**专利申请明细表**

**兹全权委托北京博思佳知识产权代理有限公司办理下列申请专利有关事项**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 发 明 名 称 | | 数据存储系统、数据存储方法及电子设备 | | |
| 第一  申请人 | 姓名或名 称 | 之江实验室 | | |
| 地址、邮编 | 默认：浙江省杭州市余杭区之江实验室南湖总部，311121  日本专利：之江实验室 中国310023浙江省杭州市余杭区文一西路1818号 | | |
| 企业机构代码或身份证号 | 12330000MB1478604D | | |
| 第二  申请人  （如果为多个申请单位共同申请，可填） | 姓名或名 称 | 华中科技大学 | | |
| 地址、邮编 | 地址：湖北省武汉市洪山区珞喻路1037号，430074 | | |
| 企业机构代码或身份证号 | 12100000441626842D | | |
| 发明人  或设计人 | 姓 名 | 张湛、张宇、赵进、吴海飞 | 第一发明人身份证号 | 410105199105240030 |
| 专利负责人 | 姓名 | 吴梅英，18662590857，wumeiying@zhejianglab.com  地址：浙江省杭州市余杭区之江实验室南湖总部主楼1402办公室，311100 | | |
| 地址、邮编 |
| 电话/E-mail |
| 技术（撰写）  联系人 | 姓名 | 张湛，18827096052，zhanzhang@zhejianglab.com  地址：浙江省杭州市余杭区之江实验室南湖总部11号楼B座B102室，311121 | | |
| 地址、邮编 |
| 电话/E-mail |
| 申请专利类型 | | 🗹发明　　　　　实用新型　　　　外观设计 | | |
| 催缴费联系人电话、Email、地址 | | 吴梅英，18662590857，wumeiying@zhejianglab.com，地址：浙江省杭州市余杭区之江实验室南湖总部主楼1402办公室，311100 | | |

注：

1、红色字体部分需要填写或选择

说 明 书 摘 要

本发明公开了一种面向图计算的内存控制器和内存数据崩溃一致性维护方法，属于计算机系统结构设计领域，包括：在内存控制器中实现细粒度的影子内存检查点维护图数据的崩溃一致性，且无需将所有有效映射数据存储在内存控制器缓存中，避免冗余的检查点提交操作；使用回滚日志维护图数据加密所需计数器数据的一致性，从而在映射数据缓存缺失时，系统能重叠执行映射数据访存操作和计数器数据访存操作，缩短系统的访存请求处理流程。本发明分别使用细粒度影子内存检查点和回滚日志维护图数据和计数器数据的崩溃一致性，降低了图数据的加密和持久化开销，可用于构建面向图计算的基于非易失内存的高效加密持久内存系统。



权 利 要 求 书

1. 一种加密持久内存系统的内存控制器设计，其特征在于，包括：由数据存储和数据处理两个模块构成，数据存储模块负责暂存设置影子内存检查点、设置回滚日志、访存请求响应和内存数据加密所需的元数据，数据处理模块负责实现维护数据崩溃一致性、数据加密及数据恢复所需的操作逻辑，与内存数据崩溃一致性（Crash Consistency）维护方法相对应。
2. 根据权利要求1所述的加密持久内存系统的内存控制器设计，其特征在于，所述数据存储模块包括：映射数据缓存、计数器缓存、日志元数据寄存器、检查点元数据寄存器。
3. 根据权利要求2所述的加密持久内存系统的内存控制器设计，其特征在于，所述映射数据缓存设计包括：

映射数据用于确定图应用的工作内存数据或检查点数据处于原内存页面（内存读写请求中目的地址所在页面）还是影子内存页面中，相邻2个原内存页面（尺寸为4KB）对应的映射数据被存储在一个64B的映射数据缓存行中；

映射数据缓存断电数据丢失，以原内存页面对应的映射数据内存地址为索引，暂存加密持久内存中较热的映射数据，如图 1所示；映射数据缓存为每一个缓存行设置标志位判断缓存行是否被修改，较冷的发生修改的映射数据将被写回加密持久内存，无需在映射数据缓存中保存所有当前有效的映射数据；映射数据缓存容量依照持久内存容量而定；

映射数据结构如图 2所示，其中检查点编号为64 bits，代表修改映射数据时的检查点周期编号；影子页面地址为64 bits，页面大小为4KB包含64个64B数据块，数据块尺寸与单个内存请求数据量相同，未发生修改的原内存页面不分配影子内存页面，对应影子页面地址为0；检查点数据位图为64 bits，位图中的每个标志位（C标志位）对应逻辑页面中的一个64B数据块，标志位取0表示数据块对应的检查点数据保存在原内存页面，取1表示保存在影子内存页面；工作数据位图位为64 bit，位图中的每个标记（W标志位）为对应位图中的每个标志位对应逻辑页面中的一个64B数据块，标志位取0表示数据块对应的最新版本数据保存在原内存页面，取1表示保存在影子内存页面。

1. 根据权利要求2所述的加密持久内存系统的内存控制器设计，其特征在于，所述计数器缓存设计包括：

原内存页面中的各个64B数据块，与影子内存页面中的对应数据块，共用同一个加密计数器，无论原内存还是影子内存中的数据块发生修改，其计数器都将自增；

单个计数器缓存行能保存整个内存页面中所有数据块的计数器**[[1]](#footnote-1)**，计数器缓存以原内存页面对应的计数器数据内存地址作为索引，缓存加密持久内存中位于原内存区域的计数器数据，如图 1所示，计数器缓存容量随映射数据缓存容量设置，保证绝大多数情况下计数器缓存命中则映射数据缓存也命中。

1. 根据权利要求2所述的加密持久内存系统的内存控制器设计，其特征在于，所述日志元数据寄存器设计包括：

计数器数据的回滚日志由日志项和日志头组成，每个日志项存储64B修改前的计数器数据块，计数器数据块的内存地址送往64B大小的日志头中，每8个连续的日志项后为其对应的日志头；

日志元数据寄存器暂存了当前可用日志项对应的日志头，当日志头寄存器装满8个计数器数据块地址（或再没有新的日志项到来时）将被写回持久内存，当意外断电发生时，日志头寄存器将借助ADR写回持久内存；

日志元数据寄存器暂存了下个可用日志内存的地址，将随着日志数据的写入而更新，当意外断电发生时，日志内存地址寄存器将借助ADR写回持久内存，用于界定计数器回滚日志的范围。

1. 根据权利要求2所述的加密持久内存系统的内存控制器设计，其特征在于，所述检查点元数据寄存器设计包括：

检查点元数据寄存器保存了当前系统检查点周期的编号，该编号将随着检查点数据的成功提交而增加，当系统崩溃时检查点编号寄存器将在ADR的支持下写回加密持久内存中的检查点元数据区域；

检查点元数据寄存器保存了映射数据在加密持久内存中的起始地址，系统可以根据内存读写请求的目的内存地址和映射数据的起始内存地址，计算出内存请求对应的映射数据内存地址；

检查点元数据寄存器缓存了影子内存空间分配位图的较热部分，位图中的每一位代表着对应影子内存页面是否已经被分配，发生修改的位图数据将被写回加密持久内存中的检查点元数据区域，当意外断电发生时，影子内存空间位图寄存器将借助ADR写回持久加密内存；

检查点元数据寄存器缓存了四个标志位，用于标记在检查点提交阶段，处理器缓存、计数器缓存、映射数据缓存中的脏数据以及处理器状态数据，是否刷新到了加密持久内存，其中处理器状态数据应被刷新到加密持久内存中的检查点元数据区域，如图 1所示。

1. 根据权利要求2所述的加密持久内存系统的内存控制器设计，其特征在于，所述数据处理模块包括：内存读写请求处理模块、检查点提交处理模块和数据恢复处理模块，这些模块分别为内存数据崩溃一致性维护方法中的内存读写请求处理方法、影子内存检查点提交方法和系统数据恢复方法，在内存控制器中的执行逻辑实现。
2. 一种内存数据崩溃一致性维护方法设计，其特征在于，包括：使用细粒度影子内存检查点和回滚日志维护图数据和加密所需的计数器数据的崩溃一致性和可恢复性，包括内存读写请求处理方法、影子内存检查点提交方法和系统数据恢复方法。
3. 根据权利要求8所述的内存数据崩溃一致性维护方法设计，其特征在于，所述内存读请求处理方法包括：

当图应用的读内存请求到达内存控制器后，系统依据请求的内存地址计算出其对应的映射数据内存地址、计数器内存地址，并分别以这两个内存地址为索引访问映射数据缓存和计数器缓存；

当计数器缓存命中时，系统将读请求对应的计数器送入加密引擎，计算解密所需的一次性数据块（One-Time-Pad，OTP）；当计数器缓存未命中时，系统从内存中读取对应的计数器后再计算OTP；

当映射数据缓存命中时，系统读取映射数据的工作数据位图中读请求对应的W标志位，并将读请求送到W标志位指向的内存页面读取图数据密文；当映射数据缓存未命中时，系统先从内存中获取映射数据，之后根据映射数据从内存中读取图数据密文，并将映射数据送入映射数据缓存；

当OTP计算完成且图数据密文被读出后，系统将二者相异或得到数据明文，完成读内存请求的处理。

1. 根据权利要求8所述的内存数据崩溃一致性维护方法设计，其特征在于，所述内存写请求处理方法包括：

当图应用的写内存请求到达内存控制器后，系统需要同时访问计数器缓存和映射数据缓存，如果缓存未能命中，则系统需要访问内存获取数据；

当系统获取到计数器数据后，如果对应计数器缓存行的脏数据标志位未被设置，则需要在修改计数器取值前先将计数器缓存行写入回滚日志，系统根据日志元数据寄存器确定可用日志空间地址将原计数器缓存行作为日志项写回内存，并将缓存行内存地址送入日志元数据寄存器中的日志头内，最后系统更新计数器取值和缓存行标志位，如果更新后的计数器取值第16位发生跳变，则将整个计数器缓存行写回持久内存；

当计数器取值更新完成后，系统根据计数器计算加密所需的OTP，再与图数据明文相异或得到密文；

当获取到映射数据后，系统将比较内存控制器寄存器中的系统当前检查点周期编号与写请求对应映射数据中的检查点周期编号，并更新映射数据；如果系统当前检查点周期编号较大，说明最近一次对映射数据对应图数据的更新，发生在之前成功提交的检查点周期内，此时将映射数据中的C标志位设置为W标志位的取值，同时设置映射数据的检查点周期编号为系统当前检查点周期编号；如果与系统当前检查点周期编号相等，则映射数据W标志位指向的图数据，其未被提交的修改只会发生在当前检查点周期内，无需更新映射数据；

当映射数据完成更新并且图数据完成加密后，系统从映射数据中读取写请求内存地址对应的C和W标志位，如果两个标志位相同，则系统翻转W标志位，并将图数据密文送往W标志位指向的内存页面，如果两个标志位不同，则系统直接将数据密文送往W标志位指向的内存页面。

1. 根据权利要求8所述的内存数据崩溃一致性维护方法设计，其特征在于，所述影子内存检查点提交方法包括：

在检查点周期的末尾，系统暂停图应用程序的执行，将处理器缓存中发生修改的缓存行写回加密持久内存，数据写回流程与内存写请求流程相同；

处理器缓存中的脏数据全部写回加密持久内存后，系统将计数器缓存中的脏数据、映射数据缓存中的脏数据和处理器状态数据，分别写回加密持久内存中的计数器数据区域、映射数据区域和检查点元数据区域；

最后系统增加寄存器中的当前系统检查点周期编号进入下个检查点周期，同时日志区域的当前计数器回滚日志可以被舍弃，日志空间在下个周期被重用。

1. 根据权利要求8所述的内存数据崩溃一致性维护方法设计，其特征在于，所述系统数据恢复方法包括：

系统从加密持久内存的日志区域读取系统发生崩溃的检查点周期内生成的计数器回滚日志，根据计数器缓存行和图数据页面的对应关系，确定检查点周期内发生修改的图数据页面的原内存地址；

系统依据发生修改的原内存页面地址，访问相应的映射数据，系统对比映射数据缓存行中的检查点周期编号，与加密持久内存中检查点元数据内的系统检查点周期编号，当映射数据中的检查点编号等于系统检查点编号时，系统将映射数据中与C标志位取值不同的W标志位更新为C标志位的取值；

系统从加密持久内存中的检查点元数据区域读取崩溃发生时的检查点周期编号，并送入检查点元数据寄存器，之后系统将增加检查点编号，并在该检查点周期内继续执行数据恢复操作；

为了避免计数器取值重用，系统对计数器回滚日志中的每一个计数器，都访问其对应的图数据密文，并利用回滚日志中计数器的取值对映射数据C标志位指向的图数据密文解密，之后系统再读取计数器在原内存区域的取值，将其取值增加到第16位发生跳变后，利用新值对图数据明文重加密，并更新计数器缓存中的计数器取值，同时按照内存写请求的处理方法更新映射数据，并将重新加密后的图数据密文写回内存；

之后系统将计数器和映射数据缓存中的脏数据刷新回内存，同时将内存中检查点元数据区域内的处理器状态数据恢复到处理器的寄存器中，增加检查点周期编号完成数据恢复，继续图应用程序的执行，并回收计数器回滚日志占用的内存空间，供之后的检查点周期使用；

如果系统在进行数据恢复的过程中再一次崩溃，由于恢复操作是在新的检查点周期内进行，并未对已经提交的检查点数据进行修改，系统只需要重新执行恢复操作即可。

说 明 书

**面向图计算的内存控制器和内存数据崩溃一致性维护方法**

**技术领域**

本发明属于计算机系统结构设计领域，更具体地，涉及面向图计算的内存控制器和内存数据崩溃一致性维护方法。

**技术背景**

随着图数据规模的爆炸式增长，图计算应用对内存容量的需求越来越高，同时大规模图计算会产生大量的图数据更新，加大了图数据持久化的开销。非易失内存（Non-Volatile Memory，NVM）可被按字节寻址并且具有内存级别的数据访问性能，以及拥有比DRAM内存更高的数据存储密度和断电数据不丢失的非易失性。利用NVM可以实现能同时作为内存与存储设备使用的持久内存，从而满足图计算应用对大容量内存系统和低持久化开销存储系统的需求。此外，图数据常常包含敏感信息，内存数据加密能够有效防止攻击者对敏感图数据的窃取，因此设计面向图计算的高效加密持久内存系统具有重要意义。

已有研究[[2]](#footnote-2)[[3]](#footnote-3)通过优化内存控制器结构，在硬件层实现细粒度的影子内存检查点来维护内存数据的崩溃一致性，设计出了对用户透明的持久内存系统，在降低用户编程难度的同时，减少了内存数据持久化过程带来的系统软件开销。此外，相比使用日志、写时拷贝实现数据检查点的方式，使用细粒度影子内存能够有效减少设置检查点过程中产生的额外写开销，在加密持久内存系统中，避免额外的写开销不仅能减少系统的数据持久化开销，还可以有效减少内存数据的加密开销。因此，通过在内存控制器中实现细粒度的影子内存检查点，以维护内存数据崩溃一致性的策略，适合用于设计面向图计算的高效加密持久内存系统。然而，已有研究都只是使用影子内存检查点实现持久内存系统，当设计面向图计算的、基于影子内存检查点的高效加密持久内存系统时，还需应对以下挑战：

1. 维护计数器数据一致性。当前常用的内存数据加密方法为基于计数器的数据加密策略，如图 3所示，每一个需要加密的数据块对应着一个计数器，数据块中明文发生更改则计数器也要随之自增，数据块在解密时需要用到与加密时相同的计数器。为加速数据加解密，计数器被缓存在内存控制器的易失缓存中，该缓存不受ADR[[4]](#footnote-4)支持，系统断电则数据丢失。因此，相比持久内存系统，面向图计算的加密持久内存系统不仅需要维护图数据之间的数据一致性，还需要维护图数据和其计数器数据的一致性，以保证图数据能被正确的加解密，数据一致性维护难度更高；
2. 避免冗余停机检查点提交操作。已有的基于细粒度影子内存检查点的持久内存系统，需要在内存控制器的映射数据缓存中维护所有有效的映射数据，当片上缓存空间不足时就需要提前进行检查点提交操作，暂停程序执行并合并原内存页面和影子内存页面，以回收映射数据缓存空间。在检查点提交阶段执行的访存请求，均位于系统运行的关键路径之上，因此这些额外的检查点提交操作，增加了系统的数据持久化开销，而在加密内存中，这些额外的访存操作还将增加系统的加密开销，进一步降低系统性能。

**发明内容**

基于上述技术背景，针对已有技术的不足和改进需求，本发明设计了面向图计算的内存控制器和内存数据崩溃一致性维护方法，其目的在于，在维护加密持久内存中图数据的保密性和崩溃一致性的前提下，通过解除所有影子内存检查点的映射数据均需存储在片上缓存中的限制，避免冗余的检查点提交操作，以减少图数据的持久化和加密开销，同时使用回滚日志维护图数据加密所需计数器数据的一致性，使映射数据访存操作和计数器数据访存操作能重叠执行，降低映射数据缓存缺失时系统的访存开销，最终使本发明能用于构建面向图计算的高效加密持久内存系统。

为了实现上述技术目的，本发明的技术方案为：

本发明实施例的第一方面设计了面向图计算的加密持久内存系统内存控制器，具体为：对内存控制器的新设计由数据存储和数据处理两个模块构成，数据存储模块负责暂存设置影子内存检查点、设置回滚日志、访存请求响应和内存数据加密所需的元数据，数据处理模块负责实现维护数据崩溃一致性及数据恢复所需的操作逻辑，与内存数据崩溃一致性维护方法相对应。

进一步地，所述数据存储模块包括：映射数据缓存、计数器缓存、日志元数据寄存器、检查点元数据寄存器。

所述映射数据缓存设计包括：

映射数据用于确定图应用的工作内存数据或检查点数据处于原内存页面（内存读写请求中目的地址所在页面）还是影子内存页面中，相邻2个原内存页面（尺寸为4KB）对应的映射数据被存储在一个64B的映射数据缓存行中；

映射数据缓存断电数据丢失，以原内存页面对应的映射数据内存地址为索引，暂存加密持久内存中较热的映射数据，如图 1所示；映射数据缓存为每一个缓存行设置标志位判断缓存行是否被修改，较冷的发生修改的映射数据将被写回加密持久内存，无需在映射数据缓存中保存所有当前有效的映射数据；映射数据缓存容量依照持久内存容量而定；

映射数据的结构如图 2所示，其中检查点编号为64 bits，代表修改映射数据时的检查点周期编号；影子页面地址为64 bits，页面大小为4KB包含64个64B数据块，数据块尺寸与单个内存请求数据量相同，未发生修改的原内存页面不分配影子内存页面，对应影子页面地址为0；检查点数据位图为64 bits，位图中的每个标志位（C标志位）对应逻辑页面中的一个64B数据块，标志位取0表示数据块对应的检查点数据保存在原内存页面，取1表示保存在影子内存页面；工作数据位图位为64 bit，位图中的每个标记（W标志位）为对应位图中的每个标志位对应逻辑页面中的一个64B数据块，标志位取0表示数据块对应的最新版本数据保存在原内存页面，取1表示保存在影子内存页面。

所述计数器缓存设计包括：

原内存页面中的各个64B数据块，与影子内存页面中的对应数据块，共用同一个加密计数器，无论原内存还是影子内存中的数据块发生修改，其计数器都将自增；

单个计数器缓存行能保存整个内存页面中所有数据块的计数器**[[5]](#footnote-5)**，计数器缓存以原内存页面对应的计数器数据内存地址作为索引，缓存加密持久内存中位于原内存区域的计数器数据，如图 1所示，计数器缓存容量随映射数据缓存容量设置，保证绝大多数情况下计数器缓存命中则映射数据缓存也命中。

所述日志元数据寄存器设计包括：

计数器数据的回滚日志由日志项和日志头组成，每个日志项存储64B修改前的计数器数据块，计数器数据块的内存地址送往64B大小的日志头中，每8个连续的日志项后为其对应的日志头；

日志元数据寄存器暂存了当前可用日志项对应的日志头，当日志头寄存器装满8个计数器数据块地址（或再没有新的日志项到来时）将被写回持久内存，当意外断电发生时，日志头寄存器将借助ADR写回持久内存；

日志元数据寄存器暂存了下个可用日志内存的地址，将随着日志数据的写入而更新，当意外断电发生时，日志内存地址寄存器将借助ADR写回持久内存，用于界定计数器回滚日志的范围。

所述检查点元数据寄存器设计包括：

检查点元数据寄存器保存了当前系统检查点周期的编号，该编号将随着检查点数据的成功提交而增加，当系统崩溃时检查点编号寄存器将在ADR的支持下写回加密持久内存中的检查点元数据区域；

检查点元数据寄存器保存了映射数据在加密持久内存中的起始地址，系统可以根据内存读写请求的目的内存地址和映射数据的起始内存地址，计算出内存请求对应的映射数据内存地址；

检查点元数据寄存器缓存了影子内存空间分配位图的较热部分，位图中的每一位代表着对应影子内存页面是否已经被分配，发生修改的位图数据将被写回加密持久内存中的检查点元数据区域，当意外断电发生时，影子内存空间位图寄存器将借助ADR写回持久加密内存；

检查点元数据寄存器缓存了四个标志位，用于标记在检查点提交阶段，处理器缓存、计数器缓存、映射数据缓存中的脏数据以及处理器状态数据，是否刷新到了加密持久内存，其中处理器状态数据应被刷新到加密持久内存中的检查点元数据区域，如图 1所示。

进一步地，所述数据处理模块包括：内存读写请求处理模块、检查点提交处理模块和数据恢复处理模块，这些模块分别为内存数据崩溃一致性维护方法中的内存读写请求处理方法、影子内存检查点提交方法和系统数据恢复方法，在内存控制器中的执行逻辑实现。

本发明实施例的第二方面设计了面向图计算的加密持久内存系统的内存数据崩溃一致性维护方法设计，具体为：分别使用影子内存检查点和回滚日志保护图数据和计数器数据的崩溃一致性，包括内存读写请求处理方法、影子内存检查点提交方法和系统数据恢复方法。

所述内存读请求处理方法包括：

当图应用的读内存请求到达内存控制器后，系统依据请求的内存地址计算出其对应的映射数据内存地址、计数器内存地址，并分别以这两个内存地址为索引访问映射数据缓存和计数器缓存；

当计数器缓存命中时，系统将读请求对应的计数器送入加密引擎，计算解密所需的一次性数据块（One-Time-Pad，OTP）；当计数器缓存未命中时，系统从内存中读取对应的计数器后再计算OTP；

当映射数据缓存命中时，系统读取映射数据的工作数据位图中读请求对应的W标志位，并将读请求送到W标志位指向的内存页面读取图数据密文；当映射数据缓存未命中时，系统先从内存中获取映射数据，之后根据映射数据从内存中读取图数据密文，并将映射数据送入映射数据缓存；

当OTP计算完成且图数据密文被读出后，系统将二者相异或得到数据明文，完成读内存请求的处理。

所述内存写请求处理方法包括：

当图应用的写内存请求到达内存控制器后，系统需要同时访问计数器缓存和映射数据缓存，如果缓存未能命中，则系统需要访问内存获取数据；

当系统获取到计数器数据后，如果对应计数器缓存行的脏数据标志位未被设置，则需要在修改计数器取值前先将计数器缓存行写入回滚日志，系统根据日志元数据寄存器确定可用日志空间地址将原计数器缓存行作为日志项写回内存，并将缓存行内存地址送入日志元数据寄存器中的日志头内，最后系统更新计数器取值和缓存行标志位，如果更新后的计数器取值第16位发生跳变，则将整个计数器缓存行写回持久内存；

当计数器取值更新完成后，系统根据计数器计算加密所需的OTP，再与图数据明文相异或得到密文；

当获取到映射数据后，系统将比较内存控制器寄存器中的系统当前检查点周期编号与写请求对应映射数据中的检查点周期编号，并更新映射数据；如果系统当前检查点周期编号较大，说明最近一次对映射数据对应图数据的更新，发生在之前成功提交的检查点周期内，此时将映射数据中的C标志位设置为W标志位的取值，同时设置映射数据的检查点周期编号为系统当前检查点周期编号；如果与系统当前检查点周期编号相等，则映射数据W标志位指向的图数据，其未被提交的修改只会发生在当前检查点周期内，无需更新映射数据；

当映射数据完成更新并且图数据完成加密后，系统从映射数据中读取写请求内存地址对应的C和W标志位，如果两个标志位相同，则系统翻转W标志位，并将图数据密文送往W标志位指向的内存页面，如果两个标志位不同，则系统直接将数据密文送往W标志位指向的内存页面。

所述影子内存检查点提交方法包括：

在检查点周期的末尾，系统暂停图应用程序的执行，将处理器缓存中发生修改的缓存行写回加密持久内存，数据写回流程与内存写请求流程相同；

处理器缓存中的脏数据全部写回加密持久内存后，系统将计数器缓存中的脏数据、映射数据缓存中的脏数据和处理器状态数据，分别写回加密持久内存中的计数器数据区域、映射数据区域和检查点元数据区域；

最后系统增加寄存器中的当前系统检查点周期编号进入下个检查点周期，同时日志区域的当前计数器回滚日志可以被舍弃，日志空间在下个周期被重用。

所述系统数据恢复方法包括：

系统从加密持久内存的日志区域读取系统发生崩溃的检查点周期内生成的计数器回滚日志，根据计数器缓存行和图数据页面的对应关系，确定检查点周期内发生修改的图数据页面的原内存地址；

系统依据发生修改的原内存页面地址，访问相应的映射数据，系统对比映射数据缓存行中的检查点周期编号，与加密持久内存中检查点元数据内的系统检查点周期编号，当映射数据中的检查点编号等于系统检查点编号时，系统将映射数据中与C标志位取值不同的W标志位更新为C标志位的取值；

系统从加密持久内存中的检查点元数据区域读取崩溃发生时的检查点周期编号，并送入检查点元数据寄存器，之后系统将增加检查点编号，并在该检查点周期内继续执行数据恢复操作；

为了避免计数器取值重用，系统对计数器回滚日志中的每一个计数器，都访问其对应的图数据密文，并利用回滚日志中计数器的取值对映射数据C标志位指向的图数据密文解密，之后系统再读取计数器在原内存区域的取值，将其取值增加到第16位发生跳变后，利用新值对图数据明文重加密，并更新计数器缓存中的计数器取值，同时按照内存写请求的处理方法更新映射数据，并将重新加密后的图数据密文写回内存；

之后系统将计数器和映射数据缓存中的脏数据刷新回内存，同时将内存中检查点元数据区域内的处理器状态数据恢复到处理器的寄存器中，增加检查点周期编号完成数据恢复，继续图应用程序的执行，并回收计数器回滚日志占用的内存空间，供之后的检查点周期使用；

如果系统在进行数据恢复的过程中再一次崩溃，由于恢复操作是在新的检查点周期内进行，并未对已经提交的检查点数据进行修改，系统只需要重新执行恢复操作即可。

总体而言，通过本发明所构思的以上技术方案，能够取得以下有益效果：

1. 本发明的内存控制器能在保证图数据一致性和保密性的前提下，使映射数据缓存可以在程序运行过程中逐出较冷的映射数据，因此不会因映射数据缓存空间不足而提前进行检查点提交操作，减少了因冗余检查点提交操作产生的、位于系统执行关键路径上的额外内存读写请求，进而降低了系统的数据持久化和加密开销，以及系统停机时间；
2. 允许映射数据在程序运行时逐出缓存，虽然可以避免额外的检查点提交操作，但是需要在映射数据缓存未命中时访问内存获取映射数据，才可确定图数据块和计数器数据块的内存地址，延长了内存读写请求的处理流程。因此，本发明通过使用回滚日志而非影子内存，来维护计数器数据的一致性，保证计数器数据的最新取值位于原内存区域，无需经映射数据重定向即可访问，从而使访问计数器内存和映射数据内存的过程相重叠，缩短了内存请求的处理流程。图 4和图 5展示映射数据缓存和计数器缓存均缺失时[[6]](#footnote-6)，内存读写请求的处理流程，从图示可以看出，允许映射数据逐出缓存并且使用回滚日志维护计数器数据一致性，对比不允许映射数据逐出缓存并且使用影子内存检查点维护计数器数据一致性，能保证内存写请求处理流程长度相同，内存读请求处理流程长度稍有延长。允许映射数据逐出缓存，能避免冗余的检查点提交操作，有效减少系统停机期间的内存读写请求数量，大大降低系统的数据加密和持久化开销，相比之下仅稍增映射数据缓存缺失时的内存读请求处理开销，代价可以被接受。

**附图说明**

图1是本发明实施例提供的面向图计算的加密持久内存系统整体结构示意图；

图2是本发明实施例提供的映射数据数据结构示意图；

图3是本发明实施例采用的基于计数器的数据加密方法处理流程示意图；

图4是本发明实施例提供的当映射数据缓存、计数器缓存均未命中时，分别允许映射数据逐出缓存且采用回滚日志维护计数器数据一致性，以及不允许映射数据逐出缓存且采用影子内存检查点维护计数器数据一致性时，内存读请求的处理流程对比；

图5是本发明实施例提供的当映射数据缓存、计数器缓存均未命中时，分别允许映射数据逐出缓存且采用回滚日志维护计数器数据一致性，以及不允许映射数据逐出缓存且采用影子内存检查点维护计数器数据一致性时，内存写请求的处理流程对比；

图6是本发明实施例提供的面向图计算的加密持久内存系统内存读请求处理方法流程图；

图7是本发明实施例提供的面向图计算的加密持久内存系统内存写请求处理方法流程图。

**具体实施方式**

下面结合具体实施例对本发明进行详细说明。以下实施例将有助于本领域的技术人员进一步理解本发明，但不以任何形式限制本发明。应当指出的是，对本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变化和改进。这些都属于本发明的保护范围。

实施例1：基于本发明实施例提供的内存控制器和内存数据崩溃一致性维护方法，而实现的面向图计算的加密持久内存系统中内存读请求的处理流程，如图 6所示，包含以下步骤：

第1步，当内存读请求到达内存控制器，系统根据读请求的目标内存地址（原内存地址），计算出被请求数据块对应计数器数据和映射数据的内存地址；

第2步，系统分别使用读请求的映射数据和计数器数据的内存地址作为索引，访问映射数据缓存和计数器缓存；

第3步，如果计数器缓存缺失，则系统从内存中读取计数器数据并送入计数器缓存。当系统获取到计数器数据后，将计数器数据、读请求的内存地址（原内存地址）和密钥送入加密引擎计算OTP；

第4步，如果映射数据缓存缺失，则系统从内存中读取映射数据并送入映射数据缓存。当系统获取到映射数据后，系统根据读请求目标地址在内存页面中的偏移量，确定目标数据块在工作数据位图中的对应W标志位。当W标志位为0时，系统从原内存页面中读取目标数据块，为1则从影子页面中读取目标数据块；

第5步，当读请求目标数据的密文由内存到达内存控制器，并且OTP也已计算完成后，系统将二者相异或得到数据明文，将明文返回给处理器完成读内存请求的处理。

实施例2：基于本发明实施例提供的内存控制器和内存数据崩溃一致性维护方法，而实现的面向图计算的加密持久内存系统中内存写请求的处理流程，如图 7所示，包含以下步骤：

第1步，当内存写请求到达内存控制器后，系统根据写请求的目标内存地址（原内存地址），计算出被写入数据块对应计数器数据和映射数据的内存地址；

第2步，系统分别使用写请求的映射数据和计数器数据的内存地址作为索引，访问映射数据缓存和计数器缓存；

第3步，如果计数器缓存缺失，则系统从内存中读取计数器数据并送入计数器缓存。当系统获取到计数器数据后，增加计数器取值，将更新后的计数器数据、写请求的内存地址（原内存地址）和密钥送入加密引擎计算OTP；

第4步，系统将更新后的计数器写回计数器缓存，如果对应计数器缓存行的脏数据标志位未被设置，则需要在修改计数器取值前先将计数器缓存行写入回滚日志。如果更新后的计数器取值第16位发生跳变，则系统将更新后的计数器缓存行写回加密持久内存；

第5步，系统将计数器缓存行修改前的取值作为回滚日志项写回内存，日志地址根据日志元数据寄存器来确定，然后系统将计数器内存地址送入日志元数据寄存器中的日志头内，同时更新可用日志空间地址寄存器。当日志元数据寄存器中的日志头写满8个日志项的计数器内存地址后，日志头将被写回可用日志空间地址寄存器指向的内存地址，之后系统更新可用日志空间地址寄存器；

第6步，当OTP计算完成后，系统将OTP和写请求中的数据明文相异或得到数据密文；

第7步，如果映射数据缓存缺失，则系统从内存中读取映射数据并送入映射数据缓存。当系统获取到映射数据后，系统将比较内存控制器寄存器中的系统当前检查点周期编号与写请求对应映射数据中的检查点周期编号。如果系统当前检查点周期编号较大，则将映射数据中的C标志位设置为W标志位的取值，同时设置映射数据的检查点周期编号为系统当前检查点周期编号；

第8步，系统根据写请求目标地址在内存页面中的偏移量，确定目标数据块在映射数据的工作数据位图和检查点数据位图中对应的W和C标志位，如果两个标志位相同，则系统翻转W标志位更新映射数据缓存行，并将数据密文送往W标志位指向的内存页面，如果两个标志位不同，则系统直接将数据密文送往W标志位指向的内存页面，完成内存写请求的处理。

以上实施例仅用于说明本发明的设计思想和特点，其目的在于使本领域内的技术人员能够了解本发明的内容并据以实施，本发明的保护范围不限于上述实施例。所以，凡依据本发明所揭示的原理、设计思路所作的等同变化或修饰，均在本发明的保护范围之内。

说 明 书 附 图



图 1



图 2



图 3



图 4



图 5



图 6



图 7

1. Chenyu Yan, Daniel Englender, Milos Prvulovic, Brian Rogers, Yan Solihin. Improving Cost, Performance, and Security of Memory Encryption and Authentication. ACM SIGARCH Computer Architecture News, 2006, 34(2): 179-190 [↑](#footnote-ref-1)
2. Jinglei Ren, Jishen Zhao, Samira Khan, Jongmoo Choi, Yongwei Wu, Onur Mutiu. ThyNVM: Enabling Software-Transparent Crash Consistency in Persistent Memory Systems. In: Proceedings of the 48th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 2015), Waikiki, HI, USA, December 5-9, 2015, ACM, 2015: 672-685 [↑](#footnote-ref-2)
3. Song Wu, Fang Zhou, Xiang Gao, Hai Jin, Jinglei Ren. Dual-Page Checkpointing: An Architectural Approach to Efficient Data Persistence for In-Memory Applications. ACM Transactions on Architecture and Code Optimization, 2018, 15(4): 1-27 [↑](#footnote-ref-3)
4. 本发明针对支持ADR的计算机系统，原有ADR保证位于内存控制器写队列中的数据可以在断电后写回NVM，诸如处理器缓存、映射数据缓存、计数器缓存中的数据将在系统断电后丢失。本发明在原有ADR支持范围的基础上，添加了对日志头寄存器、日志内存地址寄存器、检查点编号寄存器、影子内存空间位图寄存器的支持，新添加的寄存器大小为256B，新增空间较小，ADR部件无需较大改动。 [↑](#footnote-ref-4)
5. Chenyu Yan, Daniel Englender, Milos Prvulovic, Brian Rogers, Yan Solihin. Improving Cost, Performance, and Security of Memory Encryption and Authentication. ACM SIGARCH Computer Architecture News, 2006, 34(2): 179-190 [↑](#footnote-ref-5)
6. 由于本发明中映射数据缓存的数据存储密度大于计算器缓存，实验表明当计数器缓存命中时映射数据缓存也基本能够命中，因此仅在这里讨论映射数据缓存和计数器缓存均缺失时，内存读写请求的处理流程。 [↑](#footnote-ref-6)