

一种报文重传方法和装置

5 技术领域

本申请涉及芯片技术领域，尤其涉及一种报文重传方法和装置。

背景技术

以太网技术目前在全球互联的因特网中占据主导地位，但在高带宽、低延时的专
有网络中却透露出许多弊端。随着网络融合概念的兴起，基于远端直接内存访问
10 (remote direct memory access, RDMA) 的无损链路得以解决，以太网在专有网络领
域内拥有了自己的标准，同时也提出了 RDMA 承载于融合以太网(RDMA over
converged ethernet, RoCE)的概念。由于 RDMA 需要一个无损(lossless)的网络，例如，
交换机不能因为缓冲区溢出而丢包。为此，RoCE 使用优先级的流量控制(priority-based
flow control, PFC)进行流量控制。例如，一旦交换机的端口(port)的接收队列超过
15 一定阈值(shreshold)时，就会向接收端发送 PFC 暂停帧(pause frame)，通知接收端
停止继续发包。

随着 RoCE 组网规模越来越大，开启 PFC 进行流量控制会给大规模的组网带来很
多弊端，例如容易产生 PFC 风暴，导致 PFC 死锁，从而发生丢包。因此，当 RoCE 应
用于 lossless 的网络时，丢包后及时重传就很重要。现有的丢包重传技术中，可以利用
20 返回 N(Go-Back-N) 协议进行重传，丢包后会产生包序列号(packet sequence number,
PSN) 错误(Error)，会触发快速在一个超时周期内重传所丢的报文以及之后的所有
的报文，但是，重传所丢的报文之后所有的报文会浪费大量的带宽。而且，如果快速
重传的报文再次重传失败而被丢弃，当超时周期超时，会继续重传所丢的报文以及
之后的所有的报文。也就是说，再次重传时需要等待超时周期超时后才能再次重传丢
25 失的报文以及之后的报文，这样会导致延时显著增大，使得重传不及时。

发明内容

本申请实施例提供一种报文重传方法和装置，解决了报文重传不及时导致的重传
时延大的问题。

为达到上述目的，本申请实施例采用如下技术方案：

30 第一方面，提供一种报文重传方法，该方法应用于 RDMA 网络中，该方法包括：
发送端接收来自接收端的第一选择性应答 SACK 报文，第一 SACK 报文指示发送端向
接收端发送的报文中的至少一个报文发送失败；发送端向接收端重传至少一个报文，
并在重传至少一个报文中最后一个报文时，记录发送端待发送的第一个新报文的包序
列号；当发送端接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大
35 于或等于第一个新报文的包序列号时，若发送端还未接收到至少一个报文中的第一报
文对应的 ACK 报文，则发送端再次向接收端重传第一报文。

由此，相对现有技术中重传报文时，从第一个重传报文开始重传开始发送报文，
本申请通过 SACK 报文指示发送端哪些报文发送失败，可使得发送端只针对发送失败

的报文进行重传，缩短重传时延。而且，本申请重传至 SACK 报文指示的最后一个报文时，通过记录下一个新报文的 PSN，通过对比接收到的 SACK 携带的 PSN 与记录的 PSN，可以判定是否存在重传失败的报文，即是否存在“被丢弃的重传报文”。如果确定存在重传失败的报文，可以立即再次进行重传。通常，从发送端第一次重传报文开始到下一次重传报文开始，其时长通常远小于一个超时周期，该时长在一个 RTT 周期内就能再次开始重传报文，即在一个 RTT 周期内就能重传“被丢弃的重传报文”，一个 RTT 周期远小于一个超时周期，可以“被丢弃的重传报文”被及时重传，缩短了重传时延。

10 在一种可能的设计中，第一 SACK 报文中携带位图，位图包括至少一个比特，至少一个比特中的部分比特指示发送端向接收端发送的报文中的至少一个报文发送失败。当接收端接收到第一 SACK 报文时，可以根据位图准确至到那些报文发送成功，哪些报文发送失败，在重传报文时，只针对发送失败的报文进行重传，可以缩短重传时延。

15 在一种可能的设计中，发送端向接收端重传至少一个报文之前，该方法还包括：发送端暂停向接收端发送新报文；发送端向接收端重传至少一个报文后，该方法还包括：发送端继续向接收端发送新报文。现有技术存在发送失败的报文时，从发送失败的报文开始重传发送报文，包括发送失败的报文之后发送成功的报文也被重传，而本申请只针对发送失败的报文进行重传，在确定重传时，暂停向接收端发送新报文。在重传完报文后，继续向接收端发送新报文，可缩短重传时延。

20 在一种可能的设计中，当发送端接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于第一个新报文的包序列号时，若发送端还未接收到至少一个报文中的第一报文对应的 ACK 报文，则发送端再次向接收端重传第一报文包括：当发送端接收到每个 SACK 报文时，将每个 SACK 报文中携带的包序列号与记录的第一个新报文的包序列号进行对比，确定每个 SACK 报文携带的包序列号是否大于或等于第一个新报文的包序列号；当接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于第一个新报文的包序列号时，发送端确定是否存在未接收到针对至少一个报文的 ACK 报文；若发送端还未接收到至少一个报文中的第一报文对应的 ACK 报文，则发送端再次向接收端重传第一报文。由于在重传最后一个报文过程中，若记录的待发送的新报文的包序列号时，待发送的新报文是在重传的报文之后发送的，如果发送端已经接收到新报文的应答报文，但还未接收到重传的报文的应答报文时，
25 说明重传的报文第一次重传失败，需要进行第二次重传，第二次重传不需要等到超时周期超时才开始，而是可以立即开始进行第二次重传，缩短了重传时延。

30 在一种可能的设计中，发送端从开始重传至少一个报文到再次重传第一报文的时间段小于一个超时周期。或者说，本申请进行第一次重传和第二次重传的时间段远小于一个超时周期。由于本申请在确定需进行第二次重传时，立即进行第二次重传，不需要等到超时周期超时时才进行第二次重传，从第一次重传开始到第二次重传报文的时间可以为一个 RTT 周期，一个 RTT 周期远小于一个超时周期，使得报文重传及时，缩短了重传时延。也就是说，一个 RTT 周期内就能进行第二次重传报文。

第二方面，提供一种报文重传方法，该方法应用于 RDMA 网络，该方法包括：发送端接收来自接收端的第一选择性应答 SACK 报文，第一 SACK 报文指示发送端向接

收端发送的报文中的至少一个报文发送失败；发送端向接收端重传至少一个报文，并在重传至少一个报文中的每个报文时，记录发送端在重传每个报文后待发送的第一个新报文的包序列号；其中，至少一个报文包括第二报文，发送端在重传第二报文时，记录的待发送的第一个新报文的包序列号为第一新报文的包序列号；当发送端接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于记录的第一新报文的包序列号时，若还未接收到第二报文对应的 ACK 报文，则发送端再次重传第二报文。

也就是说，发送端不仅可以选择不向接收端重传发送失败的报文，而且通过记录重传每个发送失败的报文时，下一个待发送的新报文的 PSN，当记录的 PSN 大于或等于新报文的 PSN 的报文被应答，而记录的同时重传的报文却未应答，可以判定重传的报文重传失败，需再次进行重传。需再次重传时，发送端可以立即向接收端再次重传，不需要等到超时周期超时时才进行再次重传，使得报文得到及时重传，缩短了延时。况且，第一次重传报文到再次重传报文的时间可以为一个 RTT 周期，一个 RTT 周期远小于一个超时周期，使得在一个 RTT 周期内可以精确的再次重传多个被重传失败的报文，缩短了重传时延。

在一种可能的设计中，第一 SACK 报文中携带位图，位图包括至少一个比特，至少一个比特中的部分比特指示发送端向接收端发送的报文中的至少一个报文发送失败。即第一 SACK 报文中可以携带 bitmap，bitmap 包括多个比特的比特值，多个比特的比特值指示当前 QC 中的多个报文是否被正确接收的指示。

在一种可能的设计中，在发送端向接收端重传至少一个报文之前，该方法还包括：发送端为多个报文队列中的每个报文队列建立链表，链表用于记录发送端在重传报文队列中的任一报文后待发送的第一个新报文的包序列号。也可以理解为，该链表用于为重传的报文与待发送的第一个新报文建立对应关系。这是考虑到，后续执行报文发送过程中，新报文是在重传的报文之后发送的，如果发送端已经接收到的记录的待发送的第一个新报文的应答报文，却还未接收到对应的重传报文的应答报文，说明重传的报文重传失败。这样，可以使得发送端及时了解被发送失败的报文，一旦了解，立即进行报文重传，使得报文得到及时重传。

在一种可能的设计中，在重传至少一个报文中的每个报文时，记录发送端重传每个报文后待发送的第一个新报文的包序列号包括：在重传第二报文时，发送端确定第二报文所在的报文队列对应的第一链表，并在第一链表中记录发送端在重传第二报文后待发送的第一个新报文的包序列号，且包序列号为第一新报文的包序列号。由于第一新报文是在重传的第二报文之后发送的，且发送端重传第二报文时，记录的待发送的第一个新报文为即第一新报文的包序列号，当发送端接收到第一新报文的应答报文，但还未接收到第二报文的应答报文时，说明重传的第二报文重传失败，可以立即向接收端再次进行第二报文的重新传，不需要等到超时周期超时才开始进行第二次重传，可以使得报文得到及时重传，缩短重传时延。

在一种可能的设计中，发送端从开始重传至少一个报文到再次重传第二报文的时间段小于一个超时周期。或者说，本申请进行第一次重传和第二次重传的时间段远小于一个超时周期。由于本申请在确定需进行第二次重传时，立即进行第二次重传，不

需要等到超时周期超时时才进行第二次重传，从第一次重传开始到第二次重传报文的时间可以为一个 RTT 周期，一个 RTT 周期远小于一个超时周期，使得报文重传及时，缩短了重传时延。也就是说，一个 RTT 周期内就能进行第二次重传报文。

第三方面，提供一种通信装置，通信装置作为发送端，通信装置应用于远端直接内存访问 RDMA 网络中，通信装置包括：接收器，用于接收来自接收端的第一选择性应答 SACK 报文，第一 SACK 报文指示发送端向接收端发送的报文中的至少一个报文发送失败；发射器，用于向接收端重传至少一个报文；以及存储器，用于发送端在重传至少一个报文中最后一个报文时，记录发送端待发送的第一个新报文的包序列号；处理器，还用于确定发送端接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于第一个新报文的包序列号，若发送端还未接收到至少一个报文中的第一报文对应的 ACK 报文，指示发射器还用于再次向接收端重传第一报文。

在一种可能的设计中，第一 SACK 报文中携带位图，位图包括至少一个比特，至少一个比特中的部分比特指示发送端向接收端发送的报文中的至少一个报文发送失败。

在一种可能的设计中，处理器用于：当发送端接收到每个 SACK 报文时，将每个 SACK 报文中携带的包序列号与记录的第一个新报文的包序列号进行对比，确定每个 SACK 报文携带的包序列号是否大于或等于第一个新报文的包序列号；当接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于第一个新报文的包序列号时，确定是否存在未接收到针对至少一个报文的 ACK 报文；发射器，用于若发送端还未接收到至少一个报文中的第一报文对应的 ACK 报文，则再次向接收端重传第一报文。

在一种可能的设计中，发送端从开始重传至少一个报文到再次重传第一报文的时间段小于一个超时周期。

第四方面，提供一种通信装置，通信装置作为发送端，通信装置应用于远端直接内存访问 RDMA 网络中，通信装置包括：接收器，用于接收来自接收端的第一选择性应答 SACK 报文，第一 SACK 报文指示发送端向接收端发送的报文中的至少一个报文发送失败；发射器，用于向接收端重传至少一个报文；存储器，用于在重传至少一个报文中的每个报文时，记录发送端在重传每个报文后待发送的第一个新报文的包序列号；其中，至少一个报文包括第二报文，发送端在重传第二报文时，记录的待发送的第一个新报文的包序列号为第一新报文的包序列号；接收器，还用于当接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于记录的第一新报文的包序列号时，若还未接收到第二报文对应的 ACK 报文，则指示发射器，用于再次重传第二报文。

在一种可能的设计中，第一 SACK 报文中携带位图，位图包括至少一个比特，至少一个比特中的部分比特指示发送端向接收端发送的报文中的至少一个报文发送失败。

在一种可能的设计中，还包括处理器，用于：为多个报文队列中的每个报文队列建立链表，链表存储在存储器中，链表用于记录发送端在重传报文队列中的任一报文后待发送的第一个新报文的包序列号。

在一种可能的设计中，存储器用于：在重传第二报文时，确定第二报文所在的报文队列对应的第一链表，并在第一链表中记录发送端在重传第二报文后待发送的第一个

新报文的包序列号，且包序列号为第一新报文的包序列号。

在一种可能的设计中，发送端从开始重传至少一个报文到再次重传第二报文的时间段小于一个超时周期。

- 5 第五方面，本申请实施例提供了一种计算机可读存储介质，包括计算机指令，当计算机指令在电子设备上运行时，使得电子设备执行上述第一方面以及第二方面以及第一方面和第二方面及的任一项可能的实现方式中的报文重传方法。

第六方面，本申请实施例提供了一种计算机程序产品，当计算机程序产品在计算机上运行时，使得电子设备执行上述第一方面以及第二方面以及第一方面和第二方面及的任一项可能的实现方式中的报文重传方法。

- 10 第七方面，本申请实施例提供了一种系统，该系统可以包括以上任一方面的任一项可能的实现方式中的发送端和接收端。该发送端和接收端可以执行上述任一方面及任一项可能的实现方式中的报文重传方法。

附图说明

- 图 1 为一种报文重传的信号交互示意图；
15 图 2 为一种报文重传的信号交互示意图；
图 3 为本申请实施例提供的一种网络架构的示意图；
图 4 为本申请实施例提供的一种发送端和接收端的结构示意图；
图 5 为本申请实施例提供的一种 QP 对应的 bitmap 的示意图；
图 6 为本申请实施例提供的一种报文重传的流程示意图；
20 图 7 为本申请实施例提供的一种报文重传的信号交互示意图；
图 8 为本申请实施例提供的一种报文重传的流程示意图；
图 9 为本申请实施例提供的一种报文重传的信号交互示意图；
图 10 为本申请实施例提供的一种通信装置的结构示意图；
图 11 为本申请实施例提供的一种服务器的结构示意图。

25 具体实施方式

为了便于理解，示例的给出了部分与本申请实施例相关概念的说明以供参考。如下所示：

- RDMA：是一种直接内存访问技术，它将数据直接从一台计算机的内存传输到另一台计算机，无需计算机双方操作系统的介入，是为了解决网络传输中服务器端数据处理的延迟而产生的。
30

RoCE：一种允许通过以太网进行 RDMA 的网络协议。其较低的网络头是以太网头，其上网络头(包括数据)是 InfiniBand 头（支持 RDMA 的新一代网络协议）。这允许在标准以太网基础架构(交换机)上使用 RDMA。由于 RDMA 是一种新的网络技术，因此需要支持该技术的网卡和交换机。

- 35 环路延时（round trip time, RTT）：在计算机网络中，RTT 是一个重要的性能指标，表示从发送端发送数据开始，到发送端收到来自接收端的确认（接收端收到数据后便立即发送确认），总共经历的时延。

队列对（queue pair, QP）：是硬件和软件之间的一个虚拟接口。QP 是队列结构，按顺序存储着软件给硬件下发的任务，任务中包含从哪里取出多长的数据，并且发送

给哪个目的地等信息。每个 QP 间都是独立的，因此一个 QP 可以被视为某个用户独占的一种资源，一个用户也可以同时使用多个 QP。

连接上下文（queue pair context, QPC）：用于存储 QP 相关属性。QP 在硬件上的实体只是一段存储空间而已，软件通过操作系统提前申请好一片连续的空间，QPC 可以用于来承载 QP 的相关信息，硬件可以通过 QPC 了解 QP 的相关属性。而网卡及其配套的驱动程序提前约定好了 QPC 中都有哪些内容，这些内容分别占据多少空间，按照什么顺序存放。这样驱动和硬件就可以通过 QPC 这段空间来读写 QP 的状态等信息。

位图（bitmap）：用至少一个比特位（bit）来标记至少一个元素对应的值（value）。
10 应答（acknowledge, ACK）报文：数据通信中，接收端发给发送端的一种传输类控制字符，表示发来的数据已确认接收无误。

选择确认字符（selective acknowledge, SACK）：用于接收端告诉发送端哪些报文段丢失，哪些报文段重传了，哪些报文段已经提前收到等信息。

包序列号（pack sequence number, PSN）：数据包或报文的序列号。
15 下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行描述。其中，在本申请实施例的描述中，除非另有说明，“/”表示或的意思，例如，A/B 可以表示 A 或 B；本文中的“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，在本申请实施例的描述中，“多个”是指两个或两个以上。

20 以下，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。在本实施例的描述中，除非另有说明，“多个”的含义是两个或两个以上。

本申请实施例应用于 RoCE 中进行报文重传的场景中。
25 目前，RoCE 中，一种报文重传方式中，当至少一个报文发送失败时，会从至少一个报文中第一个发送失败的报文开始重新发送报文，即发送端会发送第一个重传的报文（从一个超时周期的起始时刻开始）之后所有的报文。也就是说，即使只发送失败了一个报文，也会重传该报文之后所有的报文。这样会浪费大量的带宽。而且，如果重传报文时再次出现报文发送失败的情况，需要等待该超时周期超时后，再次从重传时第一个重传失败的报文开始进行报文重传，使得重传不够及时，延时显著增加，
30 带宽下降。

如图 1 所示，发送端向接收端发送了 PSN 为 0、1、2、3、4、5……的多个报文，但是接收端向发送端反馈了 PSN 为 0 的报文的 ACK 报文后，直接向发送端反馈了 PSN 为 1 的报文的否定应答（negative acknowledgement, NACK）报文，即 NACK 报文携带的 PSN 为 1，指示 PSN 为 1 的报文接收失败，因此，发送端可以确定 PSN 为 1 的
35 报文发送失败。由此，发送端在接收到 NACK 报文时，便开启一个超时周期，从该超时周期的起始时刻开始，从 PSN 为 1 的报文开始重传 PSN 为 1 的报文以及 PSN 为 1 的报文之后的报文（PSN 为 1 的报文~超时周期开启前接收到的最后一个报文）。如果发送端在超时周期内向接收端重传报文的过程中，又存在发送失败的报文，例如图 1

中 PSN 为 1 的报文在超时周期内重传失败，那么发送端在超时周期内发送的所有报文被丢弃，等到超时周期超时后，发送端再次开始从 PSN 为 1 的报文开始重传，可以称为重传超时（retransmit timeout, RTO）后的重传，这样会使得重传不够及时，延时显著增加，带宽下降。图 1 中，发送端在超时周期超时后再次从 PSN 为 1 的报文开始重传时，发送端都接收到了 PSN 为 1、2、3 和 4 等多个报文的 ACK，表示再次重传的报文都重传成功。

另一种报文重传方式中，发送端也可以向接收端进行选择性的重传。如图 2 所示，发送端向接收端发送了 PSN 为 0、1、2、3、4、5……的多个报文，与图 1 的方案类似的，发送端接收到了携带 PSN 为 3 的 SACK 报文，该 SACK 报文指示 PSN 为 3 的报文发送成功，但是 PSN 为 1 和 2 的报文发送失败，因此终端设备需要重传 PSN 为 1 和 2 的报文。但是，与图 1 的方案不同的是，发送端可以在超时周期开启内只重传发送失败的报文，即发送端在超时周期内重传了 PSN 为 1 和 2 的报文之后，继续向接收端发送新的报文，例如新的报文为图 2 中 PSN 为 N、N+1 以及 N+2 等多个报文。但是图 2 的方案中，还与图 1 的方案类似的是，如果发送端在超时周期内还存在重传失败的报文，例如在超时周期内重传 PSN 为 1 的报文失败，发送端接收到了携带 PSN 为 2 的 SACK 报文，SACK 报文指示 PSN 为 2 的报文重传成功，但是 PSN 为 1 的报文重传失败，这时，发送端依然要等到超时周期超时时，才能继续再次重传 PSN 为 1 的报文。可以看出，图 2 中，发送端在 RTO 后再次重传了 PSN 为 1 的报文后，才开始继续向接收端发送新的报文，例如新的报文为图 2 中 PSN 为 M、M+1 以及 M+2 的报文等，且再次重传都重传成功，即发送端接收到了来自接收端的 PSN 为 1、M、M+1 以及 M+2 的报文的 ACK 报文。因此，图 2 的这种重传方式中，也存在重传不够及时，延时显著增加，带宽下降的问题。

针对该问题，本申请实施例提出一种报文重传方法，该方法应用于 RDMA 中，当发送端在确定存在发送失败的报文时，可以选择性的重传报文，即只针对发送失败的报文进行重传，而且，在确定存在发送失败的报文时，可以立即重传发送失败的报文。如果重传的报文再次出现重传失败，也不需要等到超时周期超时时才进行再次重传。当重传的报文再次出现重传失败时，发送端可以再次立即重传报文，以缩短重传延时。

其中，在判断重传的报文是否重传成功时，可以通过以下方式确定：

1) 在重传多个发送失败的报文时，如果重传到最后一个发送失败的报文时，发送端可以记录待发送的第一个新报文的 PSN。当发送端接收到一个 SACK 报文，SACK 报文中携带的报文的包序列号大于或等于上述第一个新报文的 PSN 时，若发送端还未接收到多个发送失败的报文中的第一报文对应的第二 ACK 报文，则发送端确定第一报文重传失败，发送端可以再次向接收端重传所述第一报文。也就是说，发送端不仅可以选择只向接收端重传发送失败的报文，而且根据 SACK 报文中的 PSN 与记录的待发送的新报文的 PSN 进行对比发现重传失败时，可以立即向接收端再次重传，不需要等到超时周期超时时才进行再次重传，使得报文得到及时重传，缩短了延时。

2) 在重传多个发送失败的报文时，记录发送端在重传每个报文后待发送的第一个新报文的 PSN，可以理解为建立有每个发送失败的报文与待发送的第一新报文的对应关系。当发送端接收到 SACK 报文，SACK 报文中携带的报文的 PSN 大于或等于记录

的第一新报文的 PSN 时, 若还未接收到第二报文对应的第二 ACK 报文, 则发送端再次重传第二报文; 其中, 发送失败的报文包括第二报文, 发送端在重传第二报文时, 记录的第一个新报文的 PSN 为第一新报文的 PSN。也就是说, 发送端不仅可以选择只向接收端重传发送失败的报文, 而且通过记录重传每个发送失败的报文时, 下一个待发送的新报文的 PSN, 当记录的 PSN 大于或等于新报文的 PSN 的报文被应答, 而记录的同时重传的报文却未应答, 可以判定重传的报文重传失败, 需再次进行重传。需再次重传时, 发送端可以立即向接收端再次重传, 不需要等到超时周期超时时才进行再次重传, 使得报文得到及时重传, 缩短了延时。

本申请提出的重传方法可以应用的网络架构可以如图 3 所示。

本申请的网络架构可以包括多个服务器, 服务器与服务器之间可以进行数据通信, 例如一个服务器向另一个服务器发送报文。服务器与服务器之间可以通过交换机进行数据通信。如图 3 所示, 服务器 31 向服务器 32 发送报文, 服务器 31 作为数据的发送端, 服务器 32 作为数据的接收端。本申请的网络架构可以应用于上述 RDMA 的网络, 还可以应用于使用上述 RDMA 技术的其他网络或网卡上, 例如应用于胖树结构的数据中心。

以本申请应用于 RDMA 网卡为例, 该网卡可以应用于发送端和接收端, 该 RDMA 网卡可以包括接收单元、发送单元、QPC 存储单元、BM 管理单元以及定时器单元等。如图 4 所示, 发送端可以包括接收单元 41、发送单元 42、QPC 存储单元 43、BM 管理单元 44 以及定时器单元 45, 接收端可以包括接收单元 46、发送单元 47、QPC 存储单元 48、BM 管理单元 49 以及定时器单元 50。

其中, 发送单元可以用于发送数据包或发送报文或发送数据等;

接收单元可以用于接收数据包或接收报文或接收数据等, 例如可以接收 ACK 报文或 SACK 报文等;

QPC 存储单元可以用于存储链接的上下文信息。对于发送端来讲, 上下文信息可以包括发送端预期要发送的报文的序列号, 当发送端已经发送了该报文, 发送端的发送单元可以进行 QPC 读写, 指示 QPC 存储单元更新预期要发送的报文的序列号; 对于接收端来讲, 上下文信息可以包括接收端预期要接收的报文的序列号, 以及该报文将要存储的地址信息等, 当接收端接收到预期的报文后, 接收端的接收单元可以进行 QPC 读写, 指示 QPC 存储单元更新下一个预期接收的报文的序列号以及将要存储的地址信息。应用上述网络架构中, 不同的链接可以是相同的发送端对应不同的接收端, 例如一个链接的发送端为服务器 1, 接收端为服务器 2, 另一个链接的发送端为服务器 1, 接收端为服务器 3; 该发送端作为接收端时, 接收端也可以作为发送端;

BM 管理单元, 可以理解为 bitmap 管理单元, 可以存储有多个链接对应的 QP 的 bitmap。例如如图 5 所示, QP0 对应的 bitmap 的比特值为 0101...00, QP1 对应的 bitmap 的比特值为 00001...00, QP2 对应的 bitmap 的比特值为 0100...00。以 QP0 为例, 其 bitmap 指示, 起始 PSN 的报文的比特值为 0, 表示起始 PSN 的报文未接收到, 起始 PSN+1 的报文的比特值为 1, 表示起始 PSN+1 的报文已接收到, 起始 PSN+2 的报文的比特值为 0, 表示起始 PSN+2 的报文未接收到, 起始 PSN+3 的报文的比特值为 0, 表示起始 PSN+3 的报文未接收到。

对于发送端和接收端来说，接收单元可以用于根据报文接收成功与否对 BM 管理单元进行 BM 读写操作，以更新 BM 管理单元中的 bitmap（报文接收成功时相应的 bit 值可以为 1，报文接收失败时相应的 bit 值可以为 0。），发送单元可以用于对 BM 管理单元进行 BM 读操作，以读取 BM 管理单元中的 bitmap，知道哪些报文是接收成功的，哪些报文是发送失败的，接收成功时，发送单元可以发送 ACK 报文，接收失败时，发送单元可以发送 SACK 报文。或者，当接收单元接收到重复的报文时，通过进行 BM 读操作，知道 bitmap 中该报文的 bit 值已为 1，可以丢弃接收到的重复的报文。

定时器单元，可以用于设定超时时间。如果一个报文发出后，经过超时时间，仍没有得到应答的报文，定时器单元产生超时重传触发信号，触发发送单元再次发送此报文。

应用以上网络架构，本申请提出的报文重传方法的流程可以如实施例一和实施例二，但不局限于这两种实施例。

实施例一

本申请实施例提供一种报文重传方法，该方法可以应用于 RDMA 中，如图 6 所示，该方法包括：

601、发送端向接收端发送多个报文。

当发送端的发送单元 42 接收到发送端的软件发送的 DoorBell 时，意味着发送端的发送单元 42 将要周期性地向接收端发送报文。当发送端的发送单元 42 需要向发送端发送报文时，先从发送端的 QPC 存储单元 43 中读取第一个将要发送的报文的 PSN，例如 PSN 为 0，即发送端从 PSN 为 0 的报文开始向接收端发送报文。当然，第一个将要发送的报文的 PSN 也不一定为 0，也可以为发送端与接收端协商的一个初始值。

示例性的，如图 7 所示，发送端依次向接收端发送 PSN 为 0、1、2、3、4、5... 的报文。发送端的发送单元 42 每向接收端发送一个报文时，从 QPC 存储单元 43 先读取将要发送的报文的 PSN，并将 PSN 携带在报文中。发送单元 42 每发送一个报文后，发送单元 42 指示 QPC 存储单元 43 更新预期要发送的报文的 PSN。例如发送单元 42 从 QPC 存储单元 43 中读取预期要发送的 PSN 为 0，当发送单元 42 发送了 PSN 为 0 的报文后，指示 QPC 存储单元 43 更新预期要发送的报文的 PSN，QPC 存储单元 43 更新预期要发送的报文的 PSN 为 1，以此类推。

602、接收端接收发送端发送的多个报文，当接收到发送端发送的报文时，向发送端回复 ACK 报文。相应地，发送端接收 ACK 报文。

接收端的接收单元 46 每接收到一个报文时，接收单元 46 会触发发送单元 47 向发送端发送应答报文，例如 ACK 报文，以告知发送端该报文已接收到。

对于接收端来说，期望按照 PSN 的顺序接收报文。接收端的 QPC 存储单元 48 中可以存储有接收端期望接收到的报文的 PSN，接收端的接收单元 46 每接收到一个期望的报文时，QPC 存储单元 48 会更新下一次期望接收到的报文的 PSN。如果接收单元 46 接收到的报文的 PSN 等于 QPC 存储单元 48 期望接收到的报文的 PSN，发送单元 47 会向发送端发送 ACK 报文。同时，对于每个 QC，接收端每正确接收到一个报文，将该报文对应的比特值在 bitmap 中置 1。

示例性的，如图 7 所示，接收单元 46 接收到 PSN 为 0 的报文后，会指示 QPC 存

储单元 48 更新期望接收到的报文的 PSN 为 1；接收单元接收到 PSN 为 3 的报文后，会指示 QPC 存储单元 48 更新期望接收到的报文的 PSN 为 4，以此类推。

603、接收端确定存在接收失败的报文时，向发送端发送第一 SACK 报文，第一 SACK 报文指示发送端向接收端发送的报文中的至少一个报文发送失败；相应的，发送端接收来自接收端的第一 SACK 报文。

在一些实施例中，如果接收端接收到的报文携带的 PSN 大于 QPC 存储单元 48 中存储的预期要接收的报文的 PSN，那么接收端可以确定期望的 PSN 的报文接收失败。BM 管理单元 49 可以将此次接收到的报文在 bitmap 中对应的值置为 1，将期望接收到但又未接收到的报文在 bitmap 中对应的值置为 0。并且，发送单元 47 针对此次正确接收到的报文向接收端发送第一 SACK 报文，该第一 SACK 报文中可以携带 bitmap，bitmap 包括多个比特的比特值，多个比特的比特值指示当前 QC 中的多个报文是否被正确接收的指示。多个比特包括至少一个比特，至少一个比特指示发送端向接收端发送的报文中的至少一个报文发送失败。

示例性的，参考图 7，发送端向接收端发送的 PSN 为 0、1、2、3、4、5... 的多个报文中，PSN 为 0、3、4、5 的报文发送成功，或者说，接收端接收 PSN 为 0、3、4、5 的报文成功，PSN 为 1 和 2 的报文发送失败，或者说，接收端接收 PSN 为 1 和 2 的报文失败。当接收端的接收单元 46 用于接收到 PSN 为 0 的报文后，将 bitmap 中 PSN 为 0 的报文对应的比特值置 1，QPC 存储单元 48 用于将期望接收到的下一个报文的 PSN 置为 1。但是，当接收单元 46 接收到的下一个报文的 PSN 为 3 时，PSN 为 3 与期望接收到的报文的 PSN 为 1 不符，那么接收端可以确定 PSN 为 1 和 2 的报文接收失败，BM 管理单元 49 可以将 bitmap 中 PSN 为 3 的报文对应的比特值置为 1 的同时，将 PSN 为 1 和 2 的报文对应的比特值置为 0。发送单元 48 在向发送端发送 PSN 为 3 的报文对应的应答报文时，确定将要发送的应答报文为第一 SACK 报文，并从 BM 管理单元 49 中读取 PSN 为 1、2 和 3 的报文对应的比特值携带在第一 SACK 报文中，以及将 PSN 为 3 的报文的 PSN 携带在第一 SACK 报文中。对于发送端来说，接收单元 41 接收到第一 SACK 后，发送端可以读取第一 SACK 中的比特值 001 以及 PSN，PSN 为 3，并确定 PSN 为 3 的报文之前的报文丢失，由于发送端的 BM 管理单元 44 已经记录了 PSN 为 0 的报文对应的比特值为 1，那么发送端可以确定 PSN 为 1 和 2 的报文发送失败，BM 管理单元 44 并记录 PSN 为 1、2 和 3 的报文对应的比特值为 001。

604、发送端暂停向接收端发送新报文，并向接收端重传该至少一个报文，并在重传至少一个报文中最后一个报文时，记录发送端待发送的第一个新报文的包序列号。

至少一个报文中最后一个报文可以理解为至少一个报文中 PSN 最大的报文。当发送端接收到第一 SACK 报文时，可以读取第一 SACK 报文中的 PSN 以及 bitmap，根据第一 SACK 报文中的 PSN 以及 bitmap 可以确定哪些报文发送失败需要重传。

示例性的，参考图 7，接收单元 41 在接收到第一 SACK 报文时，发送端可以读取第一 SACK 报文中的 PSN 以及 bitmap，并将 bitmap 存储到 BM 管理单元 44 中。当 PSN 为 3 时，说明 PSN 为 3 的报文发送成功。当 bitmap 为 001 时，可知 PSN 为 1 和 2 的报文发送失败。因此，发送端确定需向接收端重传 PSN 为 1 和 2 的报文。因此，当发送端在接收到第一 SACK 时，暂停向接收端发送新报文，并向接收端重传 PSN 为 1 和

2 的报文。需要说明的是，本申请中的新报文可以理解为发送端发送完重传的报文后将要发送的第一个新的报文。

当发送端向接收端重传最后一个报文，即 PSN 为 2 的报文时，可以在 QPC 存储单元 43 中记录发送端待发送的第一个新报文的 PSN，即当前最大发送 PSN。图 7 的
5 示例中，发送端重传了 PSN 为 2 的报文后，记录待发送的第一个新报文的 PSN 为 N。而后，发送单元 42 在用于重传了 PSN 为 2 的报文后，可以用于继续向接收端发送新报文，例如继续发送 PSN 为 N、N+1、N+2 的新报文。图 7 中，N 为大于或等于 6 的整数。

605、当发送端接收到每个 SACK 报文时，将每个 SACK 报文中携带的包序列号与记录的第一个新报文的包序列号进行对比，确定每个 SACK 报文携带的包序列号是否大于或等于第一个新报文的包序列号。
10

示例性的，参考图 7，当发送单元 42 用于向接收端开始重传上述至少一个报文时，每当接收单元 41 用于接收到一个 SACK 报文时，发送端便读取 SACK 报文中的 PSN，将读取的 PSN 与记录的第一个新报文的 PSN 进行对比，以确定接收到的 ACK 报文携带的 PSN 是否大于或等于第一个新报文的 PSN。例如当接收单元 41 接收到 PSN 为 2
15 的报文的 ACK 时，将 2 与记录的 N 进行对比，确定 2 小于 N。

606、当接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于第一个新报文的包序列号时，发送端确定是否存在未接收到针对至少一个报文的 ACK 报文。
20

由于记录的第一个新报文是在重传的报文之后发送的，如果发送端已经接收到了记录的第一个新报文的 SACK，但是还未接收到重传的报文的 ACK 时，说明重传的报文重传失败。

示例性的，参考图 7，发送单元 42 用于重传了 PSN 为 1 和 2 的报文后，PSN 为 1 的报文第一次重传失败，但是 PSN 为 2 的报文第一次重传成功，接收端针对 PSN 为 2
25 的报文会向发送端回复第三 SACK 报文，第三 SACK 报文携带的 PSN 为 2，指示 PSN 为 2 的报文重传成功，但是 PSN 为 1 的报文重传失败。此时，第三 SACK 携带的 PSN 为 2 是小于记录的 N 的。由于 PSN 为 2 的报文重传后发送的下一个报文为 PSN 为 N 的新报文，因此，接收端还会继续向发送端发送针对 PSN 为 N 的报文的第二 SACK 报文的，第二 SACK 报文携带的 PSN 为 N，并且指示重传的 PSN 为 1 的报文重传失败，PSN 为 2 的报文重传成功，此时第二 SACK 携带的 PSN 为 N 是等于记录的 N 的。
30 当接收单元 41 接收到 PSN 为 N 的第二 SACK 报文时，发送端会根据第二 SACK 报文确定是否存在未接收到针对 PSN 为 1 和 2 的报文的 ACK 报文，以确定 PSN 为 1 和 2 的报文是否都重传成功。

607、若发送端确定还未接收到至少一个报文中的第一报文对应的 ACK 报文，则发送端再次向接收端重传第一报文。
35

示例性的，参考图 7，假设第一报文为 PSN 为 1 的报文时，如果发送端发现接收单元 41 还未接收到 PSN 为 1 的报文的 ACK，那么发送端确定 PSN 为 1 的报文重传失败，需要再次进行重传。此时，发送端暂停向接收端发送新报文，且发送端再次重传 PSN 为 1 的报文不需要等如现有技术一样等到超时周期超时时才进行再次重传，而是

可以立即进行再次重传。如图 7 所示, 发送单元 41 立即再次重传 PSN 为 1 的报文, 而后, 继续向接收端发送新报文, 该新报文可以如图 7 中的 PSN 为 M、M+1 和 M+2 的新报文, M 为大于 N+2 的整数。

由此, 本申请中, 发送端不仅可以利用 SACK 中携带的 bitmap 准确知道发送成功的报文和发送失败的报文, 且发送端只重传发送失败的报文, 可以提高重传效率, 降低重传带宽的占用。而且, 发送端通过在重传最后一个报文时记录下一个新报文的 PSN, 通过对比接收到的 SACK 携带的 PSN 与记录的 PSN, 可以判定是否存在重传失败的报文, 即是否存在“被丢弃的重传报文”。如果确定存在重传失败的报文, 可以立即再次进行重传。通常, 从发送端第一次重传报文开始到下一次重传报文开始, 其时长通常远小于一个超时周期, 该时长在一个 RTT 周期内就能再次开始重传报文, 即在一个 RTT 周期内就能重传“被丢弃的重传报文”, 一个 RTT 周期远小于一个超时周期, 可以“被丢弃的重传报文”被及时重传, 缩短了重传时延。

实施例二

本申请实施例提供一种报文重传方法, 方法可以应用于 RDMA 中, 如图 8 所示, 该方法包括:

801、发送端向接收端发送多个报文。

步骤 801 的实现方式可以参考步骤 601。

802、接收端接收发送端发送的多个报文, 当接收到发送端发送的报文时, 向发送端回复 ACK 报文。相应地, 发送端接收 ACK 报文。

步骤 802 的实现方式可以参考步骤 602。

803、发送端为多个报文队列中的每个报文队列建立链表, 链表用于记录发送端在重传报文队列中的任一报文后待发送的第一个新报文的包序列号。

每个报文队列(即 QP)包括多个报文, 每个报文队列对应一个连接。该链表可以存储在 QPC 存储单元 43 中。

也可以理解为, 该链表用于为重传的报文与待发送的第一个新报文建立对应关系。这是考虑到, 后续执行报文发送过程中, 新报文是在重传的报文之后发送的, 如果发送端已经接收到的记录的待发送的第一个新报文的应答报文, 却还未接收到对应的重传报文的应答报文, 说明重传的报文重传失败。

804、接收端确定存在接收失败的报文时, 向发送端发送第一 SACK 报文, 第一 SACK 报文指示发送端向接收端发送的报文中的至少一个报文发送失败; 相应的, 发送端接收来自接收端的第一 SACK 报文。

步骤 804 的实现方式与步骤 603 的实现方式类似。

805、发送端暂停向接收端发送新报文, 并向接收端重传该至少一个报文, 并在重传至少一个报文中的每个报文时, 记录发送端在重传每个报文后待发送的第一个新报文的包序列号。

如果接收单元 41 接收到第一 SACK 报文, 那么发送端确定之前发送的报文存在发送失败的报文, 因此, 发送端控制发送单元 42 暂停向接收端发送新报文, 需要开始重传发送失败的报文。在重传每个报文时, 发送端需要在 QPC 存储单元 43 中记录每个重传的报文被发送后, 待发送的第一个新报文的 PSN。其中, 如果多个重传的报文

是连续发送失败的，那么这多个重传的报文对应记录的第一个新报文的 PSN 相同。如果多个重传的报文不是连续发送失败的，那么这多个重传的报文对应记录的第一个新报文的 PSN 不同。这是由于发送端接收到了在前重传的报文的 ACK 报文后，可能还未接收到在后重传的报文重传成功的应答报文时，需先发送新报文，当接收到在后重传的报文的 ACK 报文后，继续发送新报文，两次要发送的新报文不同。

5 示例性的，参考图 9，假设发送端发送的 PSN 为 0 的报文为起始报文的多个报文均属于同一个 QP。当接收单元 41 接收到了携带 PSN 为 2 的 SACK 报文（PSN 为 2 的报文发送成功）时，根据 SACK 报文中的 bitmap 确定 PSN 为 1 的报文发送失败，PSN 为 0 和 2 的报文发送成功。因此，发送单元 42 暂停发送新报文，并重传 PSN 为 1 的报文，在重传 PSN 为 1 的报文时，发送端在 QPC 存储单元 43 中的链表中记录重传 PSN 为 1 的报文后待发送的第一个新报文的 PSN 为 N；而后，接收单元 41 接收到了携带 PSN 为 3 的 SACK 报文（该 SACK 报文指示 PSN 为 0 和 3 的报文发送成功，但是 PSN 为 1 的报文发送失败）。参考图 9，发送端在 PSN 为 4 的报文之后还发送了 PSN 为 5 的报文，由于 PSN 为 4 的报文第一次也发送失败了，因此，接收单元 41 接收到了携带 PSN 为 5 的 SACK 时，该 SACK 中的 bitmap 指示 PSN 为 4 的报文也发送失败。因此，发送单元 42 暂停发送新报文，并重传 PSN 为 4 的报文，在重传 PSN 为 4 的报文时，发送端在 QPC 存储单元 43 中的链表中记录重传 PSN 为 4 的报文后待发送的第一个新报文的 PSN 为 N+2。而后，发送单元 42 继续发送 PSN 为 N+3 等包序列号的新报文。

20 也就是说，每重传一个报文时，都需要在链表中记录发送端在重传每个报文后待发送的第一个新报文的 PSN，也可以说建立有重传的报文与待发送的新报文的对应关系。

假设在上述重传的报文过程中：

25 至少一个报文包括第二报文，在重传第二报文时，发送端确定第二报文所在的报文队列对应的第一链表，并在第一链表中记录发送端在重传第二报文后待发送的第一个新报文的包序列号，且包序列号为第一新报文的包序列号。第二报文例如可以为上述重传的 PSN 为 1 和 4 的报文。

那么，该方法还可以包括：

30 806、当发送端接收到第二 SACK 报文，第二 SACK 报文中携带的报文的包序列号大于或等于记录的第一新报文的包序列号时，若还未接收到第二报文对应的 ACK 报文，则发送端再次重传第二报文。

35 当发送端在重传至少一个报文中的第一个报文开始后，每接收到一个报文的 SACK 报文时，读取 SACK 报文中的 PSN，将该 PSN 与记录链表中的 PSN 进行对比，如果发现第二 SACK 携带的 PSN 大于或等于链表中的某个报文的 PSN 时，检查是否接收到在链表中记录该报文的 PSN 时被重传的第二报文，即是否接收到链表中记录的报文对应的被重传的报文。

示例性的，重传的第二报文可以理解为图 9 中的 PSN 为 1 或 4 的重传报文。按照步骤 805 中的说明，当接收单元 41 已经接收到了 PSN 为 N 的报文的 SACK 报文，但是还没接收到重传的 PSN 为 1 的报文的 ACK 报文，发送端确定 PSN 为 1 的报文重传

失败,需再次重传 PSN 为 1 的报文。当接收单元 41 已经接收到了 PSN 为 N+2 的报文的 SACK 报文,但是还没接收到重传的 PSN 为 4 的报文的 ACK 报文,发送端确定 PSN 为 4 的报文重传失败,需再次重传 PSN 为 4 的报文。从图 9 可以看出,发送单元 42 第二次重传了 PSN 为 1 的报文时,还可以在链表中记录待发送的第一个新报文的 PSN 为 M;发送单元 42 第二次重传了 PSN 为 4 的报文时,还可以在链表中记录待发送的第一个新报文的 PSN 为 M+2,以确定第二次重传的 PSN 为 1 和 4 的报文是否重传成功。

由此,本申请中,发送端针对发送失败的报文,在每重传一个发送失败的报文时,记录待发送的下一个新报文的 PSN,当已经接收到记录的下一个新报文的 SACK 报文,但还未接收到对应的重传报文的 ACK 报文时,确定重传的报文重传失败,发送端可以立即再次重传报文,不用等到重传报文的超时周期超时时才进行再次重传,可以使得报文重传及时。况且,第一次重传报文到再次重传报文的时间可以为一个 RTT 周期,一个 RTT 周期远小于一个超时周期,使得在一个 RTT 周期内可以精确的再次重传多个被重传失败的报文,缩短了重传时延。

可以理解的是,为了实现上述功能,通信装置(例如可以为发送端)包含了执行各个功能相应的硬件和/或软件模块。结合本文中所公开的实施例描述的各示例的算法步骤,本申请能够以硬件或硬件和计算机软件的结合形式来实现。某个功能究竟以硬件还是计算机软件驱动硬件的方式来执行,取决于技术方案的特定应用和设计约束条件。本领域技术人员可以结合实施例对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

本实施例可以根据上述方法示例对通信装置进行功能模块的划分,例如,可以对各个功能划分各个功能模块,也可以将两个或两个以上的功能集成在一个处理模块中。上述集成的模块可以采用硬件的形式实现。需要说明的是,本实施例中对模块的划分是示意性的,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

在采用对应各个功能划分各个功能模块的情况下,图 10 示出了一种通信装置 100 的一种可能的组成示意图,该通信装置可以为上述实施例中涉及的发送端,如图 10 所示,该发送端可以包括:接收单元 1001、发送单元 1002、记录单元 1003 和确定单元 1004。

其中,接收单元 1001 可以用于支持发送端执行上述步骤 602、步骤 603、步骤 802、步骤 806 等,和/或用于本文所描述的技术的其他过程。

发送单元 1002 可以用于支持发送端执行上述步骤 601、步骤 604、步骤 607、步骤 801、步骤 804、步骤 805 以及步骤 806 等,和/或用于本文所描述的技术的其他过程。

记录单元 1003 可以用于支持发送端执行上述步骤 604、步骤 805 等,和/或用于本文所描述的技术的其他过程。

确定单元 1004 可以用于支持发送端执行上述步骤 605、步骤 606、步骤 607 等,和/或用于本文所描述的技术的其他过程。

还可以包括建立单元 1005,用于支持发送端执行上述步骤 803 等,和/或用于本文所描述的技术的其他过程。

图 10 中的发送单元 1002 与图 4 中发送端的发送单元 41 类似，图 10 中的接收单元 1001 与图 4 中的接收单元 41 类似，图 10 中的记录单元 1003 记录的区域可以为图 4 中的 QPC 存储单元 43、BM 管理单元 44 等。

需要说明的是，上述方法实施例涉及的各步骤的所有相关内容均可以援引到对应功能模块的功能描述，在此不再赘述。

本实施例提供的通信装置 100，用于执行上述报文重传方法，因此可以达到与上述实现方法相同的效果。

在采用集成的单元的情况下，通信装置 100 可以包括处理模块、存储模块和通信模块。其中，处理模块可以用于对通信装置 100 的动作进行控制管理，例如，可以用于支持通信装置 100 执行上述确定单元 1101 和记录单元 1003 执行的步骤。存储模块可以用于支持通信装置 100 存储程序代码和数据等，例如记录 bitmap 和 PSN 等。通信模块，可以用于支持通信装置 100 与其他设备的通信，例如与接收端的通信。

其中，处理模块可以是处理器或控制器。其可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方框，模块和电路。处理器也可以是实现计算功能的组合，例如包含一个或多个微处理器组合，数字信号处理（digital signal processing, DSP）和微处理器的组合等等。存储模块可以是存储器。通信模块具体可以为射频电路、蓝牙芯片、Wi-Fi 芯片等与其他电子设备交互的设备。

在一个实施例中，当处理模块为处理器，存储模块为存储器，通信模块包括发射器和接收器时，本实施例所涉及的通信装置可以为具有图 11 所示结构的服务器。

本申请实施例还提供一种电子设备，包括一个或多个处理器以及一个或多个存储器。该一个或多个存储器与一个或多个处理器耦合，一个或多个存储器用于存储计算机程序代码，计算机程序代码包括计算机指令，当一个或多个处理器执行计算机指令时，使得电子设备执行上述相关方法步骤实现上述实施例中的报文重传方法。

本申请的实施例还提供一种计算机存储介质，该计算机存储介质中存储有计算机指令，当该计算机指令在电子设备上运行时，使得电子设备执行上述相关方法步骤实现上述实施例中的报文重传方法。

本申请的实施例还提供了一种计算机程序产品，当该计算机程序产品在计算机上运行时，使得计算机执行上述相关步骤，以实现上述实施例中电子设备执行的报文重传方法。

另外，本申请的实施例还提供一种装置，这个装置具体可以是芯片，组件或模块，该装置可包括相连的处理器和存储器；其中，存储器用于存储计算机执行指令，当装置运行时，处理器可执行存储器存储的计算机执行指令，以使芯片执行上述各方法实施例中电子设备执行的报文重传方法。

其中，本实施例提供的电子设备、计算机存储介质、计算机程序产品或芯片均用于执行上文所提供的对应的方法，因此，其所能达到的有益效果可参考上文所提供的对应的方法中的有益效果，此处不再赘述。

本申请另一实施例提供了一种系统，该系统可以包括上述发送端和接收端，可以用于实现上述报文重传方法。

通过以上实施方式的描述，所属领域的技术人员可以了解到，为描述的方便和简

洁，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。

5 在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块或单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个装置，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

10 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是一个物理单元或多个物理单元，即可以位于一个地方，或者也可以分布到多个不同地方。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

15 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

20 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个可读取存储介质中。基于这样的理解，本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一个设备（可以是单片机，芯片等）或处理器（processor）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（read only memory, ROM）、随机存取存储器（random access memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

25 以上内容，仅为本申请的具体实施方式，但本申请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申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

权 利 要 求 书

1、一种报文重传方法，其特征在于，所述方法应用于远端直接内存访问 RDMA 网络中，所述方法包括：

5 发送端接收来自接收端的第一选择性应答 SACK 报文，所述第一 SACK 报文指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败；

所述发送端向所述接收端重传所述至少一个报文，并在重传所述至少一个报文中最后一个报文时，记录所述发送端待发送的第一个新报文的包序列号；

10 当所述发送端接收到第二 SACK 报文，所述第二 SACK 报文中携带的报文的包序列号大于或等于所述第一个新报文的包序列号时，若所述发送端根据所述第二 SACK 报文确定还未接收到所述至少一个报文中的第一报文，则所述发送端再次向所述接收端重传所述第一报文。

2、根据权利要求 1 所述的方法，其特征在于，所述第一 SACK 报文中携带位图，所述位图包括至少一个比特，所述至少一个比特中的部分比特指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败。

15 3、根据权利要求 1 或 2 所述的方法，其特征在于，所述当所述发送端接收到第二 SACK 报文，所述第二 SACK 报文中携带的报文的包序列号大于或等于所述第一个新报文的包序列号时，若所述发送端根据所述第二 SACK 报文确定还未接收到所述至少一个报文中的第一报文，则所述发送端再次向所述接收端重传所述第一报文包括：

20 当所述发送端接收到每个 SACK 报文时，将所述每个 SACK 报文中携带的包序列号与记录的所述第一个新报文的包序列号进行对比，确定所述每个 SACK 报文携带的包序列号是否大于或等于所述第一个新报文的包序列号；

当接收到所述第二 SACK 报文，所述第二 SACK 报文中携带的报文的包序列号大于或等于所述第一个新报文的包序列号时，所述发送端确定重传的所述至少一个报文是否存在未接收到针对所述至少一个报文的 ACK 报文；

25 若所述发送端还未接收到所述至少一个报文中的第一报文对应的 ACK 报文，则所述发送端再次向所述接收端重传所述第一报文。

4、根据权利要求 1-3 任一项所述的方法，其特征在于，所述发送端从开始重传所述至少一个报文到再次重传所述第一报文的时间段小于一个超时周期。

30 5、一种报文重传方法，其特征在于，所述方法应用于远端直接内存访问 RDMA 网络中，所述方法包括：

发送端接收来自接收端的第一选择性应答 SACK 报文，所述第一 SACK 报文指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败；

35 所述发送端向所述接收端重传所述至少一个报文，并在重传所述至少一个报文的每个报文时，记录所述发送端在重传每个报文后待发送的第一个新报文的包序列号；其中，所述至少一个报文包括第二报文，所述发送端在重传所述第二报文时，记录的待发送的第一个新报文的包序列号为第一新报文的包序列号；

当所述发送端接收到第二 SACK 报文，所述第二 SACK 报文中携带的报文的包序列号大于或等于记录的所述第一新报文的包序列号时，若还未接收到所述第二报文对应的 ACK 报文，则所述发送端再次重传所述第二报文。

6、根据权利要求 5 所述的方法，其特征在于，所述第一 SACK 报文中携带位图，所述位图包括至少一个比特，所述至少一个比特中的部分比特指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败。

7、根据权利要求 5 或 6 所述的方法，其特征在于，在所述发送端向所述接收端重传所述至少一个报文之前，所述方法还包括：

所述发送端为多个报文队列中的每个报文队列建立链表，所述链表用于记录所述发送端在重传报文队列中的任一报文后待发送的第一个新报文的包序列号。

8、根据权利要求 7 所述的方法，其特征在于，所述在重传所述至少一个报文中的每个报文时，记录所述发送端重传每个报文后待发送的第一个新报文的包序列号包括：
10 在重传所述第二报文时，所述发送端确定所述第二报文所在的报文队列对应的第一链表，并在所述第一链表中记录所述发送端在重传所述第二报文后待发送的第一个新报文的包序列号，且包序列号为所述第一新报文的包序列号。

9、根据权利要求 5-8 任一项所述的方法，其特征在于，所述发送端从开始重传所述至少一个报文到再次重传所述第二报文的时间段小于一个超时周期。

15 10、一种通信装置，所述通信装置作为发送端，其特征在于，所述通信装置应用于远端直接内存访问 RDMA 网络中，所述通信装置包括：

接收器，用于接收来自接收端的第一选择性应答 SACK 报文，所述第一 SACK 报文指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败；

20 发射器，用于向所述接收端重传所述至少一个报文；以及存储器，用于所述发送端在重传所述至少一个报文中最后一个报文时，记录所述发送端待发送的第一个新报文的包序列号；

25 处理器，还用于确定所述发送端接收到第二 SACK 报文，所述第二 SACK 报文中携带的报文的包序列号大于或等于所述第一个新报文的包序列号，若所述发送端还未接收到所述至少一个报文中的第一报文对应的 ACK 报文，指示所述发射器还用于再次向所述接收端重传所述第一报文。

11、根据权利要求 10 所述的通信装置，其特征在于，所述第一 SACK 报文中携带位图，所述位图包括至少一个比特，所述至少一个比特中的部分比特指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败。

30 12、根据权利要求 10 或 11 所述的通信装置，其特征在于，所述处理器用于：当所述发送端接收到每个 SACK 报文时，将所述每个 SACK 报文中携带的包序列号与记录的所述第一个新报文的包序列号进行对比，确定所述每个 SACK 报文携带的包序列号是否大于或等于所述第一个新报文的包序列号；

35 当接收到所述第二 SACK 报文，所述第二 SACK 报文中携带的报文的包序列号大于或等于所述第一个新报文的包序列号时，确定是否存在未接收到针对所述至少一个报文的 ACK 报文；

所述发射器，用于若所述发送端还未接收到所述至少一个报文中的第一报文对应的 ACK 报文，则再次向所述接收端重传所述第一报文。

13、根据权利要求 10-12 任一项所述的通信装置，其特征在于，所述发送端从开始重传所述至少一个报文到再次重传所述第一报文的时间段小于一个超时周期。

14、一种通信装置，所述通信装置作为发送端，其特征在于，所述通信装置应用于远端直接内存访问 RDMA 网络中，所述通信装置包括：

接收器，用于接收来自接收端的第一选择性应答 SACK 报文，所述第一 SACK 报文指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败；

5 发射器，用于向所述接收端重传所述至少一个报文；

存储器，用于在重传所述至少一个报文中的每个报文时，记录所述发送端在重传每个报文后待发送的第一个新报文的包序列号；其中，所述至少一个报文包括第二报文，所述发送端在重传所述第二报文时，记录的待发送的第一个新报文的包序列号为第一新报文的包序列号；

10 所述接收器，还用于当接收到第二 SACK 报文，所述第二 SACK 报文中携带的报文的包序列号大于或等于记录的所述第一新报文的包序列号时，若还未接收到所述第二报文对应的 ACK 报文，则指示所述发射器，用于再次重传所述第二报文。

15 15、根据权利要求 14 所述的通信装置，其特征在于，所述第一 SACK 报文中携带位图，所述位图包括至少一个比特，所述至少一个比特中的部分比特指示所述发送端向所述接收端发送的报文中的至少一个报文发送失败。

16、根据权利要求 14 或 15 所述的通信装置，其特征在于，还包括处理器，用于：

为多个报文队列中的每个报文队列建立链表，所述链表存储在所述存储器中，所述链表用于记录所述发送端在重传报文队列中的任一报文后待发送的第一个新报文的包序列号。

20 17、根据权利要求 16 所述的通信装置，其特征在于，所述存储器用于：

在重传所述第二报文时，确定所述第二报文所在的报文队列对应的第一链表，并在所述第一链表中记录所述发送端在重传所述第二报文后待发送的第一个新报文的包序列号，且包序列号为所述第一新报文的包序列号。

25 18、根据权利要求 14-17 任一项所述的通信装置，其特征在于，所述发送端从开始重传所述至少一个报文到再次重传所述第二报文的时间段小于一个超时周期。

19、一种计算机可读存储介质，其特征在于，包括计算机指令，当计算机指令在电子设备上运行时，使得电子设备执行上述权利要求 1-9 中的任一项所述的方法。

20、一种计算机程序产品，其特征在于，当计算机程序产品在计算机上运行时，使得电子设备执行上述权利要求 1-9 中的任一项所述的方法。