(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115242538 A
(43)申请公布日 2022.10.25

(21)申请号 202210901464.4

(22)申请日 2022.07.28

(71)申请人 天翼云科技有限公司
地址 100007 北京市东城区青龙胡同甲1号、3号2幢2层205-32室

(72)发明人 黄静

(74)专利代理机构 北京同达信恒知识产权代理有限公司 11291
专利代理师 卢亚培

(51)Int.Cl.
H04L 9/40(2022.01)
G06F 16/16(2019.01)
G06F 16/182(2019.01)

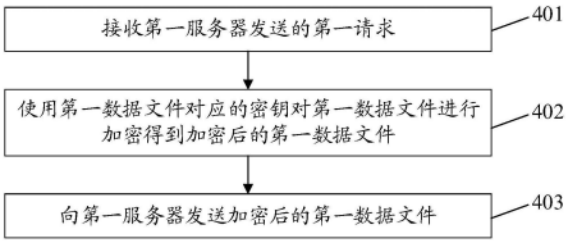
权利要求书2页 说明书11页 附图3页

(54)发明名称

一种数据传输方法及装置

(57)摘要

本申请公开了一种数据传输方法及装置，用以解决现有技术中无法保证数据在传输过程中的安全性问题。本申请提供的方法包括：接收第一服务器发送的第一请求，所述第一服务器属于第一分布式文件系统，所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件；使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件；向所述第一服务器发送所述加密后的第一数据文件。



CN 115242538 A

1.一种数据传输方法,其特征在于,包括:

接收第一服务器发送的第一请求,所述第一服务器属于第一分布式文件系统,所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件;

使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件;

向所述第一服务器发送所述加密后的第一数据文件。

2.如权利要求1所述的方法,其特征在于,所述第一分布式文件系统和所述第二分布式文件系统均部署Distcp工具,所述使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件,包括:

通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

3.如权利要求2所述的方法,其特征在于,通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件,包括:

通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

4.如权利要求1-3任一项所述的方法,其特征在于,所述第一分布式文件系统和所述第二分布式文件系统均为hadoop分布式文件系统。

5.如权利要求3所述的方法,其特征在于,所述方法还包括:

在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件之前,确定所述cipher中包括加密参数,所述加密参数用于指示所述第一数据文件在传输之前需要执行加密操作。

6.如权利要求3或5所述的方法,其特征在于,所述方法还包括:

获取所述第一数据文件对应的带宽参数,所述带宽参数用于指示所述第一数据文件的传输带宽;

所述通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件,包括:

调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件,按照所述带宽参数所指示的传输带宽进行加密,以得到加密后的第一数据文件。

7.一种数据传输方法,其特征在于,应用于第一分布式文件系统,包括:

向第二分布式文件系统发送第一请求,所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件;

接收第二分布式文件系统发送的经加密的第一数据文件;

使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

8.一种数据传输装置,其特征在于,包括:

接收模块,用于接收第一服务器发送的第一请求,所述第一服务器属于第一分布式文件系统,所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件;

处理模块,用于使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件;

发送模块，用于向所述第一服务器发送所述加密后的第一数据文件。

9.一种执行设备，其特征在于，包括：

存储器，用于存储程序指令；

处理器，用于调用所述存储器中存储的程序指令，按照获得的程序指令执行权利要求1-7中任一项所述的方法。

10.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质中存储有指令，当其在计算机上运行时，使得计算机执行如权利要求1-7中任一项所述的方法。

一种数据传输方法及装置

技术领域

[0001] 本申请涉及数据处理领域,尤其涉及一种数据传输方法及装置。

背景技术

[0002] 随着大数据Hadoop技术的不断普及,且Hadoop生态成为大数据开源的代表之一,各行各业在Hadoop上积累了大量的数据存储。但是伴随着数据安全问题,用户对放在Hadoop大数据平台上的敏感数据的安全性充满担心。现有技术中,不同Hadoop系统之间可用于对数据进行传输。然而,在Hadoop系统之间数据传输过程中容易遭到攻击而导致数据泄露,从而无法保证数据在传输过程中的安全性。

发明内容

[0003] 本申请实施例提供了一种数据传输方法及装置,用以解决现有技术中无法保证数据在传输过程中的安全性问题。

[0004] 第一方面,本申请实施例提供了一种数据传输方法,包括:

[0005] 接收第一服务器发送的第一请求,所述第一服务器属于第一分布式文件系统,所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件;使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件;向所述第一服务器发送所述加密后的第一数据文件。

[0006] 基于上述方案,可以实现在数据拷贝过程中对数据进行加密,进而将加密后的数据进行传输,保证了数据在传输过程中的安全性。

[0007] 一种可能的实现方式中,所述第一分布式文件系统和所述第二分布式文件系统均部署Distcp工具,所述使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件,包括:通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

[0008] 一种可能的实现方式中,通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件,包括:

[0009] 通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

[0010] 基于上述方案,通过Distcp工具调用cipher命令对数据进行加密,可以实现通过Hadoop中的Distcp框架本身对数据进行加密。

[0011] 一种可能的实现方式中,所述第一分布式文件系统和所述第二分布式文件系统均为hadoop分布式文件系统。

[0012] 一种可能的实现方式中,所述方法还包括:在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件之前,确定所述cipher中包括加密参数,所述加密参数用于指示所述第一数据文件在传输之前需要执行加密操作。

[0013] 基于上述方案，在Cipher命令中包括加密参数时，可以通过加密参数对数据进行加密。当Cipher命令中不包括加密参数时，可以不对数据进行加密，以数据文件流的形式直接进行传输。

[0014] 一种可能的实现方式中，所述方法还包括：确定获取所述第一数据文件对应的带宽参数，所述带宽参数用于指示所述第一数据文件的传输带宽；所述通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件，包括：调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件，按照所述带宽参数所指示的传输带宽进行加密，以得到加密后的第一数据文件。

[0015] 基于上述方案，通过任务数参数以及带宽参数，限制加密的并发和传输带宽，以达到不同网络情况下，按需控制加密数据传输速度，保障在不同网络环境下加密数据过程的稳定性。

[0016] 第二方面，本申请实施例提供了一种数据传输方法，应用于第一分布式文件系统，包括：

[0017] 向第二分布式文件系统发送第一请求，所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件；

[0018] 接收第二分布式文件系统发送的经加密的第一数据文件；

[0019] 使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0020] 一种可能的实现方式中，所述第一分布式文件系统和所述第二分布式文件系统均部署Distcp工具，所述使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件，包括：通过Distcp工具使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0021] 一种可能的实现方式中，通过Distcp工具使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件，包括：

[0022] 通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0023] 一种可能的实现方式中，所述第一分布式文件系统和所述第二分布式文件系统均为hadoop分布式文件系统。

[0024] 一种可能的实现方式中，所述方法还包括：在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件之前，确定所述cipher中包括解密参数，所述解密参数用于指示所述经加密的第一数据文件在传输之后需要执行解密操作。

[0025] 一种可能的实现方式中，所述方法还包括：确定获取所述第一数据文件对应的带宽参数，所述带宽参数用于指示所述第一数据文件的传输带宽；所述通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件，包括：调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件，按照所述带宽参数所指示的传输带宽进行解密，以得到所述第一数据文件。

[0026] 第三方面,本申请实施例提供了一种数据传输装置,包括:

[0027] 接收模块,用于接收第一服务器发送的第一请求,所述第一服务器属于第一分布式文件系统,所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件;

[0028] 处理模块,用于使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件;

[0029] 发送模块,用于向所述第一服务器发送所述加密后的第一数据文件。

[0030] 一种可能的实现方式中,所述第一分布式文件系统和所述第二分布式文件系统均部署Distcp工具,所述处理模块,在使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件时,具体用于:通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

[0031] 一种可能的实现方式中,所述处理模块,在通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件时,具体用于:

[0032] 通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

[0033] 一种可能的实现方式中,所述第一分布式文件系统和所述第二分布式文件系统均为hadoop分布式文件系统。

[0034] 一种可能的实现方式中,所述处理模块还用于:在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件之前,确定所述cipher中包括加密参数,所述加密参数用于指示所述第一数据文件在传输之前需要执行加密操作。

[0035] 一种可能的实现方式中,所述处理模块还用于:确定获取所述第一数据文件对应的带宽参数,所述带宽参数用于指示所述第一数据文件的传输带宽;所述通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件,包括:调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件,按照所述带宽参数所指示的传输带宽进行加密,以得到加密后的第一数据文件。

[0036] 第四方面,本申请实施例提供了一种数据传输装置,包括:

[0037] 发送模块,用于向第二分布式文件系统发送第一请求,所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件;

[0038] 接收模块,用于接收第二分布式文件系统发送的经加密的第一数据文件;

[0039] 处理模块,用于使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0040] 一种可能的实现方式中,所述第一分布式文件系统和所述第二分布式文件系统均部署Distcp工具,所述处理模块,在使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件时,具体用于:通过Distcp工具使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0041] 一种可能的实现方式中,所述处理模块,在通过Distcp工具使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件时,具体用于:

[0042] 通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0043] 一种可能的实现方式中，所述第一分布式文件系统和所述第二分布式文件系统均为hadoop分布式文件系统。

[0044] 一种可能的实现方式中，所述处理模块还用于：在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件之前，确定所述cipher中包括解密参数，所述解密参数用于指示所述经加密的第一数据文件在传输之后需要执行解密操作。

[0045] 一种可能的实现方式中，所述处理模块还用于：确定获取所述第一数据文件对应的带宽参数，所述带宽参数用于指示所述第一数据文件的传输带宽；所述处理模块，在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件时，具体用于：调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件，按照所述带宽参数所指示的传输带宽进行解密，以得到所述第一数据文件。

[0046] 第五方面，本申请实施例提供了一种执行设备，包括：

[0047] 存储器，用于存储程序指令；

[0048] 处理器，用于调用所述存储器中存储的程序指令，按照获得的程序指令执行第一、二方面以及第一、二方面中不同实现方式所述的方法。

[0049] 第六方面，本申请实施例提供了一种计算机可读存储介质，所述计算机可读存储介质中存储有指令，当其在计算机上运行时，使得计算机执行第一、二方面以及第一、二方面中不同实现方式所述的方法。

[0050] 另外，第二方面至第六方面中任一种实现方式所带来的技术效果可参见第一方面以及第一方面不同实现方式所带来的技术效果，此处不再赘述。

附图说明

[0051] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0052] 图1为本申请实施例提供的一种数据传输方法的使用场景示意图；

[0053] 图2为本申请实施例提供的系统架构示意图；

[0054] 图3为本申请实施例提供的服务器结构示意图；

[0055] 图4为本申请实施例提供的一种数据传输方法的流程示意图；

[0056] 图5为本申请实施例提供的另一种数据传输方法的流程示意图；

[0057] 图6为本申请实施例提供的一种数据传输装置的示意图；

[0058] 图7为本申请实施例提供的另一种数据传输装置的示意图；

[0059] 图8为本申请实施例提供的又一种数据传输装置的示意图。

具体实施方式

[0060] 下面将结合本申请实施例中附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。通常在附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。

[0061] 因此，以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围，而是仅仅表示本申请的选定实施例。基于本申请的实施例，本领域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，都属于本申请保护的范围。

[0062] 本申请的说明书和权利要求书及附图中的术语“第一”和“第二”是用于区别不同对象，而非用于描述特定顺序。此外，术语“包括”以及它们任何变形，意图在于覆盖不排他的保护。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。本申请中的“多个”可以表示至少两个，例如可以是两个、三个或者更多个，本申请实施例不做限制。

[0063] 另外，本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中字符“/”，在不做特别说明的情况下，一般表示前后关联对象是一种“或”的关系。

[0064] 为了便于理解本申请实施例提出的方案，首先对本申请涉及的技术用语进行介绍：

[0065] 1、Hadoop：是一个分布式系统基础架构，可运行在大规模集群上，实现分布式文件系统。Hadoop是一个开源分布式计算平台。Hadoop由两部分组成，一是负责存储与管理文件的分布式文件存储系统(Hadoop Distributed File System，HDFS)，二是负责处理与计算的MapReduce的计算框架。以HDFS和MapReduce为核心的Hadoop为用户提供了系统底层细节透明的分布式基础架构。基于Hadoop，能够高效地处理海量数据的分布式并行程序，将其运行于成百上千个节点组成的大规模计算机集群上。Hadoop具有高可靠性、高扩展性、高效性、高容错性等优点。

[0066] 2、HDFS：分布式文件存储系统，HDFS以流式数据访问模式来存储超大文件，具有一定的容错性，而且提供了高吞吐量的数据访问，非常适合大规模数据集上的应用。HDFS在进行文件存储时，采用文件分块存储，即HDFS在存储文件时，是将一个较大的文件平均分块并存储到不同计算器上。在读取文件时，可以从多个主机读取不同区块的文件，读取效率高。此外，HDFS还具有较高的容错性。HDFS自动将数据多份复制并且分布到物理位置的不同服务器上，并且能够自动将失败的任务重新分配。数据校验功能、后台的连续自检数据一致性功能，都为高容错提供了可能。

[0067] 3、Distcp(分布式拷贝)：是用于大规模集群内部和集群之间拷贝的工具。Distcp使用Map/Reduce实现文件分发，错误处理和恢复，以及报告生成。Distcp工具把文件和目录的列表作为map任务的输入，每个任务会完成源列表中部分文件的拷贝。Distcp可以限制带宽，使用带宽参数bandwidth对Distcp的每个map任务限流，同时控制map并发数量即可控制整个拷贝任务的带宽，防止拷贝任务将带宽打满，影响其它业务以及数据的传输。

[0068] 4、Cipher命令：Cipher命令可用于对数据进行加解密，Cipher支持4种加密算法：数据加密标准(Data Encryption Standard，DES)，DES3，高级加密标准AES(Advanced

Encryption Standard)和RSA。DES加密是一种对称加密方式，加密时采用同一个SECRET_KEY(密钥)。RSA加密是一种非对称加密方式，即PK(PUBLIC_KEY公钥)与SK(SECRET_KEY密钥)不是同一个。RSA加密方法速度慢，适合对少量数据加密的加密场景。

[0069] 5、高级加密标准AES：也叫共享密钥，是一种对称加密算法。对称加密算法是指加密和解密都是用同一个密钥。通常来说，对称加密算法效率要优于非对称加密算法，它用来代替DES(Data Encryption Standard，56位密钥)。AES有三个关键点：密钥、填充、模式。密钥分为128位(16字节)、192位(24字节)、256位(32字节)，位数越多，解密和加密的运算量越大，相应的越安全，可以折中使用192位兼顾效率和安全。在加密时，会将数据按照128位(16字节)一组分为多个明文块，然后对明文块分别加密。可以理解的是，加密时与解密时的填充必须一致，否则无法解密。

[0070] 现有技术中，通过Hadoop对数据进行传输时，是以普通数据流进行传输的。在传输过程中，如果遭到非法攻击，很容易导致数据泄露，无法保证数据的安全性。基于此，本申请提供了一种数据传输方法，可以保证数据传输过程中的安全性。

[0071] 本申请实施例提供的数据传输方法的使用场景如图1所示，包括第一分布式文件系统和第二分布式文件系统。第一分布式文件系统与第二分布式文件系统均为Hadoop分布式文件系统。第一分布式文件系统与第二分布式文件系统可以分别部署在由多个服务器构成的集群中。第一分布式文件系统与第二分布式文件系统可以部署distcp工具。当第一分布式文件系统中的服务器1请求拷贝第二分布式文件系统中的数据文件1时，可以通过distcp工具根据数据文件1对应的密钥1对数据文件1进行加密获得加密后的数据文件1，进而向第一分布式文件系统中的该服务器发送加密后的数据文件1。该服务器在接收到加密后的数据文件1后，可以根据该数据文件1的密钥1对加密后的数据文件1解密，进而获得该数据文件1。

[0072] 本申请实施例所适用的系统架构可以包括多个分布式文件系统，每个分布式文件系统包括一个或多个服务器。不同的分布式文件系统中服务器的数量可以相同，也可以不同，本申请对此不作具体限定。服务器可以为本地的服务器或者云服务器。服务器可以通过实体服务器实现，也可以通过虚拟服务器实现。服务器可以通过单个服务器实现，可以通过多个服务器组成的服务器集群实现，可以通过单个服务器或者服务器集群来实现本申请提供的数据传输方法。以系统架构包括第一分布式文件系统和第二分布式文件系统为例，每个分布式系统包括N个服务器100为例，如图2所示。以第一分布式文件系统为例，服务器100可以与第一分布式文件系统或第二分布式文件系统中的服务器相连，获取服务器发送的数据和请求。服务器100还可以与电子设备进行相连，用于接收电子设备发送的数据文件信息。电子设备也可以是个人计算机、移动设备(比如移动电话、平板电脑、个人数字助理)等等。

[0073] 作为一种举例，参见图3所示，服务器100可以包括处理器110、通信接口120。服务器100还可以包括存储器130。当然电子设备中还可以包括其它的组件，图3中未示出。

[0074] 通信接口120用于与其他服务器或者电子设备进行通信，用于接收其他服务器发送的数据文件或者请求，或者像其他服务器发送数据文件或者请求。一些场景中，通信接口120还可以用于接收电子设备发送的数据文件信息，或者向电子设备发送数据文件信息。

[0075] 处理器110是电子设备的控制中心，利用各种接口和路线连接电子设备的各个部

分，通过运行或执行存储在存储器130内的软件程序/或模块，以及调用存储在存储器130内的数据，执行电子设备的各种功能和处理数据。可选地，处理器110可以包括一个或多个处理单元。处理器110，例如可以是处理器、微处理器、控制器等控制组件，例如可以是通用中央处理器(central processing unit,CPU)，通用处理器，数字信号处理(digital signal processing,DSP)，专用集成电路(application specific integrated circuits,ASIC)，现场可编程门阵列(field programmable gate array,FPGA)或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。

[0076] 存储器130可用于存储软件程序以及模块，处理器110通过运行存储在存储器130的软件程序以及模块，从而执行各种功能应用以及数据处理。存储器130可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序等；存储数据区可存储根据业务处理所创建的数据等。存储器130作为一种非易失性计算机可读存储介质，可用于存储非易失性软件程序、非易失性计算机可执行程序以及模块。存储器130可以包括至少一种类型的存储介质，例如可以包括闪存、硬盘、多媒体卡、卡型存储器、随机访问存储器(Random Access Memory,RAM)、静态随机访问存储器(Static Random Access Memory,SRAM)、可编程只读存储器(Programmable Read Only Memory,PROM)、只读存储器(Read Only Memory,ROM)、带电可擦除可编程只读存储器(Electrically Erasable Programmable Read-Only Memory,EEPROM)、磁性存储器、磁盘、光盘等等。存储器130是能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质，但不限于此。本申请实施例中的存储器130还可以是电路或者其它任意能够实现存储功能的装置，用于存储程序指令和/或数据。

[0077] 需要说明的是，上述图2-3所示的结构仅是一种示例，本发明实施例对此不做限定。

[0078] 本申请实施例提供了一种数据传输方法，图4示例性地示出了数据传输方法的流程，该流程可以通过服务器100执行，具体还可以通过服务器100中的处理器110执行。具体流程如下：

[0079] 401，接收第一服务器发送的第一请求。

[0080] 其中，第一服务器属于第一分布式文件系统，第一请求用于请求拷贝第二分布式文件系统中的第一数据文件。第一分布式文件系统和第二分布式文件系统均为Hadoop分布式文件系统。一些实施例中，Hadoop分布式文件系统可以部署在多个服务器上。

[0081] 一些实施例中，所述第一请求包括第一数据文件的目录信息，还可以包括第一服务器用于接收第一数据文件并进行存储的存储信息或者目录信息。当接收到第一服务器发送的第一请求后，可以对第一请求进行解析，确定请求拷贝的第二分布式文件系统中的第一数据文件的存储信息或者目录信息。一些场景中，当第二分布式文件系统只部署在一个服务器上时，可以通过第一请求获取第一数据文件的目录信息，以对第一数据文件进行拷贝。另一些场景中，当第二分布式文件系统部署在多个服务器上时，可以通过第一请求获取第一数据文件的存储信息以及目录信息。通过存储信息确定所述第一数据文件所在的服务器，通过目录信息确定第一数据文件在服务器上的存储位置。

[0082] 402，使用第一数据文件对应的密钥对第一数据文件进行加密得到加密后的第一数据文件。

[0083] 一些实施例中，第一分布式文件系统和第二分布式文件系统均部署Distcp工具，在使用第一数据文件对应的密钥对第一数据文件进行加密得到加密后的第一数据文件时，可以通过如下方式实现：通过Distcp工具使用第一数据文件对应的密钥对第一数据文件进行加密得到加密后的第一数据文件。具体地，可以通过Distcp工具调用cipher命令使用第一数据文件对应的密钥对第一数据文件进行加密得到加密后的第一数据文件。在通过调用cipher命令使用第一数据文件对应的密钥对第一数据文件进行加密得到加密后的第一数据文件之前，确定cipher命令包括加密参数，加密参数用于指示第一数据文件在传输之前需要执行加密操作。作为一种举例，可以通过encryption加密参数对第一数据文件进行加密。当需要对第一数据文件进行加密时，可以通过设置encryption加密参数使用密钥对第一数据文件进行加密。一些场景中，加密算法可以使用AES加密方法，相比于其它非对称加密算法，AES加密方法具有更高的加密效率。

[0084] 一些实施例中，用于可以预先设置任务数参数以及带宽参数。任务数参数可以表示为-m参数，用于表示同步启动的任务数。带宽参数可以通过-bandwidth参数进行设置，带宽参数用于指示所述第一数据文件的传输带宽。在数字设备中，频宽通常以bps表示，即每秒可传输的位数。本申请实施例中，每个任务均对应一个相应的带宽。-m×-bandwidth用于表示数据传输过程中的总带宽。通过使用bandwidth参数对每个map任务进行限流，同时控制map并发数量，进而可以控制拷贝任务的带宽，防止拷贝任务将带宽打满，影响其它业务。在通过Distcp工具调用cipher命令使用第一数据文件对应的密钥对第一数据文件进行加密得到加密后的第一数据文件时，可以通过如下方式实现：获取第一数据文件的带宽参数，调用cipher命令使用第一数据文件对应的密钥对第一数据，按照带宽参数所指示的传输带宽进行加密，以得到加密后的第一数据文件。

[0085] 403，向第一服务器发送加密后的第一数据文件。

[0086] 基于上述方法，对distcp工具进行改造，实现在传输过程中对数据文件进行加密，可以保证数据传输过程中数据的安全性。此外，该方案还可以实现在不同网络下按需控制传输速度，进而控制数据的加解密速度，保障了在不同网络环境下整个数据加解密过程的稳定性。

[0087] 本申请实施例还提供了另一种数据传输的方法，图5示例性地示出了数据传输方法的流程，该流程可以通过第一分布式文件系统中的服务器100执行，具体还可以通过服务器100中的处理器110执行。具体流程如下：

[0088] 501，向第二分布式文件系统发送第一请求。

[0089] 其中，第一请求用于请求拷贝第二分布式文件系统中的第一数据文件。第一分布式文件系统与第二分布式文件系统均为Hadoop分布式文件系统，第一分布式文件系统与第二分布式文件系统均部署Distcp工具。

[0090] 502，接收第二分布式文件系统发送的经加密的第一数据文件。

[0091] 503，使用第一数据文件对应的密钥对经加密的第一数据文件进行解密得到第一数据文件。

[0092] 一些实施例中，在使用第一数据文件对应的密钥对经加密的第一数据文件进行解密得到第一数据文件时，可以通过如下方式实现：通过Distcp工具使用第一数据文件对应的密钥对经加密的第一数据文件进行解密得到加密后的第一数据文件。具体地，可以通过

Distcp工具调用cipher命令使用第一数据文件对应的密钥对经加密的第一数据文件进行解密得到第一数据文件。在通过调用cipher命令使用第一数据文件对应的密钥对经加密的第一数据文件进行解密得到第一数据文件之前，确定cipher命令包括解密参数，解密参数用于指示第一数据文件在传输之后需要执行解密操作。作为一种举例，可以通过decode解密参数对经加密的第一数据文件进行解密。当需要对经加密的第一数据文件进行解密时，可以通过设置decode解密参数使用密钥对经加密的第一数据文件进行解密，以获得第一数据文件。

[0093] 基于相同的技术构思，本申请实施例提供了一种数据传输方法的装置600。如图6所示。该装置600可以实现上述图4所述的数据传输方法中的任一步骤，为了避免重复，此处不再赘述。该装置600包括接收模块601、处理模块602和发送模块603。

[0094] 接收模块601，用于接收第一服务器发送的第一请求，所述第一服务器属于第一分布式文件系统，所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件；

[0095] 处理模块602，用于使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件；

[0096] 发送模块603，用于向所述第一服务器发送所述加密后的第一数据文件。

[0097] 一些实施例中，所述第一分布式文件系统和所述第二分布式文件系统均部署Distcp工具，所述处理模块602，在使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件时，具体用于：

[0098] 通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

[0099] 一些实施例中，所述处理模块602，在通过Distcp工具使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件时，具体用于：

[0100] 通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件。

[0101] 一些实施例中，所述第一分布式文件系统和所述第二分布式文件系统均为hadoop分布式文件系统。

[0102] 另一些实施例中，所述处理模块602还用于：在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件之前，确定所述cipher中包括加密参数，所述加密参数用于指示所述第一数据文件在传输之前需要执行加密操作。

[0103] 一些实施例中，所述处理模块602还用于：确定获取所述第一数据文件对应的带宽参数，所述带宽参数用于指示所述第一数据文件的传输带宽；在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件进行加密得到加密后的第一数据文件时，具体用于：

[0104] 调用cipher命令使用所述第一数据文件对应的密钥对所述第一数据文件，按照所述带宽参数所指示的传输带宽进行加密，以得到加密后的第一数据文件。

[0105] 基于相同的技术构思，本申请实施例提供了另一种数据传输方法的装置700。如图7所示。该装置700可以实现上述图5所述的数据传输方法中的任一步骤，为了避免重复，此处不再赘述。该装置700包括发送模块701、接收模块702和处理模块703。

[0106] 发送模块701,用于向第二分布式文件系统发送第一请求,所述第一请求用于请求拷贝第二分布式文件系统中的第一数据文件;

[0107] 接收模块702,用于接收第二分布式文件系统发送的经加密的第一数据文件;

[0108] 处理模块703,用于使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0109] 一种可能的实现方式中,所述第一分布式文件系统和所述第二分布式文件系统均部署Distcp工具,所述处理模块703,在使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件时,具体用于:通过Distcp工具使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0110] 一种可能的实现方式中,所述处理模块703,在通过Distcp工具使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件时,具体用于:

[0111] 通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件。

[0112] 一种可能的实现方式中,所述第一分布式文件系统和所述第二分布式文件系统均为hadoop分布式文件系统。

[0113] 一种可能的实现方式中,所述处理模块703还用于:在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件之前,确定所述cipher中包括解密参数,所述解密参数用于指示所述经加密的第一数据文件在传输之后需要执行解密操作。

[0114] 一种可能的实现方式中,所述处理模块703还用于:确定获取所述第一数据文件对应的带宽参数,所述带宽参数用于指示所述第一数据文件的传输带宽;所述处理模块,在通过所述Distcp工具调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件进行解密得到所述第一数据文件时,具体用于:调用cipher命令使用所述第一数据文件对应的密钥对所述经加密的第一数据文件,按照所述带宽参数所指示的传输带宽进行解密,以得到所述第一数据文件。

[0115] 基于相同的技术构思,本申请实施例提供了一种数据传输装置800,参见图8所示。该装置800包括存储器801和处理器802。

[0116] 存储器801,用于存储程序指令;

[0117] 处理器802,用于调用所述存储器中存储的程序指令,按照获得的程序执行上述数据传输方法。

[0118] 在本申请实施例中,处理器802可以是通用处理器、数字信号处理器、专用集成电路、现场可编程门阵列或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件,可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件处理器执行完成,或者用处理器中的硬件及软件模块组合执行完成。

[0119] 存储器801作为一种非易失性计算机可读存储介质,可用于存储非易失性软件程序、非易失性计算机可执行程序以及模块。存储器801可以包括至少一种类型的存储介质,例如可以包括闪存、硬盘、多媒体卡、卡型存储器、随机访问存储器(Random Access

Memory,RAM)、静态随机访问存储器(Static Random Access Memory,SRAM)、可编程只读存储器(Programmable Read Only Memory,PROM)、只读存储器(Read Only Memory,ROM)、带电可擦除可编程只读存储器(Electrically Erasable Programmable Read-Only Memory,EEPROM)、磁性存储器、磁盘、光盘等等。存储器801是能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质,但不限于此。本申请实施例中的存储器801还可以是电路或者其它任意能够实现存储功能的装置,用于存储程序指令和/或数据。

[0120] 基于相同的技术构思,本申请实施例提供了一种计算机可读存储介质,所述计算机可读存储介质中存储有指令,当其在计算机上运行时,使得计算机执行上述数据传输方法。

[0121] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0122] 本申请是参照根据本申请的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0123] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0124] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0125] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

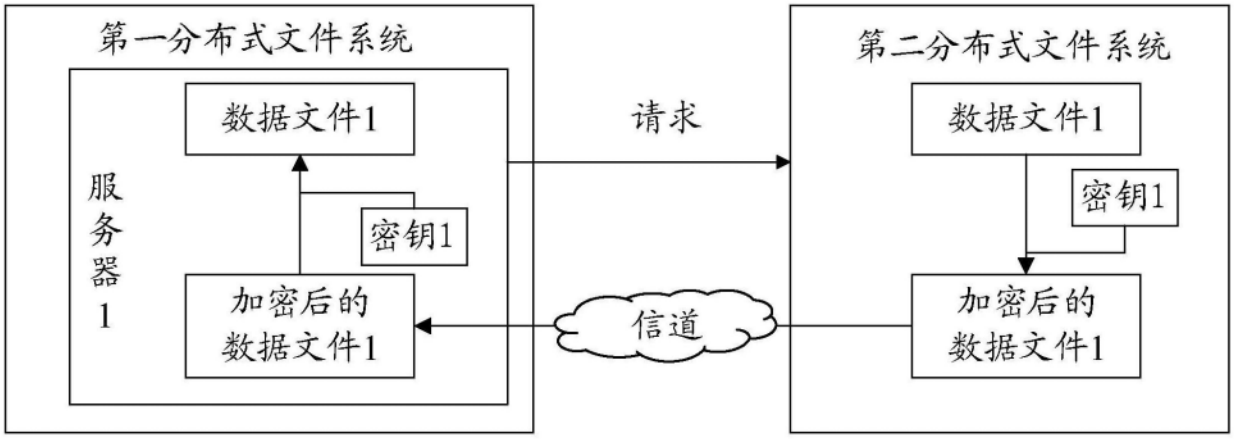

第一分布式文件系统
数据文件1
服务器1
密钥1
加密后的数据文件1
请求
信道
第二分布式文件系统
数据文件1
密钥1
加密后的数据文件1

图1

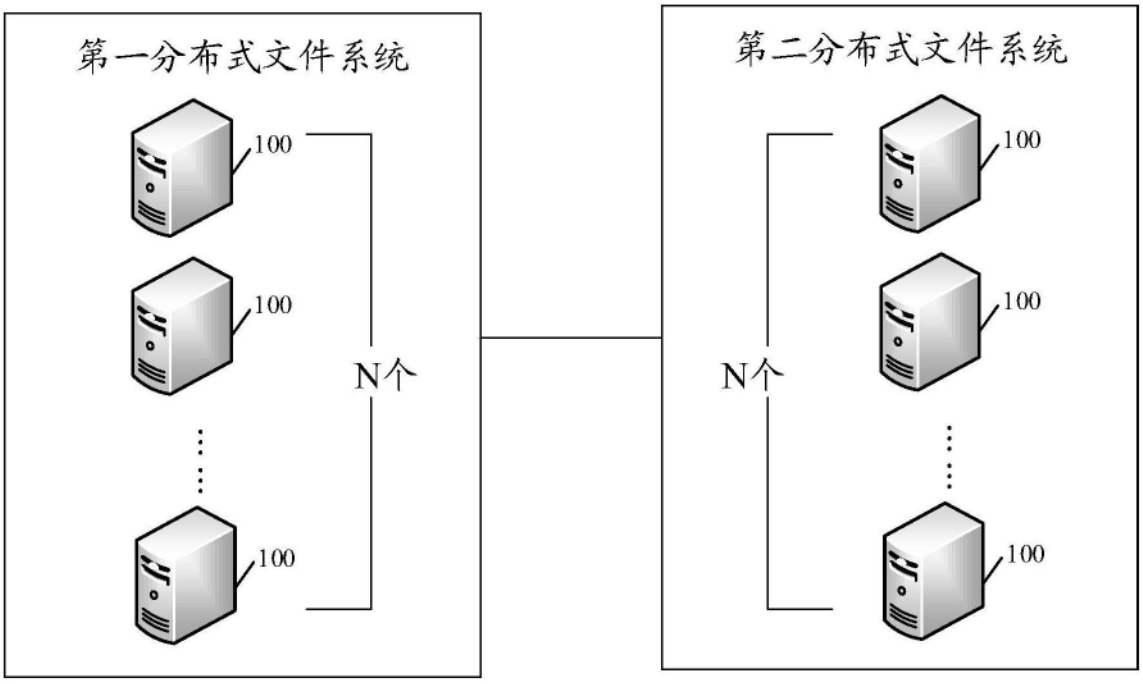

第一分布式文件系统
100
100
100
N个
第二分布式文件系统
100
100
100
N个

图2

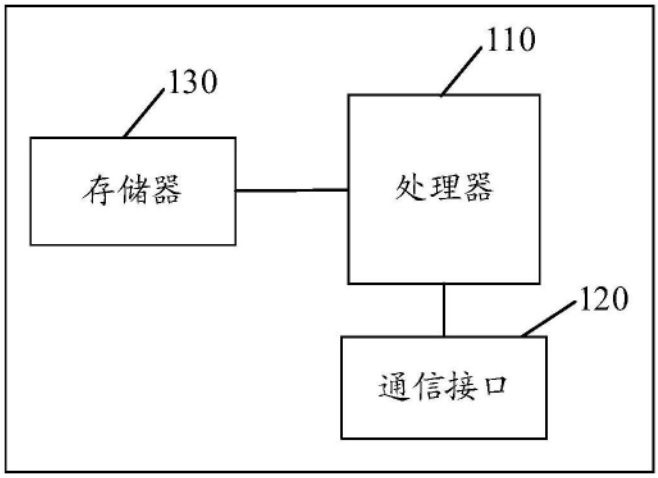

110
130
存储器
处理器
120
通信接口

图3

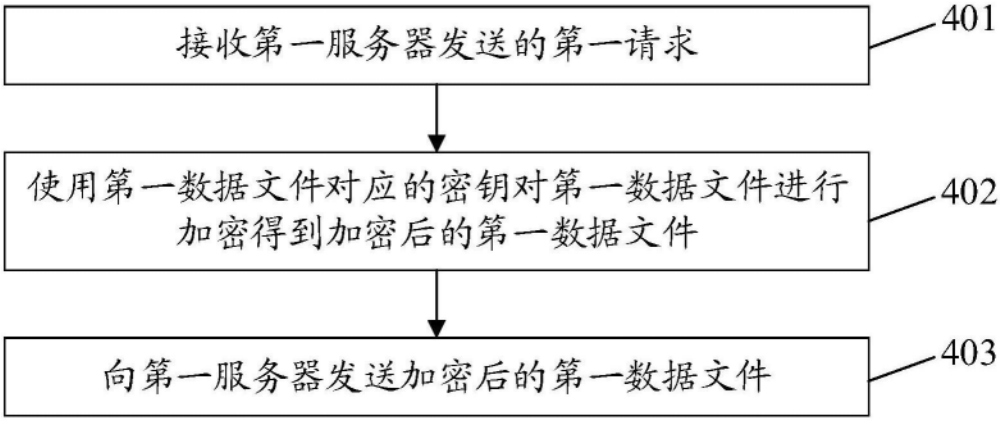

接收第一服务器发送的第一请求
401
使用第一数据文件对应的密钥对第一数据文件进行加密得到加密后的第一数据文件
402
向第一服务器发送加密后的第一数据文件
403

图4

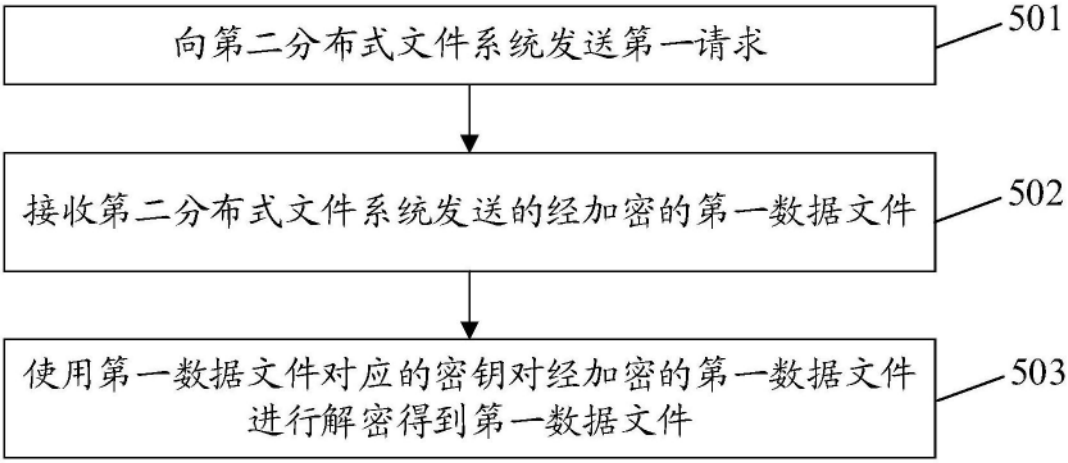

向第二分布式文件系统发送第一请求
501
接收第二分布式文件系统发送的经加密的第一数据文件
502
使用第一数据文件对应的密钥对经加密的第一数据文件进行解密得到第一数据文件
503

图5

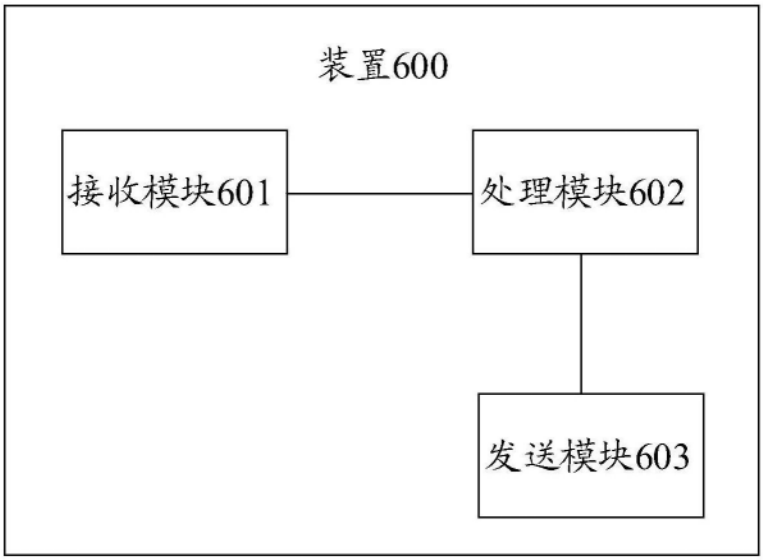

装置600
接收模块601
处理模块602
发送模块603

图6

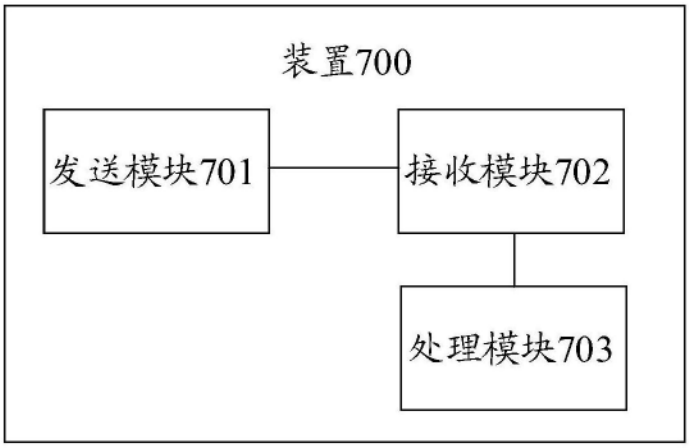

装置700
发送模块701
接收模块702
处理模块703

图7

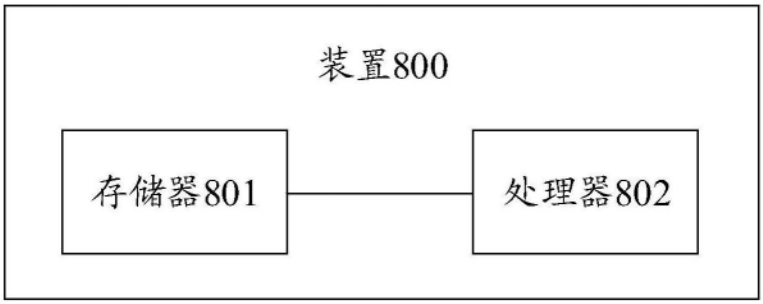

装置800
存储器801
处理器802

图8