

(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
2024 年 9 月 26 日 (26.09.2024)



(10) 国际公布号
WO 2024/192557 A1

(51) 国际专利分类号:

G06F 15/173 (2006.01)

(21) 国际申请号:

PCT/CN2023/082125

(22) 国际申请日:

2023 年 3 月 17 日 (17.03.2023)

(25) 申请语言:

中文

(26) 公布语言:

中文

(71) 申请人:清华大学(TSINGHUA UNIVERSITY) [CN/CN]; 中国北京市海淀区清华园 1 号, Beijing 100084 (CN)。

(72) 发明人:舒继武(SHU, Jiwu); 中国北京市海淀区清华园 1 号, Beijing 100084 (CN)。 颜彬(YAN, Bin); 中国北京市海淀区清华园 1 号, Beijing 100084 (CN)。 陆游游(LU, Youyou); 中国北京市海淀区清华园 1 号, Beijing 100084 (CN)。

(74) 代理人:北京东方亿思知识产权代理有限公司等(BEIJING EAST IP LTD. et al.); 中国北京

市东城区东长安街 1 号东方广场东方经贸城东 2 座 1601 室, Beijing 100738 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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,

(54) **Title:** REMOTE MEMORY ACCESS PROTECTION MECHANISM CONSTRUCTION METHOD, REMOTE MEMORY NODE, AND DEVICE

(54) 发明名称: 远端内存访问保护机制构建方法、远端内存节点及设备

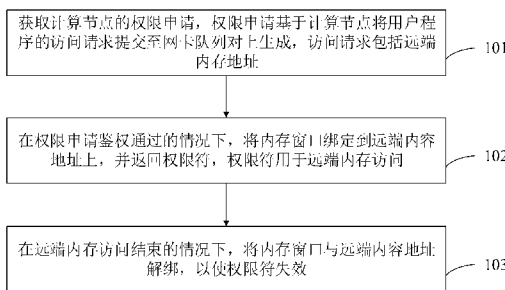


图 1

101 Obtain a permission application of a computing node, wherein the permission application is generated by submitting an access request of a user program to a network adapter queue pair by the computing node, and the access request comprises a remote memory address
102 When authentication on the permission application succeeds, bind a memory window to a remote content address, and return a permission character, wherein the permission character is used for remote memory access
103 When the remote memory access ends, unbind the memory window and the remote content address to invalidate the permission character

(57) **Abstract:** The present application provides a remote memory access protection mechanism construction method, a remote memory node, and a device. The method comprises: obtaining a permission application of a computing node, wherein the permission application is generated by submitting an access request of a user program to a network adapter queue pair by the computing node, and the access request comprises a remote memory address; when authentication on the permission application succeeds, binding a memory window to a remote content address, and returning a permission character, wherein the permission character is used for remote memory access; and when the remote memory access ends, unbinding the memory window and the remote content address to invalidate the permission character.

(57) 摘要: 本申请提供了一种远端内存访问保护机制构建方法、远端内存节点及设备, 方法包括: 获取计算节点的权限申请, 权限申请基于计算节点将用户程序的访问请求提交至网卡队列对上生成, 访问请求包括远端内存地址; 在权限申请鉴权通过的情况下, 将内存窗口绑定到远端内容地址上, 并返回权限符, 权限符用于远端内存访问; 在远端内存访问结束的情况下, 将内存窗口与远端内容地址解绑, 以使权限符失效。

HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO,
PL, PT, 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))。

远端内存访问保护机制构建方法、远端内存节点及设备

5 技术领域

本申请属于远端内存访问技术领域，尤其涉及一种远端内存访问保护机制构建方法、远端内存节点及设备。

背景技术

10 远端内存直接访问机制（Remote Direct Memory Access, RDMA）是一种高速网络技术，其具有数百 Gbps 级别的高带宽和数微秒级的低延迟。因其性能，RDMA 被各大企业广泛地运用在数据中心内。基于信息安全的需求，RDMA 往往需要通过一些保护机制来解决远端数据损坏、远端数据隐私泄露等问题，然而现有的保护机制操作均具有十分受限的性能，吞吐和
15 延迟往往远劣于 RDMA 的数据路径，导致无法利用其提供运行期的远端内存访问保护。

因此，有必要提供一种隔离非法远程访问，从而避免远端数据损坏、远端数据隐私泄露的远端内存访问保护机制构建方法。

20 发明内容

本申请实施例提供一种远端内存访问保护机制构建方法、远端内存节点及设备，能够隔离非法远程访问，从而避免远端数据损坏、远端数据隐私泄露。

第一方面，本申请实施例提供一种远端内存访问保护机制构建方法，
25 应用于远端内存架构中的远端内存节点，远端内存架构还包括计算节点，方法包括：

获取计算节点的权限申请，权限申请基于计算节点将用户程序的访问请求提交至网卡队列对上生成，访问请求包括远端内存地址；

在权限申请鉴权通过的情况下，将内存窗口绑定到远端内容地址上，

并返回权限符，权限符用于远端内存访问；

在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

5 在一些实施例中，在权限申请鉴权通过的情况下，将内存窗口绑定到远端内容地址上，并返回权限符，包括：

在权限申请鉴权通过的情况下，获取正在进行的远端内存访问对应的历史权限符；

确定权限申请与历史权限符是否存在冲突；

10 在权限申请与历史权限符不存在冲突的情况下，将内存窗口绑定到远端内容地址上，并返回权限符。

在一些实施例中，在权限申请鉴权通过的情况下，将内存窗口绑定到远端内容地址上，并返回权限符之后，方法还包括：

将权限符存储至哈希表中；

15 在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效之后，方法还包括：

将权限符从哈希表中删除。

在一些实施例中，权限符包括租约语义，租约语义对应有租期，在权限申请鉴权通过的情况下，将内存窗口绑定到远端内容地址上，并返回权限符之后，方法还包括：

20 在远端内存访问的响应时长超过租期的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

在一些实施例中，租期基于用户程序采用远端内存单边访问的方式进行更新。

25 在一些实施例中，网卡队列对具有备用队列对，其中在网卡队列对失效的情况下，备用队列对替代网卡队列对。

在一些实施例中，在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效，包括：

在远端内存访问结束的情况下，获取远端内存架构的负载压力和未绑定的内存窗口的数量；

在负载压力小于预设压力阈值，或数量小于预设数量阈值的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

在一些实施例中，方法还包括：

将内存窗口的一次绑定操作和一次解绑操作合并为一次换绑操作；

- 5 在多个待绑定的远端内存地址为连续地址的情况下，将内存窗口的多次绑定操作替换为一次批量绑定操作。

在一些实施例中，远端内存架构包括多个计算节点和多个远端内存节点。

- 10 第二方面，本申请实施例提供了一种远端内存节点，所属于远端内存架构，远端内存架构还包括计算节点，远端内存节点包括：

获取模块，用于获取计算节点的权限申请，权限申请基于计算节点将用户程序的访问请求提交至网卡队列对上生成，访问请求包括远端内存地址；

- 15 绑定模块，用于在权限申请鉴权通过的情况下，将内存窗口绑定到远端内存内容地址上，并返回权限符，权限符用于远端内存访问；

解绑模块，用于在远端内存访问结束的情况下，将内存窗口与远端内存内容地址解绑，以使权限符失效。

第三方面，本申请实施例提供了一种远端内存访问保护机制构建设备，设备包括：

- 20 处理器以及存储有程序或指令的存储器；
处理器执行程序或指令时实现上述的方法。

第四方面，本申请实施例提供了一种机器可读存储介质，机器可读存储介质上存储有程序或指令，程序或指令被处理器执行时实现上述的方法。

- 25 第五方面，本申请实施例提供了一种计算机程序产品，计算机程序产品中的指令由电子设备的处理器执行时，使得电子设备执行上述方法。

本申请实施例的远端内存访问保护机制构建方法、远端内存节点及设备，能够获取计算节点的权限申请，权限申请基于计算节点将用户程序的访问请求提交至网卡队列对上生成，访问请求包括远端内存地址；在权限申请鉴权通过的情况下，将内存窗口绑定到远端内存内容地址上，并返回权限

符，权限符用于远端内存访问；在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。这样，能够利用权限机制有效隔离了来自用户程序的非法远程访问，且在访问结束后权限符即失效，有效避免远端内存架构潜在面临的远端数据损坏、远端数据隐私泄露等问题。

5

附图说明

为了更清楚地说明本申请实施例的技术方案，下面将对本申请实施例中所需要使用的附图作简单的介绍，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

10 图 1 是本申请一个实施例提供的远端内存访问保护机制构建方法的流程示意图；

图 2 是本申请实施例提供的远端内存架构的示意图；

图 3 是本申请实施例提供的远端内存访问保护机制构建方法的远端内存布局和数据管理格式示意图；

15 图 4 是本申请实施例提供的远端内存访问保护机制构建方法的用户使用流程图；

图 5 是本申请实施例提供的远端内存访问保护机制构建方法的换绑操作的示意图；

图 6 是本申请实施例提供的远端内存访问保护机制构建方法的批量绑定操作及批量解绑操作的示意图；

20 图 7 是本申请另一个实施例提供的远端内存节点的结构示意图；

图 8 是本申请又一个实施例提供的电子设备的结构示意图。

具体实施方式

25 下面将详细描述本申请的各个方面的特征和示例性实施例，为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及具体实施例，对本申请进行进一步详细描述。应理解，此处所描述的具体实施例仅意在解释本申请，而不是限定本申请。对于本领域技术人员来说，本申请可以在不需要这些具体细节中的一些细节的情况下实施。下面对实施例的描述

仅仅是为了通过示出本申请的示例来提供对本申请更好的理解。

需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还

5 “包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还

10 存在另外的相同要素。

为了解决现有技术问题，本申请实施例提供了一种远端内存访问保护机制构建方法、远端内存节点及设备。下面首先对本申请实施例所提供的远端内存访问保护机制构建方法进行介绍。

图 1 示出了本申请一个实施例提供的远端内存访问保护机制构建方法的流程图。如图 1 所示，远端内存访问保护机制构建方法可以应用于远端内存架构中的远端内存节点，远端内存架构还可以包括计算节点，远端内存访问保护机制构建方法可以包括如下步骤：

15 的流程示意图。如图 1 所示，远端内存访问保护机制构建方法可以应用于远端内存架构中的远端内存节点，远端内存架构还可以包括计算节点，远端内存访问保护机制构建方法可以包括如下步骤：

步骤 101，获取计算节点的权限申请，权限申请基于计算节点将用户程序的访问请求提交至网卡队列对上生成，访问请求包括远端内存地址。

在步骤 101 中，用户在访问远端内存前，可以通过远端过程调用的方式，按需地向远端内存请求访问权限。即用户可以通过用户程序可以向远端内存架构中的计算节点提出访问请求，访问请求中包括了用户想要访问的远端内存地址。计算节点接收到该访问请求后可以提交至网卡队列对张，生成对应的权限申请，以便后续远端内存节点可以从网卡队列对中获取该

20 在步骤 101 中，用户在访问远端内存前，可以通过远端过程调用的方式，按需地向远端内存请求访问权限。即用户可以通过用户程序可以向远端内存架构中的计算节点提出访问请求，访问请求中包括了用户想要访问的远端内存地址。计算节点接收到该访问请求后可以提交至网卡队列对张，生成对应的权限申请，以便后续远端内存节点可以从网卡队列对中获取该

25 权限申请。

步骤 102，在权限申请鉴权通过的情况下，将内存窗口绑定到远端内容地址上，并返回权限符，权限符用于远端内存访问。

在步骤 102 中，可以理解的是，权限申请中可以携带用户程序的鉴权信息，例如用户 ID、用户密码等，远端内存节点可以基于这些鉴权信息对

该权限申请中的进行鉴权，判断该用户程序是否可以远端内存访问，若可以则说明鉴权通过，此时可以将内存窗口（Memory Window，MW）绑定到远端内容地址上，并返回权限符，权限符可以用于远端内存访问。换言之，无权限的远端访问将不被执行。

5 步骤 103，在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

在步骤 103 中，若远端内存访问结束，则可以将内存窗口与远端内容地址解绑，以使权限符失效。这样，由于权限符失效，导致无法继续执行远端内存访问，从而保护了远端数据的隐私。

10 在一些示例中，如图 2 所示，远端内存架构可以包括计算节点和远端内存节点，其中远端内存节点可以分为权限头资源池和远端内存池两部分。如图 2 序号 1 所示，当用户申请访问权限时（即远端内存节点可以获取权限申请），远端内存节点从权限头资源池分配一个权限头，从远端内存池分配一块远端内存，将两个 MW 分别绑定到权限头和远端内存上，并将两个
15 权限符返回给用户。当用户撤销访问权限（即远端内存访问结束）时，远端内存节点将对应的两个 MW 解绑，使得权限符失效。

可以理解的是，远端内存可以使用 slab 分配器进行管理，远端内存被划分进多个固定大小的块组中，用户分配和使用远端内存时，可以根据用户需求，从满足用户需要的最小非空块组中选择一块远端内存实现分配。

20 如图 3 所示，远端内存节点的主要部分划分为远端内存池，少数部分划分为权限头资源池，权限头可以包括开放权限的内存地址与长度、所使用的 MW 的位置、权限开启的时间戳和权限的租期等。

本申请实施例的远端内存访问保护机制构建方法能够获取计算节点的权限申请，权限申请基于计算节点将用户程序的访问请求提交至网卡队列
25 对上生成，访问请求包括远端内存地址；在权限申请鉴权通过的情况下，将内存窗口绑定到远端内容地址上，并返回权限符，权限符用于远端内存访问；在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。这样，能够利用权限机制有效隔离了来自用户程序的非法远程访问，且在访问结束后权限符即失效，有效避免远端内存架构潜在

面临的远端数据损坏、远端数据隐私泄露等问题。

在一些实施例中，步骤 102 可以包括如下步骤：

在权限申请鉴权通过的情况下，获取正在进行的远端内存访问对应的历史权限符；

5 确定权限申请与历史权限符是否存在冲突；

在权限申请与历史权限符不存在冲突的情况下，将内存窗口绑定到远端内容地址上，并返回权限符。

在本实施例中，在权限申请鉴权通过的情况下，可以获取正在进行的远端内存访问对应的历史权限符。换言之，历史权限符可以是指正在有效的权限符，可以根据历史权限符，判断该权限申请是否与正在进行的远端内存访问存在冲突。

例如，若历史权限符指示权限申请对应的远端内存地址正在进行私有的远端内存访问，则可以认为此时权限申请与历史权限符存在冲突，可以在冲突被化解后（即私有的远端内存访问结束后），再将内存窗口绑定到远端内容地址上，并返回权限符。

若权限申请与历史权限符不存在冲突，则可以直接将内存窗口绑定到远端内容地址上，并返回权限符。

在一些示例中，如图 4 所示，其中图 4 中的（a）部分体现了一种不冲突的情况，用户 1 向远端内存节点申请共享的访问权限，正在进行共享的远端内存访问，在远端内存访问过程中，用户 1 可以向远端内存节点拓展权限的租期，其中拓展权限租期的操作可发生多次。此时用户 1 向远端内存发起远端写和远端读，并附带权限符，若此时权限符尚未过期，则远端写和远端读操作可成功执行，若此时权限符已过期，则远端写操作和远端读操作不被执行。在此过程中，用户 2 可申请相同的共享权限，多个指向相同远端内存的共享权限可以同时存在。

其中图 4 中的（b）部分体现了一种存在冲突的情况，用户 1 向远端内存节点申请私有的访问权限，正在进行私有的远端内存访问，此时用户 2 向远端内存节点申请访问相同远端内存的权限，因权限冲突，远端内存节点延迟给予用户 2 的访问权限，直到冲突被化解。

可以理解的是，若用户 1 的用户程序崩溃后，待用户 1 的权限符的租约超时，则用户 1 的私有访问权限被强制且自动地回收，权限冲突被化解，此时远端内存节点给予用户 2 访问权限，即返回权限符给用户 2。

5 在一些实施例中，步骤 102 之后，远端内存访问保护机制构建方法还可以包括如下步骤：

将权限符存储至哈希表中；

在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效之后，方法还包括：

将权限符从哈希表中删除。

10 在本实施例中，权限符可以使用哈希表进行管理，即在返回权限符后可以将权限符存储至哈希表中，在权限符失效后，可以将权限符从哈希表中删除。

其中，哈希表的键可以为用户申请访问的远端内存地址，值为权限管理所需元数据，元数据可以包括远端内存的地址、MW 的位置、权限开启
15 时间和权限租期等。

这样，通过哈希表来管理权限符，可以快速检测存在冲突的权限申请，以便提高远端内存访问的运行效率。

20 在一些示例中，如图 2 序号 2 所示，权限符可以以哈希表的方式进行管理，权限符所覆盖的远端内存需要与 slab 分配器块组中的块对齐，权限符被存储在哈希表中，哈希表的键为用户申请访问的远端内存地址，值为权限管理所需元数据，即权限头，其中权限头可以包括远端内存的地址、MW 的位置、权限开启时间和权限租期等。

25 远端内存节点可以通过哈希表检测冲突的权限申请，远端内存节点在处理用户的权限申请请求时，以请求中所申请的远端内存地址为键查询哈希表，若检测到权限申请与哈希表中的历史权限符发生访问冲突，则权限申请会被延迟，直到冲突被化解。

在一些实施例中，权限符包括租约语义，租约语义对应有租期，步骤 102 之后，远端内存访问保护机制构建方法还可以包括如下步骤：

在远端内存访问的响应时长超过租期的情况下，将内存窗口与远端内

容地址解绑，以使权限符失效。

在本实施例中，权限符可以包括租约语义，租约语义对应有租期。在用户程序被挂起或发生崩溃的情况，远端内存访问的响应时长往往会超时，此时可以利用租约语义释放残留权限，即将内存窗口与远端内容地址解绑，
5 以使权限符失效。使用了租约机制，可以解决用户程序崩溃或被挂起所造成的权限资源泄露和系统阻塞问题。

在一些实施例中，租期基于用户程序采用远端内存单边访问的方式进行更新。

在本实施例中，租约语义可以支持拓展租期，租期基于用户程序采用远端内存单边访问的方式进行更新。示例地，将权限头上的权限租期字段通过 MW 暴露给用户访问，由用户使用远端内存单边访问的比较交换原语
10 (Compare and Swap, CAS) 完成对权限租期变量的更新。相应地，远端内存节点可以周期性地访问哈希表，检查其中权限符的权限租期变量，废除所有已超过租期的权限符。其中权限符的废除方式为解绑相关的 MW，使
15 得相关权限符失效。

在一些示例中，为避免租期反复拓展导致的共享权限与私有权限之间的公平性问题，远端内存节点在检测到一块远端内存长期活跃地处于共享或私有访问状态时，可以拒绝对该权限符的进一步租期拓展。示例地，远端内存节点可以把权限头的权限租期字段置零，用户拓展该权限的租期时，
20 比较交换原语失败并返回零值，用户由此得知权限拓展被禁止。同时，远端内存节点可以在权限头的开启时间字段中保留一个比特，将该比特置一，以标记权限符应在本轮超时后被废除。

在一些示例中，为避免租期被反复拓展而引发资源占有问题，任何权限符无法保持有效超过最大存活期限，最大存活期限可以根据实际需求进行设定，此处不作具体限定。可以使用权限头存储的权限开启时间进行判
25 别，超过最大存活期限的权限符视为超时而被废除。

在一些实施例中，网卡队列对具有备用队列对，其中在网卡队列对失效的情况下，备用队列对替代网卡队列对。

在本实施例中，远端内存访问时所需的网卡队列对可以具有若干份备

用队列对。在因为用户的非法远端访问而导致网卡队列对失效时，备用队列对可以代替失效的网卡队列对，从而快速恢复用户的远端内存访问能力。这样，可以通过使用少量冗余的备份队列对，最小化非法的远端访问引起队列对失效的额外恢复时间。

5 在一些示例中，如图 2 序号 3 所示，任何带有非法或已失效的权限符的远端内存访问不会生效。示例地，即使非法权限申请被下发至网卡队列对中，该非法权限申请附带的非法权限符会被远端内存节点侧的网卡检测，该远端内存访问不会被远端内存节点的网卡执行。

10 在一些示例中，如图 2 序号 4 所示，网卡队列对可以具有若干份冗余的备用队列对，以在网卡队列对失效时完成备用队列对的切换。远端内存架构内的每个用户程序会被分配一个固定的虚拟队列对编号，虚拟队列对编号通过虚实映射表映射到一个实际的网卡队列对上，用户程序的所有权限申请会依据映射表发送到实际的网卡队列对中，非法的远端内存访问会使所在的网卡队列对失效。此时备用队列对将替换掉失效的网卡队列对，
15 虚实映射表会发生更改，将虚拟队列编号重映射到一个有效的备用队列对上，以使远端内存访问能力能够快速恢复。

在一些实施例，步骤 103 可以包括如下步骤：

在远端内存访问结束的情况下，获取远端内存架构的负载压力和未绑定的内存窗口的数量；

20 在负载压力小于预设压力阈值，或数量小于预设数量阈值的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

在本实施例中，在远端内存访问结束的情况下，可以获取远端内存架构的负载压力和未绑定的内存窗口的数量，若负载压力小于预设压力阈值，或数量小于预设数量阈值，则可以将内存窗口与远端内容地址解绑，以使
25 权限符失效。若负载压力大于或等于预设压力阈值，且未绑定的内存窗口的数量大于或等于预设数量阈值，则可以延迟执行解绑操作。其中预设压力阈值和预设数量阈值可以根据实际需求进行设定，此处不作具体限定。

换言之，处理内存窗口的解绑操作时，若远端内存框架面临高负载压力，则可以将内存窗口解绑操作延迟执行，直至负载压力缓解或可用内

存窗口数低于临界值。

在一些实施例中，远端内存访问保护机制构建方法还可以包括如下步骤：

将内存窗口的一次绑定操作和一次解绑操作合并为一次换绑操作；

5 在多个待绑定的远端内存地址为连续地址的情况下，将内存窗口的多次绑定操作替换为一次批量绑定操作。

在本实施例中，可以将内存窗口的一次绑定操作和一次解绑操作合并为一次换绑操作。示例地，如图 5 所示，在同时处理绑定内存窗口和解绑内存窗口的请求时，可以采用内存窗口换绑的方式完成。在换绑时，将待解绑的内存窗口换绑到需要暴露访问权限的远端内存上，并返回换绑后新
10 生成的权限符。

在多个待绑定的远端内存地址为连续地址的情况下，可以将内存窗口的多次绑定操作替换为一次批量绑定操作。示例地，如图 6 所示，在将多个 MW 绑定到连续的远端内存上时，可以将多次内存窗口绑定操作减少为
15 对整片连续远端内存地址的一次内存窗口绑定操作。权限头除了可在权限资源池中分配外，还会投机地紧邻地分配在远端内存之前，绑定到远端内存和绑定到相邻权限头的两次内存窗口绑定操作减少为对该连续远端内存地址的一次绑定操作。

同理的，如图 6 所示，在连续的远端内存对应的多个 MW 解绑时，可
20 以将多次内存窗口解绑操作减少为对整片连续远端内存地址的一次内存窗口批量解绑操作。

这样，可以有效减少操作次数，从而提高远端内存访问效率。

在一些实施例中，远端内存架构包括多个计算节点和多个远端内存节点。

25 在本实施例中，远端内存架构可以包括多个计算节点和多个远端内存节点，计算节点具有较多的计算资源和受限的内存容量，远端内存节点具有受限的计算资源和较大的内存容量。远端内存架构允许计算节点和内存节点各自独立地扩展，因而提高了数据中心内部的资源利用率。

基于上述实施例提供的远端内存访问保护机制构建方法，本申请还提供
30 提供了一种远端内存节点的实施例。

图 7 示出了本申请另一个实施例提供的远端内存节点的结构示意图，为了便于说明，仅示出了与本申请实施例相关的部分。

参照图 7，远端内存节点 700 所属于远端内存架构，远端内存架构还可以包括计算节点，远端内存节点 700 可以包括：

5 获取模块 701，用于获取计算节点的权限申请，权限申请基于计算节点将用户程序的访问请求提交至网卡队列对上生成，访问请求包括远端内存地址；

绑定模块 702，用于在权限申请鉴权通过的情况下，将内存窗口绑定到远端内容地址上，并返回权限符，权限符用于远端内存访问；

10 解绑模块 703，用于在远端内存访问结束的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

在一些实施例中，绑定模块 702 还可以用于：

在权限申请鉴权通过的情况下，获取正在进行的远端内存访问对应的历史权限符；

15 确定权限申请与历史权限符是否存在冲突；

在权限申请与历史权限符不存在冲突的情况下，将内存窗口绑定到远端内容地址上，并返回权限符。

在一些实施例中，远端内存节点 700 可以包括：

存储模块，用于将权限符存储至哈希表中；

20 删除模块，用于将权限符从哈希表中删除。

在一些实施例中，权限符包括租约语义，租约语义对应有租期，解绑模块 703 还可以用于：

在远端内存访问的响应时长超过租期的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

25 在一些实施例中，租期基于用户程序采用远端内存单边访问的方式进行更新。

在一些实施例中，网卡队列对具有备用队列对，其中在网卡队列对失效的情况下，备用队列对替代网卡队列对。

在一些实施例中，解绑模块 703 还可以用于：

在远端内存访问结束的情况下，获取远端内存架构的负载压力和未绑定的内存窗口的数量；

在负载压力小于预设压力阈值，或数量小于预设数量阈值的情况下，将内存窗口与远端内容地址解绑，以使权限符失效。

5 在一些实施例中，远端内存节点 700 可以包括处理模块，处理模块可以用于：

将内存窗口的一次绑定操作和一次解绑操作合并为一次换绑操作；

在多个待绑定的远端内存地址为连续地址的情况下，将内存窗口的多次绑定操作替换为一次批量绑定操作。

10 在一些实施例中，远端内存架构包括多个计算节点和多个远端内存节点。

需要说明的是，上述装置/单元之间的信息交互、执行过程等内容，与本申请方法实施例基于同一构思，是与上述远端内存访问保护机制构建方法对应的装置，上述方法实施例中所有实现方式均适用于该装置的实施例中，其具体功能及带来的技术效果，具体可参见方法实施例部分，此处不再赘述。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、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单元、模块完成，即将装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块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。另外，各功能单元、模块的具体名称也只是为了便于相互区分，并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

图 8 示出了本申请又一个实施例提供的电子设备的硬件结构示意图。

设备可以包括处理器 801 以及存储有程序或指令的存储器 802。

处理器 801 执行程序时实现上述任意各个方法实施例中的步骤。

示例性的，程序可以被分割成一个或多个模块/单元，一个或者多个模块/单元被存储在存储器 802 中，并由处理器 801 执行，以完成本申请。一个或多个模块/单元可以是能够完成特定功能的一系列程序指令段，该指令段用于描述程序在设备中的执行过程。

5 具体地，上述处理器 801 可以包括中央处理器（CPU），或者特定集成电路（Application Specific Integrated Circuit，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 802 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 802 可包括硬盘驱动器（Hard Disk Drive，HDD）、软盘驱动器、闪存、光盘、磁光盘、磁带或通用串行总线（Universal Serial Bus，USB）
10 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 802 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 802 可在综合网关容灾设备的内部或外部。在特定实施例中，存储器 802 是非易失性固态存储器。

15 存储器可包括只读存储器（ROM），随机存取存储器（RAM），磁盘存储介质设备，光存储介质设备，闪存设备，电气、光学或其他物理/有形的存储器存储设备。因此，通常，存储器包括一个或多个编码有包括计算机可执行指令的软件的有形（非暂态）机器可读存储介质（例如，存储器设备），并且当该软件被执行（例如，由一个或多个处理器）时，其可操作来
20 执行参考根据本公开的一方面的方法所描述的操作。

处理器 801 通过读取并执行存储器 802 中存储的程序或指令，以实现上述实施例中的任意一种方法。

在一个示例中，电子设备还可包括通信接口 803 和总线 804。其中，处理器 801、存储器 802、通信接口 803 通过总线 804 连接并完成相互间的通
25 信。

通信接口 803，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。

总线 804 包括硬件、软件或两者，将在线数据流量计费设备的部件彼此耦接在一起。举例来说而非限制，总线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 804 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。

另外，结合上述实施例中的方法，本申请实施例可提供一种机器可读存储介质来实现。该机器可读存储介质上存储有程序或指令；该程序或指令被处理器执行时实现上述实施例中的任意一种方法。该机器可读存储介质可以被如计算机等机器读取。

本申请实施例另提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现上述方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

应理解，本申请实施例提到的芯片还可以称为系统级芯片、系统芯片、芯片系统或片上系统芯片等。

本申请实施例提供一种计算机程序产品，该程序产品被存储在机器可读存储介质中，该程序产品被至少一个处理器执行以实现如上述方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

需要明确的是，本申请并不局限于上文所描述并在图中示出的特定配置和处理。为了简明起见，这里省略了对已知方法的详细描述。在上述实施例中，描述和示出了若干具体的步骤作为示例。但是，本申请的方法过程并不限于所描述和示出的具体步骤，本领域的技术人员可以在领会本申请的精神后，作出各种改变、修改和添加，或者改变步骤之间的顺序。

以上所述的结构框图中所示的功能模块可以实现为硬件、软件、固件或者它们的组合。当以硬件方式实现时，其可以例如是电子电路、专用集成电路（ASIC）、适当的固件、插件、功能卡等等。当以软件方式实现时，本申请的元素是被用于执行所需任务的程序或者代码段。程序或者代码段

可以存储在机器可读介质中，或者通过载波中携带的数据信号在传输介质或者通信链路上传送。“机器可读介质”可以包括能够存储或传输信息的任何介质。机器可读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、
5 射频 (RF) 链路，等等。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。

还需要说明的是，本申请中提及的示例性实施例，基于一系列的步骤或者装置描述一些方法或系统。但是，本申请不局限于上述步骤的顺序，也就是说，可以按照实施例中提及的顺序执行步骤，也可以不同于实施例
10 中的顺序，或者若干步骤同时执行。

上面参考根据本公开的实施例的方法、装置 (系统) 和程序产品的流程图和/或框图描述了本公开的各方面。应当理解，流程图和/或框图中的每个方框以及流程图和/或框图中各方框的组合可以由计算机程序或指令实现。这些程序或指令可被提供给通用计算机、专用计算机、或其它可编程
15 数据处理装置的处理器，以产生一种机器，使得经由计算机或其它可编程数据处理装置的处理器执行的这些指令使能对流程图和/或框图的一个或多个方框中指定的功能/动作的实现。这种处理器可以是但不限于是通用处理器、专用处理器、特殊应用处理器或者现场可编程逻辑电路。还可理解，框图和/或流程图中的每个方框以及框图和/或流程图中的方框的组合，也可
20 可以由执行指定的功能或动作的专用硬件来实现，或可由专用硬件和计算机指令的组合来实现。

以上所述，仅为本申请的具体实施方式，所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的系统、模块和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。应
25 理解，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，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。

权 利 要 求 书

1. 一种远端内存访问保护机制构建方法，应用于远端内存架构中的远端内存节点，所述远端内存架构还包括计算节点，所述方法包括：

获取所述计算节点的权限申请，所述权限申请基于所述计算节点将用户程序的访问请求提交至网卡队列对上生成，所述访问请求包括远端内存地址，

在所述权限申请鉴权通过的情况下，将内存窗口绑定到所述远端内容地址上，并返回权限符，所述权限符用于远端内存访问，

在所述远端内存访问结束的情况下，将所述内存窗口与所述远端内容地址解绑，以使所述权限符失效。

2. 根据权利要求 1 所述的方法，其中，所述在所述权限申请鉴权通过的情况下，将内存窗口绑定到所述远端内容地址上，并返回权限符，包括：

在所述权限申请鉴权通过的情况下，获取正在进行的远端内存访问对应的历史权限符，

确定所述权限申请与所述历史权限符是否存在冲突，

在所述权限申请与所述历史权限符不存在冲突的情况下，将内存窗口绑定到所述远端内容地址上，并返回权限符。

3. 根据权利要求 1 或 2 所述的方法，其中，所述在所述权限申请鉴权通过的情况下，将内存窗口绑定到所述远端内容地址上，并返回权限符之后，所述方法还包括：

将所述权限符存储至哈希表中，

所述在所述远端内存访问结束的情况下，将所述内存窗口与所述远端内容地址解绑，以使所述权限符失效之后，所述方法还包括：

将所述权限符从所述哈希表中删除。

4. 根据权利要求 1 所述的方法，其中，所述权限符包括租约语义，所述租约语义对应有租期，所述在所述权限申请鉴权通过的情况下，将内存窗口绑定到所述远端内容地址上，并返回权限符之后，所述方法还包括：

在所述远端内存访问的响应时长超过所述租期的情况下，将所述内存窗口与所述远端内容地址解绑，以使所述权限符失效。

5. 根据权利要求4所述的方法，其中，所述租期基于所述用户程序采用远端内存单边访问的方式进行更新。

6. 根据权利要求1所述的方法，其中，所述网卡队列对具有备用队列对，其中在所述网卡队列对失效的情况下，所述备用队列对替代所述网卡队列对。

7. 根据权利要求1所述的方法，其中，所述在所述远端内存访问结束的情况下，将所述内存窗口与所述远端内容地址解绑，以使所述权限符失效，包括：

在所述远端内存访问结束的情况下，获取所述远端内存架构的负载压力和未绑定的内存窗口的数量，

在所述负载压力小于预设压力阈值，或所述数量小于预设数量阈值的情况下，将所述内存窗口与所述远端内容地址解绑，以使所述权限符失效。

8. 根据权利要求1所述的方法，还包括：

将所述内存窗口的一次绑定操作和一次解绑操作合并为一次换绑操作，

在多个待绑定的远端内存地址为连续地址的情况下，将所述内存窗口的多次绑定操作替换为一次批量绑定操作。

9. 根据权利要求1所述的方法，其中，所述远端内存架构包括多个计算节点和多个远端内存节点。

10. 一种远端内存节点，所属于远端内存架构，所述远端内存架构还包括计算节点，所述远端内存节点包括：

获取模块，用于获取所述计算节点的权限申请，所述权限申请基于所述计算节点将用户程序的访问请求提交至网卡队列对上生成，所述访问请求包括远端内存地址，

绑定模块，用于在所述权限申请鉴权通过的情况下，将内存窗口绑定到所述远端内容地址上，并返回权限符，所述权限符用于远端内存访问，

解绑模块，用于在所述远端内存访问结束的情况下，将所述内存窗口与所述远端内容地址解绑，以使所述权限符失效。

11. 一种电子设备，所述设备包括：处理器以及存储有程序或指令的

存储器；

所述处理器执行所述程序或指令时实现如权利要求 1-9 任意一项所述的方法。

12. 一种机器可读存储介质，所述机器可读存储介质上存储有程序或指令，所述程序或指令被处理器执行时实现如权利要求 1-9 任意一项所述的方法。

13. 一种计算机程序产品，所述计算机程序产品中的指令由电子设备的处理器执行时，使得所述电子设备执行如权利要求 1-9 任意一项所述的方法。

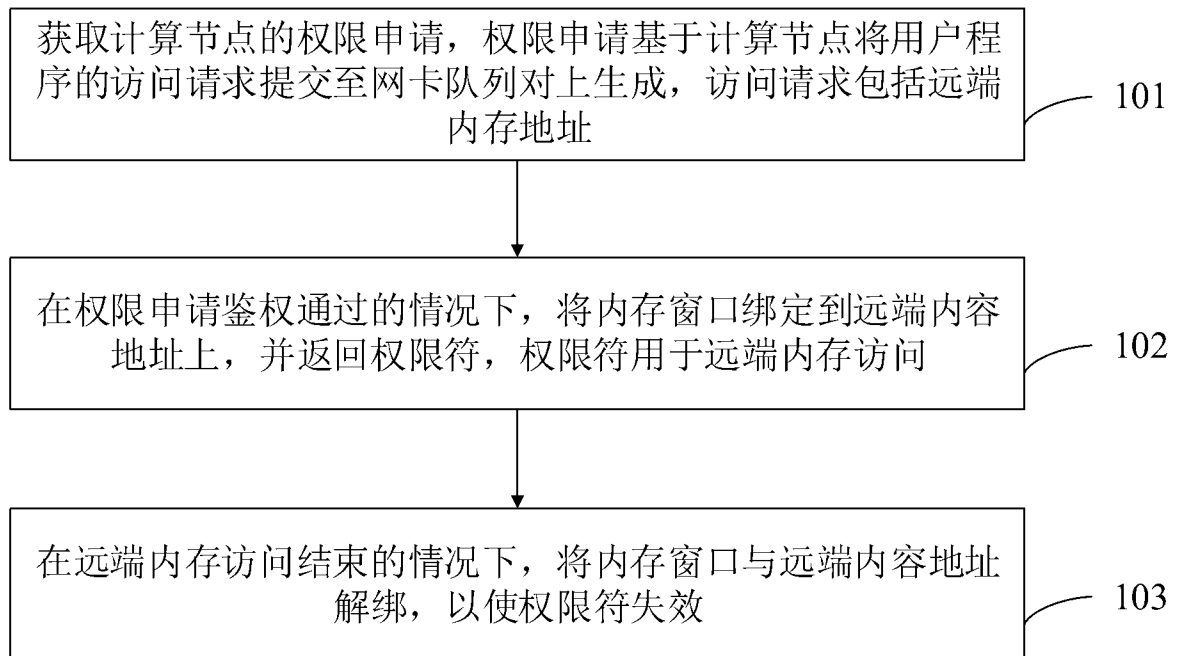
$1/4$ 

图 1

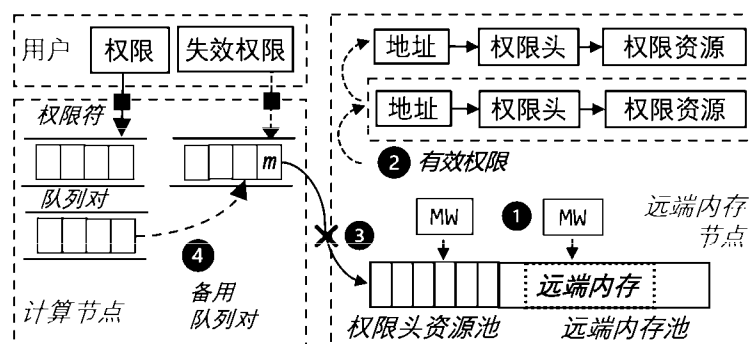


图 2

2/4

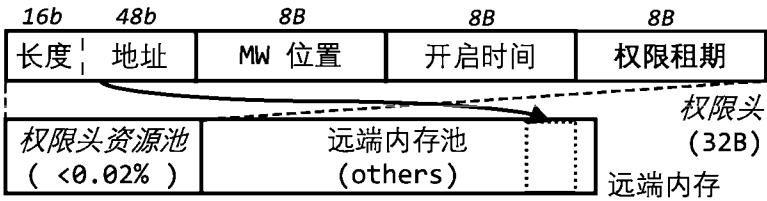
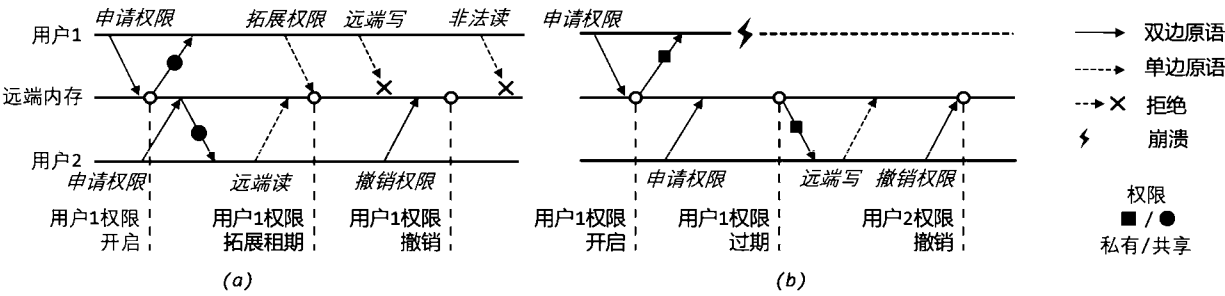


图 4

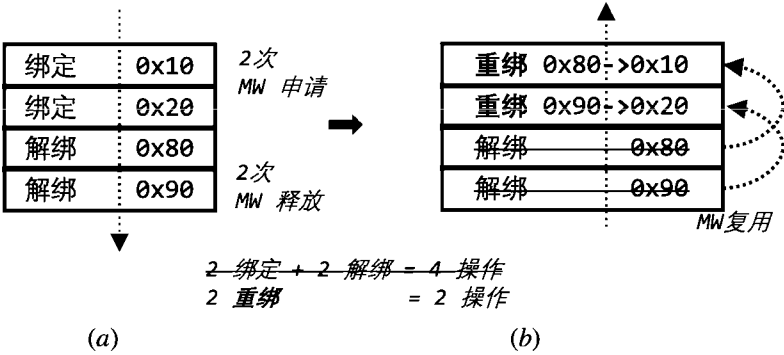


图 5

3/4

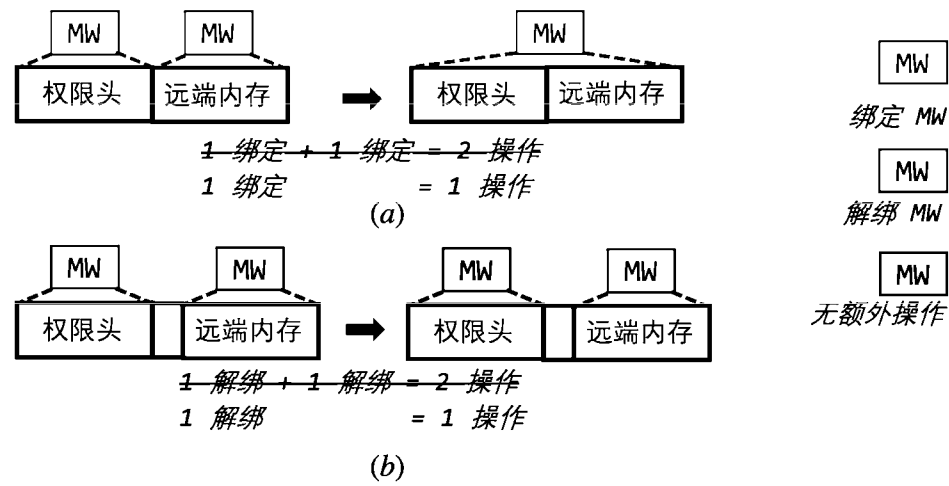


图 6

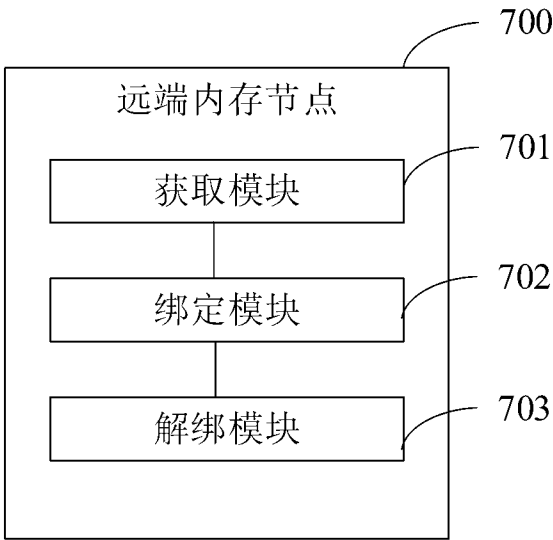


图 7

4/4

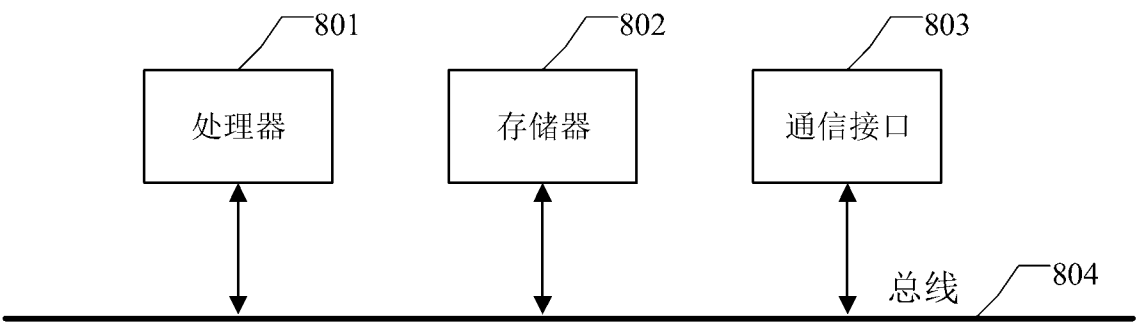


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/082125

A. CLASSIFICATION OF SUBJECT MATTER

G06F 15/173(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F, 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, VEN, CNTXT, WOTXT, USTXT, EPTXT, CNKI, IEEE: 远端内存, 远程内存, 权限, 鉴权, 授权, 网卡, 网络适配器, 绑定, 锁定, 解锁, 解绑, RDMA, remote direct memory access, authenti+, network adapter, lock+, unlock+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2019141041 A1 (CA, INC.) 09 May 2019 (2019-05-09) description, paragraphs [14]-[157], and figures 1-4	1-13
A	CN 106844048 A (SHANGHAI JIAO TONG UNIVERSITY) 13 June 2017 (2017-06-13) entire document	1-13
A	CN 109559407 A (AMAZON TECHNOLOGIES, INC.) 02 April 2019 (2019-04-02) entire document	1-13
A	CN 113014593 A (BEIJING KINGSOFT CLOUD NETWORK TECHNOLOGY CO., LTD.) 22 June 2021 (2021-06-22) entire document	1-13
A	CN 113852656 A (HUAWEI TECHNOLOGIES CO., LTD.) 28 December 2021 (2021-12-28) entire document	1-13



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

15 June 2023

Date of mailing of the international search report

21 June 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/082125

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2019141041	A1	09 May 2019	US	10523675	B2	31 December 2019
CN	106844048	A	13 June 2017	None			
CN	109559407	A	02 April 2019	US	2019312737	A1	10 October 2019
				US	11438169	B2	06 September 2022
				JP	2020113312	A	27 July 2020
				JP	7079805	B2	02 June 2022
				GB	201815400	D0	07 November 2018
				GB	2570752	A	07 August 2019
				GB	2570752	B	28 July 2021
				JP	2019061672	A	18 April 2019
				JP	6682592	B2	15 April 2020
				DE	102018007534	A1	28 March 2019
				US	2019253255	A1	15 August 2019
				US	10498538	B2	03 December 2019
CN	113014593	A	22 June 2021	None			
CN	113852656	A	28 December 2021	WO	2022001417A1	A1	06 January 2022
				EP	4155925	A1	29 March 2023

国际检索报告

国际申请号

PCT/CN2023/082125

A. 主题的分类

G06F 15/173 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

IPC: G06F, H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNABS, VEN, CNTXT, WOTXT, USTXT, EPTXT, CNKI, IEEE: 远端内存, 远程内存, 权限, 鉴权, 授权, 网卡, 网络适配器, 绑定, 锁定, 解锁, 解绑, RDMA, remote direct memory access, authori+, network adapter, lock+, unlock+

C. 相关文件

类 型*

引用文件, 必要时, 指明相关段落

相关的权利要求

X

US 2019141041 A1 (CA, INC.) 2019年5月9日 (2019 - 05 - 09)
说明书第[14]-[157]段, 图1-4

1-13

A

CN 106844048 A (上海交通大学) 2017年6月13日 (2017 - 06 - 13)
全文

1-13

A

CN 109559407 A (亚马逊技术股份有限公司) 2019年4月2日 (2019 - 04 - 02)
全文

1-13

A

CN 113014593 A (北京金山云网络技术有限公司) 2021年6月22日 (2021 - 06 - 22)
全文

1-13

A

CN 113852656 A (华为技术有限公司) 2021年12月28日 (2021 - 12 - 28)
全文

1-13

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年6月15日

国际检索报告邮寄日期

2023年6月21日

ISA/CN的名称和邮寄地址

中国国家知识产权局

中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

杨莹莹

电话号码 (+86) 010-53961527

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/082125

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)
US	2019141041	A1	2019年5月9日	US	10523675	B2	2019年12月31日
CN	106844048	A	2017年6月13日	无			
CN	109559407	A	2019年4月2日	US	2019312737	A1	2019年10月10日
				US	11438169	B2	2022年9月6日
				JP	2020113312	A	2020年7月27日
				JP	7079805	B2	2022年6月2日
				GB	201815400	D0	2018年11月7日
				GB	2570752	A	2019年8月7日
				GB	2570752	B	2021年7月28日
				JP	2019061672	A	2019年4月18日
				JP	6682592	B2	2020年4月15日
				DE	102018007534	A1	2019年3月28日
				US	2019253255	A1	2019年8月15日
				US	10498538	B2	2019年12月3日
CN	113014593	A	2021年6月22日	无			
CN	113852656	A	2021年12月28日	WO	2022001417A1	A1	2022年1月6日
				EP	4155925	A1	2023年3月29日