



## (12)发明专利

(10)授权公告号 CN 104756091 B

(45)授权公告日 2018.02.23

(21)申请号 201380051672.9

(72)发明人 S·卡皮尔 G·F·斯沃特

(22)申请日 2013.10.01

A·安格冉 W·H·布瑞治

(65)同一申请的已公布的文献号

S·加拉斯 J·G·约翰逊

申请公布号 CN 104756091 A

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038

(43)申请公布日 2015.07.01

代理人 罗亚男

(30)优先权数据

(51)Int.CI.

61/709,142 2012.10.02 US

G06F 12/14(2006.01)

13/839,525 2013.03.15 US

G06F 21/62(2006.01)

(85)PCT国际申请进入国家阶段日

(56)对比文件

2015.04.02

CN 1704922 A, 2005.12.07,

(86)PCT国际申请的申请数据

CN 102184365 A, 2011.09.14,

PCT/US2013/062859 2013.10.01

US 2006/0075236 A1, 2006.04.06,

(87)PCT国际申请的公布数据

US 2009/0289861 A1, 2009.11.26,

W02014/055512 EN 2014.04.10

审查员 陈国灿

(73)专利权人 甲骨文国际公司

权利要求书3页 说明书10页 附图7页

地址 美国加利福尼亚

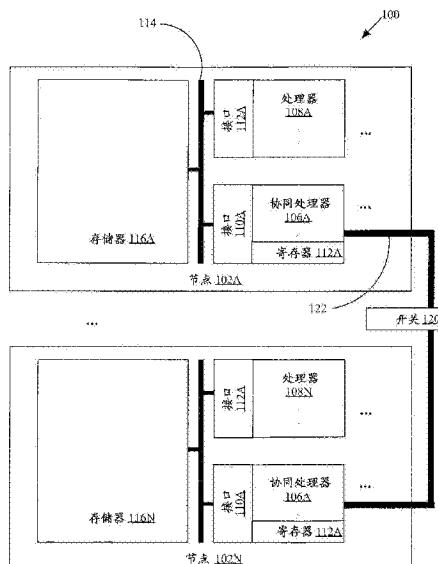
(54)发明名称

基于远程密钥的存储器缓冲区访问控制机

制

(57)摘要

本申请公开了实现可撤销的安全远程密钥的系统和方法。多个编索引的基础秘密存储在与本地存储器耦合的本地节点的协同处理器的寄存器中。当确定所选择的基础秘密过期时，基于基础秘密索引存储在寄存器中的基础秘密改变，从而使得基于过期的基础秘密生成的远程密钥无效。具有验证数据和基础秘密索引的远程密钥从请求访问该本地存储器的节点接收。验证基础秘密基于基础秘密索引从寄存器获得。协同处理器基于验证基础秘密对验证数据执行硬件验证。如果与基础秘密索引相关联的基础秘密已经在所选择的协同处理器的寄存器中改变了，则硬件验证失败。



1. 一种用于存储器访问的设备,包括:

本地节点,包括本地存储器和与本地存储器耦合的至少一个协同处理器,每个协同处理器包括寄存器,其中每个寄存器被配置为存储用于生成和验证远程密钥的多个基础秘密,其中,所述多个基础秘密中的每个基础秘密与一基础秘密索引相关联,其中,每个寄存器被配置为基于基础秘密索引来存储所述多个基础秘密;

至少一个主处理器,被配置为执行软件指令,所述软件指令使得所述至少一个主处理器基于与所选择的基础秘密相关联的所选择的基础秘密索引来改变寄存器中的所选择的基础秘密;

其中,从所述至少一个协同处理器中选择的所选择的协同处理器被配置为:

从第一节点接收访问所述本地存储器的请求,所述请求包括第一远程密钥,第一远程密钥包括第一基础秘密索引和基于第一基础秘密生成的第一验证数据;

基于第一基础秘密索引获取存储在所选择的协同处理器的寄存器中的验证基础秘密;

基于验证基础秘密对第一远程密钥中的第一验证数据执行硬件验证而不使用任何软件指令或编程API辅助来执行所述硬件验证,其中,当与第一基础秘密索引相关联的基础秘密已经在所选择的协同处理器的寄存器中改变了时,所述硬件验证失败;

在对第一远程密钥的成功认证之后,根据所述请求访问所述本地存储器。

2. 如权利要求1所述的设备,其中,第一验证数据包括使用第一基础秘密作为密钥生成的散列字段。

3. 如权利要求1所述的设备,其中,所选择的协同处理器还被配置为:

生成第二远程密钥,第二远程密钥包括第二基础秘密索引和基于第二基础秘密生成的第二验证数据,第二基础秘密基于第二基础秘密索引存储在所选择的协同处理器的寄存器中;

向第二节点发送第二远程密钥以向第二节点准予对所述本地存储器的访问,

其中,只要第二基础秘密仍然没变,则第二节点被授权访问与第二远程密钥相关联的本地存储器的一部分。

4. 如权利要求1所述的设备,其中,所选择的协同处理器还被配置为:

发送访问与第三节点相关联的远程存储器的请求,所述请求包括从第三节点接收的第三远程密钥。

5. 如权利要求1所述的设备,其中,第一远程密钥是从第一节点的软件管理程序接收的。

6. 如权利要求1所述的设备,其中,所述多个基础秘密中的至少一个基础秘密是切块式的基础秘密,其中,每个切块式的基础秘密被分割成多个切块,每个切块能够独立于所述切块式的基础秘密中的其他切块被无效。

7. 如权利要求6所述的设备:

其中,所选择的协同处理器的寄存器还被配置为存储切块验证数据;

其中,第一基础秘密是切块式的基础秘密;

其中,第一远程密钥还包括识别与第一基础秘密相关联的切块的第一切块索引;

其中,第一远程密钥的所述硬件验证还基于第一切块索引和所述切块验证数据;

其中,所述软件指令还使得所述至少一个主处理器修改与所选择的切块式的基础秘密

的所选择的切块对应的切块验证数据以指示与所选择的切块式的基础秘密的所选择的切块相关联的远程密钥无效。

8. 如权利要求7所述的设备,其中,所述切块验证数据包括与每个切块式的基础秘密相关联的位数组,其中,每个位数组的每个位值对应于针对相关联的切块生成的远程密钥的有效性。

9. 如权利要求7所述的设备,其中与给定切块索引和给定的切块式的基础秘密相关联的切块验证数据在不改变所述给定的切块式的基础秘密的情况下不能从无效变为有效。

10. 如权利要求6所述的设备:

其中,所述至少一个协同处理器的每个寄存器存储相同的切块验证数据;

其中,所述软件指令还使得所述至少一个主处理器在修改所述切块验证数据之后更新存储在寄存器上的所述切块验证数据。

11. 如权利要求1所述的设备:

其中,第一远程密钥还包括高速缓存指示符,其指示来自第一节点的请求中包含的命令应当被高速缓存;

其中,所选择的协同处理器还被配置为基于所述高速缓存指示符来有选择地高速缓存来自第一节点的命令和相关联的数据。

12. 如权利要求11所述的设备,其中,有选择地高速缓存命令还基于覆写第一远程密钥中的高速缓存指示符的本地决策。

13. 如权利要求1所述的设备,其中,所述请求包括命令,所述命令是从包括以下的组中选择的:复制命令、复制和信号命令、填充命令、存储命令、比较和交换命令、原子添加命令、原子OR命令以及中断和同步命令。

14. 如权利要求1所述的设备,

其中,所述本地节点通过与所述至少一个协同处理器可通信地耦合的互连而耦合到第一节点;

其中所述协同处理器被配置为接收通信,获取验证基础秘密,并且执行硬件验证而不通过本地节点的公共总线与本地节点的另一处理器通信。

15. 一种用于存储器访问的系统,包括:

第一节点,其包括第一本地存储器和与第一本地存储器耦合的第一多个协同处理器,第一多个协同处理器中的每个协同处理器包括寄存器,其中第一多个协同处理器的每个寄存器被配置为存储用于生成和验证远程密钥的第一多个基础秘密,第一多个基础秘密与第一节点相关联,其中,第一多个基础秘密中的每个基础秘密用第一多个基础秘密索引编索引,其中,第一多个协同处理器的每个寄存器被配置为基于第一多个基础秘密索引来存储第一多个基础秘密;

第二节点,其包括第二本地存储器和与第二本地存储器耦合的第二多个协同处理器,第二多个协同处理器的每个协同处理器包括寄存器;

其中,第一节点还包括被配置为执行软件指令的至少一个主处理器,所述软件指令使得所述至少一个主处理器基于与过期的基础秘密相关联的所选择的基础秘密索引来改变第一多个协同处理器的寄存器中的所述过期的基础秘密;

其中,从第一多个协同处理器中选择的所选择的第一节点协同处理器被配置为:

生成第一远程密钥，第一远程密钥包括第一基础秘密索引和验证数据，其中，所述验证数据是基于第一基础秘密生成的，第一基础秘密基于第一基础秘密索引存储在所选择的协同处理器的寄存器中；

向第二节点发送第一远程密钥以向第二节点准予对第一本地存储器的访问；

从第二节点接收请求，所述请求包括第一远程密钥和要求访问第一本地存储器的命令；

基于第一基础秘密索引获取存储在所选择的第一节点协同处理器的寄存器中的验证基础秘密；

基于验证基础秘密执行对第一远程密钥中的第一验证数据的硬件验证而不使用任何软件指令或编程API辅助来执行所述硬件验证，其中，当与第一基础秘密索引相关联的第一基础秘密已经在所选择的第一节点协同处理器的寄存器中改变了时，所述硬件验证失败；

在第一远程密钥的成功认证之后执行所述命令。

16. 如权利要求15所述的系统，其中，所述验证数据包括使用第一基础秘密作为密钥生成的散列字段。

17. 如权利要求15所述的系统：

其中，第一多个基础秘密中的至少一个基础秘密是切块式的基础秘密；

其中，每个切块式的基础秘密被分割成多个切块，每个切块能够独立于所述切块式的基础秘密中的其它切块被无效；

其中，所选择的第一节点协同处理器的寄存器还被配置为存储切块验证数据；

其中，第一基础秘密是切块式的基础秘密；

其中，第一远程密钥还包括识别与第一基础秘密相关联的切块的第一切块索引；

其中，第一远程密钥的所述硬件验证还基于第一切块索引和所述切块验证数据；

其中，所述软件指令还使得所述至少一个主处理器修改与所选择的切块式的基础秘密的所选择的切块对应的切块验证数据以指示与所选择的切块式的基础秘密的所选择的切块相关联的远程密钥无效。

18. 如权利要求17所述的系统，其中，所述切块验证数据包括与每个切块式的基础秘密相关联的位数组，其中，每个位数组的每个位值对应于针对相关联的切块生成的远程密钥的有效性。

19. 如权利要求17所述的系统，其中与给定切块索引和给定的切块式的基础秘密对应的切块验证数据在不改变所述给定的切块式的基础秘密的情况下不能从无效变为有效。

20. 如权利要求15所述的系统：

其中，第一远程密钥还包括高速缓存指示符，其指示来自第二节点的命令应当被高速缓存；

其中，所选择的第一节点协同处理器还被配置为基于所述高速缓存指示符来有选择地高速缓存来自第二节点的命令和相关联的数据。

## 基于远程密钥的存储器缓冲区访问控制机制

[0001] 相关申请的交叉引用;权益要求

[0002] 根据35U.S.C. §119 (e), 本申请要求于2012年10月2日提交的临时申请61/709,142的权益,其全部内容通过引用被结合于此,如同在此作了全面阐述一样。于2013年2月27日提交的苹果公司申请No.13/778,307以及于2013年5月15日提交的发明人为Sanjiv Kapil等人、发明名称为“MEMORY BUS PROTOCOL TO ENABLE CLUSTERING BETWEEN NODES OF DISTINCT PHYSICAL DOMAIN ADDRESS SPACES”的苹果公司申请No.13/838,542(代理卷号No.50277-4032)的全部内容通过引用被结合于,如同在此作了全面阐述一样。

### 技术领域

[0003] 本发明一般地涉及硬件计算设备。更具体地,本发明涉及基于远程密钥的存储器缓冲区访问控制机制。

### 背景技术

[0004] 个体处理器速度伴随着新技术而持续增大。通过利用多个处理器的节点的集群还可获得更高的性能。例如,数据库系统通常将数据库的多个部分分布在集群中的若干节点上,以便提高性能和提供可升缩性。多个节点的使用要求用于在节点之间共享数据的方法。集群可以配置为相干(coherent)存储集群或计算集群。

[0005] 相干存储集群上的节点共享物理存储器。共享物理存储器允许集群上的每个节点非常快速地通信。为了在共享的存储集群上的两个节点之间发送和接收消息,一个节点将会将数据写到该共享的存储器而另一节点将从该共享的存储器读取数据。然而,相干存储集群昂贵且共享存储器的大小有限。

[0006] 计算集群上的节点不共享物理存储器。计算集群上的节点之间的通信可以通过消息传递来执行。并且,计算节点可能需要重新组装进入的消息并将重新组装的消息存储在节点的主存储器中。通常,计算集群上的节点通过公共的总线进行通信,诸如以访问另一节点本地的存储器。共享总线结构的一个缺点是公共的总线在节点间通信排队并竞争对公共总线的使用时变成限制性能的元件。一旦公共总线饱和或接近饱和,则通过增加额外的节点仅实现非常小的性能提升。

[0007] 克服共享总线结构的缺点的一种技术涉及节点对之间的专用高速点对点通信链路。然而,需要复杂的分层通信协议来保证准确的鲁棒的通信。在通信路径上的每个节点处,接口处理器必须执行该复杂的协议以及翻译和验证源地址和目的地地址。执行这些通信任务降低性能,因为接口处理器一般比主CPU要慢得多,还因为该接口与相应节点的存储器之间的耦合较差。因此,使用共享总线结构,性能还是受限。

[0008] 本节描述的方法是能够实行的方法,但不一定是之前已经构想到或实行的方法。因此,除非另外指明,不应当认定,在本节中描述的任何方法仅因为其被包括在本节中就构成现有技术。

## 附图说明

- [0009] 图1是图示出与本文描述的基于远程密钥的存储器缓冲区访问控制机制的实施例兼容的系统的节点的框图；
- [0010] 图2是根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的远程密钥的实施例的框图；
- [0011] 图3是根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的注册表数据的实施例的框图；
- [0012] 图4图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的节点之间的地址空间的模型；
- [0013] 图5是图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制验证远程密钥的方法的实施例的流程图；
- [0014] 图6是图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制使用可切块的基础秘密来验证远程密钥的方法的实施例的流程图；
- [0015] 图7图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的节点之间的命令操作；

## 具体实施方式

[0016] 在以下描述中,为了说明的目的,阐述了大量具体细节以提供对本发明的透彻理解。但是,显而易见的是,本发明可以在没有这些具体细节的情况下实行。在其它实例中,公知的结构和设备以框图形式示出以便避免对本发明不必要的混淆。

[0017] 概览

[0018] 公开了实现可撤销的安全远程密钥的系统和方法。多个编索引的基础秘密存储在与本地存储器耦合的本地节点的协同处理器的寄存器中。这里使用的“基础秘密(base secret)”是指可用作密钥的任何数据。当确定所选择的基础秘密应当被无效时,基础秘密在寄存器中被改变,从而使得基于过期的基础秘密生成的远程密钥无效。

[0019] 当远程节点请求访问本地节点的本地存储器时,其包括远程密钥,远程密钥包括验证数据和基础秘密索引。该本地节点的协同处理器使用基础秘密索引从寄存器获得验证基础秘密。协同处理器基于验证基础秘密对远程密钥中的验证数据执行硬件验证,例如,不使用任何软件指令或编程API辅助来执行该验证。如果与基础秘密索引相关联的基础秘密已经在所选择的协同处理器的寄存器中改变了,则硬件验证失败。在密钥过期之后,针对与远程密钥相关联的物理存储器位置准予的权限被撤销,并且访问需要新密钥。

[0020] 可撤销密钥允许具有事务级安全性的可撤销存储器访问能力。物理存储器地址空间不被暴露在本地物理域外面。这种类型的访问对于节点之间的消息传送和数据共享是有用的。协同处理器被配置为执行远程密钥的硬件验证和数据有关的命令的硬件执行,提高与节点之间的进程间通信和存储器数据访问有关的消息吞吐量。

[0021] 图1是图示出与本文描述的基于远程密钥的存储器缓冲区访问控制机制的实施例兼容的系统的节点的框图。分布式共享存储系统100包括多个节点102A-102N。节点102A-102N驻留在两个或多个物理域中。在一个实施例中,每个物理域对应于节点102A-102N中的

一个。节点可以具有一个或多个处理套接口，每个处理套接口包括至少一个协同处理器106A-106N。在一个实施例中，至少一个节点102A-102N可以具有另外的处理套接口。节点102A-102N每个具有每个节点102A-102N本地的存储器116A-116N。

[0022] 这里使用的术语“存储器”可以指代与永久地址空间、非永久地址空间或其任意组合相关联的任何计算机存储介质，包括但不限于易失性存储器、非易失性存储器、软盘、磁存储介质、光存储介质、RAM、PROM、EPROM、FLASH-EPROM、任何其它存储芯片或盒，或计算机可以读取的任何其它介质。当本地存储器116A-116N是指永久地址空间时，节点102A-102N是存储节点。当本地存储器116A-116N是指非永久地址空间时，节点102A-102N是计算节点。

[0023] 每个节点102A-102N还包括至少一个主处理器108A-108N以及至少一个协同处理器106A-106N。每个节点102A-102N的主处理器108A-108N和协同处理器106A-106N被配置为访问物理域本地的本地存储器116A-116N。例如，每个处理器108A-108N可以包括到相应物理存储器116A-116N的存储器接口112A-112N，并且每个协同处理器106A-106N可以包括到相应物理存储器116A-116N的存储器接口110A-110N。存储器接口110A-112N可以经由总线114来访问相应本地存储器116A-116N。

[0024] 协同处理器106A-106N包括数字电路，数字电路被硬连线来执行一组功能或被永久编程来执行该组功能。所述功能是独立于被配置为通过运行软件指令集或程序来执行功能的通用处理器诸如主处理器108A-108N而执行的。这里使用的术语“协同处理器”是指区别的处理实体，而不一定是与CPU或其它处理器分离的区别的物理设备。例如，协同处理器可以是CPU的一个核(core)。在一个实施例中，当协同处理器是CPU的核时，节点102A-102N处理数据存储和/或维护命令的能力随着节点102A-102N具有的CPU的数目而自动缩放。

[0025] 协同处理器106A-106N可以但绝对不限于发送命令、接收命令、认证命令、条目入队、同步消息、重新组装进入的消息、以及报告错误，而没有软件干预。在一个实施例中，分布式共享存储系统100的协同处理器106A-106N被配置为接受命令块中指定的命令和地址。在命令块中，或者远程密钥或者物理地址位置可以作为地址而提供。当地址是指本地物理域外面的地址时，可以使用远程密钥。

[0026] 协同处理器106A-106N被配置为移动数据，在客户端(例如，进程、内核和管理程序)之间发送消息，并且可以被配置为执行一个或多个其它操作，而不使用任何软件指令或编程API协助。在一个实施例中，协同处理器106A-106N被配置为执行一组支持数据移动和维护命令，而不需主处理器108A-108N的支持。协同处理器106A-106N还可以硬件配置为验证远程密钥，诸如在请求中接收的远程密钥，以执行相关联的存储器116A-116N中的一个或多个命令。

[0027] 协同处理器106A-106N可以与相应的寄存器112A-112N耦合。寄存器112A-112N可以存储用于生成和验证远程密钥的基础秘密数据。远程密钥是物理存储器的块的所有者授予远程用户的凭证。在一个实施例中，远程密钥对于远程节点访问在其所属于的物理域外面的远程存储器116A-116N是必要的。例如，可以要求远程密钥来访问在所选节点外面的存储器116A-116N。远程密钥包括使用所选择的基础秘密生成的验证数据和对注册表中的基础秘密的索引。远程密钥通过使用该索引来从注册表中获取基础秘密而被认证。只要注册表中的基础秘密还没有改变，使用基础秘密发布的远程密钥就可以被协同处理器验证有效。

[0028] 在一个实施例中，协同处理器106A-106N的接口110A-110N还被配置为与和相应节点102A-102N相关联的软件管理程序交互。在一个实施例中，管理程序是在操作系统和/或其它软件代码与协同处理器106A-106N之间提供API接口的特殊的多线程驱动。管理程序通过管理程序接口向协同处理器106A-106N发布命令。

[0029] 管理程序可以被配置为配置协同处理器106A-106N，诸如同步存储在协同处理器106A-106N的寄存器112A-112N中的该组基础秘密。在一个实施例中，管理程序被配置为确定何时基础秘密已经过期并改变在本地物理域中的协同处理器的所有寄存器中的与具体基础秘密索引相关联的过期基础秘密。例如，管理程序可以被配置为在一个节点的协同处理器的所有寄存器中改变与具体基础秘密索引相关联的过期基础秘密。在基础秘密被改变之后，基于该基础秘密生成的远程密钥被无效。

[0030] 协同处理器106A-106N可以使用互连122和外部开关120可通信地耦合以与本地和非本地存储器连接，耦合到套接口、本地/非本地最后一级高速缓存以及到远程物理域，包括远程节点。在一个实施例中，协同处理器106A-106N包括包含消息传送基础设施的硬件并且不需要除了外部开关120之外的外部辅助来方便物理域之间的消息路由。所有协同处理器存储器操作在相同的本地物理域内都是高速缓存相干的。远程密钥只需要被包含与该远程密钥相关联的物理存储器的块的目的地节点认证。与远程密钥相关联的请求不需要被任何中间节点翻译或验证。这允许远程密钥管理仍然是每个节点本地的，消除了跨远程域同步密钥有效性信息的需要和开销。此本地远程密钥管理系统允许更好的伸缩性，诸如增大系统中节点的数目。

### [0031] 远程密钥结构

[0032] 图2是根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的远程密钥的实施例的框图。远程密钥是物理存储器的块的所有者授予与远程节点相关联的远程客户端的凭证。在访问本地物理地址的请求得到服务之前，远程客户端呈交在本地节点中验证有效的远程密钥。例如，远程密钥可以与数据存储或维护命令一起被发送到与该物理存储器的该块相关联的节点。

[0033] 远程密钥200包括验证数据202。验证数据202是本地节点的签名。验证数据202是使用本地节点已知的基础秘密生成的。获知用于生成远程密钥200的基础秘密对于验证远程密钥200而言是必要的。因此，仅本地节点的协同处理器能够验证远程密钥200。

[0034] 在一个实施例中，一组基础秘密被存储在驻留在包含与远程密钥200相关联的物理存储器的块的本地节点上的一个或多个本地协同处理器的注册表中。远程密钥还包括基础秘密索引214，其识别与用于生成验证数据的基础秘密相关联的基础秘密位置。例如，基础秘密索引214可以识别存储在本地协同处理器的注册表中的基础秘密数组的数组索引。

[0035] 在一个实施例中，远程密钥200包括切块索引204。切块索引204识别被切块的基础秘密中的一个切块。切块可以被单独用来无效与该切块相关联的远程密钥，而不必无效与可切块的远程密钥相关联的全部远程密钥。切块的基础秘密和未切块的基础秘密二者可以以相同的实现方式实现。切块验证数据可以硬件地存储，例如存储在与本地节点相关联的协同处理器寄存器中。

[0036] 切块的基础秘密可以用于降低所需的远程密钥无效的频率。通过切割基础秘密的至少一部分，可以使用更少的基础秘密，减小一个或多个远程密钥字段例如切块索引204的

大小。在一个实施例中,可以使用切块的基础秘密和未切块的基础秘密两者。用于生成具体远程密钥的基础秘密的类型可以被选择以最小化无效的影响。当远程密钥因基础秘密改变而被无效时,基于该基础秘密生成的每个远程密钥也被无效。单个切块的基础秘密被分割成多个切块,每个切块可以独立于与该基础秘密相关联的其它切块而被无效。

[0037] 在一个实施例中,可以使用 $2^m$ 个切块的基础秘密,并且每个切块的基础秘密被切割成 $2^n$ 个切块。切块索引可以包含识别基础秘密和该切块两者的信息。例如,切块索引可以是 $m+n$ 位,其中 $m$ 位用于表示基础秘密索引并且 $n$ 位用于表示该切块。尽管在该示例中,每个切块的基础秘密被切割成相同数目的切块,但切块的基础秘密可以被切割成不同数目的切块。

[0038] 验证数据可以包括使用散列和/或加密算法生成的散列字段,其中所选择的基础秘密被用作密钥。该算法可以应用于包含切块索引204、大小206、高速缓存指示符208、地址210、套接口ID 212和/或任何其他数据的数据。当与基础秘密索引214相关联的基础秘密被用于生成远程密钥200时,只要存储在本地节点的协同处理器的寄存器中的与基础秘密索引214相关联地存储的基础秘密保持不变,则远程密钥200有效。

[0039] 在以下非限制性示例中,通过对包括切块索引204(如果密钥是可切块的)、套接口ID 212、地址210、大小206和高速缓存指示符208的位数组应用数据加密标准(DES)算法获得验证数据202。与基础秘密索引214相关联的本地存储的基础秘密被用作密钥。

[0040] 在该非限制性示例中,验证数据包括使用密钥Basesecret[SecretNum]生成的散列签名。当协同处理器验证包含该散列签名的远程密钥时,协同处理器将基于与基础秘密索引SecretNum相关联地存储的本地存储的基础秘密来解密该散列签名。解密的信息将与远程密钥的其它信息比较,其它信息诸如是切块索引204、套接口ID 212、地址210、大小205和高速缓存指示符208。如果与基础秘密索引SecretNum相关联地存储的基础秘密已经在本地节点上改变,则该验证失败。

[0041] 远程密钥200还包括套接口ID 212。套接口ID 212识别包含与远程密钥200相关联的物理存储器的块的节点。远程密钥200还包括地址210。地址210识别与远程密钥200的物理存储器的块的物理地址。远程密钥200还包括大小206。大小206指示与远程密钥200相关联的物理存储器的块的大小。在一个实施例中,远程密钥200的固定位字段专用于对大小206进行编码,其中总范围和粒度取决于该位字段的大小。例如,大小206可以是在从大约1KB到大约1TB的范围内。

[0042] 在一个实施例中,远程密钥200包括高速缓存指示符208。高速缓存指示符208指示与远程密钥200一起发送的命令是否应当高速缓存在硬件中,诸如存储在与套接口ID 212相关联的目的地节点的任何高速缓存中。命令可以是针对与该命令相关联的数据,包括从远程节点接收的数据,执行的数据存储或维护命令。相关联的数据也可以被高速缓存。高速缓存可以是更高级别的高速缓存和/或最后一级的高速缓存,诸如L3高速缓存,但是与目的地节点相关联的任何高速缓存都可以使用。作为非限制性示例,命令可以涉及将相关联的数据的至少一部分写入本地节点的物理存储器中的写命令,并且高速缓存指示符208允许写数据在最后一级高速缓存中被修改。作为另一非限制示例,命令可以涉及从本地物理存储器读取缓冲,并且高速缓存指示符208允许本地节点响应于来自远程节点的命令从该高速缓存读取脏的和/或干净的行。处理远程密钥200和相关联的命令的协同处理器可以被配

置为基于高速缓存指示符来有选择地高速缓存命令。协同处理器可以做出本地决策来覆写远程密钥200中的高速缓存指示符208。

[0043] 远程密钥200可以被配置为具有一组，其中远程密钥200的每个字段202-214都在远程密钥200内具有已知的位置。分布式共享存储系统的协同处理器可以被配置为基于远程密钥200和其相关联的字段的已知配置来接受和读取在命令块中指定的命令和地址。

#### [0044] 寄存器数据

[0045] 图3是根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的寄存器数据的实施例的框图。寄存器300可以存储编码的数据以用于硬件地生成和验证远程密钥，例如，不必使用任何软件指令或编程API辅助来生成和/或验证远程密钥。

[0046] 在一个实施例中，寄存器300包括一组编索引的基础秘密302。基础秘密304-308用于生成发布给远程节点的远程密钥。远程密钥包括使用所选择的基础秘密生成的验证数据以及到注册表中的基础秘密的索引。当远程密钥被用于访问相关联的物理存储器地址时，存储器本地的协同处理器通过使用该索引从注册表300获取基础秘密304-308来认证远程密钥。只要该基础秘密在注册表300中还没有改变，则使用基础秘密发布的远程密钥可以被协同处理器验证有效。在一个实施例中，编索引的基础秘密302被存储在寄存器300的固定数组中，其中*i*=0。

[0047] 在一个实施例中，节点的协同处理器的所有寄存器300都包含相同的编索引的基础秘密302。在一个或多个本地主处理器上执行的软件指令可以使得这一个或多个本地主处理器确定所选择的基础秘密过期了并基于与所选择的基础秘密相关联的所选择的基础秘密索引来改变一个或多个本地寄存器300中的所选择的基础秘密。改变所选择的基础秘密使得在所选择的基础秘密在寄存器300中改变之前基于所选择的基础秘密生成的远程密钥无效。在一个实施例中，所述软件指令是软件管理程序的一部分。

[0048] 寄存器300还可以存储切块验证数据310。切块验证数据310包括用于每个切块的基础秘密的位数组。位数组的长度可以等于与被使得与切块的基础秘密相关联的切块的数目。位数组的位值可以对应于针对相关联的切块生成的远程密钥的有效性。仅当由远程密钥指示的切块基于该切块验证数据有效时，准许对本地存储器的访问。

[0049] 在一个实施例中，切块验证数据310包括与每个切块的基础秘密j, …, k相关联的有效性位数组312-314。有效性位数组312-314这样开始，数组的所有字段都指示所有切块有效。有效切块随后被指派给远程密钥。当切块和对应的远程密钥被无效时，有效性位数组312-314中对应于切块的基础秘密中的切块索引的位被翻转来指示该切块是无效的。在一个实施例中，与给定切块索引和给定的切块的基础秘密相关联的切块验证数据310在相关联的切块的基础秘密不改变的情况下不能从无效变为有效。

[0050] 在一个实施例中，可以保持单个验证位数组来跟踪所有切块的基础秘密的切块有效性。例如，当存在 $2^m$ 个切块的基础秘密并且每个切块的基础秘密被切割成 $2^n$ 个切块时，切块索引可以是m+n位长，其中m位识别切块的基础秘密索引并且n位识别切块。以这种方式，可以保持单个验证位数组VALIDATION\_DATA[ $2^{(m+n)}$ ]来表示所有切块的基础秘密的所有切块。

[0051] 在一个实施例中，节点的协同处理器的所有寄存器300包含相同的切块验证数据310。在一个或多个本地主处理器上执行的软件指令可以使得这一个或多个本地主处理器

确定所选择的切块基础秘密中的所选择的切块过期了，并改变本地寄存器300中的切块验证数据310。改变切块验证数据310使得在切块验证数据310在寄存器300中改变之前基于所选择的切块生成的远程密钥无效。在一个实施例中，所述软件指令是软件管理程序的一部分。

[0052] 物理域

[0053] 基于远程密钥的存储器缓冲区访问控制机制可以在具有多个物理域的分布式共享存储系统中实现。节点可以在分离的物理域中操作，分离的物理域具有仅本地处理器和协同处理器被准许访问的不同地址空间。

[0054] 例如，图4图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的节点之间的地址空间的模型。包括至少一个物理域402-404。在一个实施例中，每个物理域对应于具有一个或多个处理套接口的节点。每个物理域402-404包括多个本地密钥404-416，它们是提供对物理域本地的特有物理地址空间的访问权限的密钥。本地密钥与物理域的物理地址空间的大的物理上连续的部分相关联并且被分配用于物理域本地的用户进程。在图4中，PDOM1402具有密钥LKey1406、LKey2408、LKey3410。PDOM2404具有密钥LKey4412、LKey4414、LKey6416。

[0055] 远程密钥与具有LKey的存储区域的窗口相关联。每个LKey可以包括一个或多个远程密钥和相关联的存储区域。远程密钥将来自物理域中的一个给定的本地密钥的远程访问权限给与远程物理域中的另一本地密钥。由远程密钥保护的存储器的一部分的远程用户呈交该远程密钥来访问存储器的那个部分。

[0056] 每个节点可以包括至少一个协同处理器。协同处理器在接收到远程密钥时验证密钥并且如果验证成功，则继续执行该命令。图4图示出PDOM1中的LKey1406包括RKey1420和RKey2422。LKey2408包括RKey3424并且LKey3410包括RKey4426。在PDOM2404中，LKey4412包括RKey4428和RKey6430，而LKey4414包括RKey4434并且LKey6416包括RKey8436。图4另外还图示出通过PDOM1402中的Lkey1406做出的请求440，用于访问与PDOM2404中的LKey4414中的RKey4434相关联的存储器。

[0057] 远程密钥验证

[0058] 远程密钥由其保护的存储器位置的所有者发布。远程客户端被授权访问与该远程密钥相关联的关联存储器位置，直到远程密钥被撤销为止。当远程密钥被用来访问存储器位置时，存储器位置的所有者在允许访问之前验证该远程密钥。

[0059] 图5是图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制来验证远程密钥的方法的实施例的流程图。尽管图5图示出根据一个实施例的示例性步骤，但是其它实施例可以省略、增加、重新排序和/或修改所示出的步骤中的任何步骤。图5的一个或多个步骤可以由包括与远程密钥相关联的本地存储器的本地节点的所选择的协同处理器执行。

[0060] 在步骤502中，远程密钥被接收。远程密钥可以是从请求访问本地存储器的远程节点接收的。在一个实施例中，远程密钥是在包括远程密钥和命令信息的命令块中接收的。远程密钥包括验证数据。验证数据可以是基于从一组基础秘密中选择的基础秘密生成的，该组基础秘密可以被编索引。远程密钥还包括识别包含（或之前包含了）用于生成验证数据的基础秘密的基础秘密位置的基础秘密索引。在步骤504中，远程密钥中所包含的基础秘密索

引被确定。

[0061] 在步骤506中,基础秘密索引被用来获取验证基础秘密,验证基础秘密将被用于验证远程密钥。在一个实施例中,该组基础秘密被存储在所选择的协同处理器的寄存器中,并且基础秘密索引识别该组基础秘密中的该验证基础秘密。

[0062] 在步骤508中,验证基础秘密被用来验证远程密钥中包含的验证数据。在一个实施例中,所选择的协同处理器执行对基础秘密的硬件验证。远程密钥可以包括散列字段,散列字段包含使用硬件存储的一组可配置的基础秘密中的一个基础秘密生成的散列签名,并且验证远程密钥涉及使用存储在远程密钥中指定的基础秘密索引处的验证基础秘密来验证散列签名。在一个实施例中,散列签名使用验证基础秘密被解密,并且输出被与远程密钥中包含的其它数据比较。如果与远程密钥中的基础秘密索引相关联的基础秘密已经在硬件中改变,则验证失败。在决策步骤510中,如果确定远程密钥有效,则处理继续到步骤512。否则,如果远程密钥无效,则处理继续到步骤516。

[0063] 在步骤512中,本地地址被确定。本地地址可以通过翻译远程密钥来获取本地物理地址而确定。处理继续到步骤514,在该步骤中,对本地存储器的访问被准予。在一个实施例中,准予访问涉及执行与远程密钥一起接收的命令。命令可以是从远程节点接收的数据存储或维护命令,并且还可以涉及连同远程密钥一起接收的数据。所选择的协同处理器可以硬件地执行命令,例如,不使用任何软件指令或编程API辅助来执行该命令。在一个实施例中,命令在本地地址被确定和/或命令被执行之前被高速缓存。例如,远程密钥可以包括高速缓存指示符,其指示来自远程节点的命令应当被高速缓存。

[0064] 在步骤516中,确认被发送。确认可以在从远程节点接收的一个或多个命令的成功执行之后被发送。确认还可以包括指示命令没有被成功执行的一个或多个错误通知。例如,如果在决策步骤510中确定密钥是无效的,则错误通知可以被发送。

[0065] 图6是图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制使用切块的基础秘密来验证远程密钥的方法的实施例的流程图。尽管图6图示出根据一个实施例的示例性步骤,但是其它实施例可以省略、增加、重排序和/或修改所示出的步骤中的任何步骤。图6的一个或多个步骤可以被包括与远程密钥相关联的存储器的节点的所选择的协同处理器执行。

[0066] 在一个或多个实施例中,该组基础秘密包括至少一个切块的基础秘密。每个切块的基础秘密可以与一组切块相关联。切块可以被单独用于验证与该切块相关联的远程密钥,而不必验证与可切块的远程密钥相关联的所有远程密钥。切块的远程密钥和未切块的远程密钥两者可以以相同的实现方式使用。

[0067] 在步骤602中,远程密钥被接收。远程密钥可以是从请求访问本地存储器的远程节点接收的。在一个实施例中,远程密钥是在包括远程密钥和命令信息的命令块中接收的。远程密钥包括验证数据。验证数据可以基于从一组基础秘密选择的基础秘密而生成,该组基础秘密可以被编索引。远程密钥还包括识别基础秘密位置的基础秘密索引,基础秘密位置包含(或之前包含)用于生成该验证数据的基础秘密。在步骤604中,远程密钥中所包含的基础秘密索引被确定。在步骤606中,远程密钥中所包含的切块索引被确定。

[0068] 在步骤608中,基础秘密索引被用来获取将被用于验证该远程密钥的验证基础秘密。在一个实施例中,该组基础秘密被存储在所选择的协同处理器的寄存器中,并且基础秘

密索引识别该组基础秘密中的验证基础秘密。

[0069] 在步骤610中,验证基础秘密被用于验证远程密钥中包含的验证数据。在一个实施例中,所选择的协同处理器执行基础秘密的硬件验证。远程密钥可以包括散列字段,该散列字段包含使用硬件存储的一组可配置基础秘密中的一个基础秘密生成的散列签名。验证远程密钥可以涉及使用存储在远程密钥中指定的基础秘密索引处的验证基础秘密来验证散列签名。在一个实施例中,散列签名利用验证基础秘密被解密,并且输出被与远程密钥中包含的其它数据比较。在一个实施例中,散列签名包括包含切块索引的编码的信息。如果与远程密钥中的基础秘密索引相关联的基础秘密已经硬件地改变,则验证失败。在决策步骤612中,如果确定远程密钥有效,则处理继续进行到步骤614。否则,如果远程密钥无效,则处理继续进行到步骤622。

[0070] 在步骤614中,切块验证数据被访问来确定远程密钥中指示的切块是否有效。切块验证数据可以硬件地存储,诸如存储在与所选择的协同处理器相关联的注册表中。所选择的协同处理器可以硬件地确定切块是否有效,例如,不使用任何软件指令或编程API辅助来执行该验证。在一个实施例中,切块验证数据包括针对每个切块的基础秘密的位数组。位数组的长度可以等于被使得可用于相关联的基础秘密的切块的数目。位数组的位值可以对应于为相关联的切块生成的远程密钥的有效性。对本地存储器的访问仅在由远程密钥指示的切块基于切块验证数据有效时被准予。在决策步骤616中,如果确定切块有效,则处理继续到步骤618。否则,如果切块无效,则处理继续到步骤622。

[0071] 在步骤618中,本地地址被确定。本地地址可以通过翻译远程密钥来获取本地物理地址而确定。处理继续到步骤620,在该步骤,对本地存储器的访问被准予。在一个实施例中,准予访问涉及执行与远程密钥一起接收的命令。命令可以是从远程节点接收的数据存储或维护命令,并且还可以涉及连同远程密钥一起接收的数据。所选择的协同处理器可以硬件地执行命令,例如,不使用任何软件指令或编程API辅助来执行该命令。在一个实施例中,命令在本地地址被确定和/或命令被执行之前被高速缓存。例如,远程密钥可以包括高速缓存指示符,其指示来自远程节点的命令应当被高速缓存。

[0072] 在步骤622中,确认被发送。确认可以在从远程节点接收的一个或多个命令的成功执行之后被发送。确认还可以包括一个或多个错误通知,指示命令没有被成功执行。例如,如果在决策步骤612或616中确定密钥或切块无效,则错误通知可以被发送。

### [0073] 协同处理器命令执行

[0074] 协同处理器可以执行从管理程序接收的命令。在一个实施例中,协同处理器与发布命令的多线程的管理程序中的线程异步地执行命令。如果管理程序发送多个命令,则协同处理器可以高速缓存该命令。该协同处理器可以并行执行一些命令。

[0075] 协同处理器可以被设计为支持各种数据移动和维护命令而无需来自主处理器的支持。在一个实施例中,协同处理器支持数据移动命令和数据维护命令。数据移动命令可以从以下选择:Copy,Copy Immediate,CopyAndSignal,CopyAndSignal Immediate,Fill,Store,CAS和CASAndFetch,CAM/AtomicAdd/AtomicOr,和AtomicMessagePush。数据维护命令可以从以下选择:Interrupt,Sync和NoOP。

[0076] 在一个实施例中,命令可以涉及源地址和/或目的地地址。源地址(“SourceAddress”)或目的地地址(“DestAddress”)位于远程物理域,则远程密钥(“RKey”)

而不是物理地址被指定。

[0077] 图7图示出根据本文描述的基于远程密钥的存储器缓冲区访问控制机制的节点之间的命令操作。尽管图7图示出根据一个实施例的示例性步骤，但是其它实施例可以省略、增加、重排序和/或修改所示出的任何步骤。

[0078] 在步骤702中，第一物理域PDOM1中的源协同处理器接收和解码访问第二物理域PDOM 2中的远程存储器的新命令。在一个实施例中，源协同处理器从与PDOM1相关联的管理程序接收该命令。在步骤704中，源协同处理器获取物理地址处与命令相关联的本地数据。在步骤706中，源协同处理器将命令和与该命令相关联的数据连同远程密钥一起发送给第二物理域PDOM2。

[0079] PDOM2中的目的地协同处理器接收命令和与命令相关联的数据。在一个实施例中，目的地协同处理器从与PDOM2相关联的管理程序接收命令。在步骤708中，目的地协同处理器对远程密钥执行硬件验证。例如，目的地协同处理器可以执行图5-6中描述的方法的一个或多个步骤来验证远程密钥。在步骤710中，在远程密钥的成功验证之后，目的地协同处理器翻译远程密钥来获取本地物理地址。在步骤712中，目的地协同处理器执行该命令。如图7中所示，命令涉及将发送的数据写入与远程密钥相关联的本地物理地址。

[0080] 在步骤714中，PDOM2中的目的地协同处理器向PDOM1中的源协同处理器发送回确认，指示命令的完成。在一个实施例中，确认可以是指示命令没有被成功执行的错误通知。例如，错误通知可以指示验证不成功。在步骤716中，源协同处理器在接收到确认时，更新命令块中的完成状态。在步骤718中，源协同处理器将命令入队。

[0081] 在以上说明中，已经参考大量具体细节描述了本发明的实施例，这些细节根据实现方式不同可以不同。说明书和附图因此应该以说明性而非限制性的含义来考虑。本发明的范围的唯一的和排他性的指示并且申请人所意图作为本发明的范围的是从本申请产生的权利要求集合按照其发布的具体形式(包括任何后续补正)的字面和等同范围。

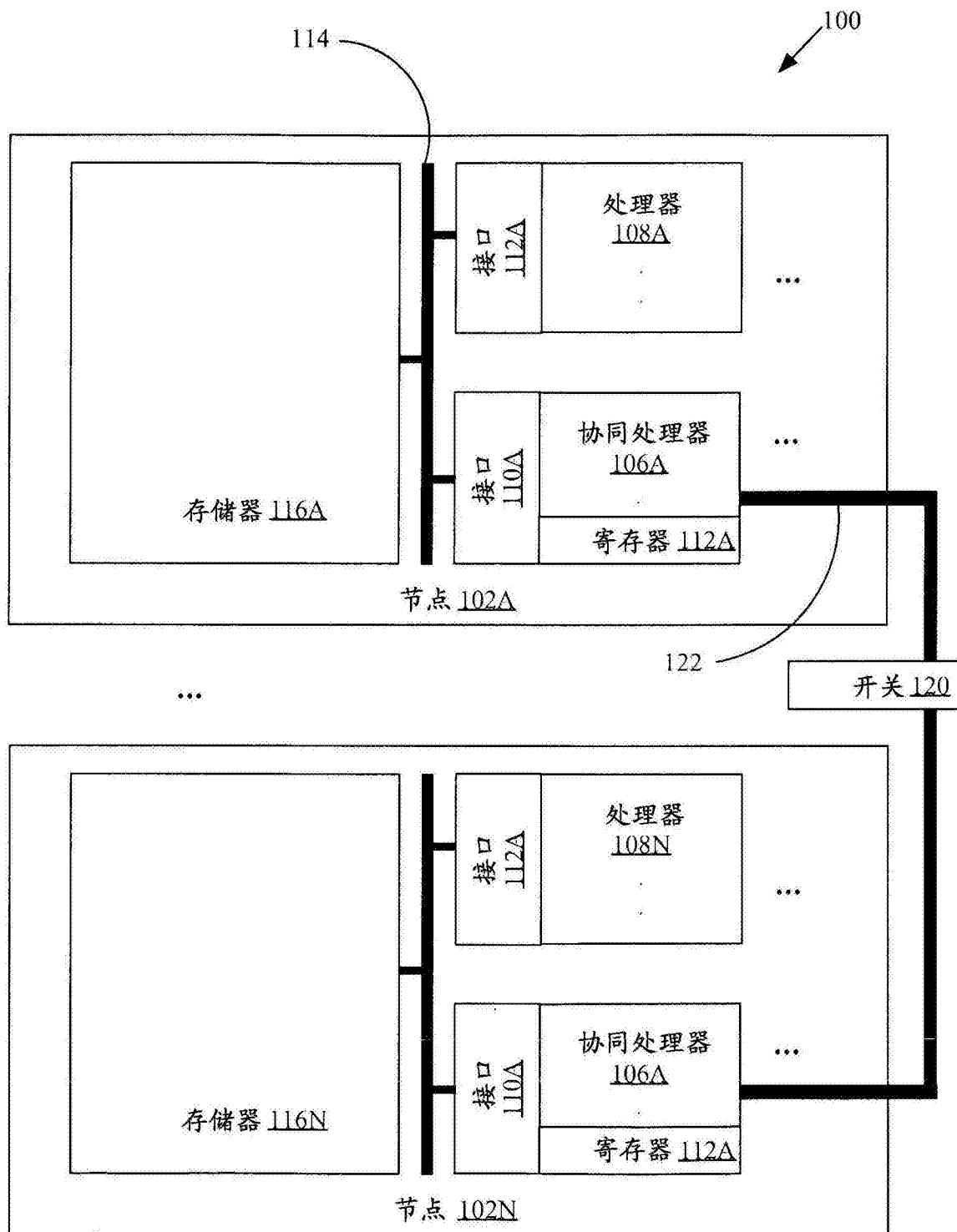


图1

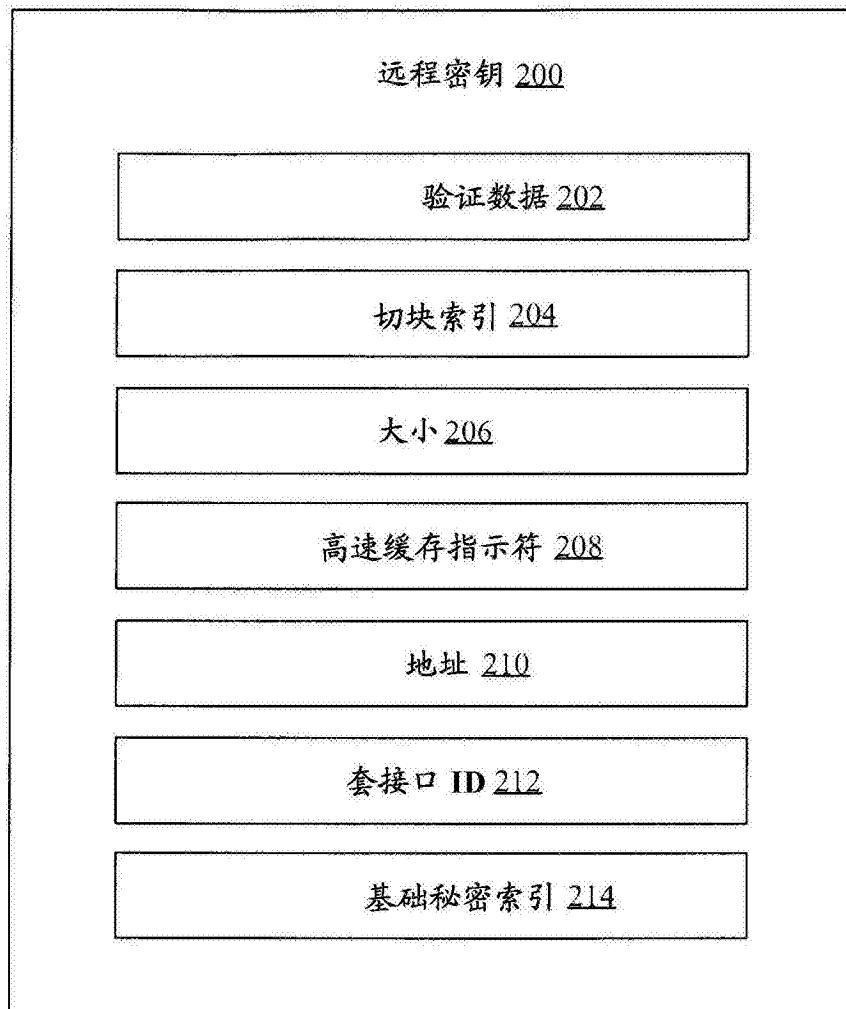


图2

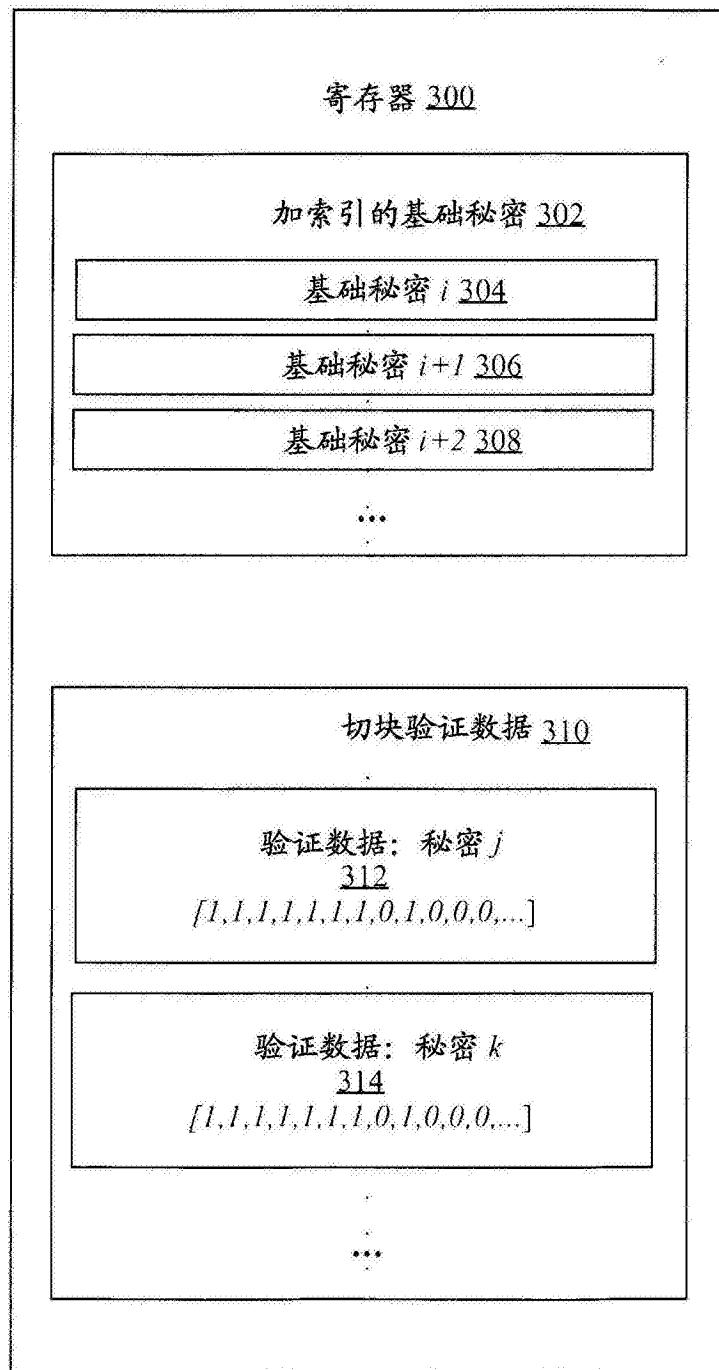


图3

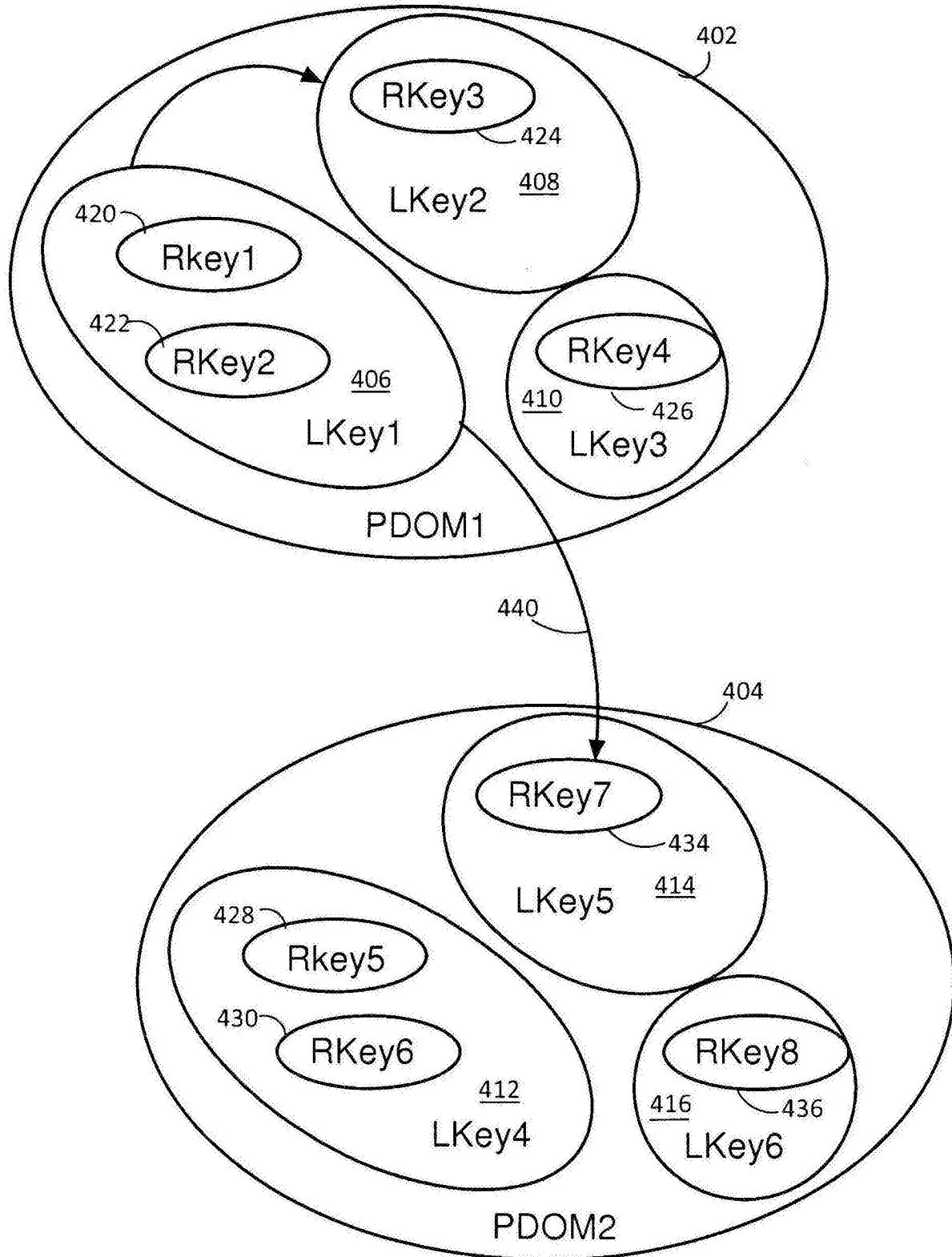


图4

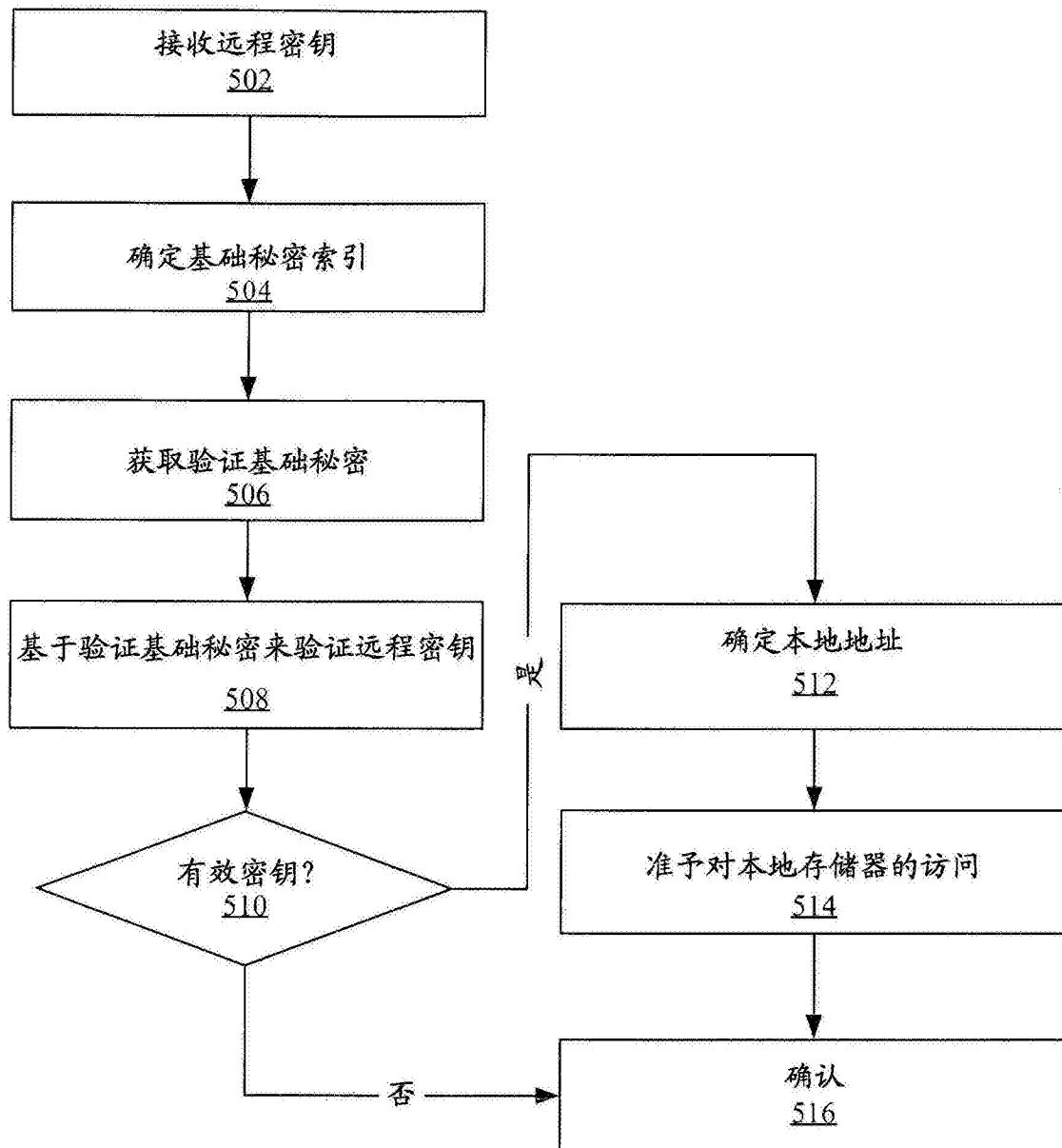


图5

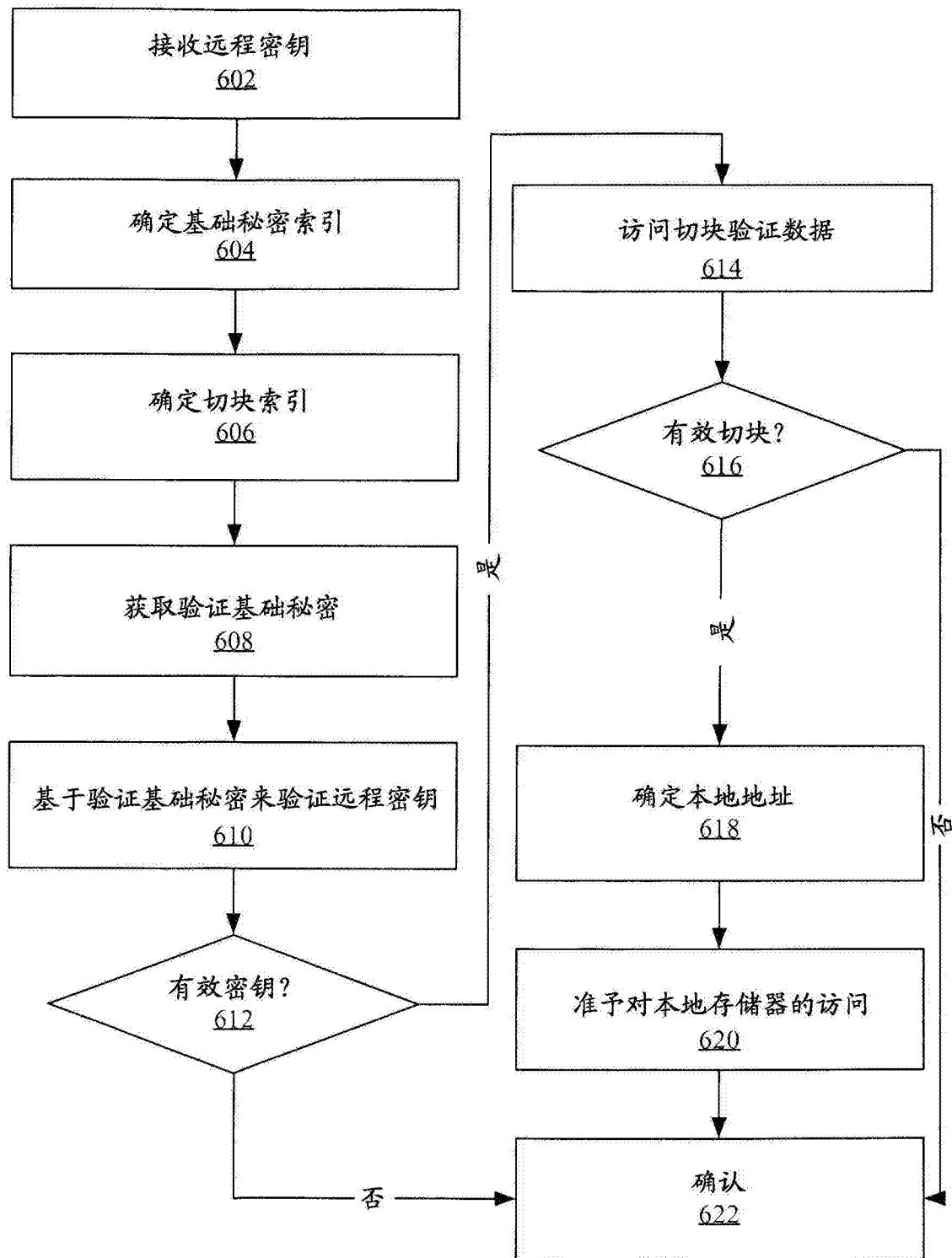


图6

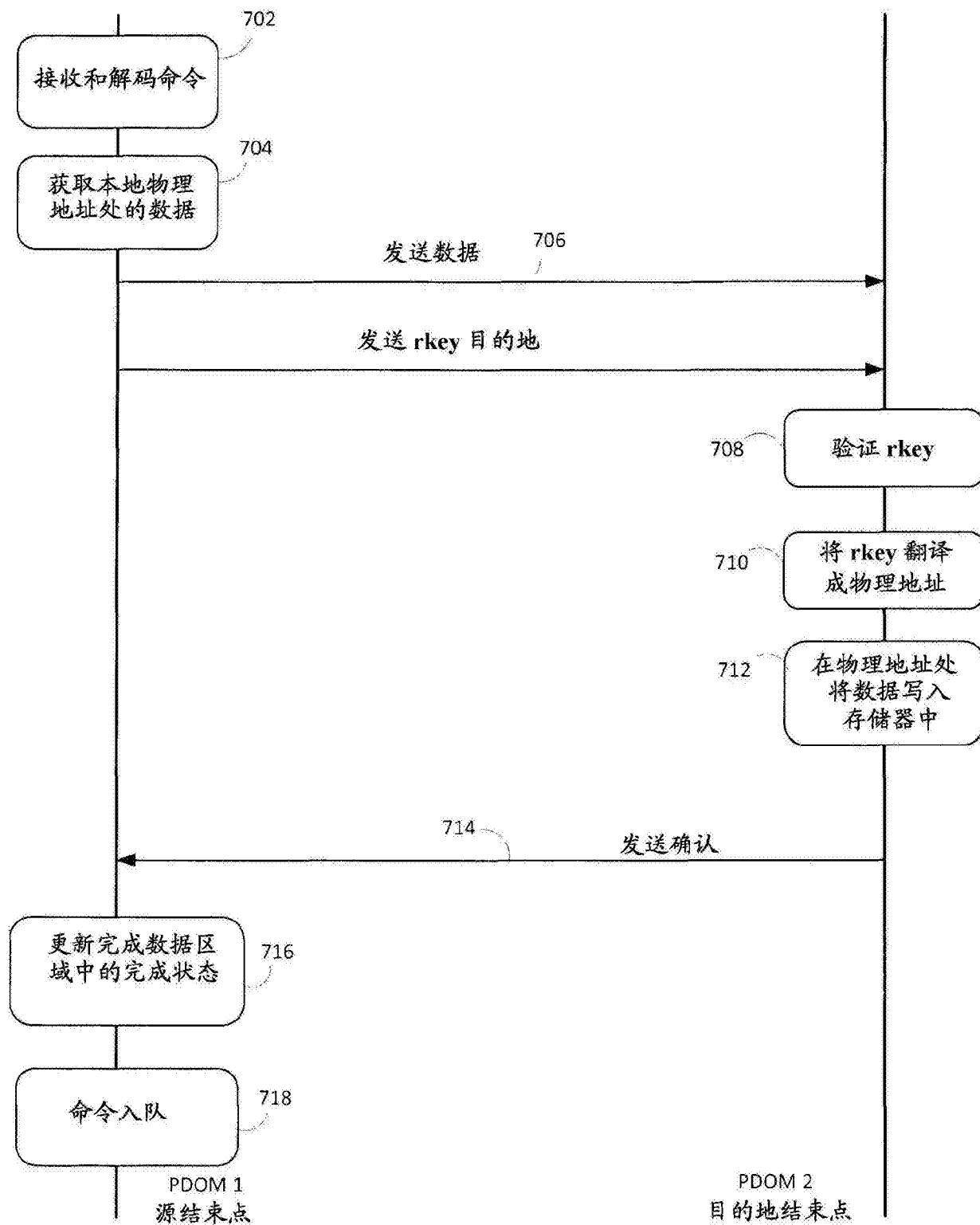


图7