接收事件 带宽探测请求报文接收事件 窗口探测请求报文接收事件	不影响报文正常发送且算法可持续触发

在本申请中，基于事件聚合与事件丢弃机制，可以显著减小算法核的吞吐需求，从而在支持大量事件信号的同时，能够有效控制算法核的成本，无需采购及部署大量高性能算法核。

在一些实施例中，算法执行模块和管控模块可以分别编译。示例性的，图 7 为本申请实施例提供的一种算法执行模块的编译示意图，图 8 为本申请实施例提供的一种管控模块的编译示意图。

5 如图 7 所示，算法模块在 SoC 上编译，用户编写的拥塞控制算法代码（含自定义.h 与.c 文件），编译时链接 PCC 框架提供的静态库生成二进制可执行文件。如图 8 所示，管控模块在主机侧编译，用户编写的管控代码（含自定义.h 与.c 文件），编译时链接 PCC 框架提供的静态库生成二进制可执行文件。

10 在一些实施例中，DPU 上包括 SoC，算法执行模块运行在 SoC 小核上，其中，SoC 小核可以基于高级精简指令集计算机（Advanced RISC Machine，ARM）架构或精简指令集计算机五级流水线（RISC-V Instruction Set Architecture，RISC-V）架构。

在本申请中，丰富的事件信号、多种事件响应反馈模式和接收端事件上报与响应能力可以支持用户设计更加灵活多样的算法，有效保障算法设计与实现的扩展性，以支撑不同场需求景、业务需求的客户选择或设计合适的算法。

在一个实施例中，如图 9 所示，提供了一种报文处理方法，以该方法应用于上述报文处理设备为例进行说明，该报文处理设备包括接收模块、发送模块、过滤模块和算法执行模块，该报文处理方法包括 S201-S204：

15 S201、接收模块接收待处理报文，并根据待处理报文的类型，触发待处理报文对应的报文处理事件。

在一些实施例中，向算法执行模块发送事件请求之前，方法还包括：过滤模块通过对过滤后的报文处理事件的初始信息进行预处理，以生成过滤后的报文处理事件的事件信息。

其中，预处理包括以下至少一项：计算处理、裁剪处理、合并处理、丢弃处理。

20 在一些实施例中，待处理报文中包括网络转发设备添加的拥塞标记信息和/或网络遥测信息，网络转发设备用于在两个报文处理设备之间转发报文。

S202、过滤模块根据预置的事件使能信息，过滤报文处理事件中的未使能事件，并向算法执行模块发送事件请求，事件请求中包括有过滤后的报文处理事件的事件信息。

S203、算法执行模块根据事件信息，执行过滤后的报文处理事件的处理算法，得到事件响应信息，并通过过滤模块将事件响应信息发送给发送模块。

25 在一些实施例中，报文处理设备中还包括管控模块，在算法执行模块根据事件信息，执行过滤后的报文处理事件的处理算法，得到事件响应信息之前，管控模块创建过滤后的报文处理事件的处理资源并通过热升级启动算法执行模块中的算法进程；算法执行模块根据算法参数和事件信息，在算法进程中运行过滤后的报文处理事件的处理算法，得到事件响应信息。

在一些实施例中，管控模块通过拥塞控制队列，获取算法执行模块在运行处理算法时的拥塞控制条目，拥塞控制

条目中包括处理算法的算法状态信息；管控模块根据算法状态信息，调整算法参数。

在一些实施例中，算法执行模块创建拥塞控制条目；算法执行模块根据拥塞控制队列的头部指针信息和尾部指针信息，确定拥塞控制队列是否处于满队列状态；若拥塞控制队列未处于满队列状态，则算法执行模块在拥塞控制队列队尾添加拥塞控制条目并翻转拥塞控制队列的标记信息；管控模块根据标记信息获取拥塞控制队列中的拥塞控制条目，并更新拥塞控制队列的头部指针信息。

在一些实施例中，若算法状态信息异常，则管控模块销毁过滤后的报文处理事件的处理资源并退出算法进程。

在一些实施例中，算法参数包括粒度配置参数、事件使能参数、反馈使能参数、算法执行参数和硬件配置参数。

S204、发送模块根据事件响应信息调节报文发送行为。

在一些实施例中，事件响应信息中包括每条报文传输的网络路径的优先级和报文发送模式；发送模块根据每条报文传输的网络路径的优先级和报文发送模式，调节报文发送行为。

其中，报文发送模式包括基于速率的发送模式、基于窗口的发送模式、基于令牌的发送模式或基于信用的发送模式。

在一些实施例中，事件响应信息用于调节该事件响应信息对应的网络路径的报文发送行为，或者，事件响应信息用于调节报文处理设备对应的全部网络路径的报文发送行为。

在一些实施例中，事件请求通过事件请求队列在算法执行模块和过滤模块之间交互，事件响应信息通过事件响应队列在算法执行模块和过滤模块之间交互。

在一些实施例中，事件请求队列包括活动子队列和提交子队列。

向算法执行模块发送事件请求，包括：过滤模块将事件请求发送给活动子队列中；过滤模块通过活动子队列对事件请求进行聚合；若提交子队列中的事件数量大于等于第一数量阈值时，过滤模块控制活动子队列中的事件请求阻塞等待，直至满足阻塞结束条件；阻塞结束条件包括：提交子队列中的事件数量小于第一数量阈值，或者，提交子队列中的事件数量未满足且活动子队列中的事件请求大于等于第二数量阈值；若提交子队列中的事件数量小于第一数量阈值时，过滤模块将活动子队列中的事件请求发送至提交子队列，以使算法执行模块读取提交子队列中的事件请求。

在一些实施例中，过滤模块根据事件聚合的类型，生成聚合后的事件请求的事件信息；其中，事件聚合的类型包括累加型和覆盖型。

在一些实施例中，若事件请求队列中的事件数量已满，过滤模块根据事件丢弃配置信息，丢弃事件请求队列中的事件请求，或者，过滤模块向接收模块发送指示信息，指示信息用于指示接收模块停止向过滤模块发送报文处理事件。

本申请实施例提供的报文处理方法，该报文处理方法应用于报文处理设备，报文处理设备包括接收模块、发送模块、过滤模块和算法执行模块。接收模块接收待处理报文，并根据待处理报文的类型，触发待处理报文对应的报文处理事件。过滤模块根据预置的事件使能信息，过滤报文处理事件中的未使能事件，并向算法执行模块发送事件请求，事件请求中包括有过滤后的报文处理事件的事件信息。算法执行模块根据事件信息，执行过滤后的报文处理事件的处理算法，得到事件响应信息，并通过过滤模块将事件响应信息发送给发送模块。发送模块根据事件响应信息调节报文发送行为。由于报文处理设备可以包括报文接收端，在报文接收端进行事件的处理，从而使得报文接收端支持可编程拥塞控制框架，具备事件上报与响应能力，从而有效提高了报文处理事件对应的处理算法的可扩展性。

应该理解的是，虽然如上的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，如上的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

基于同样的发明构思，本申请实施例还提供了一种用于实现上述所涉及的报文处理方法的报文处理装置。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似，故下面所提供的一个或多个报文处理装置实施例中的具体限定可以参见上文中对于报文处理方法的限定，在此不再赘述。

在一个实施例中，如图 10 所示，提供了一种报文处理装置 300，包括：接收模块 301、发送模块 302、过滤模块 303 和算法执行模块 304。

接收模块 301，用于接收待处理报文，并根据待处理报文的类型，触发待处理报文对应的报文处理事件。

过滤模块 302，用于根据预置的事件使能信息，过滤报文处理事件中的未使能事件，并向算法执行模块 303 发送事

件请求，事件请求中包括有过滤后的报文处理事件的事件信息。

算法执行模块 303，用于根据事件信息，执行过滤后的报文处理事件的处理算法，得到事件响应信息，并通过过滤模块 302 将事件响应信息发送给发送模块 304。

发送模块 304，用于根据事件响应信息调节报文发送行为。

- 5 上述报文处理装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

- 10 在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 11 所示。该计算机设备包括处理器、存储器、输入/输出接口(Input/Output，简称 I/O)和通信接口。其中，处理器、存储器和输入/输出接口通过系统总线连接，通信接口通过输入/输出接口连接到系统总线。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存为易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储数据。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种报文处理方法。计算机设备还可以包括 DPU 芯片。

本领域技术人员可以理解，图 11 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，提供了一种芯片，所述芯片包括上述的报文处理装置。所述芯片可以是 DPU 芯片。

- 20 在一个实施例中，提供了一种计算机设备，包括存储器和处理器，存储器中存储有计算机程序，该处理器执行计算机程序时实现上述报文处理方法。

在一个实施例中，提供了一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现上述报文处理方法。

- 25 在一个实施例中，提供了一种计算机程序产品，包括计算机程序，该计算机程序被处理器执行时实现上述报文处理方法。

- 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，该计算机程序可存储于一非易失性计算机可读存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory, ROM)、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器(ReRAM)、磁变存储器(Magnetoresistive Random Access Memory, MRAM)、铁电存储器(Ferroelectric Random Access Memory, FRAM)、相变存储器(Phase Change Memory, PCM)、石墨烯存储器等。易失性存储器可包括随机存取存储器(Random Access Memory, RAM)或外部高速缓冲存储器等。作为说明而非局限，RAM 可以是多种形式，比如静态随机存取存储器(Static Random Access Memory, SRAM)或动态随机存取存储器(Dynamic Random Access Memory, DRAM)等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等，不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等，不限于此。

- 35 以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

- 40 以上实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请的保护范围应以所附权利要求为准。

权利要求书

- 1.一种报文处理方法，其特征在于，应用于报文处理设备，所述报文处理设备包括接收模块、发送模块、过滤模块和算法执行模块，所述方法包括：
- 所述接收模块接收待处理报文，并根据所述待处理报文的类型，触发所述待处理报文对应的报文处理事件；
- 5 所述过滤模块根据预置的事件使能信息，过滤所述报文处理事件中的未使能事件，并向所述算法执行模块发送事件请求，所述事件请求中包括有过滤后的报文处理事件的事件信息；
- 所述算法执行模块根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息，并通过所述过滤模块将所述事件响应信息发送给所述发送模块；
- 所述发送模块根据所述事件响应信息调节报文发送行为。
- 10 2.根据权利要求1所述的方法，其特征在于，所述向所述算法执行模块发送事件请求之前，所述方法还包括：
- 所述过滤模块通过对所述过滤后的报文处理事件的初始信息进行预处理，以生成所述过滤后的报文处理事件的事件信息；
- 其中，所述预处理包括以下至少一项：计算处理、裁剪处理、合并处理、丢弃处理。
- 3.根据权利要求1或2所述的方法，其特征在于，所述待处理报文中包括网络转发设备添加的拥塞标记信息和/或网络遥测信息，所述网络转发设备用于在两个报文处理设备之间转发报文。
- 15 4.根据权利要求1-3任一项所述的方法，其特征在于，所述报文处理设备中还包括管控模块，在所述算法执行模块根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息之前，所述方法还包括：
- 所述管控模块创建所述过滤后的报文处理事件的处理资源并通过热升级启动所述算法执行模块中的算法进程；
- 所述算法执行模块根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息，包括：
- 20 所述算法执行模块根据算法参数和所述事件信息，在所述算法进程中运行所述过滤后的报文处理事件的处理算法，得到所述事件响应信息。
- 5.根据权利要求4所述的方法，其特征在于，在所述管控模块创建所述过滤后的报文处理事件的处理资源并通过热升级启动所述算法执行模块中的算法进程之后，所述方法还包括：
- 所述管控模块通过拥塞控制队列，获取所述算法执行模块在运行所述处理算法时的拥塞控制条目，所述拥塞控制
- 25 条目中包括所述处理算法的算法状态信息；
- 所述管控模块根据所述算法状态信息，调整所述算法参数。
- 6.根据权利要求5所述的方法，其特征在于，所述管控模块通过拥塞控制队列，获取所述算法执行模块在运行所述处理算法时的拥塞控制条目，包括：
- 所述算法执行模块创建所述拥塞控制条目；
- 30 所述算法执行模块根据所述拥塞控制队列的头部指针信息和尾部指针信息，确定所述拥塞控制队列是否处于满队列状态；
- 若所述拥塞控制队列未处于满队列状态，则所述算法执行模块在所述拥塞控制队列队尾添加所述拥塞控制条目并翻转所述拥塞控制队列的标记信息；
- 所述管控模块根据所述标记信息获取所述拥塞控制队列中的拥塞控制条目，并更新所述拥塞控制队列的头部指针
- 35 信息。
- 7.根据权利要求5或6所述的方法，其特征在于，在所述管控模块通过拥塞控制队列，获取所述算法进程在运行所述处理算法时的算法状态信息之后，所述方法还包括：
- 若所述算法状态信息异常，则所述管控模块销毁所述过滤后的报文处理事件的处理资源并退出所述算法进程。
- 8.根据权利要求4-7任一项所述的方法，其特征在于，所述算法参数包括粒度配置参数、事件使能参数、反馈使能
- 40 参数、算法执行参数和硬件配置参数。
- 9.根据权利要求1所述的方法，其特征在于，所述事件响应信息中包括每条报文传输的网络路径的优先级和报文发送模式；所述发送模块根据所述事件响应信息调节报文发送行为，包括：
- 所述发送模块根据所述每条报文传输的网络路径的优先级和所述报文发送模式，调节报文发送行为；
- 其中，所述报文发送模式包括基于速率的发送模式、基于窗口的发送模式、基于令牌的发送模式或基于信用的发
- 45 送模式。
- 10.根据权利要求1或9所述的方法，其特征在于，所述事件响应信息用于调节该事件响应信息对应的网络路径的

报文发送行为，或者，所述事件响应信息用于调节所述报文处理设备对应的全部网络路径的报文发送行为。

11.根据权利要求 1 所述的方法，其特征在于，所述事件请求通过事件请求队列在所述算法执行模块和所述过滤模块之间交互，所述事件响应信息通过事件响应队列在所述算法执行模块和所述过滤模块之间交互。

12.根据权利要求 11 所述的方法，其特征在于，所述事件请求队列包括活动子队列和提交子队列；

5 所述向所述算法执行模块发送事件请求，包括：

所述过滤模块将所述事件请求发送给所述活动子队列中；

所述过滤模块通过所述活动子队列对所述事件请求进行聚合；

若所述提交子队列中的事件数量大于等于第一数量阈值时，所述过滤模块控制所述活动子队列中的事件请求阻塞等待，直至满足阻塞结束条件；所述阻塞结束条件包括：所述提交子队列中的事件数量小于所述第一数量阈值，或者，所述提交子队列中的事件数量未满足且所述活动子队列中的事件请求大于等于第二数量阈值；

10 若所述提交子队列中的事件数量小于第一数量阈值时，所述过滤模块将所述活动子队列中的事件请求发送至所述提交子队列，以使所述算法执行模块读取所述提交子队列中的事件请求。

13.根据权利要求 12 所述的方法，其特征在于，在所述过滤模块通过所述活动子队列对所述事件请求进行聚合之后，所述方法还包括：

15 所述过滤模块根据所述事件聚合的类型，生成聚合后的事件请求的事件信息；其中，所述事件聚合的类型包括累加型和覆盖型。

14.根据权利要求 11 或 12 所述的方法，其特征在于，所述方法还包括：

若所述事件请求队列中的事件数量已满，所述过滤模块根据事件丢弃配置信息，丢弃所述事件请求队列中的事件请求，或者，所述过滤模块向所述接收模块发送指示信息，所述指示信息用于指示所述接收模块停止向所述过滤模块发送所述报文处理事件。

20 15.根据权利要求 11 或 12 所述的方法，其特征在于，所述通过热升级启动所述算法执行模块中的算法进程，包括：

管控模块将所述处理算法对应的编译文件存放在目标存储位置，以使所述报文处理设备的处理器通过在所述目标存储位置中获取并加载所述编译文件，启动所述算法进程。

25 16.一种报文处理装置，其特征在于，所述报文处理装置包括接收模块、发送模块、过滤模块和算法执行模块；

所述接收模块，用于接收待处理报文，并根据所述待处理报文的类型，触发所述待处理报文对应的报文处理事件；

所述过滤模块，用于根据预置的事件使能信息，过滤所述报文处理事件中的未使能事件，并向所述算法执行模块发送事件请求，所述事件请求中包括有过滤后的报文处理事件的事件信息；

30 所述算法执行模块，用于根据所述事件信息，执行所述过滤后的报文处理事件的处理算法，得到事件响应信息，并通过所述过滤模块将所述事件响应信息发送给所述发送模块；

所述发送模块，用于根据所述事件响应信息调节报文发送行为。

17.一种芯片，其特征在于，所述芯片包括如权利要求 16 所述的报文处理装置。

35 18.一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求 1 至 15 中任一项所述的方法的步骤。

19.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求 1 至 15 中任一项所述的方法的步骤。

20.一种计算机程序产品，包括计算机程序，其特征在于，该计算机程序被处理器执行时实现权利要求 1 至 15 中任一项所述的方法的步骤。