(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 121211430 A
(43)申请公布日 2025.12.26

(21)申请号 202511294933.0

(22)申请日 2025.09.10

(71)申请人 浙江蚂蚁密算科技有限公司
地址 310023 浙江省杭州市西湖区西溪路543号-569号(单号连续)1幢2号楼5层508室
申请人 南方科技大学

(72)发明人 侯伟星 闫守孟 宋捷 唐丹青 杜少华 张晋 慈长旭 张锋巍 黄俊杰 王晨旭 徐延楷

(74)专利代理机构 北京亿腾知识产权代理事务所(普通合伙) 11309
专利代理师 陈霁 周良玉

(51)Int.Cl.
G06F 21/44(2013.01)
G06F 21/57(2013.01)
G06F 21/46(2013.01)

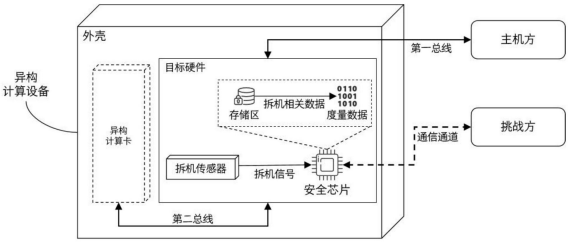
权利要求书2页 说明书12页 附图4页

(54)发明名称

一种对异构计算设备进行安全认证、密钥交换的方法及装置

(57)摘要

本说明书实施例提供了一种对异构计算设备进行安全认证的方法,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件中包含安全芯片,所述外壳关联部署有拆机传感器;所述方法由所述安全芯片执行,包括:响应于挑战方的指令,根据第一存储区中的目标数据生成度量数据,其中,所述目标数据包括,所述拆机传感器记录的拆机相关数据。利用第一私钥对所述度量数据进行签名,生成设备报告。向所述挑战方提供所述设备报告,供其进行验证。



CN 121211430 A

1.一种对异构计算设备进行安全认证的方法,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件中包含安全芯片,所述外壳关联部署有拆机传感器;所述方法由所述安全芯片执行,包括:

响应于挑战方的指令,根据第一存储区中的目标数据生成度量数据,其中,所述目标数据包括,所述拆机传感器记录的拆机相关数据;

利用第一私钥对所述度量数据进行签名,生成设备报告;

向所述挑战方提供所述设备报告,供其进行验证。

2.根据权利要求1所述的方法,其中,所述第一私钥通过以下方式获得:

响应于所述挑战方的第一请求,生成第一密钥对;

将所述第一密钥对的公钥发送至所述挑战方;将其中的私钥作为所述第一私钥。

3.根据权利要求1所述的方法,其中,向所述挑战方提供所述设备报告,供其进行验证,包括:

向所述挑战方提供所述设备报告,以使得所述挑战方基于第一证书中包含的所述第一私钥对应的第一公钥,对所述设备报告进行验签;所述第一证书为可信认证机构基于所述第一公钥签发。

4.根据权利要求1所述的方法,其中,所述拆机相关数据包括,拆机状态指示值,或拆机次数的计数值。

5.根据权利要求1所述的方法,其中,所述第一存储区为所述安全芯片的非易失性存储区NVindex。

6.根据权利要求1所述的方法,其中,所述目标硬件用于通过第一总线连接至主机方,从而在所述异构计算卡和主机方之间执行加解密操作;所述挑战方为所述主机方。

7.根据权利要求6所述的方法,其中,向所述挑战方提供所述设备报告,包括:

通过所述第一总线,将所述设备报告发送至所述主机方。

8.根据权利要求1所述的方法,其中,所述拆机传感器包括以下中的一种:机械式微动传感器、磁控传感器、光敏传感器、霍尔传感器、MEMS加速度传感器。

9.根据权利要求1所述的方法,其中,所述拆机相关数据通过以下方式生成:

获取所述拆机传感器的第一电平,所述第一电平用于指示所述外壳的拆开状态;

基于所述第一电平,生成所述拆机相关数据,存入所述第一存储区。

10.一种对异构计算设备进行密钥交换的方法,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件通过总线连接至主机方;所述目标硬件包括控制单元、加解密引擎和密钥缓冲区,控制单元上运行有目标程序;所述方法包括:

响应于收发数据请求,加解密引擎在所述密钥缓冲区查找有效密钥;

在所述密钥缓冲区存在有效密钥的情况下,加解密引擎使用所述密钥缓冲区中的密钥执行数据加密或解密操作;

在所述密钥缓冲区满足预设的密钥补充条件的情况下,目标程序与所述主机方进行密钥协商,获得若干密钥,存入所述密钥缓冲区。

11.根据权利要求10所述的方法,其中,所述密钥缓冲区的容量是根据所述目标硬件的可用存储空间,和/或第一阈值而确定。

12.根据权利要求10所述的方法,其中,所述有效密钥满足第一安全约束;所述第一安全约束用于指示,所述有效密钥的总使用次数不超过第一阈值。

13.根据权利要求10所述的方法,其中,所述有效密钥为第N次使用;所述有效密钥满足第二安全约束;所述第二安全约束用于指示,与所述有效密钥对应的初始向量IV与前序N-1次均不相同。

14.根据权利要求13所述的方法,其中,所述目标硬件还包括安全芯片;所述获得若干密钥,包括:

所述目标程序基于第一密钥交换协议,调用所述安全芯片,获取第一密钥,归入所述若干密钥。

15.根据权利要求10所述的方法,还包括:

在初始阶段,所述目标程序基于第一密钥交换协议与所述主机方进行密钥协商,获取m个密钥,存入所述密钥缓冲区;所述m是所述密钥缓冲区可存储密钥的最大数量。

16.一种对异构计算设备进行安全认证的装置,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件中包含安全芯片,所述外壳关联部署有拆机传感器;所述装置部署于所述安全芯片,包括:

生成模块,配置为,响应于挑战方的指令,根据第一存储区中的目标数据生成度量数据,其中,所述目标数据包括,所述拆机传感器记录的拆机相关数据;

签名模块,配置为,利用第一私钥对所述度量数据进行签名,生成设备报告;

提供模块,配置为,向所述挑战方提供所述设备报告,供其进行验证。

17.一种对异构计算设备进行密钥交换的装置,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件通过总线连接至主机方;所述目标硬件包括控制单元、加解密引擎和密钥缓冲区,控制单元上运行有目标程序;所述装置包括:

加解密引擎,配置为,响应于收发数据请求,加解密引擎在所述密钥缓冲区查找有效密钥;

所述加解密引擎还配置为,在所述密钥缓冲区存在有效密钥的情况下,加解密引擎使用所述密钥缓冲区中的密钥执行数据加密或解密操作;

目标程序,配置为,在所述密钥缓冲区满足预设的密钥补充条件的情况下,目标程序与所述主机方进行密钥协商,获得若干密钥,存入所述密钥缓冲区。

18.一种计算机程序产品,包括计算机程序/指令,该计算机程序/指令被处理器执行时实现权利要求1-15中任一项所述方法的步骤。

19.一种计算设备,包括存储器和处理器,其特征在于,所述存储器中存储有可执行代码,所述处理器执行所述可执行代码时,实现权利要求1-15中任一项所述的方法。

一种对异构计算设备进行安全认证、密钥交换的方法及装置

技术领域

[0001] 本说明书一个或多个实施例涉及密码学技术领域,尤其涉及一种对异构计算设备进行安全认证、密钥交换的方法及装置。

背景技术

[0002] 当前,随着电子商务、大模型推理及云计算技术的迅猛发展,个人金融数据、交易记录、身份信息等高价值数据呈指数级增长并高度集中地在各类服务器及云端平台中进行计算和存储。与此同时,数据泄露、篡改、伪造及中间人攻击等安全事件频发,导致用户隐私泄露、资金损失及企业信誉受损,已成为制约数字经济健康发展的关键瓶颈。单一的传统加密技术、访问控制及防火墙手段难以应对复杂多变的攻击面,特别是,数据在云端平台之内进行计算和内部传输中,依然面临被非法读取、窃取、执行或篡改的风险。

[0003] 因此,亟需一种技术方案,提升数据在计算和传输中的安全性。

发明内容

[0004] 本说明书的一个或多个实施例描述了一种对异构计算设备进行安全认证、密钥交换的方法,一方面能够增强异构计算设备安全认证的可靠性,另一方面能够减少因等待密钥而导致的加密通信中断,从而全面提升数据的计算安全性以及传输效率。

[0005] 根据第一方面,提供了一种对异构计算设备进行安全认证的方法,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件中包含安全芯片,所述外壳关联部署有拆机传感器;所述方法由所述安全芯片执行,包括:

[0006] 响应于挑战方的指令,根据第一存储区中的目标数据生成度量数据,其中,所述目标数据包括,所述拆机传感器记录的拆机相关数据。

[0007] 利用第一私钥对所述度量数据进行签名,生成设备报告。

[0008] 向所述挑战方提供所述设备报告,供其进行验证。

[0009] 根据一种实施方式,所述第一私钥通过以下方式获得:

[0010] 响应于所述挑战方的第一请求,生成第一密钥对。

[0011] 将所述第一密钥对的公钥发送至所述挑战方;将其中的私钥作为所述第一私钥。

[0012] 根据一种实施方式,向所述挑战方提供所述设备报告,供其进行验证,包括:

[0013] 向所述挑战方提供所述设备报告,以使得所述挑战方基于所述第一私钥对应的第一公钥和第一证书,对所述设备报告进行验签;所述第一证书为可信认证机构基于所述第一公钥签发。

[0014] 根据一种实施方式,所述拆机相关数据包括拆机值;所述拆机值为累加计数值。

[0015] 根据一种实施方式,所述第一存储区为所述安全芯片的非易失性存储区NVindex。

[0016] 根据一种实施方式,所述目标硬件用于通过第一总线连接至主机方,从而在所述异构计算卡和主机方之间执行加解密操作;所述挑战方为所述主机方。

[0017] 根据一种实施方式,向所述挑战方提供所述设备报告,包括:

[0018] 通过所述第一总线,将所述设备报告发送至所述主机方。

[0019] 根据一种实施方式,所述拆机传感器包括以下中的一种:机械式微动传感器、磁控传感器、光敏传感器、霍尔传感器、MEMS加速度传感器。

[0020] 根据一种实施方式,所述拆机相关数据通过以下方式生成:

[0021] 获取所述拆机传感器的第一电平,所述第一电平用于指示所述外壳的拆开状态。

[0022] 基于所述第一电平,生成所述拆机相关数据,存入所述第一存储区。

[0023] 根据第二方面,提供了一种对异构计算设备进行密钥交换的方法,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件通过总线连接至主机方;所述目标硬件包括控制单元、加解密引擎和密钥缓冲区,控制单元上运行有目标程序;所述方法包括:

[0024] 响应于收发数据请求,加解密引擎在所述密钥缓冲区查找有效密钥。

[0025] 在所述密钥缓冲区存在有效密钥的情况下,加解密引擎使用所述密钥缓冲区中的密钥执行数据加密或解密操作。

[0026] 在所述密钥缓冲区满足预设的密钥补充条件的情况下,目标程序与所述主机方进行密钥协商,获得若干密钥,存入所述密钥缓冲区。

[0027] 根据一种实施方式,所述密钥缓冲区的容量是根据所述目标硬件的可用存储空间,和/或第一阈值而确定。

[0028] 根据一种实施方式,所述有效密钥满足第一安全约束;所述第一安全约束用于指示,所述有效密钥的总使用次数不超过第一阈值。

[0029] 根据一种实现方式,所述有效密钥为第N次使用;所述有效密钥满足第二安全约束;所述第二安全约束用于指示,与所述有效密钥对应的初始向量IV与前序N-1次均不相同。

[0030] 根据一种实现方式,所述目标硬件还包括安全芯片;所述获得若干密钥,包括:

[0031] 所述目标程序基于第一密钥交换协议,调用所述安全芯片,获取第一密钥,归入所述若干密钥。

[0032] 根据一种实施方式,所述方法还包括,在初始阶段,所述目标程序基于第一密钥交换协议与所述主机方进行密钥协商,获取m个密钥,存入所述密钥缓冲区;所述m是所述密钥缓冲区可存储密钥的最大数量。

[0033] 根据第三方面,提供了一种对异构计算设备进行安全认证的装置,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件中包含安全芯片、以及与所述外壳关联的拆机传感器;所述装置部署于所述安全芯片,包括:

[0034] 生成模块,配置为,响应于挑战方的指令,根据第一存储区中的目标数据生成度量数据,其中,所述目标数据包括,所述拆机传感器记录的拆机相关数据。

[0035] 签名模块,配置为,利用第一私钥对所述度量数据进行签名,生成设备报告。

[0036] 提供模块,配置为,向所述挑战方提供所述设备报告,供其进行验证。

[0037] 根据第四方面,提供了一种对异构计算设备进行密钥交换的装置,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结

构,所述目标硬件通过总线连接至主机方;所述目标硬件包括控制单元、加解密引擎和密钥缓冲区,控制单元上运行有目标程序;所述装置包括:

[0038] 加解密引擎,配置为,响应于收发数据请求,加解密引擎在所述密钥缓冲区查找有效密钥。

[0039] 所述加解密引擎还配置为,在所述密钥缓冲区存在有效密钥的情况下,加解密引擎使用所述密钥缓冲区中的密钥执行数据加密或解密操作。

[0040] 目标程序,配置为,在所述密钥缓冲区满足预设的密钥补充条件的情况下,目标程序与所述主机方进行密钥协商,获得若干密钥,存入所述密钥缓冲区。

[0041] 根据第五方面,提供了一种计算机程序产品,包括计算机程序/指令,该计算机程序/指令被处理器执行时实现第一方面或第二方面所述方法的步骤。

[0042] 根据第六方面,提供了一种计算设备,包括存储器和处理器,其特征在于,所述存储器中存储有可执行代码,所述处理器执行所述可执行代码时,实现第一方面或第二方面所述的方法。

[0043] 本说明书实施例提供的方法中,可以基于安全芯片中存储的拆机相关数据,对目标设备的物理完整性进行持续且可信的监测与验证,基于签名验签完成对异构计算设备的身份认证,使得挑战方能够一次性完成物理完整性与身份可信度的双重验证,确保只有未被拆机且运行受信代码的异构计算设备才能接入系统、建立通信,从而有效防御物理拆解和篡改攻击,为高安全要求的应用场景提供可信计算根基。除了以上安全特性之外,实施例的方案中,使用密钥缓冲区结合密钥补充机制,可以为异构计算设备及与其进行加密通信的主机方,提供可靠的密钥供给,减少因密钥耗尽或频繁协商而导致的加密通信中断,在高吞吐量通信场景中实现密钥供给的低延迟与高安全性。

附图说明

[0044] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述中所需要使用的附图做简单的介绍。显而易见地,下面描述中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。

[0045] 图1A为本说明书提供的典型大模型推理场景;

[0046] 图1B为本说明书示出的异构计算系统中的总线攻击示意图;

[0047] 图1C为本说明书披露的一种异构计算系统示意图;

[0048] 图2为本说明书实施例提供的一种对异构计算设备进行安全认证的架构示意图;

[0049] 图3为根据本说明书实施例提供的一种对异构计算设备进行安全认证的方法流程图;

[0050] 图4为本说明书披露的一种对异构计算设备进行密钥交换的相关技术示意图;

[0051] 图5为本说明书实施例提供的一种对异构计算设备进行密钥交换的架构示意图;

[0052] 图6为根据本说明书实施例提供的一种对异构计算设备进行密钥交换的方法流程图;

[0053] 图7为根据本说明书实施例提供的一种对异构计算设备进行安全认证的装置示意图;

[0054] 图8为根据本说明书实施例提供的一种对异构计算设备进行密钥交换的装置示意图。

具体实施方式

[0055] 下面结合附图,对本说明书实施例提供的方案进行详细描述。

[0056] 如前所述,在大数据和云平台的背景下,数据安全保护成为各方关注的焦点。图1A示出典型的大模型推理场景。在该场景中,大模型提供商把大模型传输给服务器提供商,从而将大模型部署至云端。用户通过网络,访问云上的大模型服务。具体地,用户可以将提示词通过网络传输至云端。服务器提供商基于大模型文件和用户提示词,利用其强大的计算资源进行模型推理和计算,得出推理结果返回给用户。为了增强计算性能,服务器中一般采用“CPU+计算加速卡”的异构计算系统进行数据处理,其中计算加速卡可以是,适合于特定类型的计算任务的专用硬件。在模型推理场景中,计算加速卡通常采用GPU。

[0057] 在例如大模型推理的计算场景中,使用方(包括大模型提供商和用户)希望服务器提供商能够提供数据的跨域管控,防止攻击者窃取或篡改数据。换而言之,希望服务器提供商能够确保计算过程中数据的安全。

[0058] 为了提升数据安全性,一些服务器提供商采用可信执行环境TEE方案,来确保数据计算过程的安全。目前,已存在多种基于CPU的TEE方案,可称为CPU TEE方案。CPU TEE方案主要通过硬件级隔离与加密机制,在处理器内部构建一个与通用计算环境逻辑隔离的安全执行域,确保敏感代码与数据的机密性与完整性。典型实现如Intel SGX、AMD SEV及ARM TrustZone,均通过指令集扩展与微架构增强,在CPU中引入安全内存访问控制、实时加密引擎及可信启动机制。应用程序可调用特定指令创建受保护的安全飞地,该飞地具备独立地址空间与上下文状态,非授权软件(包括操作系统与虚拟机监控器)无法访问其内部数据。同时,CPU通过硬件根密钥构建可信启动链,支持远程认证,确保飞地加载的代码未被篡改。

[0059] 然而,对于采用“CPU+计算加速卡”的异构计算系统来说,仅仅在CPU中提供TEE环境,并不足以确保整体计算环境的安全。如图1B所示,在异构计算系统中,通常CPU和异构计算卡(即计算加速卡)通过总线相连,总线上传输的是明文数据。即使在CPU中实现了可信执行环境TEE,甚至有些异构计算卡的厂商在其异构计算卡内部也实现了TEE,攻击者依然可以通过对总线的攻击来窃取用户数据。

[0060] 为此,在本说明书实施例中,提出图1C所示的异构计算系统。在该系统中,在主机CPU和异构计算卡之间添加目标硬件,该目标硬件与异构计算卡通过防拆装置封装为整体结构,形成异构计算设备。目标硬件用于在CPU和异构计算卡之间执行加解密操作,从而使得,在与主机CPU之间连接的第一总线上传输密文,在与异构计算卡连接的第二总线上传输明文。

[0061] 为了使得上述异构计算设备能够安全工作,需要确保:目标硬件本身没有被篡改,以及异构计算设备的整体结构没有被拆卸(否则第二总线仍然有暴露的风险)。通常地,在异构计算设备启动时,对其进行安全认证,以确认其身份的安全性。

[0062] 在常规的针对外接硬件设备进行安全认证的方案中,外接硬件设备设有安全芯片,安全芯片内部存储有该硬件设备的相关身份与状态信息,这些信息可用于生成度量数据(例如,对信息数据进行哈希编码得到的摘要数据)。挑战方可通过校验安全芯片中预置

的数字证书,确认其身份的合法性,并通过对度量数据的远程和本地证明,来完成对外接硬件设备的安全认证。然而,以此方式实现的安全认证,只能确保外接硬件设备本身未被篡改。若将其应用于图1C所示的异构计算设备的安全认证,并不足以验证异构计算设备是否遭到物理拆卸。换言之,在图1C所示的异构计算设备中,即使常规安全认证通过,仍然存在异构计算设备被物理拆卸的可能,攻击者可以将外壳拆开,操纵异构计算卡与目标硬件之间的通信总线,实施总线窃听、数据篡改或其他破坏设备安全的行为。

[0063] 另一方面,在一些防拆装置中引入有拆机传感器,用于监测外壳的物理开启状态,并记录相关拆机数据。然而,常规的拆机数据记录方式,在存储或传输阶段存在被篡改的可能。攻击者可在非法拆机后,对拆机数据进行篡改,以达到欺骗挑战者的目的。

[0064] 因此,已有的安全认证方案和防拆方案,都不足以确保图1C所示的异构计算设备的安全性。

[0065] 为了解决上述技术问题,在本说明书实施例中,提出一种对异构计算设备进行安全认证的方案,能够将安全认证覆盖至异构计算设备的防拆认证,全面提升安全认证的可信度。

[0066] 图2示出根据一个实施例的对异构计算设备进行安全认证的架构示意图。如图所示,在硬件结构上,该异构计算设备包括异构计算卡(图中以虚线示出,表示该异构计算卡可根据具体需求进行合法更换)。该异构计算设备还包括目标硬件,其中包含安全芯片。目标硬件和异构计算卡之间通过第二总线连接,并通过外壳,以防拆卸方式封装为整体结构。

[0067] 拆机传感器与所述外壳关联,并可与安全芯片通信。在一些示例中,拆机传感器部署在外壳上,特别是外壳的封装部分附近,并电连接至安全芯片。在另一些示例中,如图2所示,拆机传感器位于目标硬件中,并与安全芯片通信。

[0068] 拆机传感器可持续监测外壳状态,并在检测到未授权开启或物理拆解时触发响应。具体地,拆机传感器可实时采集设备外壳的物理状态变化,例如振动、开合位移或密封破坏等信号,并通过输出电平的变化将此类信号数据(拆机信号)传递至安全芯片,安全芯片检测拆机信号,将其存入受保护的存储区,为后续异构计算设备的安全认证提供物理层面状态数据。

[0069] 参阅图2,在该实施例中,所述安全芯片可以基于拆机传感器采集到的拆机信号,生成拆机相关数据并记录于安全芯片的存储区,有效防止恶意篡改。安全芯片可以将存储区中的拆机相关数据一并纳入度量数据,使得挑战方在对所述异构计算设备执行安全认证时,可以基于度量数据评估异构计算设备的物理拆机状态。

[0070] 在图2示出的架构中,所述安全芯片可以通过通信通道与挑战方建立通信链路,例如可以是:蓝牙通道、蜂窝网络、PCIe,等等,本说明书实施例对此不做具体限定。

[0071] 异构计算设备的目标硬件还可以通过第一总线与主机方建立通信链路,在一些实践中,第一总线与第二总线可以为基于同一种总线协议的链路。在异构计算设备通过挑战方的安全认证后,基于通信链路,目标硬件可与主机方进行数据交换,在异构计算卡与主机方之间执行加解密操作。在一些可选的实施方式中,所述挑战方可以是与目标硬件进行通信的主机方。

[0072] 下面将基于上述技术框架,给出一种安全认证方法的详细描述。

[0073] 图3示出了根据一个实施例提供的一种对异构计算设备进行安全认证的方法流程

图。可以理解,该方法可以通过任何具有计算、处理能力的装置、设备、平台、设备集群来执行。所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件中包含安全芯片,所述外壳关联部署有拆机传感器。有关异构计算设备的组成结构,以及其中所包含的各个部件,在前文中已有介绍,此处不再赘述。下面将以所述安全芯片作为该实施例中方法的执行主体,详细描述各个步骤的具体执行方式。

[0074] 步骤S301,响应于挑战方的指令,根据第一存储区中的目标数据生成度量数据,其中,所述目标数据包括,所述拆机传感器记录的拆机相关数据。

[0075] 挑战方通常是一个远程主机或服务,可用于验证安全芯片的宿主设备(对应于所述目标硬件)的身份和状态是否可信。如前所述,在一些实践中,挑战方与主机方可以指代同一主体,例如,与所述异构计算设备通过总线连接的主机处理器或主机安全模块。当强调其对所述异构计算设备的验证功能时,可以称之为挑战方,而当强调其对所述异构计算设备的通信功能时,可以称之为主机方。

[0076] 在该步骤中,安全芯片接收到来自挑战方的指令(通常为安全认证指令)后,访问其第一存储区中的目标数据。该目标数据包含,基于拆机传感器所记录的拆机相关数据,还可以包含所述目标硬件的基本信息(包括固件、硬件配置、运行代码等)。所述度量数据可以是对所述目标数据进行哈希运算得到的摘要数据。所述拆机传感器可以通过GPIO与安全芯片连接,所述拆机传感器用于收集所述异构计算设备的拆机情况数据,例如,拆机传感器可以采用机械式微动传感器、磁控传感器、光敏传感器、霍尔传感器或MEMS加速度传感器实现。任何一次外壳开启或异常震动,能够拉高拆机传感器的电平信号。安全芯片可以采用中断方式检测拆机传感器的电平信号变化,一旦检测到变化,安全芯片可以将电平变化状态与其他系统信息(例如,拆机传感器序列号、时间戳等)拼接为结构化记录,记录为所述拆机相关数据。在一种可选的实现方式中,拆机传感器也可以持续输出能够指示所述外壳拆开状态的第一电平;安全芯片获取第一电平,对其变化进行判断,从而基于所述第一电平,生成所述拆机相关数据,存入所述第一存储区。

[0077] 如上所述,拆机相关数据可包括所述异构计算设备的外壳的物理开启状态、开启次数、最后一次开启时间戳等。在一个具体的应用中,所述拆机相关数据包括拆机值,用于存储所述外壳被开启的情况。举例来说,拆机值可以用于指示外壳的开启状态,其中存储的数据为布尔数值,TRUE表示外壳目前或曾经被开启,FALSE表示外壳未曾被开启;拆机值还可以用于指示外壳的开启次数,其中存储的数据为累加计数值,每当外壳被开启,拆机值自增。

[0078] 所述拆机相关数据可以被写入所述安全芯片的第一存储区中,以形成不可篡改的可信日志,不仅可用于反映异构计算设备的物理完整性与历史状态,还可为后续远程证明提供可验证的凭据。

[0079] 在一种优选的实现方式中,所述安全芯片可以是兼容TPM或符合TCM标准的可信模块,所述第一存储区为所述安全芯片的非易失性存储区NVindex。NVindex可以是由安全芯片的生产厂商在芯片出厂时预先划定的一段受保护的存储空间,作为一种可灵活配置的受保护存储空间,其访问策略(例如,只读、单次写入、需授权访问等)可在NVindex创建时,通过属性设置进行定义,并由安全芯片硬件强制执行,从而确保其中所存储的拆机相关数据

无法被非法篡改或恶意清除。如此,通过将拆机相关数据与安全芯片的NVindex存储相结合,能够极大增强异构计算设备对物理攻击的检测与防御能力。

[0080] 通过上述步骤生成度量数据之后,在步骤S303以及步骤S305,可以利用第一私钥对所述度量数据进行签名,生成设备报告;向所述挑战方提供所述设备报告,供其进行验证。

[0081] 在该步骤的一种具体实现中,所述第一私钥可以是根密钥对应的私钥,或者基于根密钥派生的私钥。其中,根密钥是通过一次性烧录,永久固化在安全芯片硬件内部的出厂密钥。安全芯片出厂后,软件、固件甚至生产厂商均无法读取或更改根密钥,只能通过调用安全芯片提供的“黑盒接口”,让安全芯片硬件基于私钥执行签名或解密操作。

[0082] 所述安全芯片可以通过其与挑战方之间的通信通道,将使用第一私钥对度量数据签名而得到的设备报告提供给所述挑战方。在挑战方即为主机方的实践中,安全芯片也可以通过目标硬件与主机方之间的总线,将设备报告发送至所述主机方。

[0083] 挑战方一侧的验证程序(可以是系统级程序或用户态程序)将该设备报告与安全芯片的生产厂商预先提供或注册的安全状态数据进行比对,从而对目标硬件进行验证,确保目标硬件中的固件、代码等没有被篡改,同时,由于设备报告是基于包含拆机相关数据的度量数据制作,在校验设备报告时,也可同时对异构计算设备的物理拆卸状态进行检测,以确定其物理拆卸状态符合安全要求。

[0084] 在上述步骤的一种具体实现中,所述第一私钥可以是所述挑战方与所述安全芯片之间,通过安全密钥交换得到的双方互信的密钥对中的私钥。安全芯片使用第一私钥对所述度量数据进行签名后,所述挑战方可以使用第一私钥所对应的公钥对其进行验签。换言之,所述安全芯片持有密钥对的第一私钥,所述挑战方持有密钥对的第一公钥。安全芯片使用第一私钥对度量数据签名,得到的设备报告可由挑战方使用第一公钥进行验签,验签结果可以表明设备报告所表示的度量数据的完整性和来源真实性,从而确定目标数据的可信度。

[0085] 具体而言,挑战方在收到设备报告后,提取设备报告中的签名块与明文数据(对应于所述目标数据),使用第一公钥执行签名验证。若验签通过,则说明所述设备报告由持有对应第一私钥的安全芯片签发,且在传输过程中未被篡改。挑战方可以将目标数据中包含的度量值(例如,目标硬件的固件信息、硬件配置信息、拆机相关数据等)与预先注册的基准值或安全名单进行比对。若所有字段均匹配或符合安全规范,则可以确定目标硬件处于可信状态,能够与其建立通信;若任意字段出现不一致或超出安全规范,则可以触发告警、拒绝通信。

[0086] 在一个优选的实现方式中,所述第一私钥的生成可以独立于上述签名验签过程,预先独立进行。所述安全芯片可以响应于所述挑战方的第一请求,生成第一密钥对;将所述第一密钥对的公钥发送至所述挑战方;将所述第一密钥对的私钥作为所述第一私钥。为了增强所述第一密钥对的可信度,可以将所述第一密钥对的生成与对所述安全芯片的身份验证结果关联,也就是说,在所述安全芯片通过了挑战方的身份验证之后,在安全芯片可信身份的基础上,派生所述第一密钥对。

[0087] 因此,在一些具体的应用中,挑战方可以根据安全芯片所内置的密钥证书,对安全芯片的身份进行验证。所述密钥证书在安全芯片出厂时,预先烧录在安全芯片内部的非易

失性存储介质中。优选的,所述密钥证书可以被预先烧录在一次可编程存储区,写入后不可被更改、不可被删除。通常来说,密钥证书与安全芯片的公钥绑定,并由可信认证机构(Certificate Authority,CA)签发,挑战方可以使用CA的根证书验证该密钥证书的有效性,从而确定对该安全芯片的身份验证结果。在一个例子中,所述密钥证书可以是背书密钥证书(Endorsement Key Certificate,EK证书)。背书密钥证书是一种基于TPM或同类安全芯片的硬件级别身份凭证,由安全芯片的制造厂商所认可的可信认证机构签发,烧录在安全芯片的非易失性存储介质,例如,一次性可编程存储区OTP、物理熔丝eFuse等,与安全芯片唯一绑定。挑战方在获取背书密钥证书后,可以通过验证证书链来确认安全芯片的身份:使用CA的根证书公钥验证背书密钥证书的签名,检查密钥证书的有效状态,读取背书密钥证书中所包含的安全芯片型号、标识符等,以确认安全芯片的身份合法性。

[0088] 在所述安全芯片的身份验证通过之后,密钥证书还可以用于派生上文所阐述的第一密钥对。具体而言,通过身份验证的安全芯片响应于挑战方的第一请求,生成第一密钥对。将所述第一密钥对的私钥(第一私钥)保存在安全芯片内部,公钥(第一公钥)发送至所述挑战方。所述挑战方获得第一公钥后,向CA申请第一公钥的公钥证书(第一证书),在后续对设备报告的验签过程中使用。所述挑战者还可使用所述安全芯片的密钥证书的公钥作为加密密钥,将所述公钥证书加密传递至所述安全芯片,安全芯片可以使用密钥证书的私钥进行解密,得到公钥证书,保存在安全芯片内部,待后续使用。

[0089] 通过上述各实施例所提供的方法,可以基于安全芯片中存储的拆机相关数据,对目标设备的物理完整性进行持续且可信的监测与验证,基于签名验签完成对异构计算设备的身份认证,使得挑战方能够一次性完成物理完整性与身份可信度的双重验证,确保只有未被拆机且运行受信代码的异构计算设备才能接入系统、建立通信,从而有效防御物理拆解和篡改攻击,为高安全要求的应用场景提供可信计算根基。

[0090] 执行上述各实施例提供的方法,挑战方与异构计算设备之间能够完成安全认证。进入数据交换阶段,进行加密通信的双方(挑战方所属主机方与目标硬件),可以建立加密信道,以进行受保护的加密通信。该过程通常依赖于通信双方之间的密钥交换机制。图4示出了异构计算设备基于常规密钥交换机制进行密钥交换的技术架构示意图。参阅附图,异构计算设备与主机方通过总线连接,双方均各自部署机密计算应用,机密计算应用之间可以基于密钥协商协议进行密钥协商。当异构计算设备与主机方之间需要发起一轮加密通信时,其中的任一通信方响应于数据收/发请求,可以通知其本地的机密计算应用提供密钥。机密计算应用需实时执行密钥协商或查询密钥库,在此过程中,该通信方因无可用有效密钥而处于等待状态。等待期间数据传输暂停,数据加密或解密操作无法执行。可见,在高频次或大数据量加密通信场景中,若每一轮加密通信都需实时协商密钥,将频繁引发通信方进入等待状态,严重影响异构计算卡与主机方之间的通信效率。

[0091] 为了解决上述技术问题,在本说明书实施例中,提出一种对异构计算设备进行密钥交换的方案,旨在为加密通信持续提供可用的有效密钥,减小甚至消除密钥等待,提高加密通信的效率。图5示出了根据一个实施例的对异构计算设备进行密钥交换的架构示意图。

[0092] 如图所示,异构计算设备包括相互连接的异构计算卡和目标硬件,二者通过外壳封装为整体结构。目标硬件与主机方通过总线连接,目标硬件包含控制单元、加解密引擎和密钥缓冲区,控制单元上运行有目标程序。目标程序可以与主机方(典型的,可以是主机方

部署的机密计算应用)执行密钥协商,获得若干密钥,存入密钥缓冲区。在该实施例中,将密钥协商与加密通信在触发时序上解耦,即所述目标程序可以在任意时间窗口(例如,满足密钥补充条件时),主动与主机方发起密钥协商,向密钥缓冲区填充密钥。而加解密引擎在需要执行数据加密或解密操作时,可直接从密钥缓冲区获取已预存的有效密钥,无需等待实时密钥协商,提升密钥供给连续性,从而可以保障高吞吐量加密通信场景下的通信效率。

[0093] 基于上述技术构思,图6示出了根据一个实施例提供的一种对异构计算设备进行密钥交换的方法流程图。可以理解,该方法可以通过任何具有计算、处理能力的装置、设备、平台、设备集群来执行。所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,有关异构计算设备的组成结构,以及其中所包含的各个部件,在前文中已有介绍,此处不再赘述。

[0094] 所述目标硬件通过总线连接至主机方,所述总线可以是任意能够提供高带宽、低延迟的点对点串行链路,也可以是具备多设备共享能力的并行链路,例如,PCIE总线、NVLink总线,等等。所述主机方可以是本地或远程的主机系统,例如,主板、CPU、嵌入式PLC,等等;也可以是其他计算加速单元,例如,另一异构计算设备;还可以是通过高速网络连接的分布式计算节点;具体的总线形态以及主机方,可以视应用环境以及应用需求而灵活多变,本说明书对此不做具体限定。

[0095] 所述目标硬件包括控制单元、加解密引擎和密钥缓冲区,控制单元上运行有目标程序。

[0096] 所述控制单元为所述目标硬件的核心处理模块,其可以被实现为诸如微控制器(MCU)、片上系统(SoC)或可编程逻辑单元,用于执行密钥交换协议、协调加解密任务以及管理密钥生命周期。

[0097] 在所述控制单元上运行的目标程序可以用于执行密钥协商,典型的,目标程序可以被实现为机密计算服务应用CCS APP,与主机方进行密钥协商。

[0098] 加解密引擎可用于高速执行对称与非对称密码算法,与主机方完成通信数据的加密或解密操作。

[0099] 密钥缓冲区可以为加解密引擎提供密钥缓存与轮转服务,其可存储多个预计算的会话密钥,以在高吞吐量通信场景中实现密钥供给的低延迟与高安全性。在具体应用中,所述密钥缓冲区的容量可以根据所述目标硬件的可用存储空间,和/或第一阈值而确定。在一个例子中,所述目标硬件可以根据其自身可用的存储资源空间(在一些实践中,也可结合异构计算卡的可用存储空间),确定所述密钥缓冲区的容量,从而使得密钥缓冲区能够在不影响异构计算设备整体性能的前提下,存储足够数量的密钥以应对突发的高频加密通信需求。在另一个例子中,所述目标硬件也可以结合具体应用需要,为所述密钥缓冲区动态评估、确定合适的容量,以使得密钥缓冲区的密钥供给策略能够灵活实时适配加密通信频率变化,例如在高吞吐时段自动扩展密钥缓冲区容量,而在低吞吐时段适当降低以释放存储资源,从而实现密钥供给与存储资源利用率的平衡。

[0100] 下面将基于上述硬件架构结合实施例,详细描述上述方法中各个步骤的具体执行方式。

[0101] 步骤S601,响应于收发数据请求,加解密引擎在所述密钥缓冲区查找有效密钥。

[0102] 在该步骤中,当目标硬件需要执行加密通信,发送或接收数据时,加解密引擎首先

被触发以获取用于本次加密通信的密钥。密钥缓冲区中存储了多个预先生成或协商的密钥。加解密引擎可以根据预设的密钥选择策略(例如轮询策略、密钥使用次数排序或基于安全等级的密钥匹配)遍历密钥缓冲区,筛选出当前处于有效状态的密钥。

[0103] 也就是说,在所述密钥缓冲区存在有效密钥的情况下,于步骤S603,加解密引擎可以使用所述密钥缓冲区中的密钥执行数据加密或解密操作。

[0104] 在使用密钥进行加密通信时,还需考虑密码学的安全性要求,保障用于加密通信的密钥(即所述密钥缓冲区中的任意有效密钥)满足安全约束。

[0105] 根据一种实现方式,所述安全约束用于指示,所述有效密钥的总使用次数不超过第一阈值。也就是说,所述密钥缓冲区中的任意一个密钥可以多次用于加密通信,然而出于安全性考虑,在密钥被使用次数超过第一阈值后,该密钥不再被视为有效密钥。所述加解密引擎需从所述密钥缓冲区获取另一个有效密钥,以用于后续的加密通信。所述第一阈值通常为预设的数值,典型的,可以根据密钥的位数动态确定该密钥所对应的第一阈值。

[0106] 根据一种实现方式,所述有效密钥为第N次使用;所述安全约束用于指示,与所述有效密钥对应的初始向量IV与前序N-1次均不相同。具体来说,在确定密钥的有效性时,可以同时考虑和密钥一同被加解密引擎使用的初始化向量IV。初始化向量IV是一种随机或伪随机生成值,用于在加密中引入随机性,确保即使对相同明文使用同一密钥加密,也可得到不同的密文,从而防止重放攻击。在加密通信中,一个密钥可以搭配不同的IV被重复使用多次,然而出于安全性考虑,需要确保每一次使用的IV唯一性。因此,在判断密钥有效性时,可以同时校验其是否始终与未曾搭配过的IV配合使用,以符合密文不可区分的安全性要求。

[0107] 相对应的,若所述密钥缓冲区当前无可用有效密钥,则加解密引擎可以向所述目标程序发送密钥更新请求,触发所述目标程序启动密钥交换流程,向所述密钥缓冲区填充新密钥。在所述密钥缓冲区成功填入新的有效密钥之前,所述加解密引擎将处于等待状态,无法执行后续的加密通信。为了减少此类情况的发生,提高加密通信的连续性和效率,可以预先设置密钥补充条件,一旦检测到密钥缓冲区满足该密钥补充条件,即可执行步骤S605,目标程序与所述主机方进行密钥协商,获得若干密钥,存入所述密钥缓冲区。

[0108] 所述密钥补充条件可以根据具体需要灵活设定,例如可以是,所述密钥缓冲区中的有效密钥余量低于预设阈值、或者预设比例的密钥的存活时间接近上限,等等。

[0109] 通常来说,所述目标程序可以基于软件实现密钥协商,例如OpenSSL。为了进一步提升密钥协商过程的安全性与可靠性,可采用硬件级保护机制。在一种实践中,所述目标硬件还可以包括安全芯片(例如,TPM/TCM),所述目标程序可以基于第一密钥交换协议,调用所述安全芯片,获取第一密钥,存入所述密钥缓冲区。所述密钥交换协议通常可以采用例如ECDH(ECC算法)协议、DH(RSA算法)协议等。在此过程中,安全芯片可确保密钥的生成、交换与存储均在其内部受信任环境中执行,私钥永不外泄,从而显著增强密钥协商的安全性。

[0110] 在一些实践中,所述密钥补充条件还可以配置为:所述异构计算设备与主机方建立加密通信信道进入初始化阶段。由于信道初始化过程中,通常尚未开始执行大规模加密通信,系统资源相对充裕,存在密钥生成窗口,因此可以在此期间重复多次执行步骤S605,向所述密钥缓冲区填充密钥。优选的,所述密钥缓冲区可以存储密钥的最大数量为m,在初始化阶段,所述目标程序可以基于密钥交换协议(例如,前文所述第一密钥交换协议)与所述主机方进行密钥协商,获取m个密钥,存入所述密钥缓冲区。如此,便可待加密通信信道建

立时,所述密钥缓冲区中已储备充足的有效密钥,可立即为后续加密通信提供密钥。

[0111] 需要知晓,上述步骤S603为所述加解密引擎从密钥缓冲区获取有效密钥的执行步骤,步骤S605为所述目标程序向密钥缓冲区适时填充密钥的执行步骤,二者相互协同、动态配合,可并行或交替执行,共同构成密钥的端到端供给机制,从而在保障密钥使用安全的前提下,持续满足高并发加密通信场景中的密钥需求。

[0112] 上述各实施例所提供的方法,使用密钥缓冲区结合密钥补充机制,可以为异构计算设备及与其进行加密通信的主机方,提供可靠的密钥供给,减少因密钥耗尽或频繁协商而导致的加密通信中断,在高吞吐量通信场景中实现密钥供给的低延迟与高安全性。由此,密钥交换双方可以便捷地使用密钥进行加密通信,满足高速、大数据量、海量数据包加密通信场景的需求。

[0113] 上述内容对本说明书的特定实施例进行了描述,其他实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行,并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要按照示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的,或者可能是有利的。

[0114] 图7为根据本说明书实施例提供的一种对异构计算设备进行安全认证的装置。该装置700部署在计算设备中,该计算设备可以通过任何具有计算、处理能力的装置、设备、平台、设备集群等来实现。该装置实施例与图3所示方法实施例相对应,其中,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件中包含安全芯片,所述外壳关联部署有拆机传感器。该装置700部署于所述安全芯片,包括:

[0115] 生成模块701,配置为,响应于挑战方的指令,根据第一存储区中的目标数据生成度量数据,其中,所述目标数据包括,所述拆机传感器记录的拆机相关数据。

[0116] 签名模块702,配置为,利用第一私钥对所述度量数据进行签名,生成设备报告。

[0117] 提供模块703,配置为,向所述挑战方提供所述设备报告,供其进行验证。

[0118] 根据另一方面的实施例,还提供了一种对异构计算设备进行密钥交换的装置,该装置可以部署在任何具有计算、处理能力的设备或平台上。图8示出根据一个实施例的对异构计算设备进行密钥交换的装置示意图,该装置实施例与图6所示方法实施例相对应,所述异构计算设备包括异构计算卡和目标硬件,所述异构计算卡和所述目标硬件通过外壳封装为整体结构,所述目标硬件通过总线连接至主机方;所述目标硬件包括控制单元、加解密引擎和密钥缓冲区,控制单元上运行有目标程序。如图8所示,所述装置800包括:

[0119] 加解密引擎801,配置为,响应于收发数据请求,加解密引擎在所述密钥缓冲区查找有效密钥。

[0120] 所述加解密引擎801还配置为,在所述密钥缓冲区存在有效密钥的情况下,加解密引擎使用所述密钥缓冲区中的密钥执行数据加密或解密操作。

[0121] 目标程序802,配置为,在所述密钥缓冲区满足预设的密钥补充条件的情况下,目标程序与所述主机方进行密钥协商,获得若干密钥,存入所述密钥缓冲区。

[0122] 根据另一方面的实施例,本说明书还提供了一种计算机程序产品,包括计算机程序/指令,该计算机程序/指令被处理器执行时实现前述结合图3或图6所述方法的步骤。

[0123] 根据又一方面的实施例,本说明书还提供了一种计算设备,包括存储器和处理器,其特征在于,所述存储器中存储有可执行代码,所述处理器执行所述可执行代码时,实现前述结合图3或图6所述方法的步骤。

[0124] 本领域技术人员应该可以意识到,在上述一个或多个示例中,本发明实施例所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时,可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。

[0125] 以上所述的具体实施方式,对本发明实施例的目的、技术方案和有益效果进行了进一步的详细说明。所应理解的是,以上所述仅为本发明实施例的具体实施方式而已,并不用于限定本发明的保护范围,凡在本发明的技术方案的基础之上所做的任何修改、等同替换、改进等,均应包括在本发明的保护范围之内。

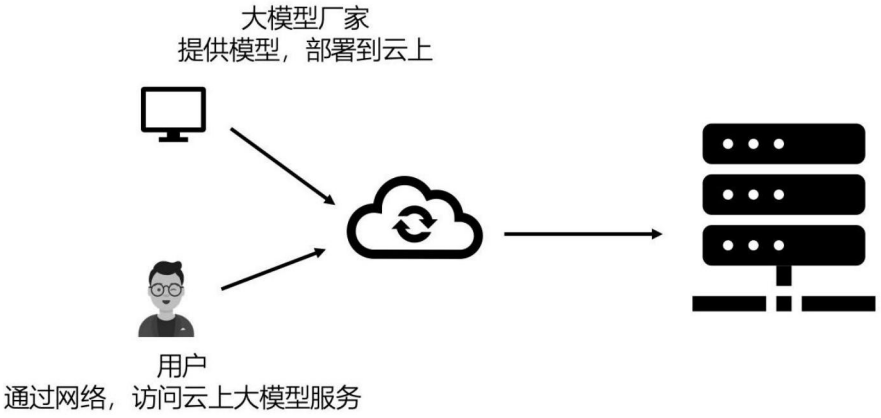

大模型厂家
提供模型，部署到云上
用户
通过网络，访问云上大模型服务

图1A

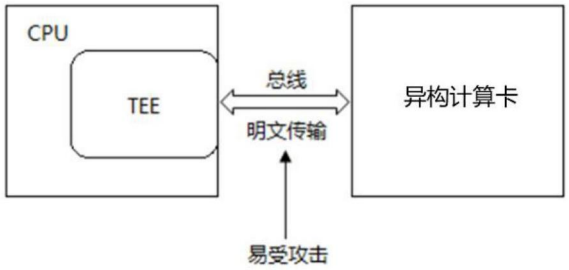

CPU
TEE
总线
明文传输
异构计算卡
易受攻击

图1B

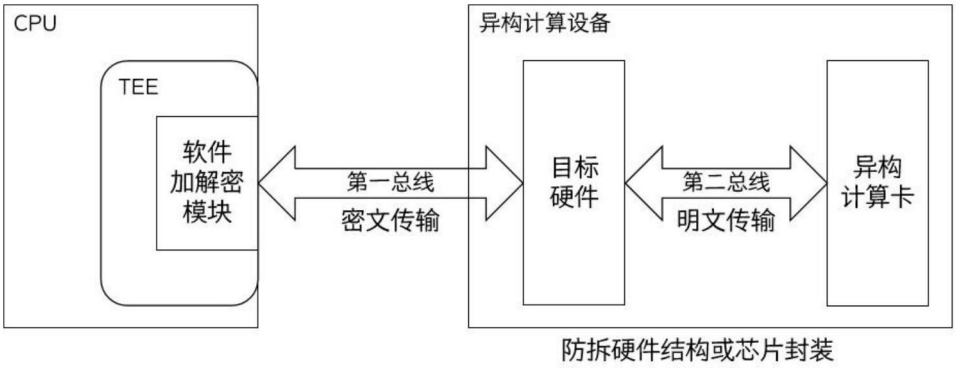

CPU
TEE
软件
加解密
模块
第一总线
密文传输
异构计算设备
目标
硬件
第二总线
明文传输
异构
计算卡
防拆硬件结构或芯片封装

图1C

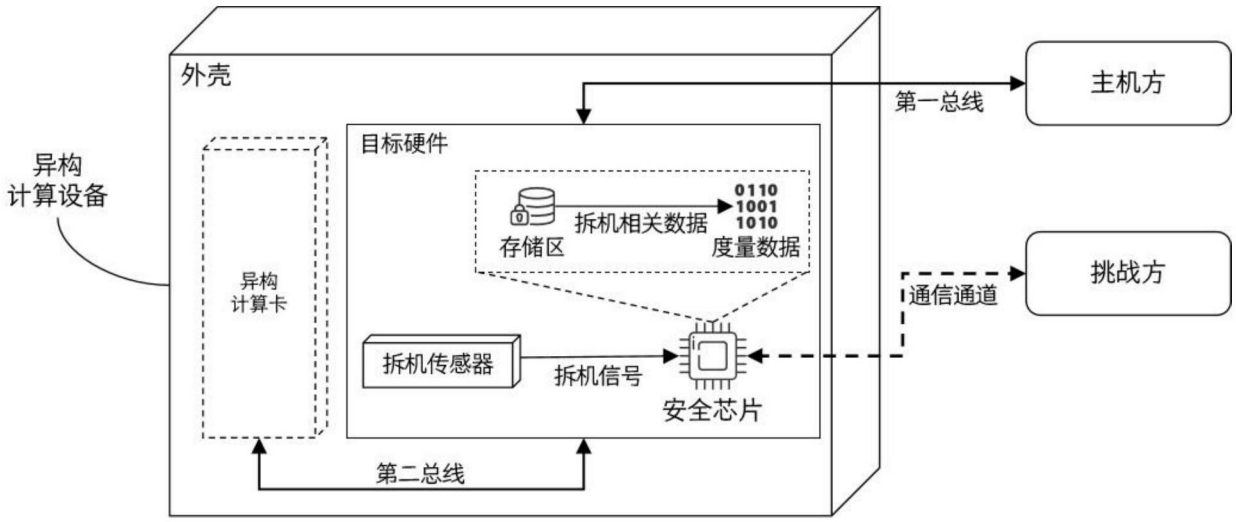

外壳
主机方
第一总线
异构
计算设备
目标硬件
异构
计算卡
0110
1001
1010
拆机相关数据
存储区
度量数据
挑战方
通信通道
拆机传感器
拆机信号
安全芯片
第二总线

图2

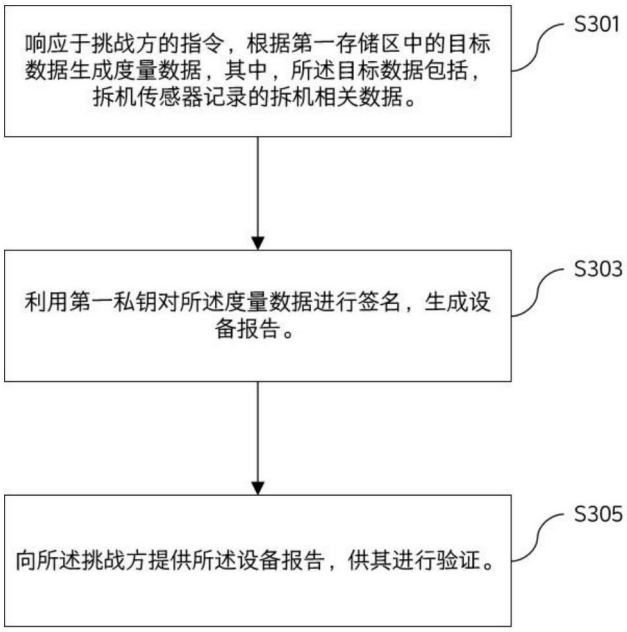

响应于挑战方的指令，根据第一存储区中的目标数据生成度量数据，其中，所述目标数据包括，拆机传感器记录的拆机相关数据。
S301
利用第一私钥对所述度量数据进行签名，生成设备报告。
S303
向所述挑战方提供所述设备报告，供其进行验证。
S305

图3

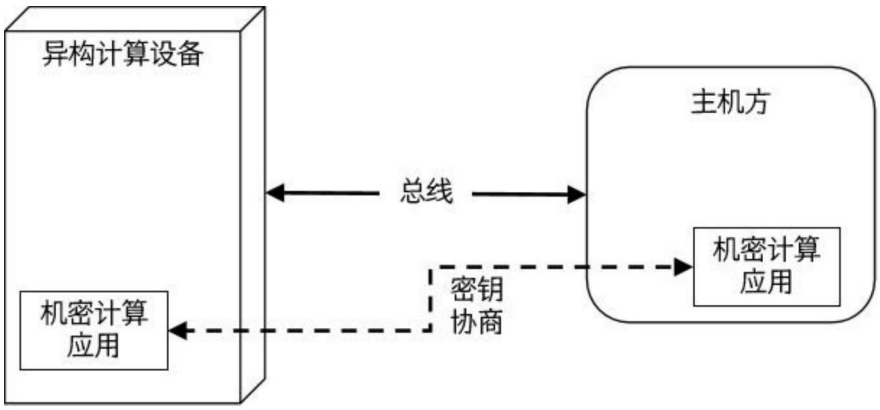

异构计算设备
主机方
总线
机密计算应用
密钥协商
机密计算应用

图4

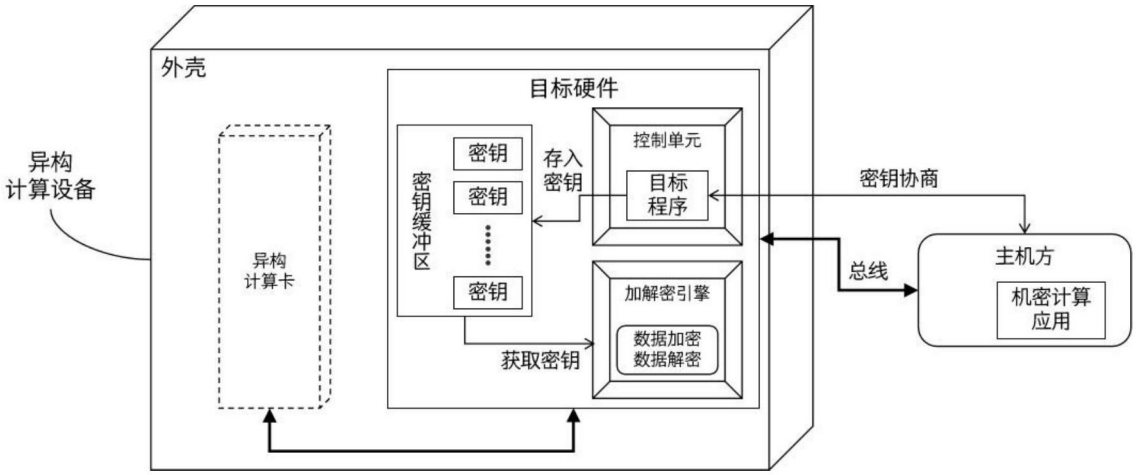

外壳
目标硬件
异构计算设备
异构计算卡
密钥缓冲区
密钥
密钥
密钥
存入密钥
控制单元
目标程序
密钥协商
总线
主机方
机密计算应用
加解密引擎
数据加密数据解密
获取密钥

图5

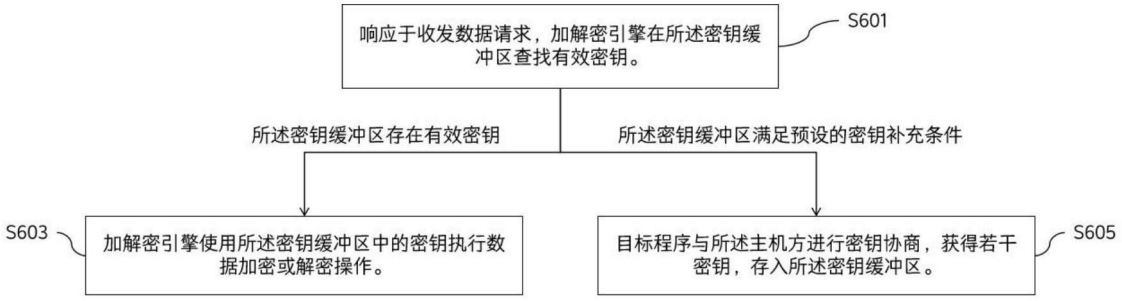

响应于收发数据请求，加解密引擎在所述密钥缓冲区查找有效密钥。
S601
所述密钥缓冲区存在有效密钥
所述密钥缓冲区满足预设的密钥补充条件
S603
加解密引擎使用所述密钥缓冲区中的密钥执行数据加密或解密操作。
目标程序与所述主机方进行密钥协商，获得若干密钥，存入所述密钥缓冲区。
S605

图6

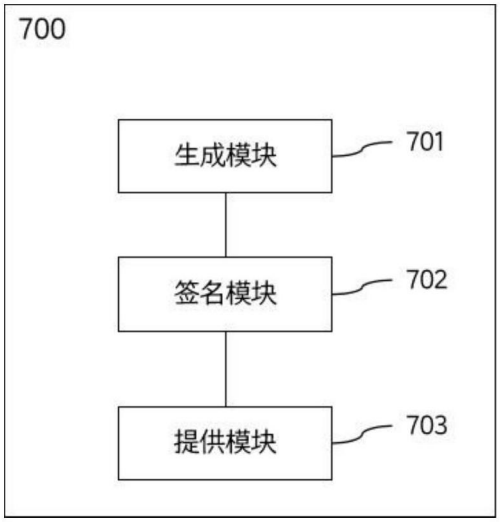


图7

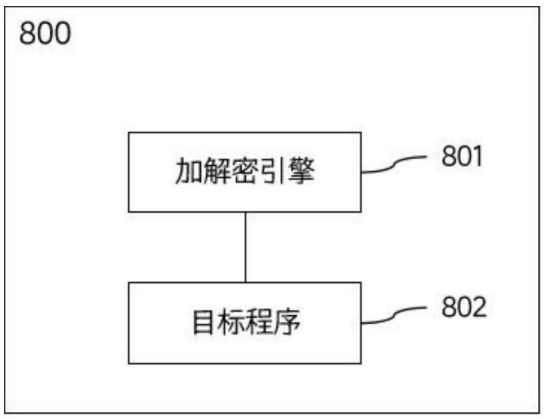


图8