(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103020004 B
(45)授权公告日 2015.09.09

(21)申请号 201210546286.4

(22)申请日 2012.12.14

(73)专利权人 杭州华为数字技术有限公司
地址 310052 浙江省杭州市滨江区滨兴路301号3幢A楼301室

(72)发明人 陈昊 徐建荣 王工艺

(74)专利代理机构 北京龙双利达知识产权代理有限公司 11329
代理人 王君 肖鹏

(51)Int.Cl.
G06F 15/17(2006.01)

(56)对比文件
US 5875352 A,1999.02.23, 全文.
CN 101887397 A,2010.11.17, 全文.
CN 102222046 A,2011.10.19, 全文.
CN 1675625 A,2005.09.28, 说明书第2页第4段第1－2行,第3页第1段第6－7行,第4页第2段第3－4行,第6页第3段第1－2行,第7页第1段第1行、第2段第1－2行、第2段第5－8行以及图1和图3.

审查员 赵会玲

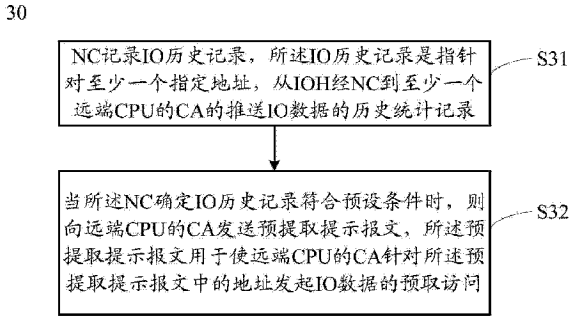
权利要求书2页 说明书19页 附图6页

(54)发明名称

高速缓存非对称一致性内存访问系统的访问方法和装置

(57)摘要

本发明实施例提供了一种CC-NUMA系统的访问方法和装置。方法包括:NC记录IO历史记录,IO历史记录是指针对至少一个指定地址,从IOH经NC到至少一个CPU的缓存CA的推送IO数据的历史统计记录;当NC确定IO历史记录符合预设条件时,则向远端CPU的CA发送预提取提示报文,预提取提示报文用于使远端CPU的CA针对预提取提示报文中的地址发起IO数据的预取访问。通过上述技术方案,通过分析该历史记录符合预定条件,从而主动发送针对该指定地址的IO数据提示报文给该远端的CPU的CA,并且由该远端CPU提前发起针对该指定地址的IO数据预取操作,由此缩短了远端IO数据访问延时,提升了系统的性能。



CN 103020004 B

1.一种高速缓存非对称一致性内存访问 CC-NUMA 系统的访问方法，其特征在于，包括：

节点控制器 NC 记录输入输出 IO 历史记录，所述 IO 历史记录是指针对至少一个指定地址，从输入输出集线器 IOH 经所述 NC 到至少一个远端中央处理器 CPU 的缓存 CA 的推送 IO 数据的历史统计记录；

当所述 NC 确定所述 IO 历史记录符合预设条件时，则向所述远端 CPU 的 CA 发送预提取提示报文，所述预提取提示报文用于使所述远端 CPU 的 CA 针对所述预提取提示报文中的地址发起 IO 数据的预取访问。

2.根据权利要求 1 所述的方法，其特征在于，所述 NC 确定所述 IO 历史记录符合预设条件，则向所述远端 CPU 的 CA 发送预提取提示报文包括：

所述 NC 确定所述 IOH 对所述指定地址的 IO 数据主动进行了更新操作，则向所述远端 CPU 的 CA 发送预提取提示报文。

3.根据权利要求 1 所述的方法，其特征在于，所述 NC 确定所述 IO 历史记录符合预设条件，则向所述远端 CPU 的 CA 发送预提取提示报文包括：

所述 NC 将所述 IO 历史记录中所述指定地址的统计指标与预设的门限值进行比较，分析比较结果后确定符合预设条件，则向所述远端 CPU 的 CA 发送预提取提示报文。

4.根据权利要求 3 所述的方法，其特征在于，所述 NC 将所述 IO 历史记录中所述指定地址的统计指标与预设的门限值进行比较，且分析比较结果后确定符合预设条件，则向所述远端 CPU 的 CA 发送预提取提示报文，包括：

所述 IO 历史记录中有关所述指定地址的统计次数大于预设的门限值，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或

所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔小于预设的门限值，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或

所述 IO 历史记录中单位时间内有关所述指定地址的统计次数大于预设的门限值，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或

确定轮询调度算法或加权的轮询调度算法的输入为所述远端 CPU 的 CA，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文。

5.根据权利要求 1 至 4 任一项所述的方法，其特征在于，所述 NC 记录 IO 历史记录，包括：

所述 NC 管理 IO 历史记录的插入、更新、替换和删除。

6.根据权利要求 5 所述的方法，其特征在于，其中根据以下条件之一，优先进行 IO 历史记录的更新或替换：

所述 IO 历史记录中有关所述指定地址的统计次数最低；或

所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔最长；或

所述 IO 历史记录中单位时间内有关所述指定地址的统计次数最低；或

根据轮询调度算法或加权的轮询调度算法的输入。

7.根据权利要求 1 至 4 任一项所述的方法，其特征在于，所述 IO 历史记录中的每一条至少包括以下内容：

有效性、指定地址、目的地和统计参数，其中所述目的地和统计参数一一对应。

8.一种高速缓存非对称一致性内存访问 CC-NUMA 系统中的装置，其特征在于，包括记

录模块、确定模块和发送模块：

所述记录模块，用于记录输入输出 IO 历史记录，所述 IO 历史记录是指针对至少一个指定地址，从输入输出集线器 IOH 经所述装置到至少一个远端中央处理器 CPU 的缓存 CA 的推送 IO 数据的历史统计记录；

所述确定模块，用于确定所述 IO 历史记录是否符合预设条件；

所述发送模块，用于当所述确定模块确定所述 IO 历史记录符合预设条件时，则向所述远端 CPU 的 CA 发送预提取提示报文，所述预提取提示报文用于使所述远端 CPU 的 CA 针对所述预提取提示报文中的地址发起 IO 数据的预取访问。

9. 根据权利要求 8 所述的装置，其特征在于：

所述确定模块确定所述 IOH 对所述指定地址的 IO 数据主动进行了更新操作，则所述发送模块向所述远端 CPU 的 CA 发送预提取提示报文。

10. 根据权利要求 8 所述的装置，其特征在于：

所述确定模块将所述 IO 历史记录中所述指定地址的统计指标与预设的门限值进行比较，分析比较结果后确定符合预设条件，则所述发送模块向所述远端 CPU 的 CA 发送预提取提示报文。

11. 根据权利要求 10 所述的装置，其特征在于：

当所述确定模块确定所述 IO 历史记录中有关所述指定地址的统计次数大于预设的门限值时，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或

当所述确定模块确定所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔小于预设的门限值时，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或

当所述确定模块确定所述 IO 历史记录中单位时间内有关所述指定地址的统计次数大于预设的门限值时，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或

所述确定模块确定轮询调度算法或加权的轮询调度算法的输入为所述远端 CPU 的 CA，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文。

12. 根据权利要求 8 至 11 任一项所述的装置，其特征在于：

所述记录模块管理 IO 历史记录的插入、更新、替换和删除。

13. 根据权利要求 12 所述的装置，其特征在于，所述记录模块根据以下条件之一，优先进行 IO 历史记录的更新或替换：

所述 IO 历史记录中有关所述指定地址的统计次数最低；或

所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔最长；或

所述 IO 历史记录中单位时间内有关所述指定地址的统计次数最低；或

根据轮询调度算法或加权的轮询调度算法的输入。

14. 根据权利要求 8 至 11 任一项所述的装置，其特征在于，所述记录模块记录的所述 IO 历史记录中的每一条至少包括以下内容：

有效性、指定地址、目的地和统计参数，其中所述目的地和统计参数一一对应。

高速缓存非对称一致性内存访问系统的访问方法和装置

技术领域

[0001] 本发明实施例涉及计算机领域，更具体地，涉及高速缓存非对称一致性内存访问（CC-NUMA，Cache Coherent-Non Uniform Memory Access）系统的访问方法和装置。

背景技术

[0002] 在基于节点（Node）控制的 CC-NUMA 系统中，随着系统规模的增长，跨节点访问的延时越来越成为系统性能提升的瓶颈。因此，如何动态地探测全系统的热点，并且尽早的将热点的缓存推送给最有可能使用该内容的远端 CPU（Central Processing Unit，中央处理器）将显著的提升现有系统被动存取的劣势。基于节点控制器（Node Controller，NC）的 CC-NUMA 系统中，IO（Input/Output，输入或输出）数据的访问延时过长，往往导致整个计算机系统的性能低下。

发明内容

[0003] 有鉴于此，本发明实施例提供一种 CC-NUMA 系统的访问方法和装置，以解决 IO 数据的访问延时过长的问题。

[0004] 第一方面，提供了一种 CC-NUMA 系统的访问方法，包括：节点控制器（NC）记录输入输出（IO）历史记录，IO 历史记录是指针对至少一个指定地址，从输入输出集线器（IOH）经 NC 到至少一个远端中央处理器（CPU）的缓存（CA）的推送 IO 数据的历史统计记录；当 NC 确定 IO 历史记录符合预设条件时，则向远端 CPU 的 CA 发送预提取提示报文，预提取提示报文用于使远端 CPU 的 CA 针对预提取提示报文中的地址发起 IO 数据的预取访问。

[0005] 在第一种可能的实现方式中，NC 确定 IOH 对指定地址的 IO 数据主动进行了更新操作，则向远端 CPU 的 CA 发送预提取提示报文

[0006] 结合第一方面的实现方式，在第二种可能的实现方式中，NC 将 IO 历史记录中指定地址的统计指标与预设的门限值进行比较，分析比较结果后确定符合预设条件，则向远端 CPU 的 CA 发送预提取提示报文。

[0007] 结合第一方面或第一方面的第二种可能的实现方式，在第三种可能的实现方式中，IO 历史记录中有关指定地址的统计次数大于预设的门限值，则向远端 CPU 的 CA 发送关于指定地址的预提取提示报文；或 IO 历史记录中有关指定地址的两次记录的计时间隔小于预设的门限值，则向远端 CPU 的 CA 发送关于指定地址的预提取提示报文；或 IO 历史记录中单位时间内有关指定地址的统计次数大于预设的门限值，则向远端 CPU 的 CA 发送关于指定地址的预提取提示报文；或确定轮询调度算法或加权的轮询调度算法的输入为远端 CPU 的 CA，则向远端 CPU 的 CA 发送关于指定地址的预提取提示报文。

[0008] 结合第一方面或第一方面的上述可能的实现方式，在第四种可能的实现方式中，NC 记录 IO 历史记录，包括：NC 管理 IO 历史记录的插入、更新、替换和删除。

[0009] 结合第一方面的第四种可能的实现方式，在第五种可能的实现方式中，根据以下条件之一，优先进行 IO 历史记录的更新或替换：IO 历史记录中有关指定地址的统计次数最

低；或 IO 历史记录中有关指定地址的两次记录的计时间隔最长；或 IO 历史记录中单位时间内有关指定地址的统计次数最低；或根据轮询调度算法或加权的轮询调度算法的输入。

[0010] 结合第一方面或第一方面的上述可能的实现方式，在第六种可能的实现方式中，IO 历史记录中的每一条至少包括以下内容：有效性（Valid）、指定地址（Address）、目的地（Destination）和统计参数（Statistical Parameters），其中目的地和统计参数一一对应。

[0011] 第二方面，提供了一种 CC-NUMA 系统中的装置，包括记录模块、确定模块和发送模块：记录模块，用于记录输入输出（IO）历史记录，IO 历史记录是指针对至少一个指定地址，从输入输出集线器（IOH）经该装置到至少一个远端中央处理器（CPU）的缓存 CA 的推送 IO 数据的历史统计记录；确定模块，用于确定 IO 历史记录是否符合预设条件；发送模块，用于当确定模块确定 IO 历史记录符合预设条件时，则向远端 CPU 的 CA 发送预提取提示报文，预提取提示报文用于使远端 CPU 的 CA 针对预提取提示报文中的地址发起 IO 数据的预取访问。

[0012] 在第一种可能的实现方式中，确定模块确定 IOH 对指定地址的 IO 数据主动进行了更新操作，则发送模块向远端 CPU 的 CA 发送预提取提示报文。

[0013] 结合第二方面的实现方式，在第二种可能的实现方式中，确定模块将 IO 历史记录中指定地址的统计指标与预设的门限值进行比较，分析比较结果后确定符合预设条件，则发送模块向远端 CPU 的 CA 发送预提取提示报文。

[0014] 结合第二方面的第二种可能的实现方式，在第三种可能的实现方式中，当确定模块确定 IO 历史记录中有关指定地址的统计次数大于预设的门限值时，则发送模块向远端 CPU 的 CA 发送关于指定地址的预提取提示报文；或当确定模块确定 IO 历史记录中有关指定地址的两次记录的计时间隔小于预设的门限值时，则发送模块向远端 CPU 的 CA 发送关于指定地址的预提取提示报文；或当确定模块确定 IO 历史记录中单位时间内有关指定地址的统计次数大于预设的门限值时，则发送模块向远端 CPU 的 CA 发送关于指定地址的预提取提示报文；或确定模块确定轮询调度算法或加权的轮询调度算法的输入为远端 CPU 的 CA，则发送模块向远端 CPU 的 CA 发送关于指定地址的预提取提示报文。

[0015] 结合第二方面或第二方面的上述可能的实现方式，在第四种可能的实现方式中，记录模块管理 IO 历史记录的插入、更新、替换和删除。

[0016] 结合第二方面的第四种可能的实现方式，在第五种可能的实现方式中，记录模块根据以下条件之一，优先进行 IO 历史记录的更新或替换：IO 历史记录中有关指定地址的统计次数最低；或 IO 历史记录中有关指定地址的两次记录的计时间隔最长；或 IO 历史记录中单位时间内有关指定地址的统计次数最低；或根据轮询调度算法或加权的轮询调度算法的输入。

[0017] 结合第二方面或第二方面的上述可能的实现方式，在第六种可能的实现方式中，记录模块记录的 IO 历史记录中的每一条至少包括以下内容：有效性（Valid）、指定地址（Address）、目的地（Destination）和统计参数（StatisticalParameters），其中目的地和统计参数一一对应。

[0018] 通过上述技术方案，可以记录针对一个或多个指定地址的推送 IO 数据的历史记录，通过分析该历史记录符合预定条件，从而推测远端 CPU 可能在将来的时间点使用该指定地址的 IO 数据，主动发送针对该指定地址的 IO 数据提示报文给该远端的 CPU 的 CA，并且

由该远端 CPU 提前发起针对该指定地址的 IO 数据预取操作，由此缩短了远端 IO 数据访问延时，提升了系统的性能。

附图说明

[0019] 为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例中所需要使用的附图作简单地介绍，显而易见地，下面所描述的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0020] 图 1 是相关技术中 CC-NUMA 系统的示意框图。

[0021] 图 2 是相关技术中 CC-NUMA 系统 IO 访问的方法的示意交互图。

[0022] 图 3 是本发明实施例的 CC-NUMA 系统的访问方法的示意流程图。

[0023] 图 4 是本发明实施例的 CC-NUMA 系统的示意框图。

[0024] 图 5 是本发明实施例的 CC-NUMA 系统的访问方法的示意交互图。

[0025] 图 6A 和图 6B 分别是一种动态事物监控器的结构示意图。

[0026] 图 7 是本发明实施例的 CC-NUMA 系统中的一种装置的示意框图。

[0027] 图 8 是本发明实施例的 CC-NUMA 系统中的另一种装置的示意框图。

具体实施方式

[0028] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明的一部分实施例，而不是全部实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都应属于本发明保护的范围。

[0029] 图 1 是相关技术中 CC-NUMA 系统 10 的示意框图。如图 1 所示，系统 10 包括多个 CPU，例如 CPU0-CPU7 和多个 NC，例如 NC0-NC3。CPU 本身具备与其他外部 CPU 互联的接口，可选的，多个 CPU 的 CC-NUMA 系统中可以通过 NC 进行互联扩展。CPU 包括缓存代理（CA，Cache Agent）。此外 CPU 与本地代理（HA，Home Agent）连接，HA 即管理内存的代理，可以是物理模块，HA 和内存（Memory，简写为 Mem）连接。IO 设备通过输入输出集线器（IOH，IO hub）由所连接的 CPU 进行输入或输出的数据访问。该 CPU 通过 NC 与其他节点进行网络互联，可以访问系统中其他 CPU 的 HA 中代理的数据。出于简洁，系统 10 中仅示出了实施例中需要使用到的 HA、CA、IO 设备等，但实际系统中的上述节点可以不只一个。

[0030] CPU 的 HA 是跟踪远端 CA 访问本 HA 管理的内存的状态的模块。以图 1 为例，HA 记录所连接的 Mem 中所有地址在远端 CA 中的状态。

[0031] 举例来说，一个远端 CA （CA1）访问 HA，并且想独占地址 Addr1 数据，HA 中之前没有接收到地址 Addr1 的访问，因此 HA 中地址 Addr1 的状态的为“未被占有”。HA 直接将 Mem 中的地址 Addr1 的数据发送给 CA1，同时在 HA 中记录下地址 Addr1 的状态为“被 CA1 独占”。

[0032] 当另一个 CA（CA2）来独占访问 HA 的地址 Addr1 数据时，HA 中 Addr1 的状态为“被 CA1 独占”，表明 CA1 占有该数据之后，可能会在 CA1 中先修改 Addr1 的数据，而并没有马上写回到 HA 代理的 Mem，因此 HA 代理的 Mem 中 Addr1 的数据可能不是最新的数据了。这时 HA 发送侦听报文到独占 Addr1 数据的 CA1。CA1 如果修改过 Addr1 的数据，CA1 会将修改过

的数据写回 Mem。之后 HA 将 CA1 更新的数据从 Mem 中读取，发送给 CA2，同时在 HA 中记录下地址 Addr1 的状态为“被 CA2 独占”。CA1 如果没有修改过 Addr1 的数据，Mem 中的数据还是最新的，CA1 可以直接失效该数据或者写回该数据到 Mem。HA 知道 CA1 中已经没有 Addr1 的数据拷贝之后，就从 Mem 中把数据送给 CA2，同时在 HA 中记录下地址 Addr1 的状态为“被 CA2 独占”。正是因为 HA 记录了代理 Mem 时的所有状态，因此在任何一个时刻，全系统所有地址的数据缓存可以保持数据的一致性，不存在同一个地址数据缓存冲突的情况。同一个地址数据缓存冲突是指同一个地址的数据在多个 CA 中有不同的值。

[0033] 为了方便说明，CPU 访问 IO 数据的流程，简单包括以下两个步骤。

[0034] S11，IO 设备更新内存中的 IO 数据。

[0035] S12，远端 CPU 发起对于该地址 IO 数据的访问。

[0036] 基于系统 IO 的系统架构，举例来说，S11 中，CPU0 上的 IO 设备通过 IOH->CPU0->NC0->NC2->CPU5->HA 的物理链路将 IO 数据更新到 CPU5 的 Mem 中；在 S12 中，CPU2 中的 CA 通过 CPU2->NC1->NC2->CPU5->HA 的物理链路将 IO 数据访问请求发送给 CPU5 的 Mem 中。HA 可以根据所记录的 Mem 中所有地址在远端 CA 中的状态，经 IO 数据更新到 Mem，或者，将 IO 数据访问请求所请求的 IO 数据反馈给远端 CA。IOH 用于将所有的不同类型的 IO 操作，翻译成统一的数据包格式发送给 CPU。其中，IOH 可以是一个物理的单元，在一种实现方式中，可以是主板上的一块芯片或者是芯片中的一个模块。

[0037] 接下来，参考图 2 来说明 CPU 访问 IO 数据的流程。图 2 是相关技术中 CC-NUMA 系统 IO 访问的方法 200 的示意交互图，包括以下内容。

[0038] 图 2 的虚线用来表证不同的 NC 域。图 2 的左侧，IO/IOH/NC 属于图 1 的 NC0 的域。图 2 的中间，CA 属于图 1 的 NC1 的域。图 2 的右侧，HA 属于图 1 的 NC2 的域。各节点经过多个 NC 之间的交叉网络连接，不同 NC 域的 CPU 互为远端。

[0039] S210，IO 设备发起更新 IO 数据的请求（MemWr）到 IOH。

[0040] S215，IOH 通过 NC0 转发数据更新的 QPI（QuickPath Interconnection，快速通路互联）请求（InvItoE）到 CPU5 的 HA。需要注意的是，这个时候 CPU5 的 HA 只需要记录下 IOH 上拥有更新的数据而本身并不需要拥有最新的数据。当别的请求访问 CPU5 的 HA 上该 IO 数据的时候，CPU5 的 HA 可以把 IOH 的最新数据通过一定的方式发送给请求者。

[0041] S220，CPU5 的 HA 向 IOH 发送关于数据更新的 QPI 请求的应答（Gnt_Cmp）。

[0042] S225，IOH 向 IO 设备发送数据请求响应（Cmp）。

[0043] S230，经过了一段时间。

[0044] 这个时间不定，最小的可以是纳秒级别，最大的可以到秒，甚至是天或年。主要根据应用的实际运行情况，当图 1 中的 CPU2 的 CA 某个线程运行到需要访问该地址时，就会发起接下来对应的请求。

[0045] 接下来的步骤中，获取指定地址的 IO 数据通常需要较长的时间，制约了全系统的性能。

[0046] S240，CPU2 的 CA 向 CPU5 的 HA 发起对于该地址 IO 数据的访问（RdData）。

[0047] S245，HA 根据记录的该 IO 数据的状态，此时最新的数据是存在放 IOH 上，发起向 IOH 的数据侦听（SnpData）。

[0048] 发起该数据侦听用于确定 IOH 中是否有最新的数据拷贝。

[0049] S250 如图 2 所示，包括三个子步骤 S250-1 至 S250-3。

[0050] S250-1，IOH 收到该侦听之后将数据直接转发（Forward）给 CPU2 的 CA，即数据请求者。

[0051] IOH 同时把状态更新到 HA 上。也就是在 S250-2 和 S250-3，IOH 分别向 HA 发送响应（RspFwdWb 和 WbIData），此时 HA 上记录的最新的数据在 CPU2 的 CA 上而不是 IOH 上。

[0052] S255，HA 向 CA 发送数据请求响应（Cmp）。

[0053] 图 2 中使用到的 QPI 协议包的具体释义可以参考下表 1。

[0054] 表 1

[0055]

协议包名称	中文名	具体含义
MemWr	内存写	由 IO 设备向 IOH 写内存数据
InvItoE	独占数据请求	请求独占数据请求，并且不需要从 HA 返回数据，只是在 HA 的目录中记录一下请求的地址。 注:后续有别的请求者 B 请求该地址数据时，需要先去侦听目录中记录的独占请求者 A（因为该独占请求者可能已经更新该地址的数据），而内存中的该地址数据可能已经不是最新的数据了。因此新的数据请求者 B 所申请的该地址数据可能就是从目录记录的独占请求者 A 那里返回的数据。
Gnt_Cmp	独占数据请求的应答	独占数据请求应答，不包含该地址的数据返回。因为独占请求者不需要读该地址的数据，而只是需要写该地址。 HA 告诉独占数据请求者 A，对于该地址的请求已经响应。同时 HA 已经在自己的目录中记录下独占数据请求者 A。以便后续进行数据一致性的操作。
Cmp	内存写操作响应	标识内存写操作完成
RdData	数据请求	缓存 B 发向 HA 的数据请求
SnpData	数据侦听	HA 查看目录中，RdData 请求地址 C 的状态，发现地址 C 现在被请求者 A 独占，而 A 可能已经改写地址 C 的数据（最新数据存在于 A，而不是内存 HA），因此需要首先侦听独

[0056]

		占请求者 A。
RspIFwdWb	回写转发标志响应	独占请求者 A 收到侦听请求（SnpData）之后，发送回写转发标志给 HA。通知 HA 2 个事件：1）最新的数据会写回 HA（通过 WbIData 数据响应包）; 2）最新的数据已经转发到后来的数据请求者 B（通过 DataC_E 响应包），同时根据转发的目的地（数据请求 B），将目录中的地址 C 的占有者从原来的 A 改成 B。 同时被侦听的 A，将自身关于地址 C 的数据清除。A 不再拥有对地址 C 数据的使用和更新的权限。如果要对地址 C 再进行操作，需要重新发送对于地址 C 的请求至 HA。
WbIData	回写数据响应	回写数据响应。 将最新的数据写回 HA。
DataC_E	转发独占数据响应	转发独占数据响应。 将最新的数据转发给数据请求者 B。
Cmp	数据请求响应	HA 在收到数据请求者 B 发来的请求之后，先进行侦听操作，等到所有的侦听操作响应全部完毕之后（HA 收到 RspIFwdWb 和 WbIData 的侦听响应之后），通知数据请求者 B，HA 这边的数据请求任务已经完成了（主要是更新目录中地址 C 的占有者为 B，占有者的状态为独占）

[0057] 如图 1 和图 2 所示，基于 NC 的 CC-NUMA 系统中 IO 访问的最大缺点在于只有当 CPU 需要远端的 IO 数据的时候才会通过 HA 去进行数据访问，而远端的 IO 数据访问延时又是非常大，该延时恰恰是整个系统性能提升的最大瓶颈。本发明实施例提供了一种上述 CC-NUMA 系统中 IO 数据访问加速的方法，能够显著的减少远端 IO 数据访问延时，提升全系统的性能。

[0058] 图 3 是本发明实施例的 CC-NUMA 系统的访问方法 30 的示意流程图，方法 30 包括以下内容。

[0059] S31，NC 记录 IO 历史记录，所述 IO 历史记录是指针对至少一个指定地址，从 IOH 经

所述 NC 到至少一个远端 CPU 的 CA 的推送 IO 数据的历史统计记录。

[0060] S32，当所述 NC 确定所述 IO 历史记录符合预设条件时，则向所述远端 CPU 的 CA 发送预提取提示报文，所述预提取提示报文用于使所述远端 CPU 的 CA 针对所述预提取提示报文中的地址发起 IO 数据的预取访问。

[0061] 本发明实施例通过 NC 记录针对一个或多个指定地址的推送 IO 数据的历史记录，通过分析该历史记录符合预定条件，从而推测远端 CPU 可能在将来的时间点使用该指定地址的 IO 数据，NC 主动发送针对该指定地址的 IO 数据提示报文给该远端的 CPU 的 CA，并且由该远端 CPU 提前发起针对该指定地址的 IO 数据预取操作，由此缩短了远端 IO 数据访问延时，提升了系统的性能。

[0062] 可选的，作为不同的实施例，所述 NC 确定所述 IOH 对所述指定地址的 IO 数据主动进行了更新操作，则向所述远端 CPU 的 CA 发送预提取提示报文。

[0063] 可选的，作为不同的实施例，NC 将所述 IO 历史记录中所述指定地址的统计指标与预设的门限值进行比较，分析比较结果后确定符合预设条件，则向所述远端 CPU 的 CA 发送预提取提示报文。

[0064] 可选的，作为不同的实施例，所述 IO 历史记录中有关所述指定地址的统计次数大于预设的门限值，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔小于预设的门限值，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或所述 IO 历史记录中单位时间内有关所述指定地址的统计次数大于预设的门限值，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或确定轮询调度算法或加权的轮询调度算法的输入为所述远端 CPU 的 CA，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文。

[0065] 可选的，作为不同的实施例，NC 管理 IO 历史记录的插入、更新、替换和删除。

[0066] 可选的，作为不同的实施例，其中根据以下条件之一，优先进行 IO 历史记录的更新或替换：所述 IO 历史记录中有关所述指定地址的统计次数最低；或所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔最长；或所述 IO 历史记录中单位时间内有关所述指定地址的统计次数最低；或根据轮询调度算法或加权的轮询调度算法的输入。

[0067] 可选的，作为不同的实施例，所述 IO 历史记录中的每一条至少包括以下内容：有效性(Valid)、指定地址(Address)、目的地(Destination)和统计参数(Statistical Parameters)，其中所述目的地和统计参数一一对应。

[0068] 图 4 是本发明实施例的 CC-NUMA 系统 40 的示意框图。系统 40 与系统 10 的节点组成相同或类似。不同之处在于本发明实施例中 NC 包括动态事物监控器(DHTM，Dynamic Hot Traffic Monitor)。由此，CPU 访问 IO 数据的流程，简单包括以下三个步骤。

[0069] S41，IO 设备更新内存中的 IO 数据。

[0070] 基于系统 40 的系统架构，举例来说，S41 中，CPU0 上的 IO 设备通过 IOH->CPU0->NC0->NC2->CPU5->HA 的物理链路将 IO 数据更新到 CPU5 的 Mem 中。

[0071] S42，NC 根据检测到的 IO 历史记录符合预设条件，推测远端 CPU 可能在将来的时间点使用该地址的 IO 数据，NC 主动发送该地址的 IO 数据提示报文给该远端的 CPU 的 CA。

[0072] NC0 的 DHTM 记录 IO 历史记录，其中 IO 历史记录是指针对至少一个指定地址，从 IOH 经 NC0 到至少一个远端 CPU 的 CA 的推送(Forward)IO 数据的历史统计记录。

[0073] NC0 的 DHTM 通过分析 IO 历史记录符合预设条件，例如 CPU2 的 CA 在一定时间内多次访问该地址的 IO 数据，推测 CPU2 中的 CA 会在将来的时间点使用到该地址的 IO 数据，因此 NC0 通过物理链路 NC0->NC1->CPU2->CA 主动发送预提取提示（PrefetchHint）报文。

[0074] S43，远端 CPU 的 CA 收到报文后，立即发起对于该地址的 IO 数据的访问。

[0075] CPU2 中的 CA 收到 PrefetchHint 报文之后，通过 CPU2->NC1->NC2->CPU5->HA 的物理链路提前将 IO 数据访问预取请求发送到 CPU5 的 Mem 中。

[0076] 本发明实施例通过 NC 记录和分析针对至少一个指定地址的 IO 历史记录，预测未来的时间点最有可能使用该指定地址的 IO 数据的远端 CA，并且主动发起预取提示报文给预测到的远端 CA，通知 CA 提前预取 IO 数据，极大地并行化了当远端 CA 需要该 IO 数据时才发起访问带来的跨节点的访问延时，从而突破了基于 NC 的 CC-NUMA 系统的 IO 访问的最大瓶颈，提升了全系统的性能。

[0077] 图 5 是本发明实施例的 CC-NUMA 系统的访问方法 500 的示意交互图。方法 500 与方法 200 的不同之处在于方法 500 的 S530 和 S535，具体内容如下。

[0078] S510，IO 设备发起更新 IO 数据的请求（MemWr）到 IOH。

[0079] S515，IOH 通过 NC0 转发数据更新的 QPI（QuickPath Interconnection，快速通路互联）请求（InvItoE）到 CPU5 的 HA。需要注意的是，这个时候 CPU5 的 HA 只需要记录下 IOH 上拥有更新的数据而本身并不需要拥有最新的数据。当别的请求访问 CPU5 的 HA 上该 IO 数据的时候，CPU5 的 HA 可以把 IOH 的最新数据通过一定的方式发送给请求者。

[0080] S520，CPU5 的 HA 向 IOH 发送关于数据更新的 QPI 请求的应答（Gnt_Cmp）。

[0081] S525，IOH 向 IO 设备发送数据请求响应（Cmp）。

[0082] S530，NC0 的动态事物监控器（DHTM）记录 IO 历史记录，且通过分析 IO 历史记录符合预设条件，推测远端 CPU2 的 CA 可能在将来的时间点使用该地址的 IO 数据，NC0 主动发送该地址的 IO 数据预取的提示报文（PrefetchHint）给该远端的 CPU2 的 CA。

[0083] 接下来的步骤中，根据预取提示报文，触发 CPU2 的 CA 预先获取到该地址的 IO 数据。节省了 CPU2 的 CA 在将来的时间点使用该地址的 IO 数据时的访问延时。

[0084] S535，NC0 接收 CPU2 的 CA 针对提示报文（PrefetchHint）的响应（Cmp）

[0085] S540，接收到数据预取的提示报文后，CPU2 的 CA 立即向 CPU5 的 HA 发起对于该地址 IO 数据的预取访问 (RdData)。

[0086] S545，HA 根据记录的该 IO 数据的状态，此时最新的数据是存在放 IOH 上，发起向 IOH 的数据侦听（SnpData）。

[0087] S550 如图 5 所示，包括三个子步骤 S550-1 至 S550-3。

[0088] S550-1，IOH 收到该侦听之后将数据直接推送（Forward）给 CPU2 的 CA，即数据请求者。

[0089] IOH 同时把状态更新到 HA 上。也就是在 S550-2 和 S550-3，IOH 分别向 HA 发送响应（RspFwdWb 和 WbIData），此时 HA 上记录的最新的数据在 CPU2 的 CA 上而不是 IOH 上。

[0090] S555，HA 向 CA 发送数据请求响应（Cmp）。

[0091] 方法 500 的 S530 和 S535 使用到的 QPI 协议包参见下表 2，其他使用到的 QPI 协议包的具体释义可以参考表 1。

[0092] 表 2

[0093]

协议包名称	中文名	具体含义
PrefetchHint	预取提示报文	NC 推测缓存 B 很有可能在将来的时间内使用到地址 C 的数据，因此提前发送预取提示报文给缓存 B
Cmp	预取提示报文响应	缓存 B 在收到对于地址 C 的缓存提示报文之后，响应 NC 的预取提示报文。同时会立即发送对于地址 C 的数据请求至 HA。

[0094] 本发明实施例通过 NC 记录和分析 IO 历史记录，预测未来的时间点最有可能使用该地址的 IO 数据的远端 CA，并且主动发起预取提示报文给对应的远端 CA，通知 CA 提前预取 IO 数据，极大地并行化了当远端 CA 需要该 IO 数据时才发起访问带来的跨节点的访问延时，从而突破了基于 NC 的 CC-NUMA 系统的 IO 访问的最大瓶颈，提升了全系统的性能。

[0095] 通常 NC 由专用芯片制得，本发明实施例的 NC 可以包括一个动态事物监控器，用于记录 IO 历史记录且确定 IO 历史记录符合预设条件。可选地，可以通过软件、硬件或两者的结合实现满足上述要求的 NC。优选地，以硬件结构实现本发明实施例中的 NC 的动态事物监控器（DHTM），由此作为提高计算机系统性能的优选手段，。

[0096] 图 6A 是一种动态事物监控器的结构示意图。该结构具有 N 条记录，其中 N 为非负整数。根据系统的跨节点的数目可以调整记录的条数。

[0097] 每条记录的内容是对于指定地址，从 IOH 经该 NC 到远端 CPU 的 CA 的推送 IO 数据的历史记录，包括若干个字段，其中至少包括以下核心字段 {Valid, Address, Destination, Statistical Parameters}。

[0098] 各个字段的具体含义如下：

[0099] 有效位（Valid）：表示该条记录是否有效。

[0100] 举例来说，一个表中有 64 条记录，一旦使用了该条记录，则该记录以有效位为 1 来标识该条记录是有效的，其余空余的表项记录的有效位为 0，标识这条记录可以被占用。

[0101] 地址（Address）：标识历史请求的地址。其中具体的一个地址，上下文中也称为指定地址。

[0102] 目的地（Destination，A 个）：标识访问指定地址（Address）的请求者，可以是一个也可以是多个请求者，也就是具体的 CPU 的 CA。可以以 CPU 的 CA 的全局域 ID 表示。A 为正整数。CA 既是 IOH 推送数据的目的地，也是 NC 发送预取提示报文的目的地。

[0103] 统计参数（Statistical Parameters，B 个）：用于统计的参数，B 为正整数。统计参数可以包含对该地址历史请求的次数，即命中计数参数（Hit count）；或记录第一次记录该地址请求到目前为止的时间（Time）等等。一个目的地只能对应一种统计参数，可以成对表示（Destination/Statistical Parameters）。如果推送的目的地不同，命中次数以及所有相关的统计数据不会进行累加而是重新标记。

[0104] 图 6B 是一种动态事物监控器的扩展结构示意图。如图 6B 所示，可以包括多个目的地，例如目的地 1（Destination1）至目的地 A（DestinationA），每个 Destination 对应的统计参数的具体内容可以相同，也可以不同，例如标识为统计参数 1（StatisticalParameters1）至统计参数 B（StatisticalParametersB）中的一个。DestinationX/StatisticalParametersX 表示在拥有多个 Destination/Statistical Parameters 字段的动态事务监控器的扩展结构中的某一对（Destination/Statistical Parameters）字段。具体地如何选择其中的一对字段，则根据相应的统计算法。

[0105] 如果用于预取提示报文的推送的目的地在 2 个或者更多个目的地之间交替推送，那么动态事务监控器的硬件结构效率就会很低。因此图 6B 所示意的动态事务监控器的扩展结构可以规避以上问题。

[0106] 图 6A 的结构也称为只有一对 (Destination/Statistical Parameters) 字段的普通结构，简称普通结构；图 6B 的结构简称为多对 (Destination/StatisticalParameters) 的扩展结构，简称扩展结构。

[0107] 接下来，参考表 3，具体说明 NC 如何记录 IO 历史记录，且确定该 IO 历史记录符合预设条件的方法。NC 管理 IO 历史记录时具体包括以下事件：记录的插入、更新、替换和删除等。表 3 的表项中可以包括事件、事件发生的前提条件、动作，以及策略等。

[0108] 表 3

[0109]

事件	事件发生的前提条件	动作	策略
记录插入	普通结构: 1）有从 IOH 到 NC 的推送事务时。 注：也就是说，当有从 CA 发向 IOH 的 IO 数据请求时，IOH 需要将数据直接推送给数据请求者 CA。 2）如果硬件资源中有空余的表项，可占用该表项。 3）所有当前 Valid 标识的记录中，没有对应的地址请求。	把当前推送 IO 数据请求（即由 CA 发起的 IO 数据请求）的所有标识位，更新到动态事务监控器的各个字段中。 ➢ Valid = 1 ➢ Destination = IO 数据请求者的全局标识 ID ➢ Address = 该 IO 数据的地址 ➢ Statistical Parameters = 根据算法统计的值，例如: ■ Hit_count = 1。标识第一次请求推送。 ■ Time=0，并且从当前 cycle 开始计时。标识请求推送的开始时间。 ■ 等等。	
	扩展结构，如果下列条件均满足: 1）如果 Valid 等于 1 2）请求包 Address 等于表项中的 Address。 3）请求包 Destination	把当前的数据请求更新到扩展动态事务监控器的各个字段中。 ➢ Valid 和 Address 保持不变。 ➢ DestinationX = IO 数	

[0110]

	不等于表项中已记录的所有Destination。 4）有空余的（Destination/Statistical Parameters）字段表项。	据的请求者的全局标识ID ➢ Statistical Parameters X= 根据算法统计的值。 该统计算法和之前已经更新的 Statistical Parameters 列表项所用的算法一致。	
记录更新	普通结构: 1）有从 IOH 到 NC 的推送事务时。 2）所有当前 Valid 标识的记录中，有对应的地址请求。 3）当前 Forward 请求 Address 等于表项中 Address。	（如果当前推送请求的目的地不等于记录中 Destination 字段），则 ➢ Destination = 新的推送请求的目的地的全局标识ID ➢ Statistical Parameters 可更新如下 ■ Hit_count = 1 ■ Time=0，重新开始计时 ■ 等等 （如果当前推送请求的目的地等于记录中 Destination 字段） ➢ Destination = 保持不变 ➢ Statistical Parameters 可更新如下 ■ Hit_count = Hit_count + 1 ■ Time 继续计时 ■ 等等。	
	扩展结构	（如果当前推送请求的目的地不等于记录中所有 Destination 字段）则根据一定替换策略选择	这个时候可以有多种替换策略，可以根据当前

[0111]

		某对(Destination/Statistical Parameters) 进行更新 ➢ DestinationX = 新的 Forward 请求的目的地的全局标识 ID ➢ Statistical Parameters X 更新的值和之前已经更新的 Statistical Parameters 列表项所用的算法一致。 （如果当前推送请求的目的地等记录着某个 Destination 字段），则更新对应的(Destination/Statistical Parameters)字段 ➢ Destination X= 保持不变 ➢ Statistical Parameter-sX 可更新如下 ■ Hit_count = Hit_count + 1 ■ Time 继续计时 ■ 等等。	系统选择其中一种进行。例如: ➢ 根据轮询调度或加权的轮询调度（Round-Robin/ Weighted Round-Robin）算法 Entry（输入）先替换 ➢ 历史命中最少的 Entry 先替换 ➢ 历史命中计时最久 Entry 先替换 ➢ 历史单位时间命中率最低的 Entry 先替换 ➢ 等等。
记录替换	普通结构: 1） 有从 IOH 到 NC 的推送事务时 2） 所有当前 Valid 标识的记录中，没有对应的地址	把当前的推送请求的所有标识位，更新到命中记录的各个字段中。 ➢ Destination = 新的 Fwd 请求的目的地的全局标识 ID	这个时候可以有多种替换策略，可以根据当前系统选择其

[0112]

	请求 3） 记录表中的Valid全部使用完毕。	➢ Address = 该IO数据的地址 ➢ Statistical Parameters可更新如下 ■ Hit_count = 1 ■ Time=0，重新开始计时 ■ Etc	中一种进行。例如: ➢ 根据Round-Robin/Weighted Round-Robin算法Entry先替换 ➢ 历史命中最少的Entry先替换 ➢ 历史命中计时最久Entry先替换 ➢ 历史单位时间命中率最低的Entry先替换 ➢ 等等
	扩展结构	动作和只有一对(Destination/Statistical Parameters)字段的普通结构动作一致。	
记录删除	普通结构: 复位或者配置重置等。	清除所有记录中的Valid位为0	
	扩展结构	动作和只有一对(Destination/Statistical Parameters)字段的普通结构动作一致。	

[0113] 记录更新或记录替换中的策略可以根据统计算法的不同而不同，包括但不限于表3中的示例，出于简洁，此处不再详细举例。其中，轮询调度或加权的轮询调度(Round-Robin/Weighted Round-Robin)算法的基本原理是，根据输入(Entry)轮流确定调

用的对象。加权轮询调度是指，每个调度对象给予不同的权重，有些对象的权重较高，有些对象的权重较低。在本发明实施例中，所调用的对象即为目的地(Destination)。

[0114] 由于推送预取提示报文的操作是基于分析历史记录，且这些记录符合预设条件而确定的，因此并不是每次的推送的目的地都是成功的。表 3 策略中的命中率是指将预取提示报文成功推送给目的地的次数占总的推送次数的百分比，可以通过统计获得。接下来，NC 确定 IO 历史记录符合预设条件时，可以向远端 CPU 的 CA 发送预提取提示报文。其中 NC 发送预提取提示报文的推送策略可以有多种，下面简单介绍其中的两种。

[0115] ➢NC 确定 IOH 有主动性的对于某个指定地址的 IO 数据的更新操作，其中该指定地址记录在动态事务监控器的硬件中。

[0116] ➢动态事务监控器的硬件对于某个指定地址的统计指标大于预设的门限值。该统计指标为统计参数的具体数值。

[0117] 统计算法中涉及如何计算统计指标大于预设的门限值，可以有多种选择。例如：根据 Round-Robin/Weighted Round-Robin 算法选择推送；历史命中最多的优先推送；历史命中计时间隔最短的优先推送；历史单位时间命中率最高的优先推送；以及其他根据不同的统计算法得出的优先推送等等。上面列出的推送策略只是基于 NC 的 CC-NUMA 系统 IO 加速方法的一种、多种或者是多种组合具体的实现，本发明实施例对此不做限定。但凡 NC 通过记录和分析 IO 历史记录，向远端 CA 发送了预提取提示报文，即落入本发明实施例保护范围。

[0118] 本发明实施例通过 NC 记录和分析 IO 历史记录，预测未来的时间点最有可能使用该地址的 IO 数据的远端 CA，并且主动发起预取提示报文给对应的远端 CA，通知 CA 提前预取 IO 数据，极大地并行化了当远端 CA 需要该 IO 数据时才发起访问带来的跨节点的访问延时，从而突破了基于 NC 的 CC-NUMA 系统的 IO 访问的最大瓶颈，提升了全系统的性能。

[0119] 图 7 是本发明实施例的 CC-NUMA 系统中的一种装置 70 的示意框图。装置 70 包括记录模块 71、确定模块 72 和发送模块 73。

[0120] 记录模块 71 记录输入输出 IO 历史记录，所述 IO 历史记录是指针对至少一个指定地址，从输入输出集线器 IOH 经所述装置到至少一个远端中央处理器 CPU 的缓存 CA 的推送 IO 数据的历史统计记录。

[0121] 确定模块 72 确定所述 IO 历史记录是否符合预设条件。

[0122] 发送模块 73 当确定模块 72 确定所述 IO 历史记录符合预设条件时，则向所述远端 CPU 的 CA 发送预提取提示报文，所述预提取提示报文用于使所述远端 CPU 的 CA 针对所述预提取提示报文中的地址发起 IO 数据的预取访问。

[0123] 本发明实施例通过 CC-NUMA 系统中的装置记录针对一个或多个指定地址的推送 IO 数据的历史记录，通过分析该历史记录符合预定条件，从而推测远端 CPU 可能在将来的时间点使用该指定地址的 IO 数据，该装置主动发送针对该指定地址的 IO 数据提示报文给该远端的 CPU 的 CA，并且由该远端 CPU 提前发起针对该指定地址的 IO 数据预取操作，由此缩短了远端 IO 数据访问延时，提升了系统的性能。

[0124] 装置 70 可以执行方法 30 或 40，结构例如图 6A 或图 6B 中所示的动态事物监控器，作为不同的实现方式可以是 NC，也可以包括在 CC-NUMA 的 NC 中，还可以独立存在。其中，NC 可以是一块专用的芯片或者现场可编程门阵列(FPGA，Field Programmable Gate Array)

设备等。

[0125] 可选的，作为不同的实施例，当所述确定模块确定所述 IO 历史记录中有关所述指定地址的统计次数大于预设的门限值时，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或当所述确定模块确定所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔小于预设的门限值时，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或当所述确定模块确定所述 IO 历史记录中单位时间内有关所述指定地址的统计次数大于预设的门限值时，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或所述确定模块确定轮询调度算法或加权的轮询调度算法的输入为所述远端 CPU 的 CA，则所述发送模块向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文。

[0126] 可选的，作为不同的实施例，所述记录模块管理 IO 历史记录的插入、更新、替换和删除。

[0127] 可选的，作为不同的实施例，所述记录模块根据以下条件之一，优先进行 IO 历史记录的更新或替换：所述 IO 历史记录中有关所述指定地址的统计次数最低；或所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔最长；或所述 IO 历史记录中单位时间内有关所述指定地址的统计次数最低；或根据轮询调度算法或加权的轮询调度算法的输入。

[0128] 可选的，作为不同的实施例，所述记录模块记录的所述 IO 历史记录中的每一条至少包括以下内容：有效性（Valid）、指定地址（Address）、目的地（Destination）和统计参数（Statistical Parameters），其中所述目的地和统计参数一一对应。记录的具体内容参考表 3。

[0129] 图 8 是本发明实施例的 CC-NUMA 系统中的另一种装置 80 的示意框图。装置 80 包括处理器 81、存储器 82。

[0130] 存储器 82 用于存储处理器 81 执行本发明实施例的方法的可执行程序。此外，存储器 82 记录输入输出 IO 历史记录，所述 IO 历史记录是指针对至少一个指定地址，从输入输出集线器 IOH 经所述装置到至少一个远端中央处理器 CPU 的缓存 CA 的推送 IO 数据的历史统计记录。

[0131] 处理器 81 确定所述 IO 历史记录是否符合预设条件；当确定所述 IO 历史记录符合预设条件时，则向所述远端 CPU 的 CA 发送预提取提示报文，所述预提取提示报文用于使所述远端 CPU 的 CA 针对所述预提取提示报文中的地址发起 IO 数据的预取访问。

[0132] 本发明实施例通过 CC-NUMA 系统中的装置记录针对一个或多个指定地址的推送 IO 数据的历史记录，通过分析该历史记录符合预定条件，从而推测远端 CPU 可能在将来的时间点使用该指定地址的 IO 数据，该装置主动发送针对该指定地址的 IO 数据提示报文给该远端的 CPU 的 CA，并且由该远端 CPU 提前发起针对该指定地址的 IO 数据预取操作，由此缩短了远端 IO 数据访问延时，提升了系统的性能。

[0133] 装置 80 可以执行方法 30 或 40，结构例如图 6A 或图 6B 中所示的动态事物监控器，作为不同的实现方式可以是 NC，也可以包括在 CC-NUMA 的 NC 中，还可以独立存在。其中，NC 可以是一块专用的芯片或者现场可编程门阵列（FPGA，Field Programmable Gate Array）设备等。

[0134] 可选的，作为不同的实施例，当所述处理器确定所述 IO 历史记录中有关所述指定地址的统计次数大于预设的门限值时，则所述处理器向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或当所述处理器确定所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔小于预设的门限值时，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或当所述处理器确定所述 IO 历史记录中单位时间内有关所述指定地址的统计次数大于预设的门限值时，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文；或所述处理器确定轮询调度算法或加权的轮询调度算法的输入为所述远端 CPU 的 CA，则向所述远端 CPU 的 CA 发送关于所述指定地址的预提取提示报文。

[0135] 可选的，作为不同的实施例，所述处理器管理 IO 历史记录的插入、更新、替换和删除。

[0136] 可选的，作为不同的实施例，所述处理器根据以下条件之一，优先进行 IO 历史记录的更新或替换：所述 IO 历史记录中有关所述指定地址的统计次数最低；或所述 IO 历史记录中有关所述指定地址的两次记录的计时间隔最长；或所述 IO 历史记录中单位时间内有关所述指定地址的统计次数最低；或根据轮询调度算法或加权的轮询调度算法的输入。

[0137] 可选的，作为不同的实施例，所述处理器记录的所述 IO 历史记录中的每一条至少包括以下内容：有效性(Valid)、指定地址(Address)、目的地(Destination)和统计参数(Statistical Parameters)，其中所述目的地和统计参数一一对应。记录的具体内容参考表 3。

[0138] 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

[0139] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置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。

[0140] 在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0141] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0142] 另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

[0143] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说

对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(ROM，Read-Only Memory)、随机存取存储器(RAM，Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0144] 以上所述，仅为本发明的具体实施方式，但本发明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发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应所述以权利要求的保护范围为准。

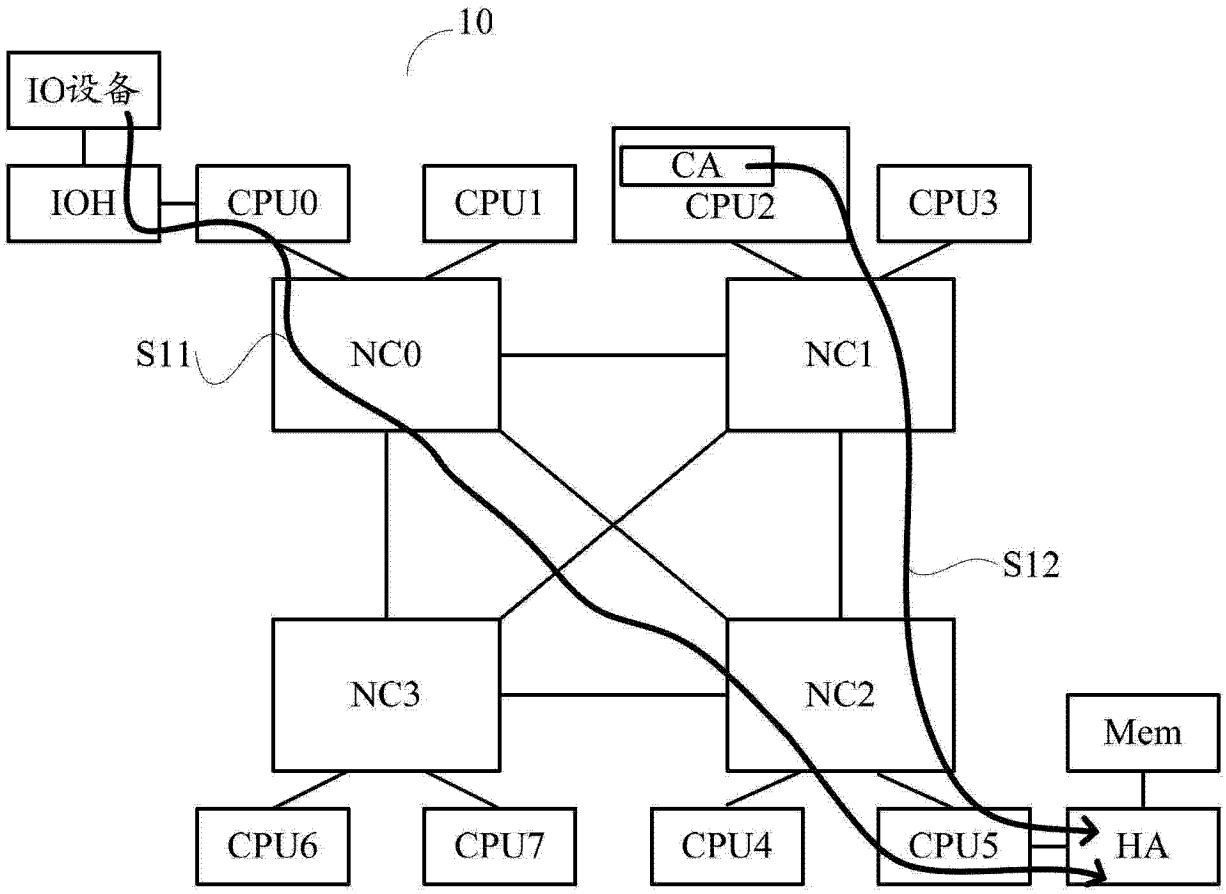

10
IO设备
IOH
CPU0
CPU1
CA
CPU2
CPU3
S11
NC0
NC1
S12
NC3
NC2
Mem
CPU6
CPU7
CPU4
CPU5
HA

图 1

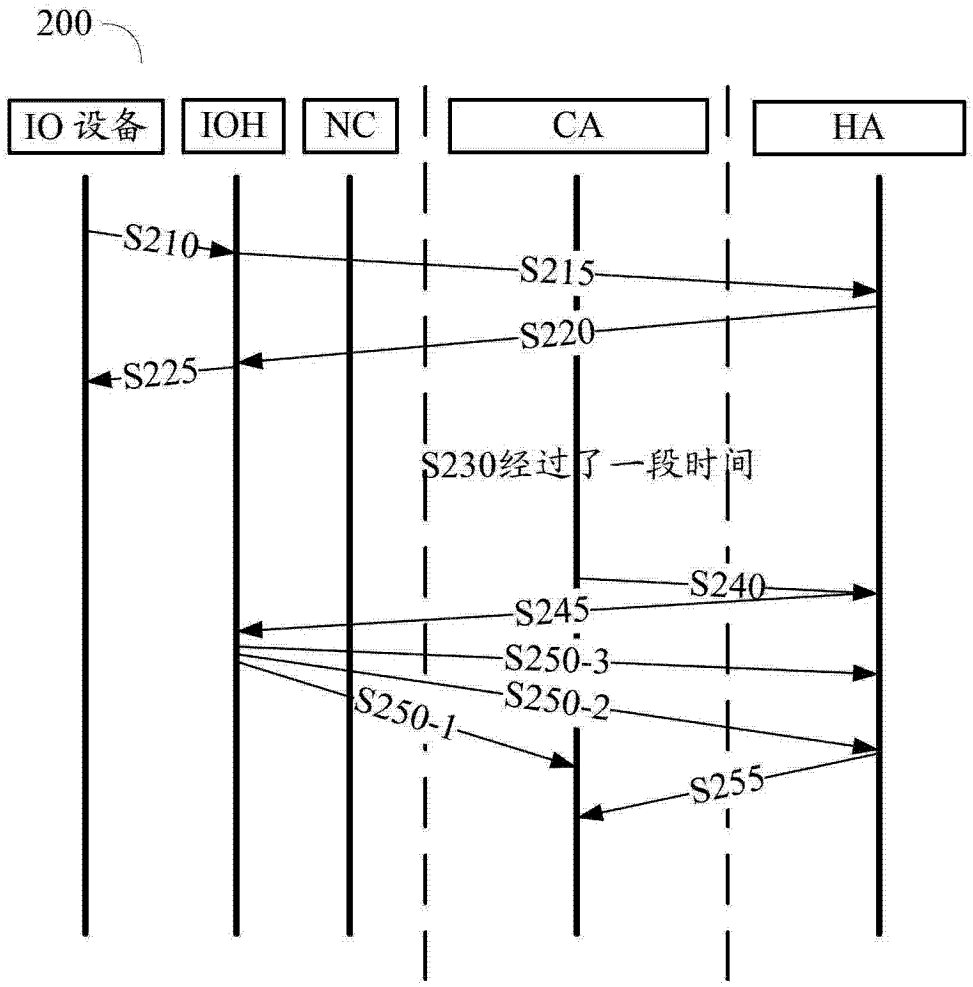

200
IO 设备
IOH
NC
CA
HA
S210
S215
S220
S225
S230经过了一段时间
S240
S245
S250-3
S250-2
S250-1
S255

图 2

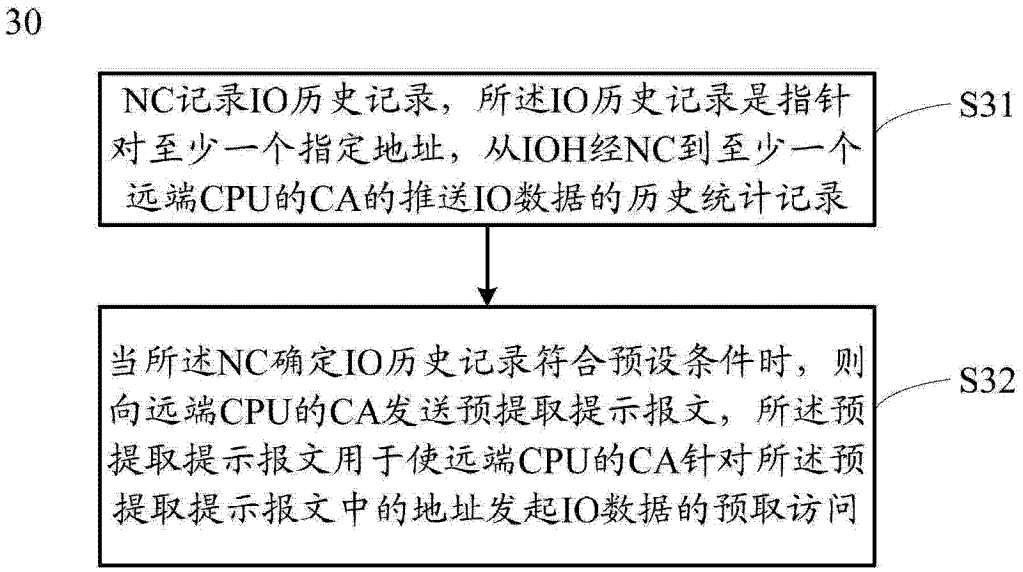

30
NC记录IO历史记录，所述IO历史记录是指针对至少一个指定地址，从IOH经NC到至少一个远端CPU的CA的推送IO数据的历史统计记录
S31
当所述NC确定IO历史记录符合预设条件时，则向远端CPU的CA发送预提取提示报文，所述预提取提示报文用于使远端CPU的CA针对所述预提取提示报文中的地址发起IO数据的预取访问
S32

图 3

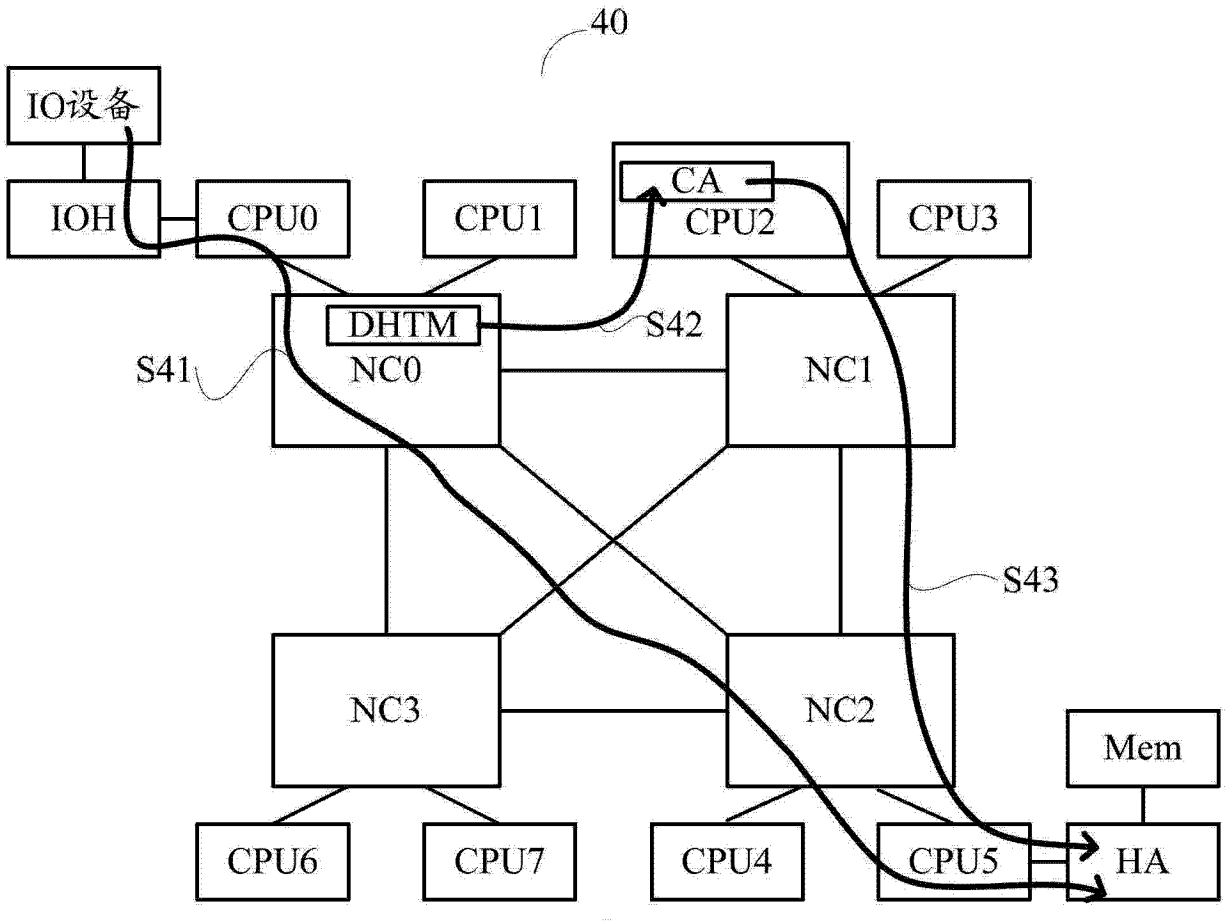

40
IO设备
IOH
CPU0
CPU1
CA
CPU2
CPU3
DHTM
NC0
S41
S42
NC1
S43
NC3
NC2
Mem
CPU6
CPU7
CPU4
CPU5
HA

图 4

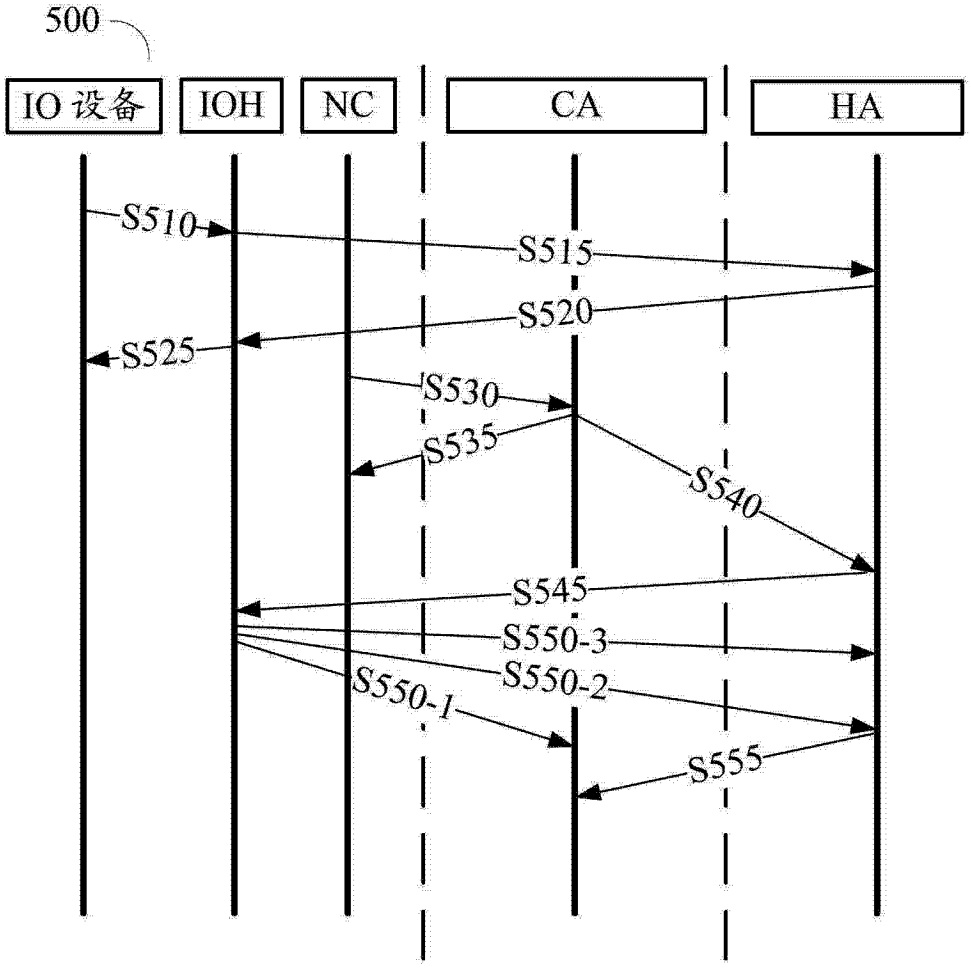


图 5

字段 / 索引	有效性	地址	目的地	统计参数
0				
1				
· · ·				
N-1				

图 6A

字段 / 索引	有效性	地址	目的地1	统计参数1	目的地2	统计参数2	· · ·	目的地A	统计参数B
0									
1									
· · ·									
N-1									

图 6B

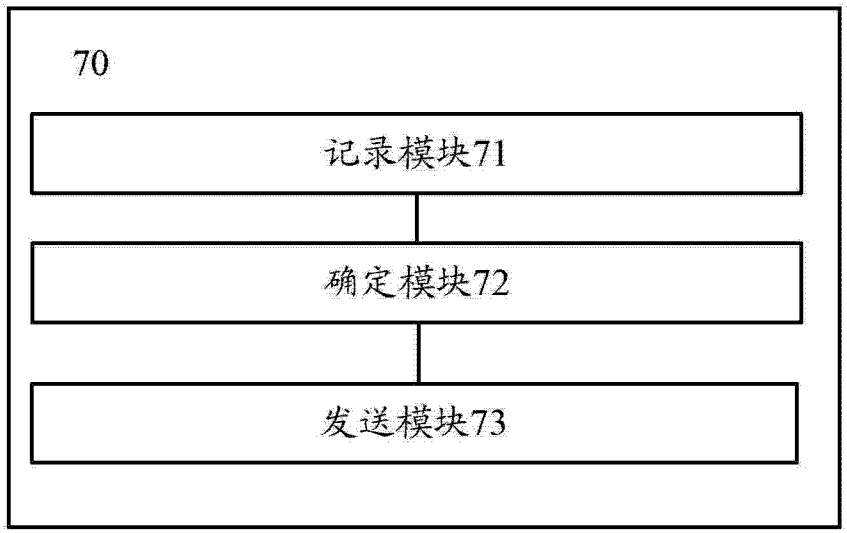


图 7

图 8