(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2019 年 7 月 4 日 (04.07.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/129023 A1

(51) 国际专利分类号:
G06F 3/06 (2006.01)

(21) 国际申请号: PCT/CN2018/123749

(22) 国际申请日: 2018 年 12 月 26 日 (26.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711444963.0 2017年12月27日 (27.12.2017) CN

(71) 申请人: 华 为 技 术 有 限 公 司 (**HUAWEI TECHNOLOGIES CO., LTD.**) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 张 胜 光 (**ZHANG, Shengguang**); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 胡海燕 (**HU, Haiyan**); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 郑新钱 (**ZHENG, Xinqian**); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限责任公司 (**BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.**); 中国北京市海淀区学院路蓟门里和景园A座1单元102室, Beijing 100088 (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,

(54) **Title:** NOF BASED READING CONTROL METHOD, DEVICE AND SYSTEM

(54) 发明名称: 基于NOF的读取控制方法、装置及系统

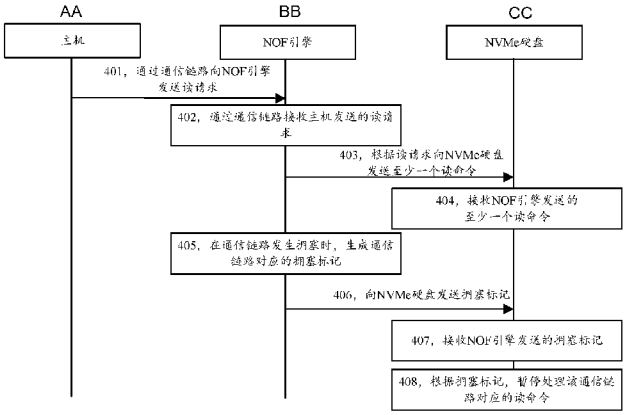


图 4

401 Send a reading request to a NOF engine through a communication link
402 Receive, through the communication link, a reading request sent by a host
403 Send at least one reading command to an NVMe hard disk according to the reading request
404 Receive the at least one reading command sent by the NOF engine
405 Generate a congestion label corresponding to the communication link when the communication link is congested
406 Send the congestion label to the NVMe hard disk
407 Receive the congestion label sent by the NOF engine
408 According to the congestion label, suspend the processing of the reading command corresponding to the communication link
AA Host
BB NOF engine
CC NVMe hard disk

(57) **Abstract:** A NOF based reading control method, device and system, belonging to the field of networked storage. The method comprises: a NOF engine receiving, through a communication link, a reading request sent by a host (402), sending at least one reading command to an NVMe hard disk according to the reading request (403), generating a congestion label corresponding to the communication link when the communication link is congested (405), and sending the congestion label to the NVMe hard disk (406), wherein the congestion label is used for instructing the NVMe hard disk to suspend processing of the reading command corresponding to the

WO 2019/129023 A1

[见续页]

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, 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))。

communication link. In a scenario of network congestion, the NVMe hard disk can suspend the processing of reading commands which have been acquired but not processed yet, solving the problem that an entire reading process fails due to data of this portion of reading commands being directly discarded due to network congestion after being read; when the NVMe hard disk suspends the processing of this portion of reading commands, invalid occupation of a transmission bandwidth between a NOF engine and an NVMe hard disk can be avoided, achieving the effect of reducing waste.

(57) 摘要：一种基于NOF的读取控制方法、装置及系统，属于网络化存储领域。该方法包括：NOF引擎通过通信链路接收主机发送的读请求（402），根据读请求向NVMe硬盘发送至少一个读命令（403），在通信链路发生拥塞时，生成通信链路对应的拥塞标记（405），并向NVMe硬盘发送拥塞标记（406），拥塞标记用于指示NVMe硬盘暂停处理通信链路对应的读命令。在发生网络拥塞的场景下，NVMe硬盘能够对已经获取且尚未处理的读命令暂停处理，解决了这部分读命令的数据在读取后，因为网络拥塞而被直接丢弃，导致整个读取过程失败的问题，达到了NVMe硬盘暂停对这部分读命令处理后，能够避免对NOF引擎与NVMe硬盘之间的传输带宽的无效占用，减少浪费的效果。

说 明 书

基于 NOF 的读取控制方法、装置及系统

本申请要求于 2017 年 12 月 27 日提交的申请号为 201711444963.0、申请名称为“基于 NOF 的读取控制方法、装置及系统”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及网络化存储领域，特别涉及一种基于网络连接的非易失性存储（Non-Volatile Memory express Over Fabric，NOF）的读取控制方法、装置及系统。

背景技术

非易失性存储（Non-Volatile Memory express，NVMe）总线协议组织在 2016 年 6 月，发布了 NOF1.0 协议。NOF 协议赋予 NVMe 设备接入网络的通信能力。

典型的基于 NOF 协议的应用模型包括：主机（Host）和控制器（Controller），控制器包括 NOF 引擎和与其相连的多个 NVMe 硬盘。其中，主机和控制器之间通过网络相连，NOF 引擎和 NVMe 硬盘之间通过 PCIe 总线相连。以主机从控制器读取数据为例，主机通过网络向 NOF 引擎发送一个或多个读请求。NOF 引擎将一个或多个读请求转发给 NVMe 硬盘，NVMe 硬盘将每个读请求对应的数据反馈给 NOF 引擎，NOF 引擎再将数据通过网络发送给主机。

但是由于主机和控制器之间的网络可能会发生拥塞，NOF 引擎通过网络向主机发送数据时，主机可能会因为网络拥塞而无法正常接收数据，导致读请求响应失败。目前尚未有技术来解决 NOF 应用场景下的网络拥塞问题。

发明内容

本申请实施例提供了一种基于 NOF 的读取控制方法、装置及系统，可以用于解决 NOF 引用场景下的网络拥塞问题。所述技术方案如下：

根据本申请的第一方面，提供了一种基于 NOF 的读取控制方法，所述方法包括：

NOF 引擎通过通信链路接收主机发送的读请求；所述 NOF 引擎根据所述读请求向 NVMe 硬盘发送至少一个读命令；所述 NOF 引擎在所述通信链路发生拥塞时，生成所述通信链路对应的拥塞标记；所述 NOF 引擎向所述 NVMe 硬盘发送所述拥塞标记，所述拥塞标记用于指示所述 NVMe 硬盘暂停处理所述通信链路对应的所述读命令。对应地，NVMe 硬盘接收 NOF 引擎发送的至少一个读命令，所述读命令是所述 NOF 引擎根据主机通过通信链路发送的读请求产生的；所述 NVMe 硬盘接收 NOF 引擎发送的拥塞标记，所述拥塞标记是所述 NOF 引擎在所述通信链路发送拥塞时发送的；所述 NVMe 硬盘根据所述拥塞标记，暂停处理所述通信链路对应的所述读命令。

由于在发生网络拥塞的场景下，NVMe 硬盘能够对已经获取且尚未处理的读命令暂停

处理，解决了这部分读命令的数据在读取后，因为网络拥塞而被直接丢弃，导致整个读取过程失败的问题，达到了 NVMe 硬盘暂停对这部分读命令处理后，能够避免对 NOF 引擎与 NVMe 硬盘之间的传输带宽（PCI 带宽）的无效占用，减少浪费的效果。

在第一方面的第一种可能的实现方式中，所述 NOF 引擎向 NVMe 硬盘发送所述拥塞标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的提交队列的条目（Submission Queue Entry，SQE）命令，所述 SQE 命令包括读命令和/或写命令；所述 NVMe 硬盘向所述 NOF 引擎发送命令读取请求，所述命令读取请求用于读取与所述通信链路对应的提交队列中的 SQE 命令；所述 NOF 引擎在接收到所述 NVMe 硬盘对所述提交队列的命令读取请求时，向所述 NVMe 硬盘发送携带有所述拥塞标记的 SQE 命令。所述 NVMe 硬盘接收所述 NOF 引擎发送的携带有所述拥塞标记的 SQE 命令。

本实现方式通过利用提交队列中的 SQE 命令来传递拥塞标记，使得在尽量不增加额外信令的前提下，实现 NOF 引擎与 NVMe 硬盘之间的拥塞通知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。由于提交队列中有很大概率存在 NVMe 硬盘还未读取的 SQE 命令，而利用 SQE 命令中的保留域来传递拥塞标记，几乎没有增加额外信令和数据量，从而减少对 NOF 引擎与 NVMe 硬盘之间的传输带宽的占用。

在第一方面的第二种可能的实现方式中，所述 NOF 引擎向 NVMe 硬盘发送所述拥塞标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；所述 NOF 引擎向所述 NVMe 硬盘发送指向第一自定义地址的寄存器写命令，所述第一自定义地址是用于存储所述拥塞标记的地址，所述寄存器写命令携带有所述提交队列的标识和所述拥塞标记。所述 NVMe 硬盘接收所述 NOF 引擎发送的指向第一自定义地址的寄存器写命令，所述第一自定义地址是用于存储所述拥塞标记的地址，所述寄存器写命令携带有与所述通信链路对应的提交队列的标识和所述拥塞标记；所述 NVMe 硬盘根据所述寄存器写命令对所述第一自定义地址进行写入操作。

本实现方式通过利用指向第一自定义地址的寄存器写命令来传递拥塞标记，使得在不改变 SQE 命令的正常读写机制时，也能够实现 NOF 引擎与 NVMe 硬盘之间的拥塞通知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。

在第一方面的第三种可能的实现方式中，所述 NOF 引擎向 NVMe 硬盘发送所述拥塞标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；所述 NOF 引擎向所述 NVMe 硬盘发送具有自定义内容的管理命令，所述自定义内容携带有所述提交队列的标识和所述拥塞标记。所述 NVMe 硬盘接收所述 NOF 引擎发送的具有第一自定义内容的管理命令，所述第一自定义内容携带有提交队列的标识和所述拥塞标记，所述提交队列是与所述通信链路对应的队列。

本实现方式通过利用具有第一自定义内容的管理命令来传递拥塞标记，使得利用 NVMe 协议提供的管理命令，即可实现 NOF 引擎与 NVMe 硬盘之间的拥塞通知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。

在第一方面的第四种可能的实现方式中，所述 NVMe 硬盘根据所述拥塞标记，暂停处理所述通信链路对应的读命令，包括：

所述 NVMe 硬盘根据所述拥塞标记执行如下操作中的至少一种操作：

暂停对所述提交队列的命令读取请求；

将从所述提交队列已经取回且尚未完成的读命令设置为挂起状态；

将所述提交队列对应的调度权重降低至第一权重。

本实现方式将提交队列对应的读请求进行暂停处理，能够节省 NVMe 硬盘的处理资源，以及 NOF 引擎和 NVMe 硬盘之间的传输带宽。

结合第一方面、第一方面的第一种可能的实现方式、第一方面的第二种可能的实现方式、第一方面的第三种可能的实现方式中的任意一种实现方式，在第一方面的第五种可能的实现方式中，所述向所述 NVMe 硬盘发送添加有所述拥塞标记的 SQE 命令之后，还包括：

所述 NOF 引擎通过所述通信链路向所述主机间隔发送空闲帧；所述 NOF 引擎在接收到所述空闲帧的确认信号时，生成所述通信链路对应的拥塞解除标记；所述 NOF 引擎向所述 NVMe 硬盘发送所述拥塞解除标记，所述拥塞解除标记用于指示所述 NVMe 硬盘恢复对所述通信链路对应的读命令的处理。所述 NVMe 硬盘接收所述 NOF 引擎发送的拥塞解除标记，所述拥塞解除标记是所述 NOF 引擎在与主机之间的通信链路拥塞解除时发送的；所述 NVMe 硬盘根据所述拥塞解除标记，恢复对所述通信链路对应的读命令的处理。

本实现方式通过 NOF 引擎在与主机之间的通信链路发生拥塞时，向 NVMe 硬盘发送拥塞标记，由 NVMe 硬盘根据拥塞标记暂停处理该通信链路对应的读命令。使得在发生网络拥塞的场景下，NVMe 硬盘对已经获取且尚未处理的读命令暂停处理，解决了这部分读命令的数据在读取后，因为网络拥塞而无法顺利发送给主机，导致整个读取过程失败的问题，达到了在网络拥塞场景下减少无效数据的发送，从而避免网络更加拥塞的效果。

结合第一方面的第五种可能的实现方式，在第一方面的第六种可能的实现方式中，

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；所述 NOF 引擎在接收到所述 NVMe 硬盘对所述提交队列的命令读取请求时，向所述 NVMe 硬盘发送携带有所述拥塞解除标记的 SQE 命令。所述 NVMe 硬盘向所述 NOF 引擎发送命令读取请求，所述命令读取请求用于读取与所述通信链路对应的提交队列中的 SQE 命令；所述 NVMe 硬盘接收所述 NOF 引擎发送的携带有所述拥塞解除标记的 SQE 命令。

本实现方式通过利用提交队列中的 SQE 命令来传递拥塞解除标记，使得在尽量不增加额外信令的前提下，实现 NOF 引擎与 NVMe 硬盘之间的拥塞解除通知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。由于提交队列中有很大概率存在 NVMe 硬盘还未读取的 SQE 命令，而利用 SQE 命令中的保留域来传递拥塞标记，几乎没有增加额外信令和数据量。

结合第一方面的第五种可能的实现方式，在第一方面的第七种可能的实现方式中，

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；所述 NOF 引擎向所述 NVMe 硬盘发送指向第二自定义地址的寄存器写命令，所述第二自定义地址是用于存储所述拥塞解除标记的地址，所述寄存器写命令携带有所述提交队列的标识和所述拥塞解除标记。所述 NVMe 硬盘接收所述 NOF 引擎发

送的指向第二自定义地址的寄存器写命令，所述第二自定义地址是用于存储所述拥塞解除标记的地址，所述寄存器写命令携带有与所述通信链路对应提交队列的标识和所述拥塞解除标记；所述 NVMe 硬盘根据所述寄存器写命令对所述第二自定义地址进行写入操作。

本实现方式通过利用指向第二自定义地址的寄存器写命令来传递拥塞解除标记，使得在不改变 SQE 命令的正常读写机制时，实现 NOF 引擎与 NVMe 硬盘之间的拥塞解除通知机制，进而控制 NVMe 硬盘恢复对该通信链路对应的读命令进行处理。

结合第一方面的第五种可能的实现方式，在第一方面的第八种可能的实现方式中，

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；所述 NOF 引擎向所述 NVMe 硬盘发送具有第二自定义内容的管理命令，所述第二自定义内容携带有所述提交队列的标识和所述拥塞解除标记。所述 NVMe 硬盘接收所述 NOF 引擎发送的具有第二自定义内容的管理命令，所述第二自定义内容携带有与所述通信链路对应提交队列的标识和所述拥塞解除标记。

本实现方式通过利用具有第二自定义内容的管理命令来传递拥塞解除标记，使得利用 NVMe 协议提供的管理命令，即可实现 NOF 引擎与 NVMe 硬盘之间的拥塞解除通知机制，进而控制 NVMe 硬盘恢复对该通信链路对应的读命令进行处理。

结合第一方面的第五种可能的实现方式，在第一方面的第九种可能的实现方式中，所述 NVMe 硬盘根据所述拥塞解除标记执行如下操作中的至少一种操作：

恢复对所述提交队列的命令读取请求；

将从所述提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态；

将所述提交队列对应的调度权重提高至第二权重。

本实现方式通过 NOF 引擎在通信链路的拥塞接触时，向 NVMe 硬盘发送拥塞解除标记，由 NVMe 硬盘根据拥塞解除标记恢复处理该通信链路对应的读命令。使得在网络拥塞已经解除的场景下，NVMe 硬盘对已经获取且尚未处理的读命令恢复处理，保证了主机在读取数据时的成功率。

根据本申请的第二方面，提供了一种基于 NOF 的读取控制方法，应用于 NOF 引擎中，NOF 引擎中设置有与通信链路对应的数据缓存，所述方法包括：

所述 NOF 引擎通过所述通信链路接收主机发送的读请求，所述读请求用于请求读取第一数据量的数据，所述第一数据量超过所述通信链路对应的数据缓存大小；

所述 NOF 引擎将所述读请求拆分为 K 个读命令，所述读命令用于请求读取第二数据量的数据，所述第二数据量小于所述第一数据量；

所述 NOF 引擎向 NVMe 硬盘发送 m 个所述读命令，$m \leqslant K$，m 与所述第二数据量的乘积不大于所述通信链路对应的数据缓存大小；

所述 NOF 引擎在所述通信链路发生拥塞时，暂停向所述 NVMe 硬盘发送剩余的其它读命令。

本实现方式可以在 NOF 引擎侧有数据缓存的情况下，由 NOF 引擎控制向 NVMe 硬盘发送的读命令的数量，使得 NVMe 硬盘正在处理的读命令的读取数据量不超过数据缓存的缓存大小。即便通信链路发生了网络拥塞，NOF 引擎暂停向 NVMe 硬盘继续下发读命令即可解决拥塞问题，不需要 NVMe 硬盘感知拥塞过程，减少 NOF 引擎和 NVMe 硬盘之间的信令交互以及传输带宽的占用。

另外，由于本实现方式是为每个通信链路设置各自对应的数据缓存，而不是多个通信链路共用同一个数据缓存，因此即便某一个通信链路发生了网络拥塞，导致该通信链路对应的数据缓存被占用，也不会出现将 NOF 引擎中所有的数据缓存都被该通信链路对应的读取数据所占用，导致其它通信链路的读取过程也收到影响的问题。达到了每个通信链路都分别独立，即便一条或多条通信链路发生拥塞，也不会影响剩余通信链路的正常收发过程的效果。

在第二方面的第一种可能的实现方式中，所述 NOF 引擎在所述通信链路发生拥塞时，暂停向所述 NVMe 硬盘发送剩余的其它读命令之后，还包括：

所述 NOF 引擎接收所述 NVMe 硬盘根据所述 m 个读命令读取的数据；

所述 NOF 引擎将所述数据存储在所述通信链路对应的所述数据缓存中；

所述 NOF 引擎在所述通信链路解除拥塞时，将所述数据缓存中的所述数据通过所述通信链路发送至所述主机。

在第二方面的第二种可能的实现方式中，所述 NOF 引擎在所述通信链路解除拥塞时，将所述数据缓存中的所述数据通过所述通信链路发送至所述主机之后，还包括：

所述 NOF 引擎继续向所述 NVMe 硬盘发送 n 个所述读命令，$n \leqslant K-m$，n 与所述第二数据量的乘积不大于所述数据缓存的缓存大小。

根据本申请的第三方面，提供了一种基于 NOF 的读取控制装置，所述装置包括用于执行上述第一方面及第一方面的各种可选实现方式中任一实现方式中，由 NOF 引擎实现的读取控制方法的单元或手段。

根据本申请的第四方面，提供了一种基于 NOF 的读取控制装置，所述装置包括用于执行上述第一方面及第一方面的各种可选实现方式中任一实现方式中，由 NVMe 硬盘实现的读取控制方法的单元或手段。

根据本申请的第五方面，提供了一种基于 NOF 的读取控制装置，所述装置设置有为通信链路对应的数据缓存，所述装置包括用于执行上述第二方面及第二方面的各种可选实现方式的任一实现方式中，由 NOF 引擎实现的读取控制方法的单元或手段。

根据本申请的第六方面，提供了一种 NOF 引擎，所述 NOF 引擎包括：处理器和存储器，所述存储器存储有至少一个指令，所述至少一个指令被所述处理器执行以实现如上述第一方面及第一方面的各种可选实现方式的任一实现方式中，由 NOF 引擎实现的读取控制方法。

根据本申请的第七方面，提供了一种 NVMe 硬盘，所述 NVMe 硬盘包括：控制器和存储器，所述存储器存储有至少一个指令，所述至少一个指令被所述控制器执行以实现如上述第一方面及第一方面的各种可选实现方式的任一实现方式中，由 NVMe 硬盘实现的读取控制方法。

根据本申请的第八方面，提供了一种 NOF 引擎，所述 NOF 引擎包括：处理器和存储器，所述存储器存储有至少一个指令，所述至少一个指令被所述处理器执行以实现如上述第二方面及第二方面的各种可选实现方式的任一实现方式中，由 NOF 引擎实现的读取控制方法。

根据本申请的第九方面，提供了一种计算机可读存储介质，所述存储介质存储有至少一个指令，所述至少一个指令被执行以实现如上述第一方面及第一方面的各种可选实现方

式的任一实现方式中，由 NOF 引擎实现的读取控制方法。

根据本申请的第十方面，提供了一种计算机可读存储介质，所述存储介质存储有至少一个指令，所述至少一个指令被执行以实现如上述第一方面及第一方面的各种可选实现方式的任一实现方式中，由 NVMe 硬盘实现的读取控制方法。

根据本申请的第十一方面，提供了一种计算机可读存储介质，所述存储介质存储有至少一个指令，所述至少一个指令被执行以实现如上述第二方面及第二方面的各种可选实现方式的任一实现方式中，由 NOF 引擎实现的读取控制方法。

附图说明

图 1 是本申请一个示例性实施例提供的 NOF 存储系统 100 的结构示意图；

图 2 是本申请另一个示例性实施例提供的 NOF 存储系统 100 的结构示意图；

图 3 是一种典型的 NOF 存储系统发生拥塞时的原理示意图；

图 4 是本申请一个示例性实施例提供的基于 NOF 的读取控制方法的方法流程图；

图 5 是本申请另一个示例性实施例提供的基于 NOF 的读取控制方法的方法流程图；

图 6 是本申请另一个示例性实施例提供的基于 NOF 的读取控制方法的方法流程图；

图 7 是本申请另一个示例性实施例提供的基于 NOF 的读取控制方法的方法流程图；

图 8 是本申请另一个示例性实施例提供的基于 NOF 的读取控制方法的方法流程图；

图 9 是本申请另一个示例性实施例提供的基于 NOF 的读取控制方法的方法流程图；

图 10 是本申请另一个示例性实施例提供的基于 NOF 的读取控制装置的框图；

图 11 是本申请另一个示例性实施例提供的基于 NOF 的读取控制装置的框图；

图 12 是本申请另一个示例性实施例提供的 NOF 引擎的框图；

图 13 是本申请另一个示例性实施例提供的 NVMe 硬盘的框图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

图 1 示出了本申请一个示例性实施例提供的 NOF 存储系统 100 的结构示意图。该 NOF 存储系统 100 包括：主机 120 和 NOF 控制器 140。

主机 120 可以是一个或多个。主机 120 用于采用远程直接数据存取（Remote Direct Memory Access，RDMA）从 NOF 控制器 140 读数据，或者向 NOF 控制器 140 写数据。图 1 中以主机 120 为多个主机示例说明，比如，主机 120 包括主机 0 至主机 n，n 为正整数。多个主机 120 可以同属于同一使用者，也可以属于不同的使用者。

在不同的实施例中，主机 120 可以是业务服务器、网络主机、计算节点等。多个主机 120 中可以存在类型不相同的主机，本实施例对主机 120 的具体类型不加以限定，只需要主机 120 支持 NOF 协议即可。

主机 120 通过网络 160 与多个 NOF 控制器 140 相连。可选地，该网络 160 为光纤通道架构（Fabric）网络，Fabric 网路是由多台基于光纤通道（Fibre Channel，FC）协议的交换机互联组成的网络。该 Fabric 网络包括多个交换网关和通信线路。该通信线路可以是光缆

或电缆，当通信线路是光缆时，交换网关可以是光交换网关；当通信线路是电缆时，交换网关可以是路由器和/或交换机。示意性的，每个主机120在网络160上的接入带宽是100GB，每个NOF控制器140在网络160上的接入带宽是100GB。

NOF控制器140可以是一个或多个。每个NOF控制器140具有各自的设备标识（Identify，ID），不同的NOF控制器140通过不同的设备ID进行区分。同一个物理载体上可以运行一个或多个NOF控制器140。图1中以NOF控制器140为多个NOF控制器示例说明，比如，NOF控制器140包括NOF控制器0至NOF控制器m，m为正整数。

每个NOF控制器140中包括：NOF引擎142和至少一个NVMe硬盘144，NVMe硬盘144是支持NVMe协议的固态硬盘（Solid State Drives，SSD）。一方面，NOF引擎142通过网络160与主机120相连；另一方面，NOF引擎142通过PCIe总线与至少一个NVMe硬盘144相连。每个NVMe硬盘144都以NOF引擎142作为访问对象。NOF引擎142负责在网络协议与PCIe协议之间的协议转换，以及数据转发处理。

结合参考图2，一个主机120和一个NOF控制器140之间可以建立一个或多个通信链路。该通信链路可以是队列对（Queue Pair，QP）连接。该QP连接是承载在以太网物理线路上的传输链路。同一个以太网物理线路上可以承载一条或多条QP连接，每个QP连接具有唯一的通信链路标识，该通信链路标识可以标识为QPID，不同的QP连接采用不同的QPID进行区分。示意性，主机120中存在n个队列QP0~QPn，NOF引擎142中存在对应的n个队列QP0~QPn。主机120中的QP0与NOF引擎142中的QP0构成一个QP连接0，主机120中的QP1与NOF引擎142中的QP1构成一个QP连接1，依次类推，主机120中的QPn与NOF引擎142中的QPn构成一个QP连接n。

可选地，队列对还可以称为RDMA队列对。队列对中的队列也称为完成队列（Completion Queue，CQ）。在NOF引擎142中的CQ队列用于缓存来自主机120的RDMA读请求或写请求。

在NOF引擎142中，还存在与每个CQ队列（或者说通信链路）对应的提交队列（Submission Queue，SQ）。SQ队列用于存储NOF引擎142需要发送给NVMe硬盘144的输入/输出（input/output，I/O）命令，也称SQE命令。SQE命令包括读命令和/或写命令。可选地，CQ队列和SQ队列存在一一对应关系。示意性的，完成队列QP0与提交队列SQ0对应，完成队列QP1与提交队列SQ1对应，依次类推，完成队列QPn与提交队列SQn对应。

可选地，在NOF引擎142中，还存在与每个通信链路对应的数据缓存（Data Buffer）。该数据缓存用于在某一个通信链路发生网络拥塞时，对来自NVMe硬盘144的与该通信链路对应的读数据进行缓存。也即，本申请实施例中不会让全部的通信链路共用同一个数据缓存，而是存在至少两个通信链路采用各自对应的数据缓存。可选地，每个通信链路采用与自身对应的数据缓存，任意两个通信链路对应的数据缓存是不同的；或者，与同一个主机对应的通信链路采用与自身对应的数据缓存，任意两个主机对应的数据缓存是不同的。

由于主机120和NOF控制器140通过网络160相连，而网络160可能会发生拥塞，所以在主机120从NOF控制器140读取数据的过程中，可能会由于网络拥塞而无法正常读取数据。比如在图1中，主机1的接入带宽是100GB，当主机0向主机1发送近似100GB的

数据，NOF 引擎 0 向主机 1 发送近似 100GB 的数据，NOF 引擎 1 向主机 1 发送近似 100GB 的数据时，主机 1 的接入带宽无法同时承受 300GB 数据的传输，会产生拥塞现象。

图 3 示出了一种典型的 NOF 存储系统发生拥塞时的原理示意图。NOF 引擎 142 接收到主机的读命令后，读命令会沿传输路径：完成队列 QP0→提交队列 SQ0→NVMe 硬盘 144 进行传输，由 NVMe 硬盘 144 对该读命令进行处理后，将数据反馈给 NOF 引擎 142。假设 NOF 引擎 142 共接收到 32 个读命令，尚在完成队列 QP0 中缓存的有 16 个读命令，尚在提交队列 SQ0 中缓存的有 8 个读命令，NVMe 硬盘 144 正在处理的有 4 个读命令，NVMe 硬盘 144 已经处理完 4 个读命令。对于 NVMe 硬盘 144 已经处理完的 4 个读命令，这 4 个读命令会由 NOF 引擎 142 发送至主机。

若此时网络发生了拥塞，则尚未处理的 28 个读请求的数据，很可能无法顺利反馈给主机，导致主机发送的 32 个读命令无法被正确响应。

图 4 示出了本申请一个示例性实施例提供的基于 NOF 的读取控制方法的方法流程图。本实施例以该方法应用于图 1 或图 2 所示的 NOF 存储系统中来举例说明。该方法包括：

步骤 401，主机通过通信链路向 NOF 引擎发送读请求；

通信链路是主机和 NOF 引擎之间的通信链路。可选地，通信链路是主机和 NOF 引擎之间的 QP 连接。一个主机和一个 NOF 引擎之间可以有一个或多个 QP 连接。

可选地，每个 QP 连接具有唯一的通信链路标识，该通信链路标识可以标识为 QPID。一个 QPID 指向一个主机，多个 QPID 可以分别指向不同的主机，也可以指向同一个主机。

主机通过 QP 连接向 NOF 引擎发送读请求，该读请求的封装形式可以是 RDMA 发送命令消息包（RDMA_SEND command capsule），该读请求中携带有待读取数据的地址信息、待读取数据的数据量大小，和序列号（Processing Sequence Number，PSN）。该序列号是按照读请求的先后顺序连续递增的。

可选地，主机发送的读请求是一个或多个。

步骤 402，NOF 引擎通过通信链路接收主机发送的读请求；

NOF 引擎将接收到的读请求缓存至与该通信链路对应的完成队列中。

步骤 403，NOF 引擎根据读请求向 NVMe 硬盘发送至少一个读命令；

可选地，当读请求所请求的数据量较小时，NOF 引擎可以将完成队列中的一个读请求转换为一个读命令后，缓存至提交队列中；当读请求所请求的数据量较大时，NOF 引擎可以将完成队列中的一个读请求转换为多个读命令后，缓存至提交队列中。比如，读请求所请求的数据量是 1MB，被拆分为 8 个读命令，每个读命令所请求的数据量是 128KB。

可选地，该读命令具有与读请求相同的 PSN 号。

步骤 404，NVMe 硬盘接收 NOF 引擎发送的至少一个读命令；

NVMe 硬盘定期对提交队列进行读取，当提交队列中存在未处理的读命令时，将该读命令进行读取并处理。

当 NVMe 硬盘根据读命令读取到数据后，会将携带有数据和 PSN 的数据帧反馈给 NOF 引擎。由 NOF 引擎通过通信链路将该数据帧发送给主机。也即，该数据帧用于反馈已处理的读命令对应的数据。

可选地，主机和 NOF 引擎之间的信息交互符合 RDMA 协议。

步骤 405，NOF 引擎在通信链路发生拥塞时，生成通信链路对应的拥塞标记；

当主机正确接收该数据帧时，会向 NOF 引擎反馈具有相同 PSN 的确认（Acknowledgement，ACK）信号；当主机未正确接收该数据帧时，会向 NOF 引擎反馈具有相同 PSN 的否认（Non-Acknowledgement，NACK）信号。当主机未收到该数据帧时，不会向 NOF 引擎反馈任何信号，这时会在 NOF 引擎产生定时器超时事件，该定时器是 NOF 引擎从开始发送数据帧时进行计时的定时器。

根据 RDMA 协议，当 NOF 引擎收到该数据帧的 NACK 或者超时未接收到确认信号，则确定该通信链路发生了拥塞，NOF 引擎生成该通信链路对应的拥塞标记。该拥塞标记可以记为 FC。

可选地，该拥塞标记可以显式或隐式指示该通信链路的标识，比如，该拥塞标记中携带有该通信链路的标识，该标识是 QPID。

步骤 406，NOF 引擎向 NVMe 硬盘发送拥塞标记；

由于 NVMe 硬盘可能已经读取了一部分与该通信链路对应的读命令进行处理，而这些读命令如果被 NVMe 硬盘继续处理的话，很可能因为网络拥塞而失败，所以 NOF 引擎向 NVMe 硬盘发送拥塞标记，该拥塞标记用于指示 NVMe 硬盘暂停处理通信链路对应的读命令。

可选地，NOF 引擎暂停接收来自该通信链路的读请求。

步骤 407，NVMe 硬盘接收 NOF 引擎发送的拥塞标记；

步骤 408，NVMe 硬盘根据拥塞标记，暂停处理该通信链路对应的读命令。

当 NVMe 硬盘中存在已经获取且尚未处理的读命令时，暂停处理该通信链路对应的读命令。

综上所述，本实施例提供的基于 NOF 的读取控制方法，通过 NOF 引擎在与主机之间的通信链路发生拥塞时，向 NVMe 硬盘发送拥塞标记，由 NVMe 硬盘根据拥塞标记暂停处理该通信链路对应的读命令。使得在发生网络拥塞的场景下，NVMe 硬盘对已经获取且尚未处理的读命令暂停处理，解决了这部分读命令的数据在读取后，因为网络拥塞而无法顺利发送给主机，导致整个读取过程失败的问题，达到了在网络拥塞场景下减少无效数据的发送，从而避免网络更加拥塞的效果。

在基于图 4 实施例的可选实施例中，网络拥塞可能在一段时间后发生缓解，此时 NOF 引擎可以向 NVMe 硬盘发送拥塞解除标记。也即，步骤 408 之后还可以包括如下步骤 409 至步骤 413，如下图 5 所示：

步骤 409，NOF 引擎通过通信链路向主机间隔发送空闲帧；

可选地，NOF 引擎在发送某一个通信链路的拥塞标记后，每隔预定时间间隔通过该通信链路向主机发送空闲帧。该空闲帧是不包括数据但包括序列号的探测帧。该序列号可以是 NOF 引擎在已经使用过的序列号中选择出的任一序列号。

步骤 410，主机在正确接收到空闲帧时，向 NOF 引擎发送确认信号；

假如网络拥塞已经解除，主机可以正确接收到空闲帧，则主机向 NOF 引擎发送该空闲帧对应的 ACK。该 ACK 中携带有空闲帧的序列号。

步骤 411，NOF 引擎在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标

记；

可选地，该拥塞解除标记可以显式或隐式指示该通信链路的标识，比如，该拥塞标记中携带有该通信链路的标识，该标识是 QPID。

步骤 412，NOF 引擎向 NVMe 硬盘发送拥塞解除标记；

拥塞解除标记用于指示 NVMe 硬盘恢复对通信链路对应的读命令的处理。

对应地，NVMe 硬盘接收 NOF 引擎发送的拥塞解除标记。

步骤 413，NVMe 硬盘根据拥塞解除标记，恢复对该通信链路对应的读命令的处理。

当 NVMe 硬盘中存在已经获取且尚未处理的读命令时，恢复对该通信链路对应的读命令的处理。

综上所述，本实施例提供的基于 NOF 的读取控制方法，通过 NOF 引擎在通信链路的拥塞接触时，向 NVMe 硬盘发送拥塞解除标记，由 NVMe 硬盘根据拥塞解除标记恢复处理该通信链路对应的读命令。使得在网络拥塞已经解除的场景下，NVMe 硬盘对已经获取且尚未处理的读命令恢复处理，保证了主机在读取数据时的成功率。

由于 NOF 引擎与 NVMe 硬盘之间存在多种信息交互方式，上述步骤 406 中，NOF 引擎向 NVMe 硬盘发送拥塞标记（和/或拥塞解除标记）也可以采用如下三种不同的方式中的任意一种：

第一， 通过提交队列中的 SQE 命令进行发送；

第二， 通过寄存器写命令进行发送；

第三， 通过管理（Admin）命令进行发送。

针对第一种可选的实现方式，上述步骤 406 至步骤 413 可实现成为如下步骤，如图 6 所示：

步骤 506，NOF 引擎确定与通信链路对应的提交队列；

提交队列用于存储与通信链路对应的提交队列的条目（Submission Queue Entry，SQE）SQE 命令，SQE 命令包括读命令和/或写命令；

NOF 引擎中存储有通信链路与提交队列之间的对应关系。表一示意性的示出了该对应关系。

表一

通信链路	提交队列
QP ID 0	SQ0
QP ID 1	SQ1
QP ID 2	SQ2
QP ID 3	SQ3

示意性的，当通信链路 0 发生拥塞时，NOF 引擎能够根据该对应关系，确定出与通信链路 0 对应的提交队列 SQ0。该 SQ0 用于存储与通信链路 0 对应的 SQE 命令，SQE 命令包括读命令和/或写命令。可选地，该 SQE 命令是 NOF 引擎根据该通信链路 0 上收到读写请求生成或转发的。

步骤 507，NVMe 硬盘向 NOF 引擎发送命令读取请求，命令读取请求用于读取与通信

链路对应的提交队列中的 SQE 命令。

NVMe 硬盘会每隔预定时间间隔向 NOF 引擎发送命令读取请求，或者，NVMe 硬盘在处理资源空闲时，向 NOF 引擎发送命令读取请求。

该命令读取请求用于读取与通信链路对应的提交队列中的 SQE 命令。

需要说明的是，该读取过程是 NVMe 硬盘正常运行过程中的例行过程，NVMe 硬盘事先并不知晓哪一个通信链路发生了拥塞。

步骤 508，NOF 引擎在接收到 NVMe 硬盘对提交队列的命令读取请求时，向 NVMe 硬盘发送携带有拥塞标记的 SQE 命令；

示意性的，假设通信链路 0 对应的提交队列是 SQ0，NOF 引擎在接收到 NVMe 硬盘对 SQ0 的命令读取请求时，向 NVMe 硬盘发送携带有拥塞标记的 SQE 命令。

可选地，本步骤分为两种情况：

一、 当提交队列中存在待处理的 SQE 命令时，NOF 引擎在该待处理的 SQE 命令的保留域（Reserved）中添加拥塞标记后，将携带有拥塞标记的 SQE 命令发送给 NVMe 硬盘；

二、 当提交队列中不存在待处理的 SQE 命令时，也即提交队列为空，则 NOF 引擎生成一个携带有拥塞标记且不进行实际读写的 SQE 命令（简称空命令），将该携带有拥塞标记的空命令发送给 NVMe 硬盘。

步骤 509，NVMe 硬盘接收 NOF 引擎发送的携带有拥塞标记的 SQE 命令；

步骤 510，NVMe 硬盘在接收到携带有拥塞标记的 SQE 命令后，根据拥塞标记暂停处理该通信链路对应的读命令。

对于已发生拥塞的通信链路所对应的提交队列，NVMe 硬盘根据拥塞标记执行如下操作中的至少一种操作：

1、暂停对该提交队列的命令读取请求；

NVMe 硬盘暂停对该提交队列进行周期性的命令读取过程，从而减少 NVMe 硬盘的冗余读取。

2、将从该提交队列已经取回且尚未完成的读命令设置为挂起状态；

NVMe 硬盘对从该提交队列已经取回且尚未完成的读命令暂不处理，全部设置为挂起状态。

3、将该提交队列对应的调度权重降低至第一权重。

NVMe 硬盘可以对每个提交队列设置有调度权重（也称调度优先级）。第一权重可以是最低调度权重或者较低调度权重。

可选地，当该提交队列对应的通信链路发生拥塞时，NVMe 硬盘将该提交队列对应的调度权重，降低至最低调度权重。因此，其它提交队列对应的 SQE 命令会被优先调度。

本实施例以 NOF 引擎将三种操作全部执行为例来说明。

步骤 511，NOF 引擎通过通信链路向主机间隔发送空闲帧；

示意性的，在通信链路 0 发生拥塞后，NOF 引擎可以每隔预定时间间隔，通过通信链路 0 向主机发送空闲帧。该空闲帧是不包括数据但包括序列号的探测帧。该序列号可以是 NOF 引擎自定义的序列号，还可以是最近一个未收到确认信号的序列号。

步骤 512，NOF 引擎在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记；

假如网络拥塞已经解除，主机可以正确接收到空闲帧，则主机向 NOF 引擎发送该空闲帧对应的 ACK。该 ACK 中携带有空闲帧的序列号。

NOF 引擎在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记。可选地，该拥塞解除标记可以显式或隐式指示该通信链路的标识，比如，该拥塞标记中携带有该通信链路的标识，该标识是 QPID。又比如，NOF 引擎通过与该通信链路对应的提交队列的标识，隐式地向 NVMe 硬盘指示该通信链路。

步骤 513，NOF 引擎确定与通信链路对应的提交队列，提交队列用于存储与通信链路对应的 SQE 命令；

本步骤可以参考步骤 506，不再赘述。

步骤 514，NOF 引擎在接收到 NVMe 硬盘对提交队列的命令读取请求时，向 NVMe 硬盘发送携带有拥塞解除标记的 SQE 命令；

示意性的，假设通信链路 0 对应的提交队列是 SQ0，由于 NVMe 硬盘已经停止读取该提交队列，NOF 引擎可以主动向 NVMe 硬盘发送携带有拥塞解除标记的 SQE 命令。

可选地，本步骤分为两种情况：

当提交队列中存在待处理的 SQE 命令时，NOF 引擎在该待处理的 SQE 命令的保留域（Reserved）中添加拥塞解除标记后，将携带有拥塞解除标记的 SQE 命令发送给 NVMe 硬盘；

当提交队列中不存在待处理的 SQE 命令时，也即提交队列为空，则 NOF 引擎生成一个携带有拥塞解除标记且不进行实际读写的 SQE 命令（简称空命令），将该携带有拥塞标记的空命令发送给 NVMe 硬盘。

步骤 515，NVMe 硬盘接收 NOF 引擎发送的携带有拥塞解除标记的 SQE 命令；

步骤 516，NVMe 硬盘根据拥塞解除标记，恢复对通信链路对应的读命令的处理。

对于拥塞已经解除的通信链路所对应的提交队列，NVMe 硬盘根据拥塞解除标记执行如下操作中的至少一种操作：

一、恢复对提交队列的命令读取请求；

NVMe 硬盘恢复对该提交队列进行周期性的命令读取过程，从而继续对该提交队列中的 SQE 命令进行处理。

二、将从提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态；

NVMe 硬盘还将从该提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态。

三、将该提交队列对应的调度权重提高至第二权重。

NVMe 硬盘可以对每个提交队列设置有调度权重（也称调度优先级）。第二权重可以是该提交队列在调整为第一权重之前的权重，或者，第二权重是预设的最高权重或者较高权重。也即，第二权重高于第一权重。

可选地，当该提交队列对应的通信链路的拥塞解除时，NVMe 硬盘将该提交队列对应的调度权重，调高至第二权重。因此，该提交队列对应的 SQE 命令会被优先调度或继续调度。

综上所述，本实施例提供的读取控制方法，通过利用提交队列中的 SQE 命令来传递拥塞标记，使得在尽量不增加额外信令的前提下，实现 NOF 引擎与 NVMe 硬盘之间的拥塞通

知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。由于提交队列中有很大概率存在 NVMe 硬盘还未读取的 SQE 命令，而利用 SQE 命令中的保留域来传递拥塞标记，几乎没有增加额外信令和数据量。

本实施例提供的读取控制方法，通过利用提交队列中的 SQE 命令来传递拥塞解除标记，使得在尽量不增加额外信令的前提下，实现 NOF 引擎与 NVMe 硬盘之间的拥塞解除通知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。由于提交队列中有很大概率存在 NVMe 硬盘还未读取的 SQE 命令，而利用 SQE 命令中的保留域来传递拥塞标记，几乎没有增加额外信令和数据量。

针对第二种可选的实现方式，上述步骤 406 至步骤 413 可实现成为如下步骤，如图 7 所示：

步骤 606，NOF 引擎确定与通信链路对应的提交队列，提交队列用于存储与通信链路对应的 SQE 命令，SQE 命令包括读命令和/或写命令；

NOF 引擎中存储有通信链路与提交队列之间的对应关系。示意性的，当通信链路 0 发生拥塞时，NOF 引擎能够根据该对应关系，确定出与通信链路 0 对应的提交队列 SQ0。该 SQ0 用于存储与通信链路 0 对应的 SQE 命令，SQE 命令包括读命令和/或写命令。可选地，该 SQE 命令是 NOF 引擎根据该通信链路 0 上收到读写请求生成或转发的。

步骤 607，NOF 引擎向 NVMe 硬盘发送指向第一自定义地址的寄存器写命令，第一自定义地址是用于存储所述拥塞标记的地址，该寄存器写命令携带有提交队列的标识和拥塞标记；

示意性的，假设通信链路 0 对应的提交队列是 SQ0，NOF 引擎向 NVMe 硬盘发送指向第一自定义地址的寄存器写命令。第一自定义地址是用于存储拥塞标记的地址，该第一自定义地址是 NVMe 硬盘的寄存器中的存储地址。该寄存器写命令携带有提交队列的标识和拥塞标记。每个提交队列可以有各自对应的第一自定义地址，也可以多个提交队列共用同一个第一自定义地址。

也即，该第一自定义地址是 NOF 引擎和 NVMe 硬盘之间事先设置的一个寄存器地址，位于 NVMe 硬盘内，该第一自定义地址用来存储拥塞标记。

步骤 608，NVMe 硬盘接收 NOF 引擎发送的指向第一自定义地址的寄存器写命令；

步骤 609，NVMe 硬盘根据寄存器写命令对第一自定义地址进行写入操作；

NVMe 硬盘根据该寄存器写命令，将提交队列的标识和拥塞标记写入第一自定义地址中。之后，NVMe 硬盘可以根据该提交队列的标识和拥塞标记，获知该提交队列对应的通信链路已经发生了拥塞。

需要说明的是，当多个提交队列共同同一个第一自定义地址时，若 NVMe 硬盘需要写入第 i 个指向第一自定义地址的寄存器写命令，$i \geq 2$，则 NVMe 硬盘可以将第一自定义地址中的已有数据进行擦除，然后再根据第 i 个指向第一自定义地址的寄存器写命令进行写入，以避免已有数据对本次写入过程的影响。

步骤 610，NVMe 硬盘在向第一自定义地址进行写入后，根据拥塞标记暂停处理该通信链路对应的读命令。

对于已发生拥塞的通信链路所对应的提交队列，NVMe 硬盘根据拥塞标记执行如下操

作中的至少一种操作：

1、暂停对该提交队列的命令读取请求；

NVMe 硬盘暂停对该提交队列进行周期性的命令读取过程，从而减少 NVMe 硬盘的冗余读取。

2、将从该提交队列已经取回且尚未完成的读命令设置为挂起状态；

NVMe 硬盘对从该提交队列已经取回且尚未完成的读命令暂不处理，全部设置为挂起状态。

3、将该提交队列对应的调度权重降低至第一权重。

NVMe 硬盘可以对每个提交队列设置有调度权重（也称调度优先级）。第一权重可以是最低调度权重或者较低调度权重。

可选地，当该提交队列对应的通信链路发生拥塞时，NVMe 硬盘将该提交队列对应的调度权重，降低至最低调度权重。因此，其它提交队列对应的 SQE 命令会被优先调度。

本实施例以 NOF 引擎将三种操作全部执行为例来说明。

步骤 611，NOF 引擎通过通信链路向主机间隔发送空闲帧；

示意性的，在通信链路 0 发生拥塞后，NOF 引擎可以每隔预定时间间隔，通过通信链路 0 向主机发送空闲帧。该空闲帧是不包括数据但包括序列号的探测帧。该序列号可以是 NOF 引擎自定义的序列号，还可以是最近一个未收到确认信号的序列号。

步骤 612，NOF 引擎在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记；

假如网络拥塞已经解除，主机可以正确接收到空闲帧，则主机向 NOF 引擎发送该空闲帧对应的 ACK。该 ACK 中携带有空闲帧的序列号。

NOF 引擎在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记。可选地，该拥塞解除标记可以显式或隐式指示该通信链路的标识，比如，该拥塞标记中携带有该通信链路的标识，该标识是 QPID。又比如，NOF 引擎通过与该通信链路对应的提交队列的标识，隐式地向 NVMe 硬盘指示该通信链路。

步骤 613，NOF 引擎向 NVMe 硬盘发送指向第二自定义地址的寄存器写命令；

第二自定义地址是用于存储拥塞解除标记的地址，该寄存器写命令携带有提交队列的标识和拥塞解除标记。可选地，第二自定义地址与第一自定义地址是不同的地址，或者，第二自定义地址与第一自定义地址是相同的地址。每个提交队列可以对应有各自的第二自定义地址，也可以多个提交队列共用同一个第二自定义地址。

示意性的，假设通信链路 0 对应的提交队列是 SQ0，NOF 引擎可以主动向 NVMe 硬盘发送指向第二自定义地址的寄存器写命令。

步骤 614，NVMe 硬盘接收 NOF 引擎发送的指向第二自定义地址的寄存器写命令；

步骤 615，NVMe 硬盘根据寄存器写命令对第二自定义地址进行写入操作；

NVMe 硬盘根据该寄存器写命令，将提交队列的标识和拥塞解除标记写入第二自定义地址中。之后，NVMe 硬盘可以根据该提交队列的标识和拥塞解除标记，获知该提交队列对应的通信链路的拥塞已经解除。

需要说明的是，当多个提交队列共同同一个第二自定义地址时，若 NVMe 硬盘需要写入第 i 个指向第二自定义地址的寄存器写命令，$i \geq 2$，则 NVMe 硬盘可以将第二自定义地址

中的已有数据进行擦除，然后再根据第 i 个指向第二自定义地址的寄存器写命令进行写入，以避免已有数据对本次写入过程的影响。

同理，对第一自定义地址和第二自定义地址是同一个地址的场景下，若 NVMe 硬盘需要写入第 i 个指向第二自定义地址的寄存器写命令，i≥2，则 NVMe 硬盘也可以将第二自定义地址中的已有数据进行擦除，然后再根据第 i 个指向第二自定义地址的寄存器写命令进行写入，以避免已有数据对本次写入过程的影响。

步骤 616，NVMe 硬盘根据拥塞解除标记，恢复对通信链路对应的读命令的处理。

对于拥塞已经解除的通信链路所对应的提交队列，NVMe 硬盘根据拥塞解除标记执行如下操作中的至少一种操作：

一、恢复对提交队列的命令读取请求；

NVMe 硬盘恢复对该提交队列进行周期性的命令读取过程，从而继续对该提交队列中的 SQE 命令进行处理。

二、将从提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态；

NVMe 硬盘还将从该提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态。

三、将该提交队列对应的调度权重提高至第二权重。

NVMe 硬盘可以对每个提交队列设置有调度权重（也称调度优先级）。第二权重可以是该提交队列在调整为第一权重之前的权重，或者，第二权重是预设的最高权重或者较高权重。也即，第二权重高于第一权重。

可选地，当该提交队列对应的通信链路的拥塞解除时，NVMe 硬盘将该提交队列对应的调度权重，调高至第二权重。因此，该提交队列对应的 SQE 命令会被优先调度或继续调度。

综上所述，本实施例提供的读取控制方法，通过利用指向第一自定义地址的寄存器写命令来传递拥塞标记，使得在不改变 SQE 命令的正常读写机制时，也能够实现 NOF 引擎与 NVMe 硬盘之间的拥塞通知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。

本实施例提供的读取控制方法，通过利用指向第二自定义地址的寄存器写命令来传递拥塞解除标记，使得在不改变 SQE 命令的正常读写机制时，实现 NOF 引擎与 NVMe 硬盘之间的拥塞解除通知机制，进而控制 NVMe 硬盘恢复对该通信链路对应的读命令进行处理。

针对第三种可选的实现方式，上述步骤 406 至步骤 413 可实现成为如下步骤，如图 8 所示：

步骤 706，NOF 引擎确定与通信链路对应的提交队列，提交队列用于存储与通信链路对应的提交队列的条目（Submission Queue Entry，SQE）SQE 命令，SQE 命令包括读命令和/或写命令；

NOF 引擎中存储有通信链路与提交队列之间的对应关系。示意性的，当通信链路 0 发生拥塞时，NOF 引擎能够根据该对应关系，确定出与通信链路 0 对应的提交队列 SQ0。该 SQ0 用于存储与通信链路 0 对应的 SQE 命令，SQE 命令包括读命令和/或写命令。可选地，该 SQE 命令是 NOF 引擎根据该通信链路 0 上收到读写请求生成或转发的。

步骤 707，NOF 引擎向 NVMe 硬盘发送具有第一自定义内容的管理命令，第一自定义内容携带有提交队列的标识和拥塞标记；

示意性的，假设通信链路 0 对应的提交队列是 SQ0，NOF 引擎向 NVMe 硬盘发送具有第一自定义内容的管理命令。第一自定义内容携带有提交队列的标识和拥塞标记。由于管理命令

步骤 708，NVMe 硬盘接收 NOF 引擎发送的具有第一自定义内容的管理命令，第一自定义内容携带有提交队列的标识和拥塞标记；

步骤 709，NVMe 硬盘在收到具有第一自定义内容的管理命令后，根据拥塞标记暂停处理该通信链路对应的读命令。

可选地，NVMe 硬盘从该管理命令中，解析出提交队列的标识和拥塞标记。

对于已发生拥塞的通信链路所对应的提交队列，NVMe 硬盘根据拥塞标记执行如下操作中的至少一种操作：

1、暂停对该提交队列的命令读取请求；

NVMe 硬盘暂停对该提交队列进行周期性的命令读取过程，从而减少 NVMe 硬盘的冗余读取。

2、将从该提交队列已经取回且尚未完成的读命令设置为挂起状态；

NVMe 硬盘对从该提交队列已经取回且尚未完成的读命令暂不处理，全部设置为挂起状态。

3、将该提交队列对应的调度权重降低至第一权重。

NVMe 硬盘可以对每个提交队列设置有调度权重（也称调度优先级）。第一权重可以是最低调度权重或者较低调度权重。

可选地，当该提交队列对应的通信链路发生拥塞时，NVMe 硬盘将该提交队列对应的调度权重，降低至最低调度权重。因此，其它提交队列对应的 SQE 命令会被优先调度。

本实施例以 NOF 引擎将三种操作全部执行为例来说明。

步骤 710，NOF 引擎通过通信链路向主机间隔发送空闲帧；

示意性的，在通信链路 0 发生拥塞后，NOF 引擎可以每隔预定时间间隔，通过通信链路 0 向主机发送空闲帧。该空闲帧是不包括数据但包括序列号的探测帧。该序列号可以是 NOF 引擎自定义的序列号，还可以是最近一个未收到确认信号的序列号。

步骤 711，NOF 引擎在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记；

假如网络拥塞已经解除，主机可以正确接收到空闲帧，则主机向 NOF 引擎发送该空闲帧对应的 ACK。该 ACK 中携带有空闲帧的序列号。

NOF 引擎在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记。可选地，该拥塞解除标记可以显式或隐式指示该通信链路的标识，比如，该拥塞标记中携带有该通信链路的标识，该标识是 QPID。又比如，NOF 引擎通过与该通信链路对应的提交队列的标识，隐式地向 NVMe 硬盘指示该通信链路。

步骤 712，NOF 引擎向 NVMe 硬盘发送具有第二自定义内容的管理命令；

第二自定义地址是用于存储拥塞解除标记的地址，该寄存器写命令携带有提交队列的标识和拥塞解除标记。可选地，第二自定义地址与第一自定义地址是不同的地址，或者，

第二自定义地址与第一自定义地址是相同的地址。每个提交队列可以对应有各自的第二自定义地址，也可以多个提交队列共用同一个第二自定义地址。

示意性的，假设通信链路 0 对应的提交队列是 SQ0，NOF 引擎可以主动向 NVMe 硬盘发送具有第二自定义内容的管理命令。

步骤 713，NVMe 硬盘接收 NOF 引擎发送的具有第二自定义内容的管理命令；

步骤 714，NVMe 硬盘根据拥塞解除标记，恢复对通信链路对应的读命令的处理。

对于拥塞已经解除的通信链路所对应的提交队列，NVMe 硬盘根据拥塞解除标记执行如下操作中的至少一种操作：

一、恢复对提交队列的命令读取请求；

NVMe 硬盘恢复对该提交队列进行周期性的命令读取过程，从而继续对该提交队列中的 SQE 命令进行处理。

二、将从提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态；

NVMe 硬盘还将从该提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态。

三、将该提交队列对应的调度权重提高至第二权重。

NVMe 硬盘可以对每个提交队列设置有调度权重（也称调度优先级）。第二权重可以是该提交队列在调整为第一权重之前的权重，或者，第二权重是预设的最高权重或者较高权重。也即，第二权重高于第一权重。

可选地，当该提交队列对应的通信链路的拥塞解除时，NVMe 硬盘将该提交队列对应的调度权重，调高至第二权重。因此，该提交队列对应的 SQE 命令会被优先调度或继续调度。

综上所述，本实施例提供的读取控制方法，通过利用具有第一自定义内容的管理命令来传递拥塞标记，使得利用 NVMe 协议提供的管理命令，即可实现 NOF 引擎与 NVMe 硬盘之间的拥塞通知机制，进而控制 NVMe 硬盘暂停对该通信链路对应的读命令进行处理。

本实施例提供的读取控制方法，通过利用具有第二自定义内容的管理命令来传递拥塞解除标记，使得利用 NVMe 协议提供的管理命令，即可实现 NOF 引擎与 NVMe 硬盘之间的拥塞解除通知机制，进而控制 NVMe 硬盘恢复对该通信链路对应的读命令进行处理。

上述三个实施例，可以适用于 NOF 引擎 142 中不存在数据缓存的场景，也可以适用于 NOF 引擎 142 中存储数据缓存的场景。可选地，对于 NOF 引擎 142 中存在数据缓存的场景，还可以提供有如下实施例：

图 9 示出了本申请一个示例性实施例提供的基于 NOF 的读取控制方法的流程图。本实施例以该方法应用于图 1 或图 2 所示的 NOF 存储系统中来举例说明，设 NOF 引擎 142 中设置有数据缓存，该数据缓存的缓存大小为 n MB，NOF 引擎 142 支持的通信链路数量为 m，则分配给每个通信链路的数据缓存的缓存大小是总缓存大小 n 的 1/m。该方法包括：

步骤 801，NOF 引擎通过通信链路接收主机发送的读请求，读请求用于请求读取第一数据量的数据，第一数据量超过数据缓存的缓存大小；

示意性的，假设 NOF 引擎 142 中设置有数据缓存的缓存大小 n 为 1024MB，NOF 引擎 142 支持的 QP 连接数量 m 为 1024 个 QP 连接，则分配给每个 QP 连接的数据缓存为

1MB=1024KB 的数据缓存空间。

主机通过通信链路向 NOF 引擎发送读请求。可选地，该读请求用于读取第一数据量的数据，该第一数据量超过该通信链路对应的数据缓存的缓存大小。例如，第一数据量超过 1MB。

可选地，主机通过 RDMA 协议向 NOF 引擎发送读请求，NOF 引擎通过 RDMA 协议接收主机发送的读请求。

步骤 802，NOF 引擎将读请求拆分为 K 个读命令；

NOF 引擎将第一数据量的读请求拆分为 K 个第二数据量的读命令，每个读命令用于请求读取第二数据量的数据，第二数据量小于第一数据量。可选地，每个读命令对应的第二数据量是相同的，或者，存在至少两个读命令对应的第二数据量是不同的。该第二数据量是可以自定义的。

例如，NOF 引擎将 2MB 的读请求拆分为 16 个 128KB 的读命令。

步骤 803，NOF 引擎向 NVMe 硬盘发送 m 个读命令，m≤K，m 与第二数据量的乘积不大于数据缓存的缓存大小；

NOF 引擎会控制向 NVMe 硬盘发送的读命令个数，使得 NVMe 硬盘正在处理的读命令对应的读取数据量不超过数据缓存的缓存大小。

例如，由于每个 QP 连接的数据缓存大小为 1024KB，NOF 引擎每次向 NVMe 硬盘发送的读命令个数不超过 8 个，8*128KB=1024KB。

可选地，NOF 引擎将需要发送至 NVMe 硬盘的读命令缓存在提交队列中，由 NVMe 硬盘自行读取。NOF 引擎将暂时不需要发送至 NVMe 硬盘的读命令，先缓存在本地的完成队列中，暂不发送至提交队列中。

当 NVMe 硬盘根据读命令将读取到的数据反馈给 NOF 引擎后，NOF 引擎先将读取到的数据进行缓存，然后再以数据帧的形式发送给主机。

步骤 804，NOF 引擎在通信链路发生拥塞时，暂停向 NVMe 硬盘发送剩余的其它读命令；

当 NOF 引擎收到数据帧的否认信号，或者，超时未收到数据帧的确认信号时，确定该条通信链路发生了网络拥塞。该过程可以参考步骤 405 的相关描述细节，本实施例不再赘述。

此时，NOF 引擎暂停向 NVMe 硬盘发送剩余的其它读命令。由于已经向 NVMe 硬盘发送的读命令有限，即便 NVMe 硬盘将已接收的读命令进行处理和响应，NOF 引擎侧的数据缓存也可以将读取到的数据进行缓存，不会产生缓存溢出。

需要说明的是，NVMe 硬盘侧完全不感知网络拥塞过程。

步骤 805，NOF 引擎接收 NVMe 硬盘根据 m 个读命令读取的数据；

在网络拥塞期间，NVMe 硬盘可以继续对该通信链路对应的读命令进行处理，NOF 引擎正常接收 NVMe 硬盘根据 n 个读命令读取的数据。

步骤 806，NOF 引擎将数据存储在通信链路对应的数据缓存中；

NOF 引擎可以将数据存储在通信链路对应的数据缓存中，暂时不向主机发送。

步骤 807，NOF 引擎在通信链路解除拥塞时，将数据缓存中的数据通过通信链路发送至主机；

可选地，NOF 引擎在发送某一个通信链路的拥塞标记后，每隔预定时间间隔通过该通信链路向主机发送空闲帧。该空闲帧是不包括数据但包括序列号的探测帧。该序列号可以是 NOF 引擎自定义的序列号，还可以是最近一个未收到确认信号的序列号。假如网络拥塞已经解除，主机可以正确接收到空闲帧，则主机向 NOF 引擎发送该空闲帧对应的 ACK。该 ACK 中携带有空闲帧的序列号。

NOF 引擎在通信链路解除拥塞时，将数据缓存中的数据通过通信链路发送至主机。

需要说明的是，NVMe 硬盘侧也完全不感知网络拥塞解除过程。

步骤 808，NOF 引擎继续向 NVMe 硬盘发送 n 个读命令，n≤K-m，n 与第二数据量的乘积不大于数据缓存的缓存大小。

当存在该通信链路对应的待处理的读命令时，NOF 引擎继续向 NVMe 硬盘发送后续的 n 个读命令。每个批次向 NVMe 硬盘发送的 n 个读命令对应的读取数据量，不超过该通信链路对应的数据缓存的缓存大小。

综上所述，本实施例提供的基于 NOF 的读取控制方法，可以在 NOF 引擎侧有数据缓存的情况下，由 NOF 引擎控制向 NVMe 硬盘发送的读命令的数量，使得 NVMe 硬盘正在处理的读命令的读取数据量不超过数据缓存的缓存大小。即便通信链路发生了网络拥塞，NOF 引擎暂停向 NVMe 硬盘继续下发读命令即可解决拥塞问题，不需要 NVMe 硬盘感知拥塞过程，减少 NOF 引擎和 NVMe 硬盘之间的信令交互以及传输带宽的占用。

另外，由于本实施例是为每个通信链路设置各自对应的数据缓存，而不是多个通信链路共用同一个数据缓存，因此即便某一个通信链路发生了网络拥塞，导致该通信链路对应的数据缓存被占用，也不会出现将 NOF 引擎中所有的数据缓存都被该通信链路对应的读取数据所占用，导致其它通信链路的读取过程也收到影响的问题。达到了每个通信链路都分别独立，即便一条或多条通信链路发生拥塞，也不会影响剩余通信链路的正常收发过程的效果。

本实施例提供的基于 NOF 的读取控制方法，还可以在通信链路的拥塞解除后，NOF 引擎继续向 NVMe 硬盘下发读命令即可恢复正常处理流程，不需要 NVMe 硬盘感知拥塞解除过程，减少 NOF 引擎和 NVMe 硬盘之间的信令交互以及传输带宽的占用。

需要说明的是，上述各个方法实施例均以一条通信链路发生网络拥塞或拥塞解除时，如何对该条通信链路的读命令进行处理来进行示意性说明。但实际上发生拥塞的通信链路可以不止一条，此乃本领域技术人员在阅读上述实施例后易于思及的内容，本文不再赘述。

图 10 示出了本申请一个实施例提供的基于 NOF 的读取控制装置的结构框图。该装置可以实现成为 NOF 引擎本身或 NOF 引擎中的功能模块。该装置包括：接收单元 920、发送单元 940 和处理单元 960。

接收单元 920，用于实现上述方法实施例中步骤 401、步骤 410、步骤 507、步骤 512、步骤 612、步骤 711、步骤 801、步骤 805 中的至少一个接收步骤，以及其它明示或隐含的接收步骤。

发送单元 940，用于实现上述方法实施例中步骤 403、步骤 406、步骤 409、步骤 412、步骤 508、步骤 511、步骤 514、步骤 607、步骤 612、步骤 707、步骤 712、步骤 803、步骤 807、步骤 808 中的至少一个发送步骤，以及其它明示或隐含的发送步骤。

处理单元960，用于实现上述方法实施例中步骤411、步骤506、步骤513、步骤606、步骤706、步骤802、步骤804、步骤806中的至少一个处理步骤，以及其它明示或隐含的处理步骤。

上述接收单元920可以通过处理器控制接收机（或通信芯片或物理网络接口或虚拟网络接口）来实现，上述发送单元940可以通过处理器控制发送机（或通信芯片或物理网络接口或虚拟网络接口）来实现，上述处理单元960可以通过处理器、存储器和存储器中的至少一个程序或指令来实现。

图11示出了本申请一个实施例提供的基于NOF的读取控制装置的结构框图。该装置可以实现成为NVMe硬盘本身或NVMe硬盘中的功能模块。该装置包括：接收单元1020、发送单元1040和处理单元1060。

接收单元1020，用于实现上述方法实施例中步骤404、步骤407、步骤413、步骤509、步骤515、步骤608、步骤614、步骤708、步骤713中的至少一个接收步骤，以及其它明示或隐含的接收步骤。

发送单元1040，用于实现上述方法实施例中步骤507对应的发送步骤，以及其它明示或隐含的发送步骤。

处理单元106，用于实现上述方法实施例中步骤408、步骤413、步骤510、步骤516、步骤609、步骤610、步骤615、步骤616、步骤709、步骤714中的至少一个处理步骤，以及其它明示或隐含的处理步骤。

上述接收单元1020可以通过处理器控制接收机（或通信芯片或物理网络接口或虚拟网络接口）来实现，上述发送单元1040可以通过处理器控制发送机（或通信芯片或物理网络接口或虚拟网络接口）来实现，上述处理单元1060可以通过处理器、存储器和存储器中的至少一个程序或指令来实现。

图12示出了本申请一个实施例提供的NOF引擎的结构框图。该NOF引擎包括：处理器1220和存储器1240。存储器1240存储有至少一个指令，所述至少一个指令被所述处理器执行以实现如上述任一方法实施例中，由NOF引擎执行的步骤。

可选地，该NOF引擎还包括：网络接口1260和PCI接口1280。其中，网络接口1260用于结合物理线缆与主机相连，PCI接口1280用于与NVMe硬盘相连。

图13示出了本申请一个实施例提供的NVMe硬盘的结构框图。该NVMe硬盘包括：控制器1320和存储器1340。存储器1340存储有至少一个指令，所述至少一个指令被所述处理器执行以实现如上述任一方法实施例中，由NVMe硬盘执行的步骤。可选地，存储器1340也可以集成在控制器1320中。

可选地，该NVMe硬盘还包括：至少一个非易失性存储单元1360，该非易失性存储单元1360用于存储数据。

本申请实施例还提供了一种计算机可读存储介质，所述存储介质存储有至少一个指令，所述至少一个指令被执行以实现如任一方法实施例中，由NOF引擎执行的步骤。

本申请实施例还提供了一种计算机可读存储介质，所述存储介质存储有至少一个指令，所述至少一个指令被执行以实现如任一方法实施例中，由 NVMe 硬盘执行的步骤。

本申请实施例还提供了一种计算机程序产品，所述程序产品存储有至少一个指令，所述至少一个指令被执行以实现如任一方法实施例中，由 NOF 引擎执行的步骤。

本申请实施例还提供了一种计算机程序产品，所述程序产品存储有至少一个指令，所述至少一个指令被执行以实现如任一方法实施例中，由 NVMe 硬盘执行的步骤。

上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

以上所述仅为本申请的可选实施例，并不用以限制本申请，凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权 利 要 求 书

1、一种基于网络连接的非易失性存储 NOF 的读取控制方法，其特征在于，所述方法包括：

NOF 引擎通过通信链路接收主机发送的读请求；

所述 NOF 引擎根据所述读请求向 NVMe 硬盘发送至少一个读命令；

所述 NOF 引擎在所述通信链路发生拥塞时，生成所述通信链路对应的拥塞标记；

所述 NOF 引擎向所述 NVMe 硬盘发送所述拥塞标记，所述拥塞标记用于指示所述 NVMe 硬盘暂停处理所述通信链路对应的所述读命令。

2、根据权利要求 1 所述的方法，其特征在于，所述 NOF 引擎向 NVMe 硬盘发送所述拥塞标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的提交队列的条目 SQE 命令，所述 SQE 命令包括读命令和/或写命令；

所述 NOF 引擎在接收到所述 NVMe 硬盘对所述提交队列的命令读取请求时，向所述 NVMe 硬盘发送携带有所述拥塞标记的 SQE 命令。

3、根据权利要求 1 所述的方法，其特征在于，所述 NOF 引擎向 NVMe 硬盘发送所述拥塞标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；

所述 NOF 引擎向所述 NVMe 硬盘发送指向第一自定义地址的寄存器写命令，所述第一自定义地址是用于存储所述拥塞标记的地址，所述寄存器写命令携带有所述提交队列的标识和所述拥塞标记。

4、根据权利要求 1 所述的方法，其特征在于，所述 NOF 引擎向 NVMe 硬盘发送所述拥塞标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；

所述 NOF 引擎向所述 NVMe 硬盘发送具有第一自定义内容的管理命令，所述第一自定义内容携带有所述提交队列的标识和所述拥塞标记。

5、根据权利要求 1 至 4 任一所述的方法，其特征在于，所述向所述 NVMe 硬盘发送添加有所述拥塞标记的 SQE 命令之后，还包括：

所述 NOF 引擎通过所述通信链路向所述主机间隔发送空闲帧；

所述 NOF 引擎在接收到所述空闲帧的确认信号时，生成所述通信链路对应的拥塞解除标记；

所述 NOF 引擎向所述 NVMe 硬盘发送所述拥塞解除标记，所述拥塞解除标记用于指示所述 NVMe 硬盘恢复对所述通信链路对应的读命令的处理。

6、根据权利要求 5 所述的方法，其特征在于，所述 NOF 引擎向所述 NVMe 硬盘发送所述拥塞解除标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；

所述 NOF 引擎在接收到所述 NVMe 硬盘对所述提交队列的命令读取请求时，向所述 NVMe 硬盘发送携带有所述拥塞解除标记的 SQE 命令。

7、根据权利要求 5 所述的方法，其特征在于，所述 NOF 引擎向所述 NVMe 硬盘发送所述拥塞解除标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；

所述 NOF 引擎向所述 NVMe 硬盘发送指向第二自定义地址的寄存器写命令，所述第二自定义地址是用于存储所述拥塞解除标记的地址，所述寄存器写命令携带有所述提交队列的标识和所述拥塞解除标记。

8、根据权利要求 5 所述的方法，其特征在于，所述 NOF 引擎向所述 NVMe 硬盘发送所述拥塞解除标记，包括：

所述 NOF 引擎确定与所述通信链路对应的提交队列，所述提交队列用于存储与所述通信链路对应的 SQE 命令；

所述 NOF 引擎向所述 NVMe 硬盘发送具有第二自定义内容的管理命令，所述第二自定义内容携带有所述提交队列的标识和所述拥塞解除标记。

9、一种基于 NOF 的读取控制方法，其特征在于，所述方法包括：

NVMe 硬盘接收 NOF 引擎发送的至少一个读命令，所述读命令是所述 NOF 引擎根据主机通过通信链路发送的读请求产生的；

所述 NVMe 硬盘接收 NOF 引擎发送的拥塞标记，所述拥塞标记是所述 NOF 引擎在所述通信链路发送拥塞时发送的；

所述 NVMe 硬盘根据所述拥塞标记，暂停处理所述通信链路对应的所述读命令。

10、根据权利要求 9 所述的方法，其特征在于，所述 NVMe 硬盘接收 NOF 引擎发送的拥塞标记，包括：

所述 NVMe 硬盘向所述 NOF 引擎发送命令读取请求，所述命令读取请求用于读取与所述通信链路对应的提交队列中的 SQE 命令；

所述 NVMe 硬盘接收所述 NOF 引擎发送的携带有所述拥塞标记的 SQE 命令。

11、根据权利要求 9 所述的方法，其特征在于，所述 NVMe 硬盘接收 NOF 引擎发送的拥塞标记，包括：

所述 NVMe 硬盘接收所述 NOF 引擎发送的指向第一自定义地址的寄存器写命令，所述

第一自定义地址是用于存储所述拥塞标记的地址，所述寄存器写命令携带有与所述通信链路对应的提交队列的标识和所述拥塞标记；

所述 NVMe 硬盘根据所述寄存器写命令对所述第一自定义地址进行写入操作。

12、根据权利要求 9 所述的方法，其特征在于，所述 NOF 引擎向 NVMe 硬盘发送所述拥塞标记，包括：

所述 NVMe 硬盘接收所述 NOF 引擎发送的具有第一自定义内容的管理命令，所述第一自定义内容携带有提交队列的标识和所述拥塞标记，所述提交队列是与所述通信链路对应的队列。

13、根据权利要求 9 至 12 任一所述的方法，其特征在于，所述 NVMe 硬盘根据所述拥塞标记，暂停处理所述通信链路对应的读命令，包括：

所述 NVMe 硬盘根据所述拥塞标记执行如下操作中的至少一种操作：

暂停对所述提交队列的命令读取请求；

将从所述提交队列已经取回且尚未完成的读命令设置为挂起状态；

将所述提交队列对应的调度权重降低至第一权重。

14、根据权利要求 9 所述的方法，其特征在于，所述 NVMe 硬盘根据所述拥塞标记，暂停处理所述通信链路对应的读命令之后，还包括：

所述 NVMe 硬盘接收所述 NOF 引擎发送的拥塞解除标记，所述拥塞解除标记是所述 NOF 引擎在与主机之间的通信链路拥塞解除时发送的；

所述 NVMe 硬盘根据所述拥塞解除标记，恢复对所述通信链路对应的读命令的处理。

15、根据权利要求 14 所述的方法，其特征在于，所述 NVMe 硬盘接收所述 NOF 引擎发送的拥塞解除标记，包括：

所述 NVMe 硬盘向所述 NOF 引擎发送命令读取请求，所述命令读取请求用于读取与所述通信链路对应的提交队列中的 SQE 命令；

所述 NVMe 硬盘接收所述 NOF 引擎发送的携带有所述拥塞解除标记的 SQE 命令。

16、根据权利要求 14 所述的方法，其特征在于，所述 NVMe 硬盘接收所述 NOF 引擎发送的拥塞解除标记，包括：

所述 NVMe 硬盘接收所述 NOF 引擎发送的指向第二自定义地址的寄存器写命令，所述第二自定义地址是用于存储所述拥塞解除标记的地址，所述寄存器写命令携带有与所述通信链路对应提交队列的标识和所述拥塞解除标记；

所述 NVMe 硬盘根据所述寄存器写命令对所述第二自定义地址进行写入操作。

17、根据权利要求 14 所述的方法，其特征在于，所述 NVMe 硬盘接收所述 NOF 引擎发送的拥塞解除标记，包括：

所述 NVMe 硬盘接收所述 NOF 引擎发送的具有第二自定义内容的管理命令，所述第二

自定义内容携带有与所述通信链路对应提交队列的标识和所述拥塞解除标记。

18、根据权利要求 15 至 17 任一所述的方法，其特征在于，所述 NVMe 硬盘根据所述拥塞解除标记，恢复处理所述通信链路对应的读命令，包括：

所述 NVMe 硬盘根据所述拥塞解除标记执行如下操作中的至少一种操作：

恢复对所述提交队列的命令读取请求；

将从所述提交队列已经取回且尚未完成的读命令从挂机状态切换为执行状态；

将所述提交队列对应的调度权重提高至第二权重。

19、一种基于 NOF 的读取控制方法，其特征在于，应用于 NOF 引擎中，所述 NOF 引擎中设置有与通信链路对应的数据缓存，所述方法包括：

所述 NOF 引擎通过所述通信链路接收主机发送的读请求，所述读请求用于请求读取第一数据量的数据，所述第一数据量超过所述通信链路对应的数据缓存大小；

所述 NOF 引擎将所述读请求拆分为 K 个读命令，所述读命令用于请求读取第二数据量的数据，所述第二数据量小于所述第一数据量；

所述 NOF 引擎向 NVMe 硬盘发送 m 个所述读命令，m≤K，m 与所述第二数据量的乘积不大于所述通信链路对应的数据缓存大小；

所述 NOF 引擎在所述通信链路发生拥塞时，暂停向所述 NVMe 硬盘发送剩余的其它读命令。

20、根据权利要求 19 所述的方法，其特征在于，所述 NOF 引擎在所述通信链路发生拥塞时，暂停向所述 NVMe 硬盘发送剩余的其它读命令之后，还包括：

所述 NOF 引擎接收所述 NVMe 硬盘根据所述 m 个读命令读取的数据；

所述 NOF 引擎将所述数据存储在所述通信链路对应的所述数据缓存中；

所述 NOF 引擎在所述通信链路解除拥塞时，将所述数据缓存中的所述数据通过所述通信链路发送至所述主机。

21、根据权利要求 20 所述的方法，其特征在于，所述 NOF 引擎在所述通信链路解除拥塞时，将所述数据缓存中的所述数据通过所述通信链路发送至所述主机之后，还包括：

所述 NOF 引擎继续向所述 NVMe 硬盘发送 n 个所述读命令，n≤K-m，n 与所述第二数据量的乘积不大于所述通信链路对应的数据缓存大小。

22、一种基于 NOF 的读取控制装置，其特征在于，所述装置包括：

接收单元，用于通过通信链路接收主机发送的读请求；

发送单元，用于根据所述读请求向 NVMe 硬盘发送至少一个读命令；

处理单元，用于在所述通信链路发生拥塞时，生成所述通信链路对应的拥塞标记；

所述发送单元，用于向所述 NVMe 硬盘发送所述拥塞标记，所述拥塞标记用于指示所述 NVMe 硬盘暂停处理所述通信链路对应的所述读命令。

23、一种基于 NOF 的读取控制装置，其特征在于，所述装置包括：

接收单元，用于接收 NOF 引擎发送的至少一个读命令，所述读命令是所述 NOF 引擎根据主机通过通信链路发送的读请求产生的；

所述接收单元，用于接收 NOF 引擎发送的拥塞标记，所述拥塞标记是所述 NOF 引擎在所述通信链路发送拥塞时发送的；

处理单元，用于根据所述拥塞标记，暂停处理所述通信链路对应的所述读命令。

24、一种基于 NOF 的读取控制装置，其特征在于，所述装置中设置有与通信链路对应的数据缓存，所述装置包括：

接收单元，用于通过所述通信链路接收主机发送的读请求，所述读请求用于请求读取第一数据量的数据，所述第一数据量超过所述通信链路对应的数据缓存大小；

处理单元，用于将所述读请求拆分为 K 个读命令，所述读命令用于请求读取第二数据量的数据，所述第二数据量小于所述第一数据量；

发送单元，用于向 NVMe 硬盘发送 m 个所述读命令，m≤K，m 与所述第二数据量的乘积不大于所述通信链路对应的数据缓存大小；

处理单元，还用于在所述通信链路发生拥塞时，暂停向所述 NVMe 硬盘发送剩余的其它读命令。

25、一种 NOF 引擎，其特征在于，所述 NOF 引擎包括：处理器和存储器，所述存储器存储有至少一个指令，所述至少一个指令被所述处理器执行以实现如下步骤：

通过通信链路接收主机发送的读请求；

根据所述读请求向 NVMe 硬盘发送至少一个读命令；

在所述通信链路发生拥塞时，生成所述通信链路对应的拥塞标记；

向所述 NVMe 硬盘发送所述拥塞标记，所述拥塞标记用于指示所述 NVMe 硬盘暂停处理所述通信链路对应的所述读命令。

26、一种 NVMe 硬盘，其特征在于，所述 NVMe 硬盘包括：控制器和存储器，所述存储器存储有至少一个指令，所述至少一个指令被所述控制器执行以实现如下步骤：

接收 NOF 引擎发送的至少一个读命令，所述读命令是所述 NOF 引擎根据主机通过通信链路发送的读请求产生的；

接收 NOF 引擎发送的拥塞标记，所述拥塞标记是所述 NOF 引擎在所述通信链路发送拥塞时发送的；

根据所述拥塞标记，暂停处理所述通信链路对应的所述读命令。

27、一种 NOF 引擎，其特征在于，所述 NOF 引擎包括：处理器和存储器，所述存储器包括与通信链路对应的数据缓存，所述存储器存储有至少一个指令，所述至少一个指令被所述处理器执行以实现如下步骤：

通过所述通信链路接收主机发送的读请求，所述读请求用于请求读取第一数据量的数据，所述第一数据量超过所述通信链路的数据缓存大小；

将所述读请求拆分为 K 个读命令，所述读命令用于请求读取第二数据量的数据，所述第二数据量小于所述第一数据量；

向 NVMe 硬盘发送 m 个所述读命令，m≤K，m 与所述第二数据量的乘积不大于所述通信链路的数据缓存大小；

在所述通信链路发生拥塞时，暂停向所述 NVMe 硬盘发送剩余的其它读命令。

28、一种计算机可读存储介质，其特征在于，所述存储介质存储有至少一个指令，所述至少一个指令被执行时实现如下步骤：

通过通信链路接收主机发送的读请求；

根据所述读请求向 NVMe 硬盘发送至少一个读命令；

在所述通信链路发生拥塞时，生成所述通信链路对应的拥塞标记；

向所述 NVMe 硬盘发送所述拥塞标记，所述拥塞标记用于指示所述 NVMe 硬盘暂停处理所述通信链路对应的所述读命令。

29、一种计算机可读存储介质，其特征在于，所述存储介质存储有至少一个指令，所述至少一个指令被执行时实现如下步骤：

接收 NOF 引擎发送的至少一个读命令，所述读命令是所述 NOF 引擎根据主机通过通信链路发送的读请求产生的；

接收 NOF 引擎发送的拥塞标记，所述拥塞标记是所述 NOF 引擎在所述通信链路发送拥塞时发送的；

根据所述拥塞标记，暂停处理所述通信链路对应的所述读命令。

30、一种计算机可读存储介质，其特征在于，所述存储介质存储有至少一个指令，所述至少一个指令被执行时实现如下步骤：

通过通信链路接收主机发送的读请求，所述读请求用于请求读取第一数据量的数据，所述第一数据量超过所述数据缓存的缓存大小；

将所述读请求拆分为 K 个读命令，所述读命令用于请求读取第二数据量的数据，所述第二数据量小于所述第一数据量；

向 NVMe 硬盘发送 m 个所述读命令，m≤K，m 与所述第二数据量的乘积不大于所述数据缓存的缓存大小；

在所述通信链路发生拥塞时，暂停向所述 NVMe 硬盘发送剩余的其它读命令。

附 图

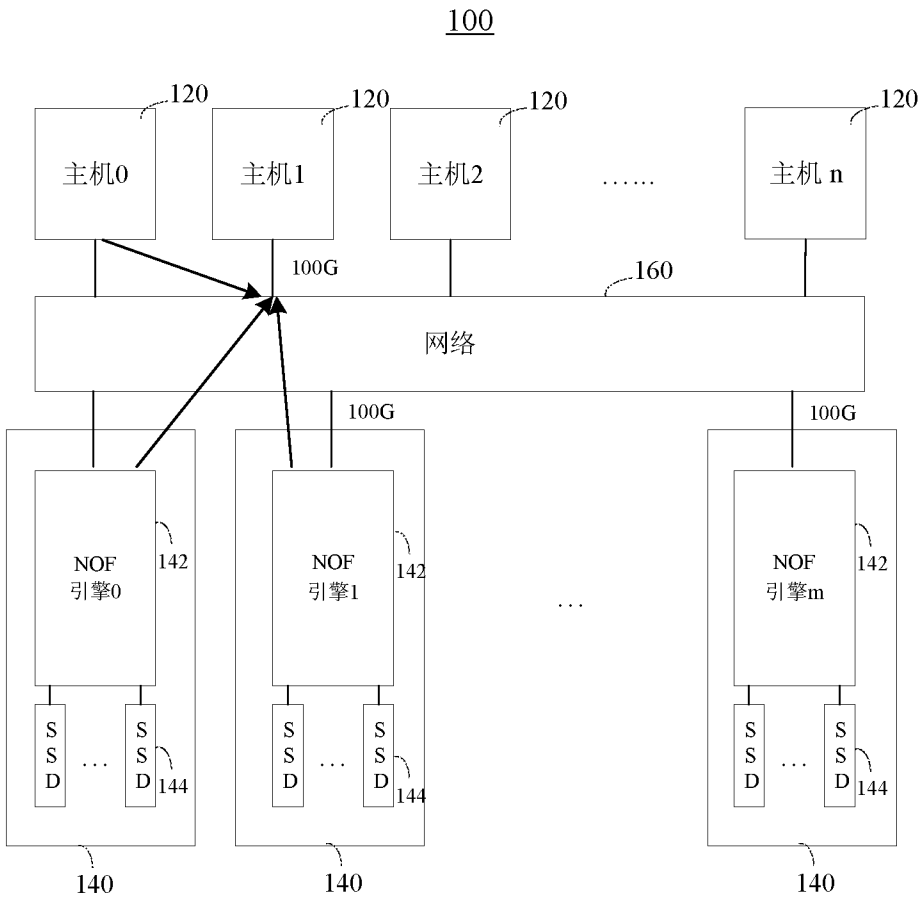

100
120
主机0
主机1
主机2
……
主机 n
100G
160
网络
NOF
引擎0
NOF
引擎1
NOF
引擎m
142
SSD
144
140

图 1

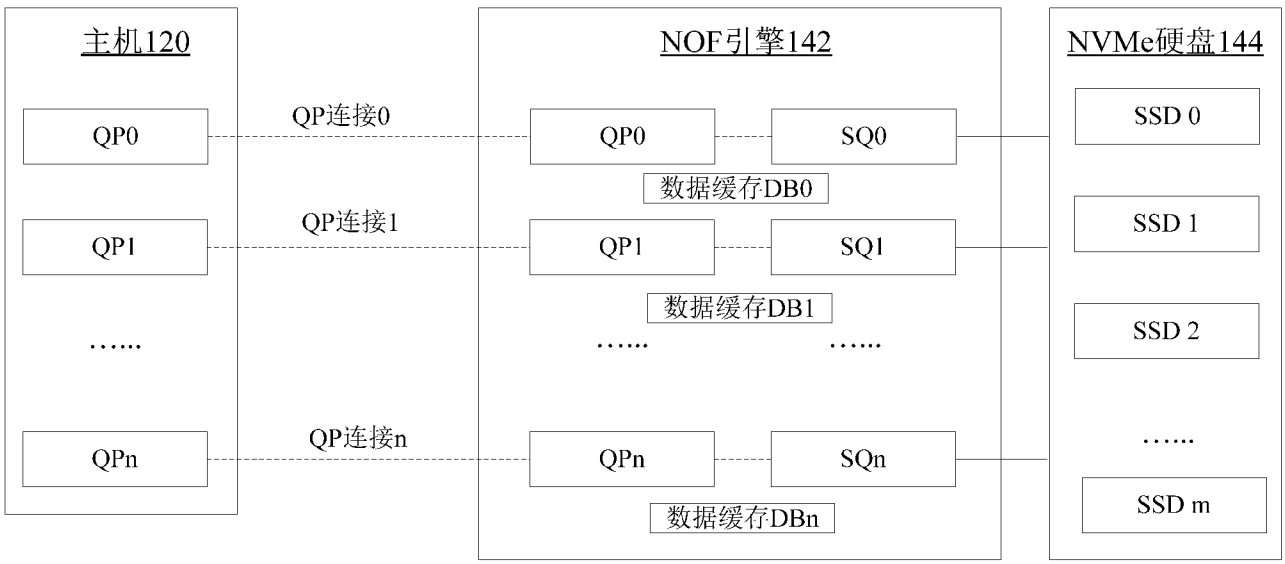

主机120
NOF引擎142
NVMe硬盘144
QP0
QP1
……
QPn
QP连接0
QP连接1
QP连接n
SQ0
SQ1
SQn
数据缓存DB0
数据缓存DB1
数据缓存DBn
SSD 0
SSD 1
SSD 2
SSD m

图 2

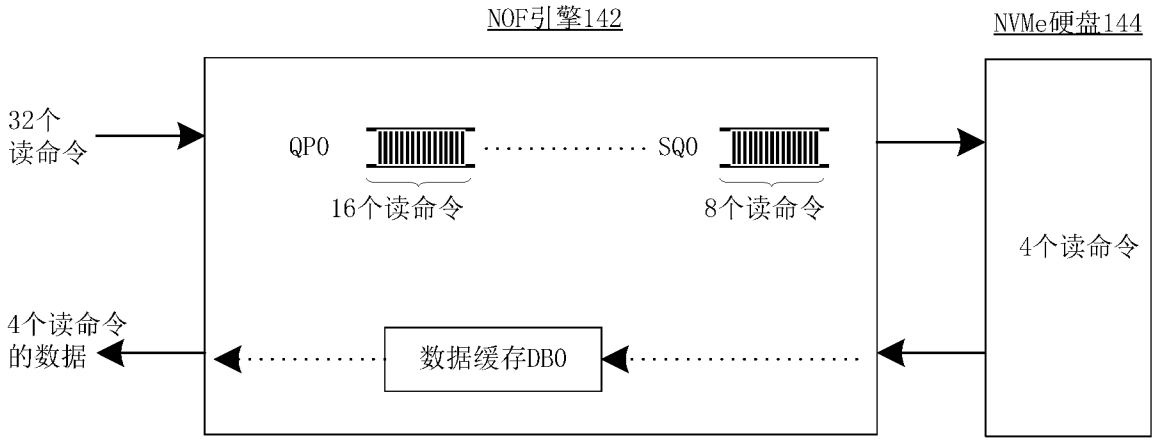

NOF引擎142
NVMe硬盘144
32个
读命令
QP0
SQ0
16个读命令
8个读命令
4个读命令
4个读命令
的数据
数据缓存DB0

图 3

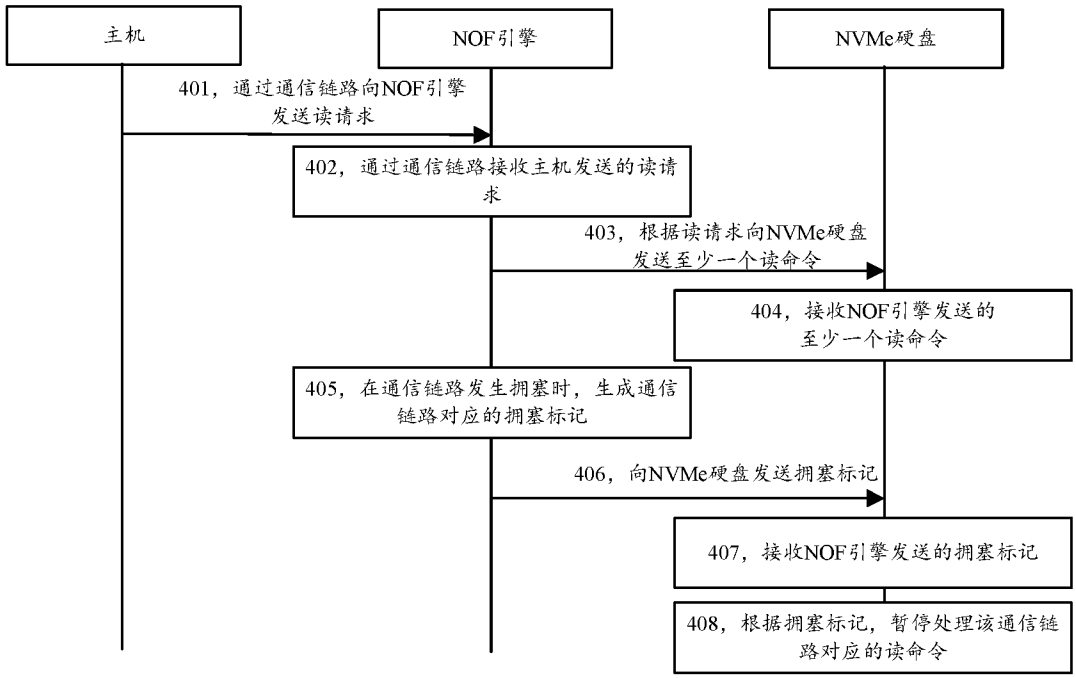

主机
NOF引擎
NVMe硬盘
401，通过通信链路向NOF引擎发送读请求
402，通过通信链路接收主机发送的读请求
403，根据读请求向NVMe硬盘发送至少一个读命令
404，接收NOF引擎发送的至少一个读命令
405，在通信链路发生拥塞时，生成通信链路对应的拥塞标记
406，向NVMe硬盘发送拥塞标记
407，接收NOF引擎发送的拥塞标记
408，根据拥塞标记，暂停处理该通信链路对应的读命令

图 4

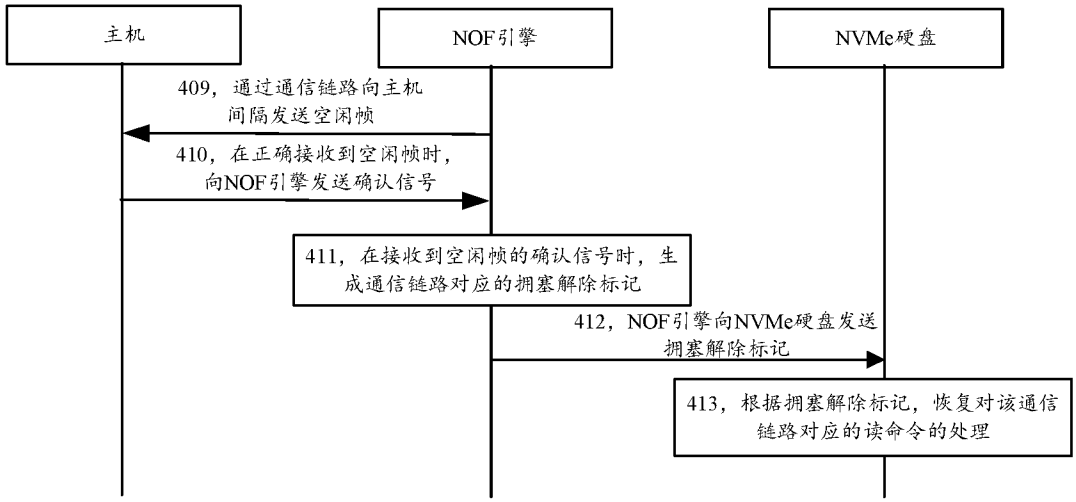

主机
NOF引擎
NVMe硬盘
409，通过通信链路向主机间隔发送空闲帧
410，在正确接收到空闲帧时，向NOF引擎发送确认信号
411，在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记
412，NOF引擎向NVMe硬盘发送拥塞解除标记
413，根据拥塞解除标记，恢复对该通信链路对应的读命令的处理

图 5

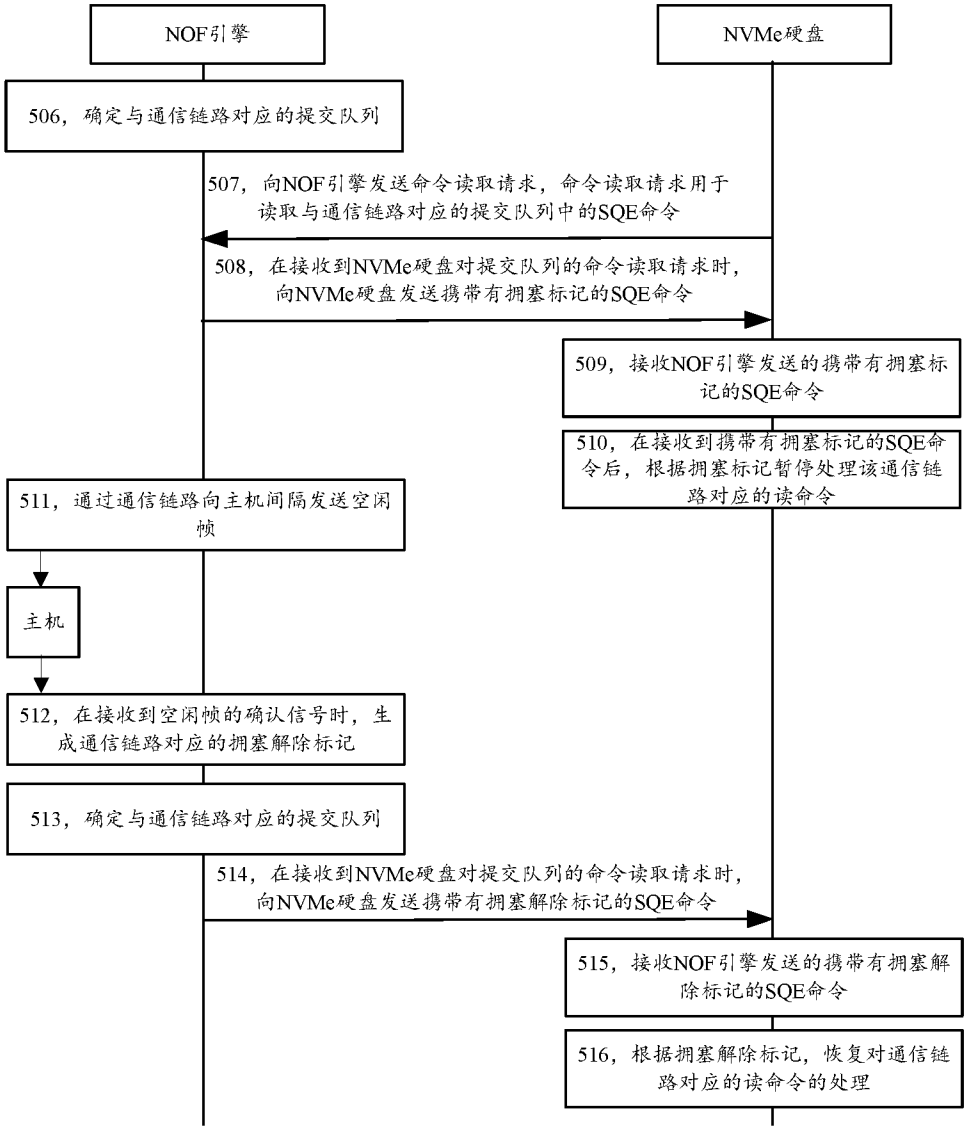

NOF引擎
NVMe硬盘
506，确定与通信链路对应的提交队列
507，向NOF引擎发送命令读取请求，命令读取请求用于读取与通信链路对应的提交队列中的SQE命令
508，在接收到NVMe硬盘对提交队列的命令读取请求时，向NVMe硬盘发送携带有拥塞标记的SQE命令
509，接收NOF引擎发送的携带有拥塞标记的SQE命令
510，在接收到携带有拥塞标记的SQE命令后，根据拥塞标记暂停处理该通信链路对应的读命令
511，通过通信链路向主机间隔发送空闲帧
主机
512，在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记
513，确定与通信链路对应的提交队列
514，在接收到NVMe硬盘对提交队列的命令读取请求时，向NVMe硬盘发送携带有拥塞解除标记的SQE命令
515，接收NOF引擎发送的携带有拥塞解除标记的SQE命令
516，根据拥塞解除标记，恢复对通信链路对应的读命令的处理

图 6

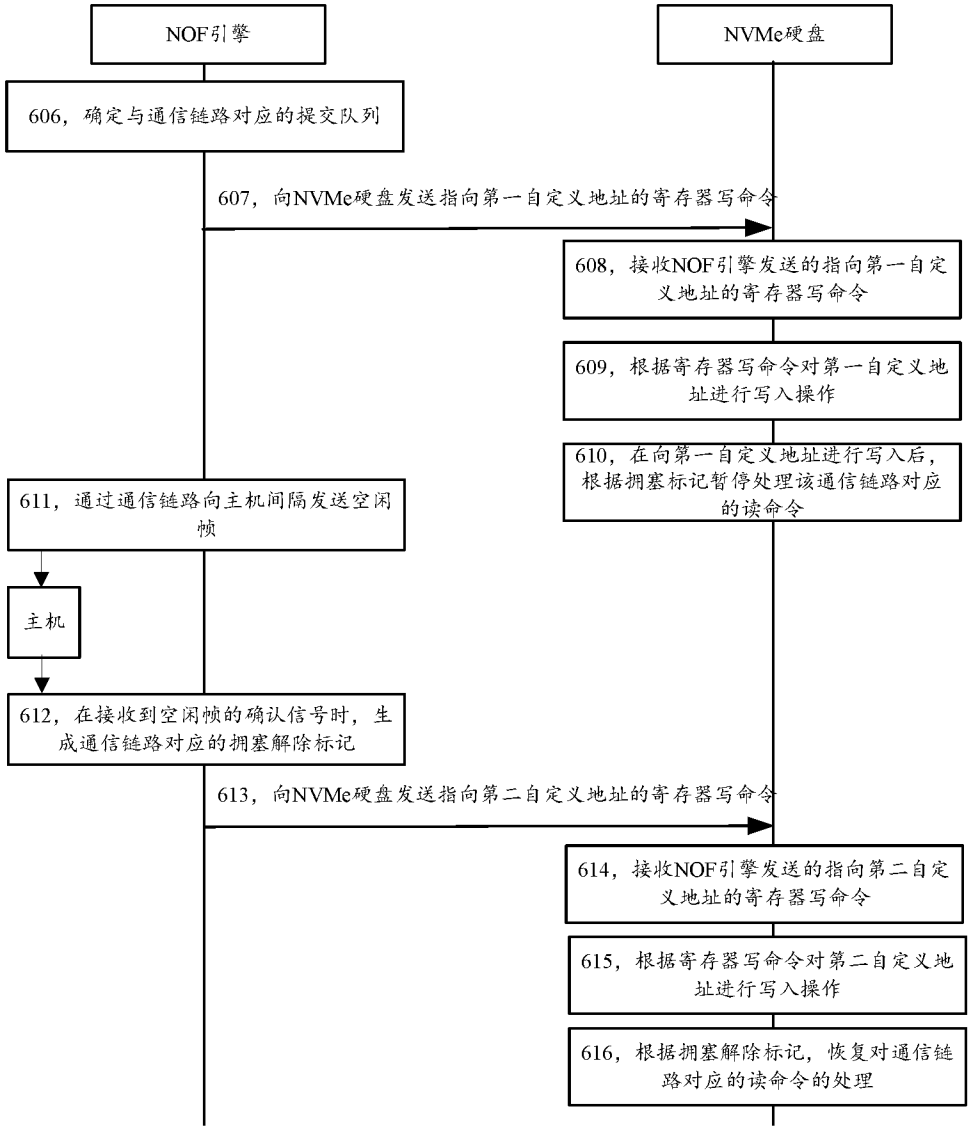

NOF引擎
NVMe硬盘
606，确定与通信链路对应的提交队列
607，向NVMe硬盘发送指向第一自定义地址的寄存器写命令
608，接收NOF引擎发送的指向第一自定义地址的寄存器写命令
609，根据寄存器写命令对第一自定义地址进行写入操作
610，在向第一自定义地址进行写入后，根据拥塞标记暂停处理该通信链路对应的读命令
611，通过通信链路向主机间隔发送空闲帧
主机
612，在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记
613，向NVMe硬盘发送指向第二自定义地址的寄存器写命令
614，接收NOF引擎发送的指向第二自定义地址的寄存器写命令
615，根据寄存器写命令对第二自定义地址进行写入操作
616，根据拥塞解除标记，恢复对通信链路对应的读命令的处理

图 7

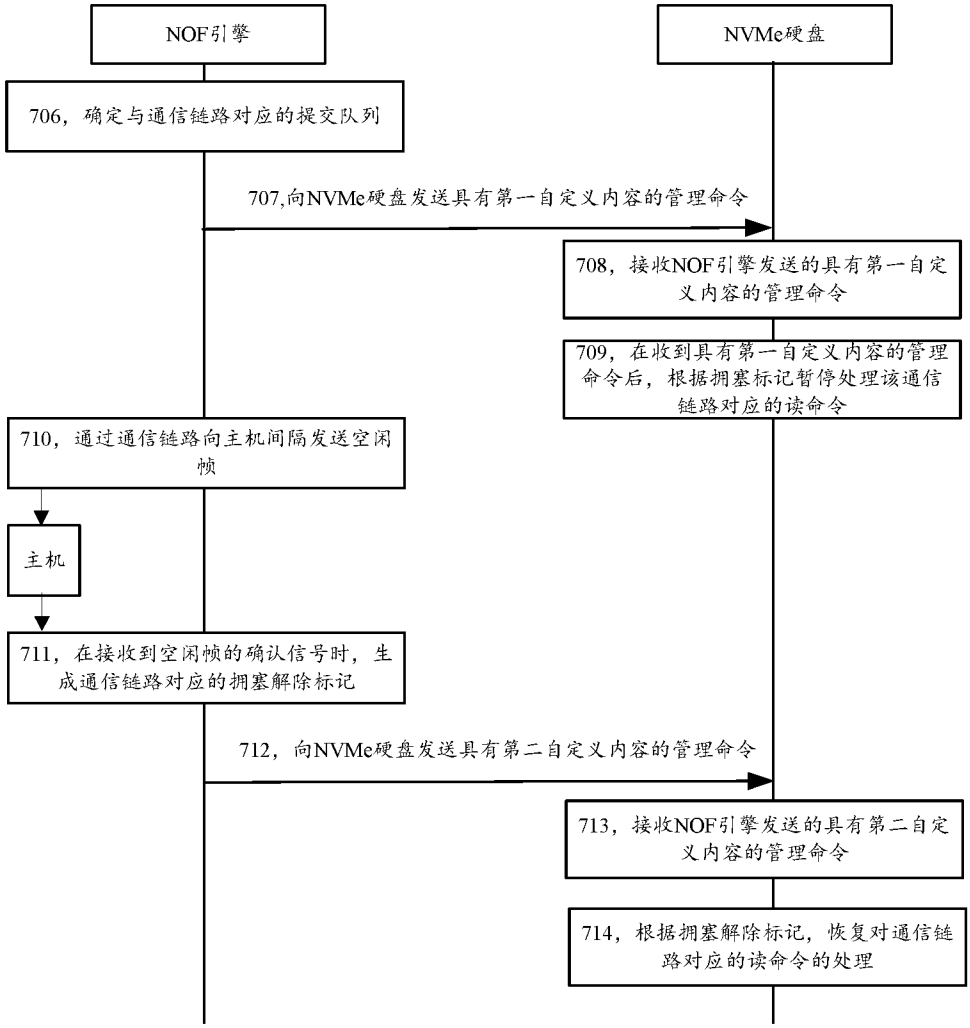

NOF引擎
NVMe硬盘
706，确定与通信链路对应的提交队列
707,向NVMe硬盘发送具有第一自定义内容的管理命令
708，接收NOF引擎发送的具有第一自定义内容的管理命令
709，在收到具有第一自定义内容的管理命令后，根据拥塞标记暂停处理该通信链路对应的读命令
710，通过通信链路向主机间隔发送空闲帧
主机
711，在接收到空闲帧的确认信号时，生成通信链路对应的拥塞解除标记
712，向NVMe硬盘发送具有第二自定义内容的管理命令
713，接收NOF引擎发送的具有第二自定义内容的管理命令
714，根据拥塞解除标记，恢复对通信链路对应的读命令的处理

图 8

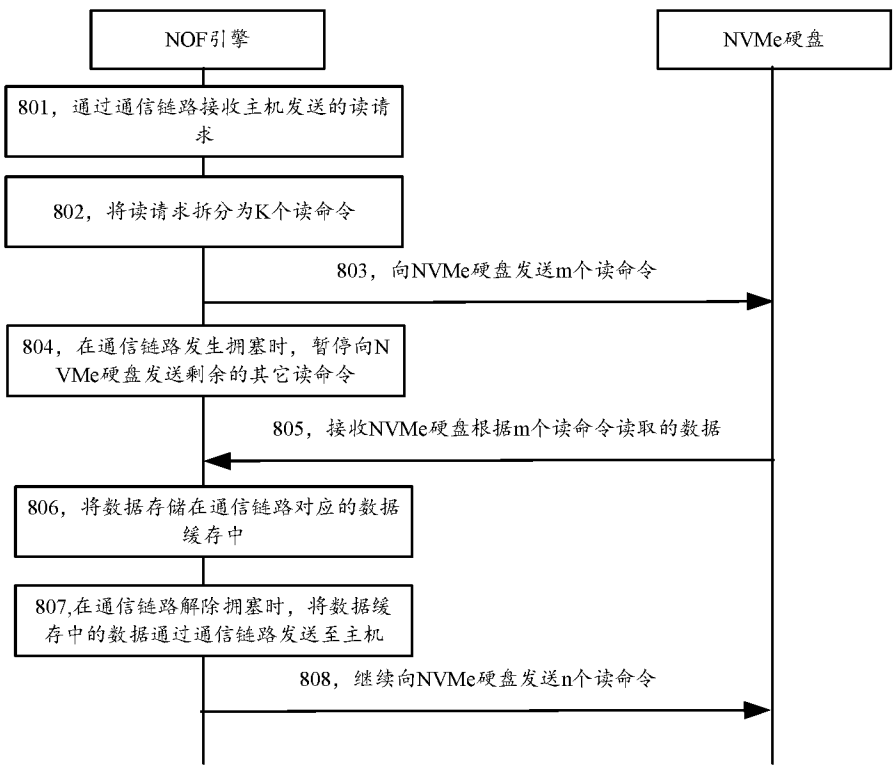

NOF引擎
NVMe硬盘
801，通过通信链路接收主机发送的读请求
802，将读请求拆分为K个读命令
803，向NVMe硬盘发送m个读命令
804，在通信链路发生拥塞时，暂停向NVMe硬盘发送剩余的其它读命令
805，接收NVMe硬盘根据m个读命令读取的数据
806，将数据存储在通信链路对应的数据缓存中
807,在通信链路解除拥塞时，将数据缓存中的数据通过通信链路发送至主机
808，继续向NVMe硬盘发送n个读命令

图 9

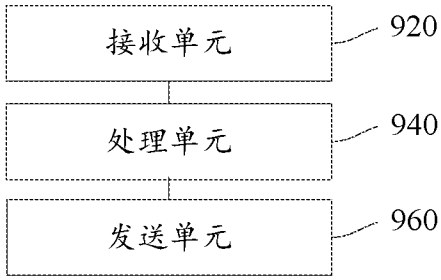


图 10

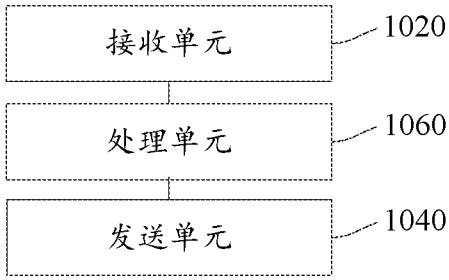


图 11

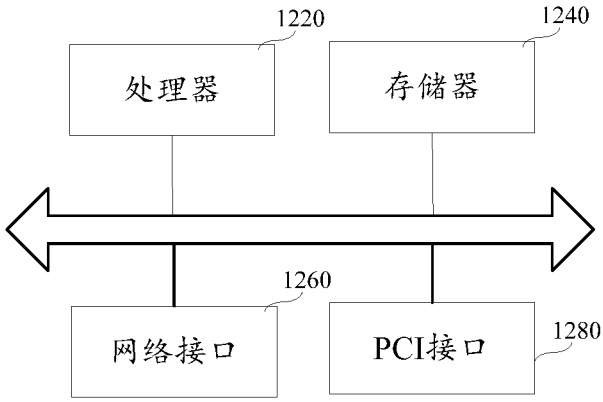


图 12

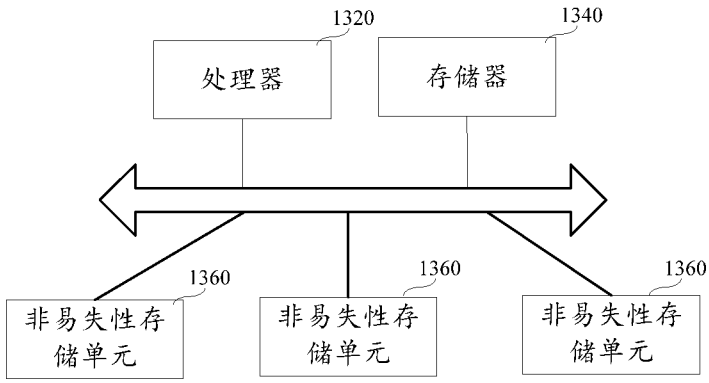


图 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/123749

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/06(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)

G06F

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)

CNPAT, WPI, EPODOC, CNKI, IEEE: 非易失性存储, 引擎, 硬盘, 命令, 拥堵, 拥塞, 标识, 标记, 暂停, NOF, engine, NVMe, SSD, command, plug, congestion, label, mark, pause

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106775434 A (HUAWEI TECHNOLOGIES CO., LTD.) 31 May 2017 (2017-05-31) description, paragraphs [0122]-[0265]	1-30
A	CN 102664803 A (HANGZHOU H3C TECHNOLOGIES CO., LTD.) 12 September 2012 (2012-09-12) entire document	1-30
A	CN 103583022 A (CITRIX SYSTEMS, INC.) 12 February 2014 (2014-02-12) entire document	1-30
A	CN 104536701 A (RAMAXEL TECHNOLOGY (SHENZHEN) CO., LTD.) 22 April 2015 (2015-04-22) entire document	1-30
A	WO 2007027481 A2 (CISCO TECHNOLOGY, INC.) 08 March 2007 (2007-03-08) entire document	1-30

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 February 2019	**27 March 2019**

Name and mailing address of the ISA/CN	Authorized officer
National Intellectual Property Administration, PRC (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/123749

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106775434	A	31 May 2017	WO	2017084400	A1	26 May 2017
				EP	3276481	A1	31 January 2018
				US	2018067685	A1	08 March 2018
CN	102664803	A	12 September 2012		None		
CN	103583022	A	12 February 2014	US	2012250512	A1	04 October 2012
				WO	2012134962	A1	04 October 2012
				EP	2692098	A1	05 February 2014
CN	104536701	A	22 April 2015		None		
WO	2007027481	A2	08 March 2007	EP	1920557	A2	14 May 2008
				US	2007047446	A1	01 March 2007
				CN	101427497	A	06 May 2009

国际检索报告

国际申请号
PCT/CN2018/123749

A. 主题的分类

G06F 3/06(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)

G06F

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））

CNPAT, WPI, EPODOC, CNKI, IEEE: 非易失性存储，引擎，硬盘，命令，拥堵，拥塞，标识，标记，暂停，NOF, engine, NVMe, SSD, command, plug, congestion, label, mark, pause

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
A	CN 106775434 A (华为技术有限公司) 2017年 5月 31日 (2017-05-31) 说明书第[0122]-[0265]段	1-30
A	CN 102664803 A (杭州华三通信技术有限公司) 2012年 9月 12日 (2012-09-12) 全文	1-30
A	CN 103583022 A (思杰系统有限公司) 2014年 2月 12日 (2014-02-12) 全文	1-30
A	CN 104536701 A (记忆科技深圳有限公司) 2015年 4月 22日 (2015-04-22) 全文	1-30
A	WO 2007027481 A2 (CISCO TECHNOLOGY, INC.) 2007年 3月 8日 (2007-03-08) 全文	1-30

☐ 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

“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	国际检索报告邮寄日期
2019年 2月 28日	2019年 3月 27日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 胡贝贝 电话号码 86-(10)-53961415

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/123749

检索报告引用的专利文件			公布日（年/月/日）	同族专利			公布日（年/月/日）
CN	106775434	A	2017年 5月 31日	WO	2017084400	A1	2017年 5月 26日
				EP	3276481	A1	2018年 1月 31日
				US	2018067685	A1	2018年 3月 8日
CN	102664803	A	2012年 9月 12日		无		
CN	103583022	A	2014年 2月 12日	US	2012250512	A1	2012年 10月 4日
				WO	2012134962	A1	2012年 10月 4日
				EP	2692098	A1	2014年 2月 5日
CN	104536701	A	2015年 4月 22日		无		
WO	2007027481	A2	2007年 3月 8日	EP	1920557	A2	2008年 5月 14日
				US	2007047446	A1	2007年 3月 1日
				CN	101427497	A	2009年 5月 6日