国 家 知 识 产 权 局

250014

山东省济南市历下区经十路 17703 号华特广场 B510 室 济南圣达知识产权代理有限公司

闫圣娟(0531-68605722)

发文日：

2026 年 05 月 20 日

申请号：202610699735.0 **发文序号：**2026052001446580

专 利 申 请 受 理 通 知 书

根据专利法第 28 条及其实施细则第 43 条、第 44 条的规定，申请人提出的专利申请已由国家知识产权局受理。现将确定的申请号、申请日等信息通知如下：

申请号：2026106997350

申请日：2026 年 05 月 20 日

申请人：山东省计算中心（国家超级计算济南中心）,齐鲁工业大学（山东省科学院）

发明人：谭立状,褚夫明,史慧玲,张玮,张志远

发明创造名称：一种基于位图的 RDMA 网络丢包检测系统及方法

经核实，国家知识产权局确认收到文件如下：

权利要求书 1 份 1 页,权利要求项数 ： 10 项

说明书 1 份 7 页

说明书附图 1 份 3 页

说明书摘要 1 份 1 页

发明专利请求书 1 份 5 页

实质审查请求书 文件份数：1 份

申请方案卷号：2026703545

提示：

1.申请人收到专利申请受理通知书之后，认为其记载的内容与申请人所提交的相应内容不一致时，可以向国家知识产权局请求更正。

2.申请人收到专利申请受理通知书之后，再向国家知识产权局办理各种手续时，均应当准确、清晰地写明申请号。

国家知识产权局 专利申查业务章 1101081359734

审 查 员：自动受理

联系电话：010-62356655

审查部门：初审及流程管理部

纸件申请，回函请寄：100088 北京市海淀区蓟门桥西土城路 6 号 国家知识产权局专利局受理处收

电子申请，应当通过专利业务办理系统以电子文件形式提交相关文件。除另有规定外，以纸件等其他形式提交的文件视为未提交。

权 利 要 求 书

1. 一种基于位图的 RDMA 网络丢包检测系统，其特征在于，包括：
发送端和接收端，发送端与接收端通信连接；
发送端被配置为：建立发送窗口，并初始化与发送窗口覆盖的包序列号范围相对应的发送位图；发送数据分组，并根据数据分组的包序列号更新发送位图中对应分组的发送状态；接收来自接收端的快速反馈消息；根据快速反馈消息，定位发送位图中对应的分组状态项，并将未确认的对应分组状态更新为待重传状态；对待重传状态对应的数据分组执行重传；
接收端被配置为：建立接收窗口，并初始化与接收窗口覆盖的包序列号范围相对应的接收位图；接收数据分组，根据数据分组的包序列号更新接收位图中对应位置的置位状态；基于接收位图中未置位位置与后续已置位位置之间形成的状态缺口，识别疑似丢包分组；基于乱序容忍深度、持续时间阈值、窗口停滞阈值和路径标识中的至少一种信息，对疑似丢包分组进行乱序容忍判定；当判定为真实丢包时，生成包含缺口位置信息或区间信息的快速反馈消息并发送至发送端。

2. 根据权利要求 1 的系统，其特征在于，当接收位图中存在未置位位置、且该位置的后续位置已置位时，形成状态缺口；状态缺口为单个包序列号对应的单点缺口，或者为多个连续未置位位置组成的缺口区间。

3. 根据权利要求 1 的系统，其特征在于，乱序容忍判定包括：当疑似丢包分组对应的乱序深度未超过预设最大乱序容忍深度、持续时间未超过预设持续时间阈值且接收窗口停滞程度未超过预设窗口停滞阈值时，维持等待状态而不立即判定为丢包；当乱序深度超过预设最大乱序容忍深度，或者持续时间超过预设持续时间阈值，或者窗口停滞的程度超过预设窗口停滞阈值时，将疑似丢包分组判定为丢包事件。

4. 根据权利要求 3 的系统，其特征在于，乱序深度为当前已接收最大包序列号与缺口起始包序列号之间的差值；持续时间为状态缺口首次被发现的时刻至当前判定时刻之间的时间差；窗口停滞为接收窗口基准包序列号因状态缺口未被填补而无法继续推进的持续时间，或者在接收窗口未推进期间已接收的后续分组数量。

5. 根据权利要求 1 的系统，其特征在于，接收端支持单路径模式和多路径模式；在单路径模式下，采用全局位图模式统一维护窗口内所有分组状态；在多路径模式下，采用路径级位图模式分别为每条子路径维护对应的路径级接收位图，并结合路径标识区分跨路径乱序与真实丢包。

6. 根据权利要求 1 的系统，其特征在于，接收位图采用分段位图、变长位图或两者结合的方式进行扩展，以适配高带宽时延积传输；其中，分段位图将逻辑大窗口划分为多个连续片段，每个片段对应一个子位图；变长位图根据往返时延、链路带宽、带宽时延积、拥塞状态或业务负载动态调整位图长度。

7. 根据权利要求 1 的系统，其特征在于，快速反馈消息包括以下字段中的一种或多种：队列对标识、缺口起始包序列号、缺口长度、当前接收边界、当前窗口基准、时间戳、路径标识和统计计数。

8. 一种基于位图的 RDMA 网络丢包检测方法，应用于接收端，其特征在于，包括：
建立接收窗口，并初始化与接收窗口覆盖的包序列号范围相对应的接收位图；接收数据分组，根据数据分组的包序列号更新接收位图中对应位置的置位状态；基于接收位图中未置位位置与后续已置位位置之间形成的状态缺口，识别疑似丢包分组；基于乱序容忍深度、持续时间阈值、窗口停滞阈值和路径标识中的至少一种信息，对疑似丢包分组进行乱序容忍判定；当判定为真实丢包时，生成包含缺口位置信息或区间信息的快速反馈消息并发送至发送端。

9. 一种基于位图的 RDMA 网络丢包检测方法，应用于发送端，其特征在于，包括：
建立发送窗口，并初始化与发送窗口覆盖的包序列号范围相对应的发送位图；发送数据分组，并根据数据分组的包序列号更新发送位图中对应分组的发送状态；接收来自接收端的快速反馈消息；根据快速反馈消息，定位发送位图中对应的分组状态项，并将未确认的对应分组状态更新为待重传状态；对待重传状态对应的数据分组执行重传。

10. 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，该计算机程序被处理器执行时实现权利要求 8 或 9 所述的一种基于位图的 RDMA 网络丢包检测方法。

100001
2023.03

说 明 书

一种基于位图的 RDMA 网络丢包检测系统及方法

技术领域

[0001] 本发明涉及远程直接内存访问网络传输技术领域，尤其涉及一种基于位图的 RDMA 网络丢包检测系统及方法。

背景技术

[0002] 本部分的陈述仅仅是提供了与本发明相关的背景技术信息，不必然构成在先技术。

[0003] 远程直接内存访问（RDMA）网络是指采用远程直接内存访问技术的通信网络，该技术允许网络接口在较少主机内核干预的条件下直接完成本地与远端内存之间的数据访问，具有低时延、高吞吐和低 CPU 开销等特点，已广泛应用于数据中心存储、分布式计算、人工智能训练、高性能计算和内存池化等场景。随着网络规模扩大及链路速率提升，RDMA 运行环境逐步呈现高带宽、高并发、大窗口、多路径和一定比例乱序等特征。可靠连接类 RDMA 通常依赖包序列号、确认反馈与重传机制来保障数据正确交付。

[0004] 然而，现有方案存在以下不足：

（1）丢包识别粒度较粗：传统顺序确认和粗粒度超时机制难以及时定位具体缺失分组，只能在超时或异常累计后触发恢复，恢复时延较高。

[0005] （2）难以区分乱序与真实丢包：在链路抖动或多路径转发场景下，后续分组可能先于前序分组到达，容易将短时乱序误判为丢包，触发不必要的反馈与重传。

[0006] （3）大窗口场景扩展性不足：高带宽时延积条件下窗口覆盖范围大，发送窗口和接收窗口覆盖的包序列号范围较大，传统按包维护或稀疏状态维护方式在查找、更新和滑动方面复杂度较高。

[0007] （4）多路径场景误判率高：当同一逻辑业务流的数据包经由不同子路径传输时，各路径时延差异会导致跨路径乱序，仅维护全局顺序状态时容易将跨路径乱序误识别为丢包。

[0008] （5）反馈与重传协同效率不足：对疑似缺失分组的反馈不够精细，重传触发粗放，难以实现单包或区间级的精准恢复。

[0009] 上述问题的根源在于传统机制缺乏对窗口内分组状态的紧凑维护、对乱序与丢包的区分能力以及对反馈与重传的精细协同设计，这给高吞吐、低延迟传输带来了技术挑战。

发明内容

[0010] 为了解决现有技术的不足，本发明提供了一种基于位图的 RDMA 网络丢包检测系统及方法，在发送端建立发送位图，在接收端建立接收位图；接收端根据接收位图中未置位与后续已置位形成的状态缺口识别疑似丢包分组，结合乱序容忍深度、持续时间、窗口停滞及路径标识进行判定，区分乱序与真实丢包；判定为丢包时生成包含缺口位置或区间信息的快速反馈消息并发送至发送端；发送端依据反馈消息定位发送位图中对应分组并执行单包、区间或策略聚合重传。本发明适用于高带宽、多路径及大窗口 RDMA 网络，可提高丢包检测精度，降低误判率，缩短恢复时延，减少冗余重传。

[0011] 一方面，提供了一种基于位图的 RDMA 网络丢包检测系统，包括：

发送端和接收端，发送端与接收端通信连接；

发送端被配置为：建立发送窗口，并初始化与发送窗口覆盖的包序列号范围相对应的发送位图；发送数据分组，并根据数据分组的包序列号更新发送位图中对应分组的发送状态；接收来自接收端的快速反馈消息；根据快速反馈消息，定位发送位图中对应的分组状态项，并将未确认的对应分组状态更新为待重传状态；对待重传状态对应的数据分组执行重传；

接收端被配置为：建立接收窗口，并初始化与接收窗口覆盖的包序列号范围相对应的接收位图；接收数据分组，根据数据分组的包序列号更新接收位图中对应位置的置位状态；基于接收位图中未置位位置与后续已置位位置之间形成的状态缺口，识别疑似丢包分组；基于乱序容忍深度、持续时间阈值、窗口停滞阈值和路径标识中的至少一种信息，对疑似丢包分组进行乱序容忍判定；当判定为真实丢包时，生成包含缺口位置信息或区间信息的快速反馈消息并发送至发送端。

[0012] 进一步地，当接收位图中存在未置位位置、且该位置的后续位置已置位时，形成状态缺口；状态缺口为单个包序列号对应的单点缺口，或者为多个连续未置位位置组成的缺口区间。

[0013] 进一步地，乱序容忍判定包括：当疑似丢包分组对应的乱序深度未超过预设最大乱序容忍深度、持续时间未超过预设持续时间阈值且接收窗口停滞程度未超过预设窗口停滞阈值时，维持等待状态而不立即判定为丢包；当乱序深度超过预设最大乱序容忍深度，或者持续时间超过预设持续时间阈值，或者窗口停滞的程度超过预设窗口停滞阈值时，将疑似丢包分组判定为丢包事件。

[0014] 进一步地，乱序深度为当前已接收最大包序列号与缺口起始包序列号之间的差值；持续时间为状态缺口首次被发现的时刻至当前判定时刻之间的时间差；窗口停滞为接收窗口基准包序列号因状态缺口未被

填补而无法继续推进的持续时间，或者在接收窗口未推进期间已接收的后续分组数量。

[0015] 进一步地，接收端支持单路径模式和多路径模式；在单路径模式下，采用全局位图模式统一维护窗口内所有分组状态；在多路径模式下，采用路径级位图模式分别为每条子路径维护对应的路径级接收位图，并结合路径标识区分跨路径乱序与真实丢包。

[0016] 进一步地，接收位图采用分段位图、变长位图或两者结合的方式进行扩展，以适配高带宽时延积传输；其中，分段位图将逻辑大窗口划分为多个连续片段，每个片段对应一个子位图；变长位图根据往返时延、链路带宽、带宽时延积、拥塞状态或业务负载动态调整位图长度。

[0017] 进一步地，快速反馈消息包括以下字段中的一种或多种：队列对标识、缺口起始包序列号、缺口长度、当前接收边界、当前窗口基准、时间戳、路径标识和统计计数。

[0018] 再一方面，还提供了一种基于位图的 RDMA 网络丢包检测方法，应用于接收端，包括：

建立接收窗口，并初始化与接收窗口覆盖的包序列号范围相对应的接收位图；接收数据分组，根据数据分组的包序列号更新接收位图中对应位置的置位状态；基于接收位图中未置位位置与后续已置位位置之间形成的状态缺口，识别疑似丢包分组；基于乱序容忍深度、持续时间阈值、窗口停滞阈值和路径标识中的至少一种信息，对疑似丢包分组进行乱序容忍判定；当判定为真实丢包时，生成包含缺口位置信息或区间信息的快速反馈消息并发送至发送端。

[0019] 再一方面，还提供了一种基于位图的 RDMA 网络丢包检测方法，应用于发送端，包括：

建立发送窗口，并初始化与发送窗口覆盖的包序列号范围相对应的发送位图；发送数据分组，并根据数据分组的包序列号更新发送位图中对应分组的发送状态；接收来自接收端的快速反馈消息；根据快速反馈消息，定位发送位图中对应的分组状态项，并将未确认的对应分组状态更新为待重传状态；对待重传状态对应的数据分组执行重传。

[0020] 再一方面，还提供了一种计算机可读存储介质，其上存储有计算机程序，程序被处理器执行时，执行第一方面方法。

[0021] 上述技术方案具有如下优点或有益效果：

（1）采用与接收窗口等宽的位图对窗口内分组到达状态进行直接索引和紧凑表示，能够以比特级粒度精准定位缺失分组，克服了传统顺序确认和超时机制丢包识别粗、恢复时延高的问题。

[0022] （2）通过引入基于乱序容忍深度、持续时间、窗口停滞及路径标识的多维度乱序容忍判定机制，在短时乱序发生时维持缺口等待状态而不立即触发重传，有效区分了乱序与真实丢包，显著降低了误报率和不必要的重传开销。

[0023] （3）在多路径场景下，通过维护路径级位图并结合路径标识，能够准确判断跨路径乱序与真实丢包，避免了仅依赖全局顺序状态导致的误判。在满足快速反馈条件时，接收端生成携带缺口起始包序列号和缺口长度的精细化快速反馈消息，无需等待粗粒度超时即可通知发送端，从而将丢包恢复时延缩短至微秒级。发送端依据该反馈中的缺口起始包序列号、缺口长度或缺口区间信息，定位发送位图中对应的分组状态项；若对应分组尚未被确认，则将其由已发送未确认状态或疑似丢失状态更新为待重传状态，并执行单包重传、区间重传或策略聚合重传，并对重复、过期及已确认的重传请求进行抑制和去重，减少了冗余数据传输，提高了网络带宽利用率。

[0024] （4）对于高带宽时延积和大窗口场景，采用分段位图与变长位图相结合的扩展方式，能够动态适配链路带宽、往返时延及拥塞状态，保证了系统在大窗口条件下的可扩展性和状态维护效率。

[0025] 综上，本发明为高带宽、高并发、多路径及大窗口 RDMA 网络提供了一种高效、精准、低误判的丢包检测与快速恢复方案。

附图说明

[0026] 构成本发明的一部分的说明书附图用来提供对本发明的进一步理解，本发明的示意性实施例及其说明用于解释本发明，并不构成对本发明的不当限定。

[0027] 图 1 为本发明提供的一种基于位图的 RDMA 网络丢包检测系统总体架构图；

图 2 为本发明提供的一种基于位图的 RDMA 网络丢包检测系统执行流程图；

图 3 为本发明提供的一种接收端位图状态管理与缺口识别结构图；

图 4 为本发明提供的一种发送端反馈处理与重传协同流程图；

图 5 为本发明提供的一种多路径场景下路径级位图维护示意图；

图 6 为本发明提供的一种分段位图或变长位图在高带宽时延积场景下的状态扩展示意图。

具体实施方式

[0028] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施例作进一步详细描述。本领域技术人员应当理解，此处所描述的具体实施方式仅用于解释本发明，并不用于限定本发明。

[0029] 应该指出，以下详细说明都是例示性的，旨在对本发明提供进一步的说明。除非另有指明，本文使

用的所有技术和科学术语具有与本发明所属技术领域的普通技术人员通常理解的相同含义。

[0030] 实施例一

如图 1 所示，本实施例提供了一种基于位图的 RDMA 网络丢包检测系统，该系统包括发送端和接收端，且发送端与接收端之间通信连接。其中，发送端被配置为：建立发送窗口，并初始化与发送窗口覆盖的包序列号范围相对应的发送位图；发送数据分组，并根据数据分组的包序列号更新发送位图中对应分组的发送状态；接收来自接收端的快速反馈消息；根据快速反馈消息，定位发送位图中对应的分组状态项，并将未确认的对应分组状态更新为待重传状态；对待重传状态对应的数据分组执行重传；

接收端被配置为：建立接收窗口，并初始化与接收窗口覆盖的包序列号范围相对应的接收位图；接收数据分组，根据数据分组的包序列号更新接收位图中对应位置的置位状态；基于接收位图中未置位位置与后续已置位位置之间形成的状态缺口，识别疑似丢包分组；基于乱序容忍深度、持续时间阈值、窗口停滞阈值和路径标识中的至少一种信息，对疑似丢包分组进行乱序容忍判定；当判定为真实丢包时，生成包含缺口位置信息或区间信息的快速反馈消息并发送至发送端。

[0031] 本实施例中，系统整体采用控制管理层、发送端协议处理层、接收端协议处理层和数据转发层协同的分层架构。其中，控制管理层与发送端协议处理层、接收端协议处理层之间分别通过配置/查询通道进行交互；发送端协议处理层和接收端协议处理层之间通过 RDMA 数据流和确认/快速反馈通道进行交互；数据转发层用于承载业务数据的底层转发。

[0032] 在发送端协议处理层中设置有发送状态管理模块和重传协同模块。发送状态管理模块内部设置发送位图，用于维护发送窗口内分组的发送状态、确认状态和待重传状态；重传协同模块用于根据反馈结果触发精细化重传。

[0033] 在接收端协议处理层中设置有接收状态管理模块和反馈生成模块。接收状态管理模块内部设置接收位图，并可选择地设置路径级位图，用于维护接收窗口内分组的到达状态、缺口状态和路径相关状态；反馈生成模块用于依据位图状态生成确认信息或快速反馈信息。

[0034] 本实施例中，发送位图和接收位图为基础能力模型，路径级位图为增强能力模型。系统可根据多路径和高带宽时延积场景启用路径级位图、分段位图或变长位图。

[0035] 下面结合附图对本实施例提供的基于位图的 RDMA 网络丢包检测系统的执行流程作进一步说明，如图 2 所示，图 2 为系统执行流程图。

[0036] 具体包括，S101，初始化位图。针对目标 RDMA 连接中的队列对，发送端建立发送窗口，确定发送窗口的起始包序列号、窗口长度和滑动步长，并初始化与发送窗口覆盖的包序列号范围相对应的发送位图；同时，接收端建立接收窗口，确定接收窗口的起始包序列号、窗口长度和滑动步长，并初始化与接收窗口覆盖的包序列号范围相对应的接收位图。其中，发送位图中的每一状态位或状态项与一个或多个连续的包序列号建立映射关系，用于记录发送端侧已发送分组的确认状态、疑似丢失状态、待重传状态、已重传待确认状态或已完成状态。接收位图中的每一比特位与一个或多个连续的包序列号建立映射关系，初始状态下所有比特位均置为 0，用于记录接收端侧各数据分组的到达状态。

[0037] S102，数据分组发送与接收状态更新。在数据传输过程中，发送端根据数据分组的发送、确认反馈、快速反馈或重传处理结果，实时更新发送位图中的对应状态项；接收端则根据数据分组的到达情况更新接收位图中的对应置位状态。通过发送端和接收端分别维护各自的位图结构，为后续的状态缺口识别、乱序容忍判定、快速反馈生成、重传定位以及窗口推进建立协同基础。

[0038] 预设映射规则用于根据当前窗口基准包序列号、数据分组包序列号以及预设序列号映射粒度，计算数据分组在位图中的偏移位置。具体地，定义以下参数：BasePSN 表示当前接收窗口的起始包序列号，PSN 表示当前收到的数据分组的包序列号，G 表示映射粒度（即接收位图中一个比特位对应的连续包序列号个数），L 表示接收位图长度。接收端采用如下步骤计算分组在接收位图中的目标位置：

首先判断该 PSN 是否落在当前接收窗口内，即是否满足 BasePSN<= PSN<BasePSN +WindowSize；若满足，计算该分组相对于窗口起点的距离 Delta = PSN − BasePSN；然后根据映射粒度计算位图偏移 Offset = floor(Delta / G)；最后将该 Offset 对应的 bit 作为目标位置 RBM[Offset]。

[0039] 发送端在发送数据分组时，获取数据分组的包序列号，并根据发送窗口的窗口基准包序列号与预设映射规则将该数据分组映射至发送位图中的对应状态项。发送完成后，将对应状态项更新为“已发送未确认状态”，表示该数据分组已经发送但尚未收到确认反馈。

[0040] 接收端在接收到有效数据分组时，获取该数据分组的包序列号，并根据接收窗口的窗口基准包序列号与预设映射规则，将该数据分组映射至接收位图中的对应位置。若该位置此前未置位，则将其置位，表示该数据分组已到达；若该位置已置位，则将该数据分组标记为重复到达分组，并更新重复统计信息。重复统计信息包括重复分组计数、重复分组对应的包序列号、重复到达时间戳、所属队列对标识以及路径标识中的一种或多种。

[0041] 完成接收位图更新后，接收端进一步检测从当前接收窗口起始位置起是否存在连续完整的已置位区间，即从接收窗口基准包序列号开始的连续多个位图位置均处于已置位状态。若存在连续完整区间，则接收端将该区间作为可推进区间，推进接收窗口基准包序列号，并对已完成区间对应的历史位图状态进行释放、清零或复用，以记录后续到达的数据分组。

[0042] 相应地，当发送端接收到接收端返回的确认反馈时，根据确认反馈中的确认边界、确认区间或包序列号信息，更新发送位图中对应分组的确认状态，并推进发送窗口。

[0043] S103：状态缺口识别与疑似丢包分组判定。具体而言，接收端持续扫描当前接收窗口对应的接收位图，当发现接收位图中存在未置位位置，且在该未置位位置对应的包序列号增大方向上存在至少一个后续已置位位置时，确定该未置位位置处形成状态缺口。状态缺口可以为单个未置位位置对应的单点缺口，也可以为多个连续未置位位置组成的缺口区间。每一个状态缺口对应一个或多个疑似未到达的数据分组，并被识别为疑似丢包分组。

[0044] 如图 3 所示，接收端维护接收位图 RBM，该位图对应当前接收窗口覆盖的包序列号范围。设当前接收窗口基准包序列号为$\mathrm{BasePSN}$，接收窗口长度为L，预设映射粒度为G，则当前接收窗口覆盖的包序列号范围为$[\mathrm{BasePSN},\ \mathrm{BasePSN}+\mathrm{L}-1]$。当接收端接收到包序列号为 PSN 的数据分组时，若该 PSN 落入当前接收窗口范围内，则根据预设映射规则计算其在接收位图中的偏移位置，例如：$\mathrm{Offset}=\mathrm{floor}((\mathrm{PSN}-\mathrm{BasePSN})/\mathrm{G})$。接收端根据偏移位置更新接收位图 RBM 中对应位置的置位状态。

[0045] 若从 BasePSN 对应的位置开始，RBM 中存在连续已置位的前缀区间，则该前缀区间被认定为连续完整区间，接收端可据此推进接收窗口基准包序列号，并对已完成区间对应的位图空间进行释放、清零或复用。

[0046] 若在接收位图 RBM 中存在未置位位置，而该未置位位置之后已经存在已置位位置，则说明后续数据分组已经到达，而该未置位位置对应的数据分组尚未到达，接收端据此形成状态缺口。为便于后续处理，接收端可为状态缺口建立缺口记录项，缺口记录项包括缺口起始包序列号、缺口长度、首次发现时间、当前缺口状态、所属路径标识、乱序深度、窗口停滞信息以及已触发反馈标志中的一种或多种。

[0047] 该状态缺口用于后续乱序容忍判定和快速反馈生成。当该状态缺口被判定为真实丢包事件时，接收端根据缺口记录项生成快速反馈消息，并发送至发送端，以使发送端根据快速反馈消息定位其发送位图中对应的分组状态项，并执行后续的重传处理。

[0048] S104，为了避免将链路抖动、多路径传输或短时乱序导致的分组后到误判为真实丢包，接收端结合乱序容忍深度、缺口持续时间、窗口停滞程度和路径标识中的至少一种信息，对 S103 中识别出的疑似丢包分组进行乱序容忍判定。

[0049] 具体而言，乱序容忍深度是指在状态缺口尚未被填补的情况下，衡量该缺口被跨越程度的一种度量。例如，可采用状态缺口之后已经到达的后续分组数量作为乱序深度；或者，采用当前已接收最大包序列号与缺口起始包序列号之间的差值作为乱序深度；缺口持续时间是指从状态缺口首次被发现的时刻到当前判定时刻之间的时间差；窗口停滞程度是指接收窗口基准包序列号因状态缺口未被填补而无法继续推进的持续时间，或者在接收窗口未推进期间已经接收的后续分组数量；路径标识是指用于区分数据分组所属子路径、转发路径或等效路径的信息，用于判断该缺口是否可能由跨路径乱序引起。

[0050] 接收端根据预设最大乱序容忍深度、预设丢包判定等待时间和预设窗口停滞阈值进行判断：当乱序容忍深度未超过预设最大乱序容忍深度、缺口持续时间未超过预设丢包判定等待时间，且窗口停滞程度未超过预设窗口停滞阈值时，维持状态缺口的等待状态，不立即判定为真实丢包。当乱序容忍深度超过预设最大乱序容忍深度，或者缺口持续时间超过预设丢包判定等待时间，或者窗口停滞程度超过预设窗口停滞阈值，或者基于路径标识判断同一路径上的后续分组已经到达而状态缺口仍未被填补时，将状态缺口对应的疑似丢包分组判定为真实丢包事件。

[0051] 例如，设置最大乱序容忍深度为 8 个分组、丢包判定等待时间设置为 50 微秒、窗口停滞阈值设置为 80 微秒时。若某一缺口之后仅有 3 个后续分组到达，且该缺口持续时间为 20 微秒、窗口停滞时间为 20 微秒，则接收端维持该缺口等待状态；若该缺口之后已有 9 个后续分组到达，或者该缺口持续时间超过 50 微秒，或者窗口基准因该缺口停滞超过 80 微秒，则接收端将该疑似丢包分组升级为真实丢包事件。在多路径场景下，若后续到达分组来自不同路径，则可适当放宽等待阈值；若同一路径上的后续分组已经连续到达而该缺口仍未被填补，则可提前判定为丢包事件。

[0052] 当疑似丢包分组被判定为真实丢包事件时，接收端生成快速反馈消息。快速反馈消息至少包括缺口位置信息或缺口区间信息。快速反馈消息还可以包括队列对标识、缺口起始包序列号、缺口长度、当前接收边界、当前窗口基准包序列号、时间戳、路径标识、乱序深度、缺口状态码和统计计数中的一种或多种。

[0053] 需要说明的是，当接收位图中形成连续完整区间并推进接收窗口时，接收端生成常规确认信息，以使发送端正常推进发送窗口；而当检测到满足丢包判定条件的状态缺口时，接收端生成快速反馈消息，并将该消息发送至发送端，以使发送端根据快速反馈消息定位发送位图中的对应分组状态项并执行重传。

[0054] S105，发送端接收到接收端发送的快速反馈消息后，解析快速反馈消息中的缺口位置信息或缺口区间信息，并根据缺口起始包序列号、缺口长度或缺口区间，定位发送位图中对应的分组状态项。

[0055] 若对应分组状态项处于已发送未确认状态或疑似丢失状态，且尚未被确认，则发送端将该分组状态项更新为待重传状态，并配合重传协同模块对待重传分组进行定位和调度；若对应分组状态项已经处于已确认状态或已完成状态，则发送端忽略该快速反馈消息对应的重传请求，以避免对已经确认或已经完成的分组进行冗余重传。

[0056] 重传协同模块根据发送位图中的待重传状态执行单包重传、区间重传或策略聚合重传。其中，单包重传是指仅对单个缺失分组执行重传；区间重传是指对连续多个缺失分组执行重传；策略聚合重传是指发送端根据缺口相邻关系、反馈到达时间、重传优先级、拥塞控制状态或重传窗口限制，将多个离散缺口或多个快速反馈请求合并为一次或少数几次重传操作，以减少重复调度和重传开销。

[0057] 执行重传后，发送端可将对应分组状态由待重传状态更新为已重传待确认状态。当后续接收到接收端返回的确认反馈后，发送端根据确认边界、确认区间或包序列号信息，将发送位图中对应分组状态更新为已确认状态或已完成状态，并推进发送窗口。

[0058] 此外，发送端还进行如下协同处理：当接收到确认反馈时，根据确认边界或确认区间，将发送位图中对应分组状态由已发送未确认状态或已重传待确认状态更新为已确认状态或已完成状态，并释放或复用相应窗口状态；当检测到分组确认超时或发送窗口长时间未推进时，可将对应的已发送未确认状态更新为疑似丢失状态或待重传状态，以触发超时重传。

[0059] 同时，发送端对重复反馈、过期反馈、已确认分组对应的重传请求以及设定时间窗口内针对同一缺口的重复触发请求进行抑制和去重；在发送端处于拥塞控制约束状态时，发送端将重传分组与正常发送分组统一纳入当前队列对或业务流的发送速率、发送窗口、拥塞窗口或令牌桶约束中，使重传流量与正常发送流量共同受拥塞控制策略限制，避免重传流量破坏已有拥塞控制机制。

[0060] 可以理解地，系统循环执行 S102 至 S105，从而实现连续的丢包检测、快速反馈、精准重传和窗口推进。

[0061] 如图 4 所示，当接收端识别到某缺口满足快速反馈条件后生成快速反馈消息（FFM），该快速反馈消息可以携带以下字段中的一种或多种：队列对（QP）标识、当前接收边界、缺口起始包序列号（PSN）、缺口长度、缺口状态码、时间戳、路径标识以及已累计重复次数或乱序深度中的一种或多种。

[0062] 发送端在收到快速反馈消息后，检索发送位图中对应的分组状态。若缺口对应分组尚未被确认且未处于已完成重传状态，则将其置为待重传状态，并执行重传动作。为避免重传泛滥，可采用如下控制策略：

（1）对已经收到确认信息的分组不执行重传；

（2）对同一缺口在设定时间窗口内仅允许一次有效重传触发；

（3）对多个相邻缺口进行合并处理，采用区间重传；

（4）在拥塞控制生效时，将重传流量纳入整体发送速率约束中，避免重传流量破坏已有拥塞控制策略。

其中，“拥塞控制生效时”是指目标 RDMA 连接已启用拥塞控制或速率控制机制，且发送端因接收到拥塞指示、快速反馈、确认停滞、重传超时，或因发送速率、发送窗口被限制，而处于受拥塞控制约束的发送状态。具体实现上，发送端将重传分组与正常发送分组统一纳入当前 QP 或业务流的发送速率、发送窗口、拥塞窗口约束中，使重传流量与正常发送流量共同受拥塞控制机制限制。

[0063] 通过上述机制，系统能够根据快速反馈、确认停滞、定时器事件或管理面触发结果，执行单包重传、区间重传或策略聚合重传，并对重复反馈、过期反馈和已确认分组对应的重传请求进行抑制和去重，从而实现精细化重传协同。

[0064] 可以理解地，在后续数据分组持续传输过程中，重复执行上述步骤，以实现连续的位图更新、缺口识别、反馈生成及重传处理。

[0065] 对于不同的网络场景，本发明提供了相应的优化实施方式。

[0066] 在单路径场景中，接收端采用全局位图模式统一维护窗口内所有分组状态；

如图 5 所示，在多路径 RDMA 场景中，同一逻辑业务流的数据分组可能经由多个不同子路径转发到达接收端。由于不同子路径上的传播时延、排队时延和拥塞状态不同，分组到达顺序可能发生改变。为此，本实施例进一步提出路径级位图机制：针对每一条子路径维护一个对应的路径级接收位图PBM_k，各路径级位图与相同的逻辑包序列号空间对应，但分别记录不同路径上的分组到达状态。

[0067] 例如，当接收端收到带有路径标识PathID的数据分组时，不仅更新全局位图 GBM，还更新与PathID对应的路径级位图PBM_PathID。若某一分组在全局位图中形成缺口，但该缺口的相邻后续分

组来自其他路径，则优先判定为跨路径乱序，而非立即认定为丢失。在路径级位图能力不可用、路径标识缺失或设备未启用该功能时，可退化为仅使用全局位图模式执行检测。

[0068] 通过上述路径级位图机制，本发明有效区分了多路径传输中的跨路径乱序与真实丢包，进一步降低了误判率。

[0069] 如图 6 所示，在高带宽时延积场景中，由于发送窗口和接收窗口可能覆盖大量包序列号，若采用固定长度的单片位图，容易造成位图过长、访问延迟增加以及状态存储占用过大等问题。为此，本实施例提出采用分段位图、变长位图或两者结合的方式进行扩展，以提升窗口覆盖能力和状态维护扩展性。

[0070] 分段位图将一个逻辑大窗口划分为多个连续片段，每个片段对应一个子位图，数据分组根据其包序列号先定位所属片段，再定位片段内偏移位置，从而降低单次访问复杂度，便于硬件流水线并行处理。变长位图则根据往返时延、链路带宽、带宽时延积、拥塞状态或业务负载动态调整位图长度：当网络负载较低或窗口较小时，可采用较短位图以节约资源；当处于高带宽时延积场景时，可在基础位图长度的基础上增加扩展位图长度ΔL，以维持足够的检测覆盖范围。对于超大窗口场景，还可先根据窗口总长度动态调整片段数量，再在各片段内部维持固定或自适应长度位图结构，从而实现更好的扩展性。

[0071] 在一些实施方式中，系统还支持管理与监测能力，用于运维部署和策略控制。可配置参数包括但不限于：基础位图长度、扩展位图长度ΔL、位图片段大小、窗口起始包序列号 PSN、窗口滑动步长、最大乱序容忍深度、丢包判定等待时间、快速反馈触发阈值、单队列 QP 状态上限、单设备最大并发队列 QP 数量以及路径级位图实例数。

[0072] 同时可提供统计信息：例如当前启用的位图模式、当前基础位图长度、当前扩展位图长度ΔL、当前实际位图长度、分段数量、丢包事件次数、策略聚合重传次数、重传次数、重复分组次数、误判修正次数、不同路径的统计信息以及资源占用情况等。

[0073] 此外，作为本系统的可选增强功能，还可设置管理与监测模块，该模块位于控制管理层，分别与发送端协议处理层和接收端协议处理层双向通信，用于完成参数配置、状态查询、能力发现、事件记录、统计导出和告警管理。具体地，管理与监测模块支持对位图长度、位图片段大小、窗口起始包序列号、窗口滑动步长、快速反馈触发阈值、丢包判定等待时间、最大乱序容忍深度、路径级位图实例数、单队列对状态上限以及单设备最大并发队列对数量等参数进行配置和查询。同时，该模块可提供统计信息，包括当前启用的位图模式、当前位图长度和分段数量、丢包事件次数、策略聚合重传次数、重传次数、重复分组次数、误判修正次数、不同路径的统计信息以及资源占用情况等。

[0074] 需要说明的是，该管理与监测模块并非实现丢包检测与重传所必需的核心模块，其功能主要用于系统运维和策略调优，在实际部署中可根据需求选择是否启用。

[0075] 一种示例性应用场景。

[0076] 为了使本发明的技术方案更加清楚，下面以 RDMA 可靠连接数据传输场景为例进行说明，但本发明不限于此。

[0077] 在该场景中，发送端按目标队列对（QP）建立发送窗口，并初始化发送位图。该发送位图用于记录已发送分组的确认状态、疑似丢失状态、重传状态或待处理状态，发送端通过发送位图维护已发送分组、已确认分组、疑似丢失分组、待重传分组以及已完成分组等信息。接收端按同一 QP 建立接收窗口，并初始化接收位图，用于记录各数据分组的到达状态。

[0078] 当发送端连续发送$PSN = 1000$至$PSN = 1015$的数据分组后，将发送位图中对应状态项更新为已发送未确认状态。接收端收到$PSN = 1000$、1001、1002、1004、1005后，将接收位图中对应位置置位，并识别出$PSN = 1003$对应位置未置位而后续位置已置位，从而形成状态缺口。若该缺口在乱序容忍判定后被升级为丢包事件，接收端生成包含缺口起始$PSN = 1003$和缺口长度的快速反馈消息。

[0079] 发送端接收到该快速反馈消息后，根据缺口位置或区间信息检索发送位图中对应的状态项。若$PSN = 1003$的状态为已发送未确认状态或疑似丢失状态，且尚未被确认，则将其更新为待重传状态，并配合重传协同模块对该待处理分组进行定位和单包重传；重传完成后，将其更新为已重传待确认状态。接收端收到重传的$PSN = 1003$后，将接收位图中对应位置置位，并在形成连续完整区间后推进接收窗口，同时向发送端返回确认反馈。发送端收到确认反馈后，将发送位图中对应分组状态更新为已确认状态或已完成状态，并推进发送窗口。

[0080] 通过上述过程，接收位图负责记录到达状态、识别状态缺口并触发快速反馈，发送位图负责记录确认状态、疑似丢失状态、重传状态和待处理状态，并根据确认反馈或快速反馈更新状态，与重传协同模块共同完成待处理分组定位、精准重传和窗口推进。

[0081] 实施例二

本实施例提供了一种基于位图的 RDMA 网络丢包检测方法，应用于接收端，包括：

建立接收窗口，并初始化与接收窗口覆盖的包序列号范围相对应的接收位图；接收数据分组，根据数据分组的包序列号更新接收位图中对应位置的置位状态；基于接收位图中未置位位置与后续已置位位置之间形成的状态缺口，识别疑似丢包分组；基于乱序容忍深度、持续时间阈值、窗口停滞阈值和路径标识中的至少一种信息，对疑似丢包分组进行乱序容忍判定；当判定为真实丢包时，生成包含缺口位置信息或区间信息的快速反馈消息并发送至发送端。

[0082] 实施例三

本实施例提供了一种基于位图的 RDMA 网络丢包检测方法，应用于发送端，其特征在于，包括：

建立发送窗口，并初始化与发送窗口覆盖的包序列号范围相对应的发送位图；发送数据分组，并根据数据分组的包序列号更新发送位图中对应分组的发送状态；接收来自接收端的快速反馈消息；根据快速反馈消息，定位发送位图中对应的分组状态项，并将未确认的对应分组状态更新为待重传状态；对待重传状态对应的数据分组执行重传。

[0083] 实施例四

本实施例还提供了一种计算机可读存储介质，用于存储计算机指令，计算机指令被处理器执行时，完成实施例一的方法。

[0084] 以上仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

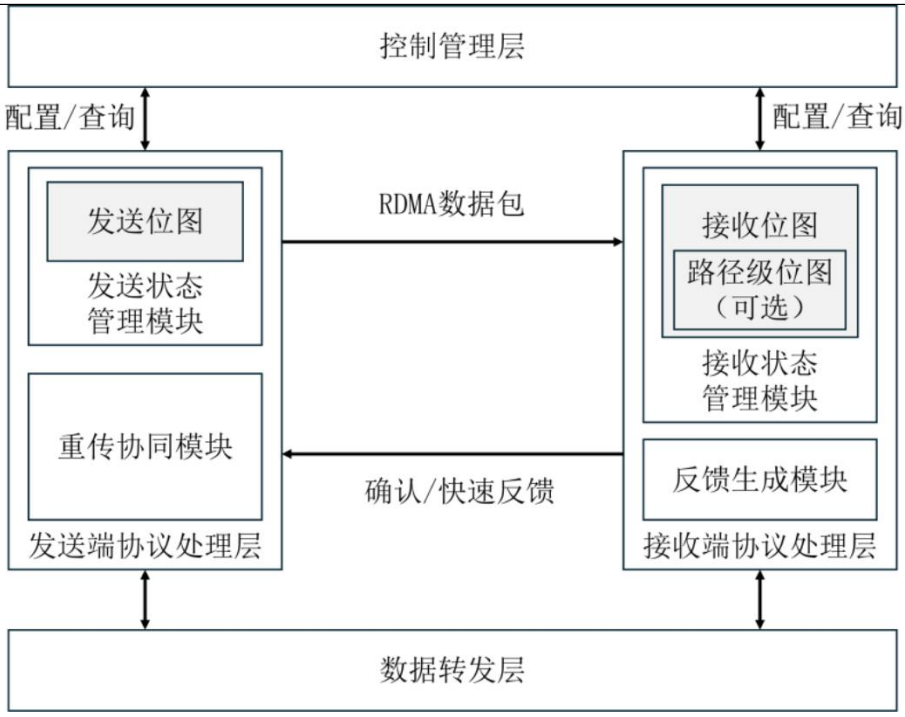

图 1

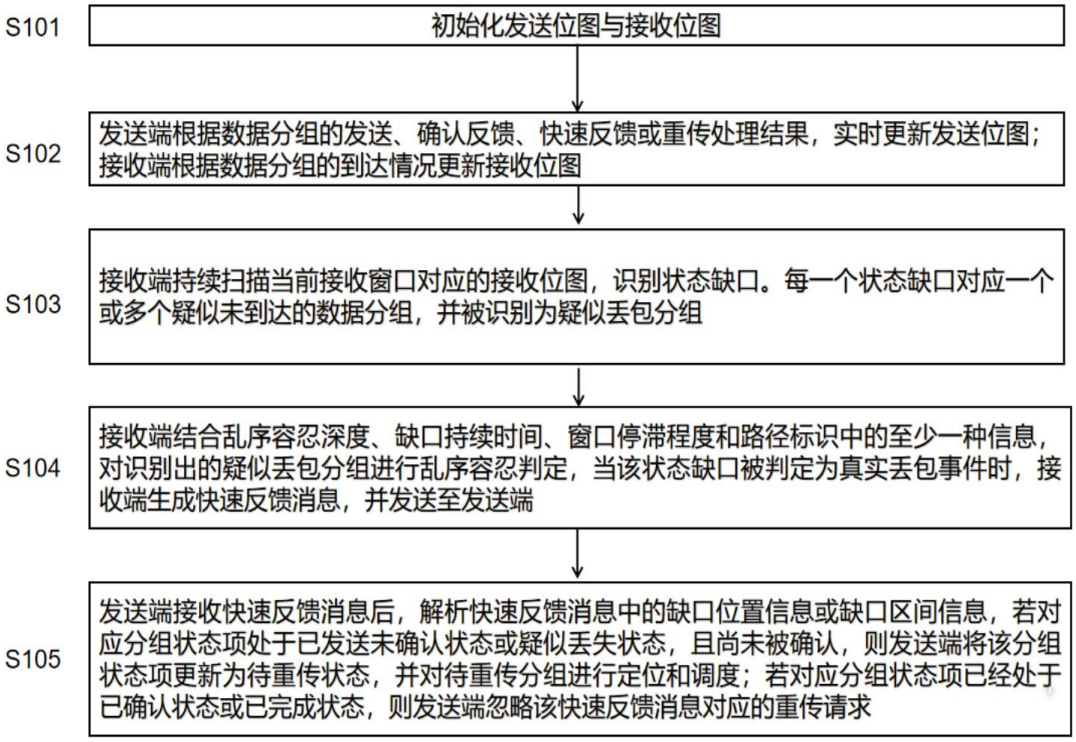

图 2

100003
2023.03

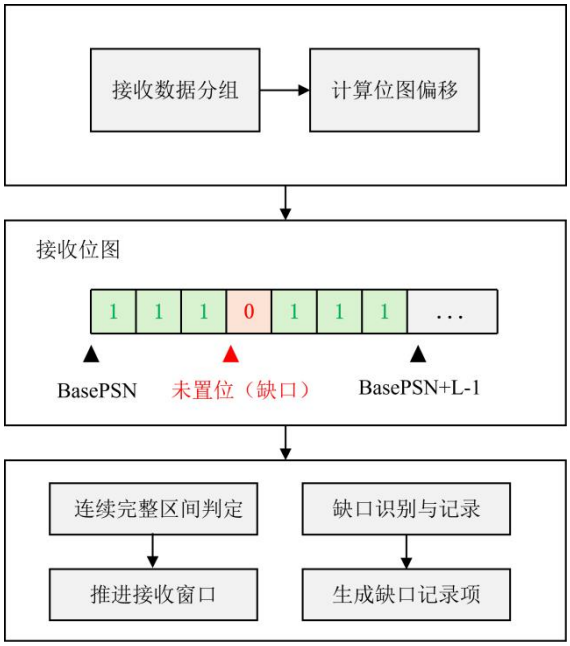

图 3

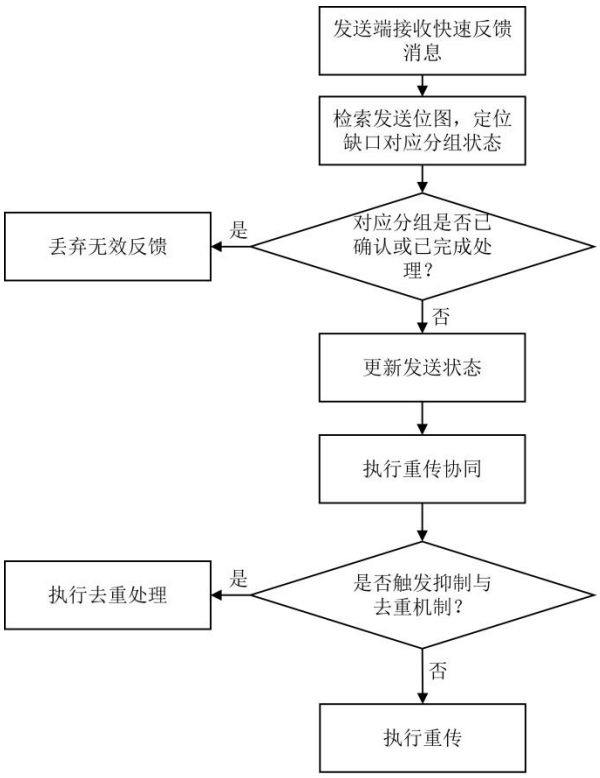

图 4

100003
2023.03

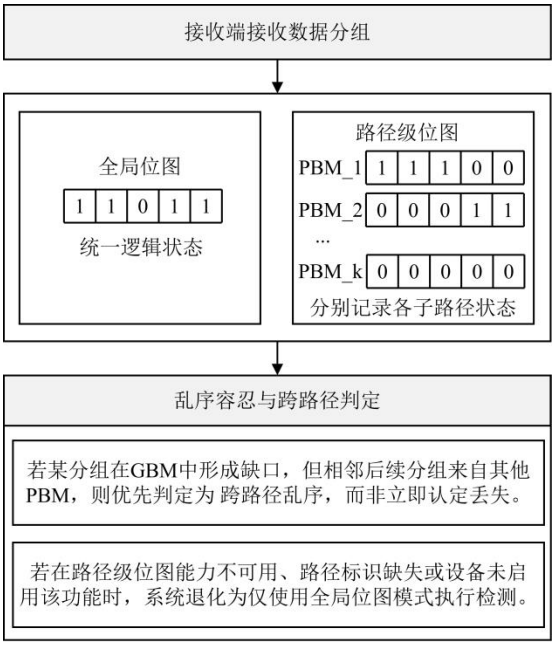

图 5

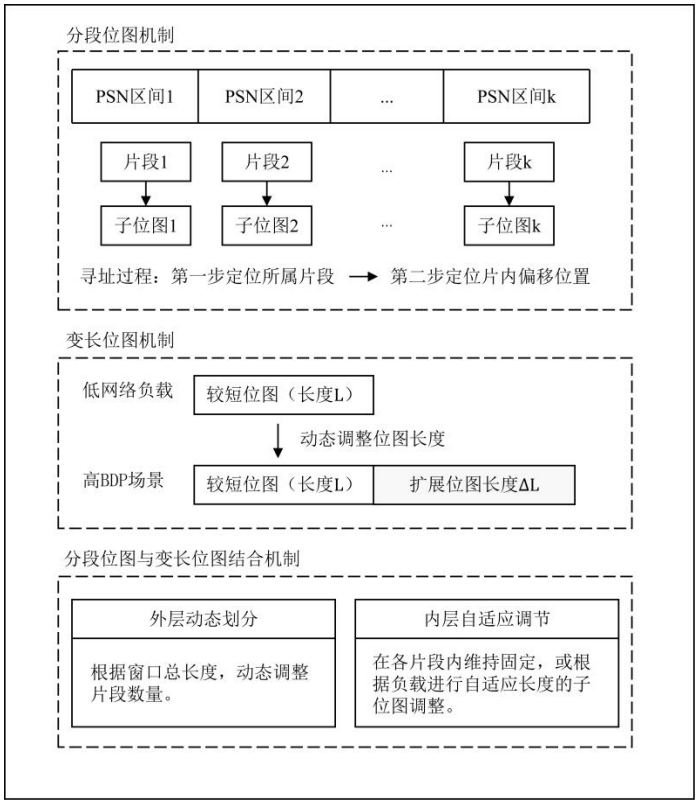

图 6

100003
2023.03

说 明 书 摘 要

本发明公开了一种基于位图的 RDMA 网络丢包检测系统及方法，属于远程直接内存访问网络传输领域。为解决丢包识别粒度粗、乱序误判率高、反馈与重传协同效率不足的问题，本发明在发送端建立发送位图，在接收端建立接收位图；接收端根据接收位图中未置位与后续已置位形成的状态缺口识别疑似丢包分组，结合乱序容忍深度、持续时间、窗口停滞及路径标识进行判定，区分乱序与真实丢包；判定为丢包时生成包含缺口位置或区间信息的快速反馈消息并发送至发送端；发送端依据反馈消息定位发送位图中对应分组并执行单包、区间或策略聚合重传。本发明适用于高带宽、多路径及大窗口 RDMA 网络，可提高丢包检测精度，降低误判率，缩短恢复时延，减少冗余重传。

100004
2023.03