



(12) 发明专利申请

(10) 申请公布号 CN 115129669 A

(43) 申请公布日 2022. 09. 30

(21) 申请号 202210798969.2

(22) 申请日 2022.07.08

(71) 申请人 中国建设银行股份有限公司
地址 100033 北京市西城区金融大街25号

(72) 发明人 夏松 郭玉章

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 王云晓

(51) Int.Cl.
G06F 16/16 (2019.01)
G06K 9/62 (2022.01)

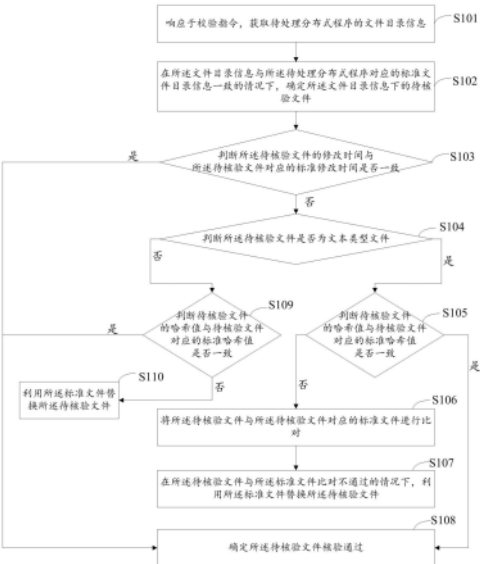
权利要求书2页 说明书9页 附图7页

(54) 发明名称

文件处理方法及装置、存储介质及电子设备

(57) 摘要

本发明提供了一种文件处理方法及装置、存储介质及电子设备,该方法包括:响应于校验指令,获取待处理分布式程序的文件目录信息;在文件目录信息与待处理分布式程序对应的标准文件目录信息一致的情况下,确定文件目录信息下的待核验文件;若待核验文件的修改时间与所述待核验文件对应的标准修改时间不一致、待核验文件为文本类型文件且待核验文件的校验值与标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对;在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。应用本发明提供的方法,能够快速准确地处理发生不一致异常的文件。



1. 一种文件处理方法,其特征在于,包括:

响应于校验指令,获取待处理分布式程序的文件目录信息;

在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件;

判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;

在所述待核验文件的修改时间与所述标准修改时间不一致的情况下,判断所述待核验文件是否为文本类型文件;

在所述待核验文件为文本类型文件的情况下,判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;

在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对;

在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。

2. 根据权利要求1所述的方法,其特征在于,所述判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致,包括:

在所述待核验文件满足预设的分片条件的情况下,对所述待核验文件进行切分,获得各个分片;计算出每一所述分片的分片校验值,并判断每一所述分片的分片校验值与预先记录的所述分片对应的标准分片校验值是否一致;

在所述待核验文件不满足预设的分片条件的情况下,则计算所述待核验文件的文件校验值,并判断所述待核验文件的文件校验值与预先记录的所述待核验文件对应的标准文件校验值是否一致。

3. 根据权利要求1所述的方法,其特征在于,所述将所述待核验文件与所述待核验文件对应的标准文件进行比对,包括:

分别对所述待核验文件以及所述标准文件进行预处理,以剔除所述待核验文件以及所述标准文件中的无效信息;

计算出预处理后的所述待核验文件与预处理后的所述标准文件之间的匹配度;

在所述匹配度满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对通过;

在所述匹配度不满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对不通过。

4. 根据权利要求1所述的方法,其特征在于,还包括:

在所述待核验文件不为文本类型文件的情况下,若待核验文件的校验值与所述待核验文件对应的标准校验值不一致,则利用所述待核验文件对应的标准文件替换所述待核验文件。

5. 根据权利要求1所述的方法,其特征在于,还包括:

在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息不一致的情况下,确定所述文件目录信息异常,并输出所述文件目录信息中与所述标准文件目标信息不一致的部分。

6. 一种文件处理装置,其特征在于,包括:

获取单元,用于响应于校验指令,获取待处理分布式程序的文件目录信息;

确定单元,用于在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件;

第一判断单元,用于判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;

第二判断单元,用于在所述待核验文件的修改时间与所述标准修改时间不一致的情况下,判断所述待核验文件是否为文本类型文件;

第三判断单元,用于在所述待核验文件为文本类型文件的情况下,判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;

比对单元,用于在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对;

替换单元,用于在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。

7. 根据权利要求6所述的装置,其特征在于,所述第三判断单元,包括:

第一执行子单元,用于在所述待核验文件满足预设的分片条件的情况下,对所述待核验文件进行切分,获得各个分片;计算出每一所述分片的分片校验值,并判断每一所述分片的分片校验值与预先记录的所述分片对应的标准分片校验值是否一致;

第二执行子单元,用于在所述待核验文件不满足预设的分片条件的情况下,则计算所述待核验文件的文件校验值,并判断所述待核验文件的文件校验值与预先记录的所述待核验文件对应的标准文件校验值是否一致。

8. 根据权利要求6所述的装置,其特征在于,所述比对单元,包括:

预处理子单元,用于分别对所述待核验文件以及所述标准文件进行预处理,以剔除所述待核验文件以及所述标准文件中的无效信息;

计算子单元,用于计算出预处理后的所述待核验文件与预处理后的所述标准文件之间的匹配度;

第一确定子单元,用于在所述匹配度满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对通过;

第二确定子单元,用于在所述匹配度不满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对不通过。

9. 一种存储介质,其特征在于,所述存储介质包括存储指令,其中,在所述指令运行时控制所述存储介质所在的设备执行如权利要求1~5任意一项所述的文件处理方法。

10. 一种电子设备,其特征在于,包括存储器,以及一个或者一个以上的指令,其中一个或一个以上指令存储于存储器中,且经配置以由一个或者一个以上处理器执行如权利要求1~5任意一项所述的文件处理方法。

文件处理方法及装置、存储介质及电子设备

技术领域

[0001] 本发明涉及数据处理技术领域,特别涉及一种文件处理方法及装置、存储介质及电子设备。

背景技术

[0002] 随着计算机科学的快速发展,分布式技术也得到了越来越广泛的应用,数据的一致性、可靠性是分布式存储的基本要求,但由于磁盘、网络等硬件制约,故障场景频繁,导致分布式程序的部署面临着程序文件一致性的问题。

[0003] 目前,为了避免分布式程序文件不一致的问题,通常是采用信息摘要算法去比对文件的一致性,然而信息摘要算法存在哈希碰撞等一系列问题,会导致不同的数据文件产生相同的哈希散列,进而导致无法及时处理已发生不一致异常的文件,容易造成运维风险。

发明内容

[0004] 本发明所要解决的技术问题是提供一种文件处理方法,能够快速准确地处理发生不一致异常的文件。

[0005] 本发明还提供了一种文件处理装置,用以保证上述方法在实际中的实现及应用。

[0006] 一种文件处理方法,包括:

[0007] 响应于校验指令,获取待处理分布式程序的文件目录信息;

[0008] 在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件;

[0009] 判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;

[0010] 在所述待核验文件的修改时间与所述标准修改时间不一致的情况下,判断所述待核验文件是否为文本类型文件;

[0011] 在所述待核验文件为文本类型文件的情况下,判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;

[0012] 在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对;

[0013] 在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。

[0014] 上述的方法,可选的,所述判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致,包括:

[0015] 在所述待核验文件满足预设的分片条件的情况下,对所述待核验文件进行切分,获得各个分片;计算出每一所述分片的分片校验值,并判断每一所述分片的分片校验值与预先记录的所述分片对应的标准分片校验值是否一致;

[0016] 在所述待核验文件不满足预设的分片条件的情况下,则计算所述待核验文件的文

件校验值,并判断所述待核验文件的文件校验值与预先记录的所述待核验文件对应的标准文件校验值是否一致。

[0017] 上述的方法,可选的,所述将所述待核验文件与所述待核验文件对应的标准文件进行比对,包括:

[0018] 分别对所述待核验文件以及所述标准文件进行预处理,以剔除所述待核验文件以及所述标准文件中的无效信息;

[0019] 计算出预处理后的所述待核验文件与预处理后的所述标准文件之间的匹配度;

[0020] 在所述匹配度满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对通过;

[0021] 在所述匹配度不满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对不通过。

[0022] 上述的方法,可选的,还包括:

[0023] 在所述待核验文件不为文本类型文件的情况下,若待核验文件的校验值与所述待核验文件对应的标准校验值不一致,则利用所述待核验文件对应的标准文件替换所述待核验文件。

[0024] 上述的方法,可选的,还包括:

[0025] 在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息不一致的情况下,确定所述文件目录信息异常,并输出所述文件目录信息中与所述标准文件目标信息不一致的部分。

[0026] 一种文件处理装置,包括:

[0027] 获取单元,用于响应于校验指令,获取待处理分布式程序的文件目录信息;

[0028] 确定单元,用于在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件;

[0029] 第一判断单元,用于判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;

[0030] 第二判断单元,用于在所述待核验文件的修改时间与所述标准修改时间不一致的情况下,判断所述待核验文件是否为文本类型文件;

[0031] 第三判断单元,用于在所述待核验文件为文本类型文件的情况下,判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;

[0032] 比对单元,用于在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对;

[0033] 替换单元,用于在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。

[0034] 上述的装置,可选的,所述第三判断单元,包括:

[0035] 第一执行子单元,用于在所述待核验文件满足预设的分片条件的情况下,对所述待核验文件进行切分,获得各个分片;计算出每一所述分片的分片校验值,并判断每一所述分片的分片校验值与预先记录的所述分片对应的标准分片校验值是否一致;

[0036] 第二执行子单元,用于在所述待核验文件不满足预设的分片条件的情况下,则计算所述待核验文件的文件校验值,并判断所述待核验文件的文件校验值与预先记录的所述

待核验文件对应的标准文件校验值是否一致。

[0037] 上述的装置,可选的,所述比对单元,包括:

[0038] 预处理子单元,用于分别对所述待核验文件以及所述标准文件进行预处理,以剔除所述待核验文件以及所述标准文件中的无效信息;

[0039] 计算子单元,用于计算出预处理后的所述待核验文件与预处理后的所述标准文件之间的匹配度;

[0040] 第一确定子单元,用于在所述匹配度满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对通过;

[0041] 第二确定子单元,用于在所述匹配度不满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对不通过。

[0042] 一种存储介质,所述存储介质包括存储指令,其中,在所述指令运行时控制所述存储介质所在的设备执行如上述的文件处理方法。

[0043] 一种电子设备,包括存储器,以及一个或者一个以上的指令,其中一个或一个以上指令存储于存储器中,且经配置以由一个或者一个以上处理器执行如上述的文件处理方法。

[0044] 与现有技术相比,本发明包括以下优点:

[0045] 本发明提供了一种文件处理方法及装置、存储介质及电子设备,该方法包括:响应于校验指令,获取待处理分布式程序的文件目录信息;在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件;判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;在所述待核验文件的修改时间与所述标准修改时间不一致的情况下,判断所述待核验文件是否为文本类型文件;在所述待核验文件为文本类型文件的情况下,判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对;在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。应用本发明实施例提供的方法,能够快速准确地处理发生不一致异常的文件。

附图说明

[0046] 为了更清楚地说明本发明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,下面描述中的附图仅仅是本发明的实施例,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,还可以根据提供的附图获得其他的附图。

[0047] 图1为本发明提供的一种文件处理方法的方法流程图;

[0048] 图2为本发明提供的一种将待核验文件与标准文件进行比对的过程的流程图;

[0049] 图3为本发明提供的一种文件处理过程的流程图;

[0050] 图4为本发明提供的一种收集标准信息的过程的流程图;

[0051] 图5为本发明提供的一种初始化信息校验过程的流程图;

[0052] 图6为本发明提供的一种日常信息校验过程的流程图;

[0053] 图7为本发明提供的一种文件处理装置的结构示意图;

[0054] 图8为本发明提供的一种电子设备的结构示意图。

具体实施方式

[0055] 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,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。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前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0056] 在本申请中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0057] 本发明实施例提供了一种文件处理方法,该方法可以应用于电子设备,所述方法的方法流程图如图1所示,具体包括:

[0058] S101:响应于校验指令,获取待处理分布式程序的文件目录信息。

[0059] 在本实施例中,该待处理分布式程序可以为预先部署好的各个分布式程序中的一个或多个。

[0060] 可选的,每个待处理分布式程序具有一个文件目录信息。

[0061] S102:在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件。

[0062] 在本实施例中,可以将文件目标目录信息与标准文件目录信息进行比对,该标准文件目录信息可以是预先收集的,文件目录信息下的待核验文件的数量为一个或多个,标准文件目录信息记录于预设的配置文件中。

[0063] S103:判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;若否,则执行S104,若是,则执行S108。

[0064] 在本实施例中,修改时间可以是待核验文件最后一次发生变动的时间,标准修改时间可以是预先收集的,标准修改时间记录于预设的配置文件中。

[0065] S104:判断所述待核验文件是否为文本类型文件;若是,则执行S105;若否,则执行S109。

[0066] 在本实施例中,待核验文件可以是文本类型文件或非文本类型文件中的一种,可以根据待核验文件的文件后缀或者文件内容确定待核验文件是否为文本类型文件。

[0067] S105:判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;若否,则执行S106,若是,则执行S108。

[0068] 在本实施例中,可以计算出待核验文件的校验值,计算待核验的校验值时所采用的算法可以是但不限于MD5(Message-DigestAlgorithm 5,信息-摘要算法5)等信息摘要算法。

[0069] 可选的,待核验文件对应的标准校验值可以是预先收集的;在待核验文件为本类型文件的情况下,其对应的标准校验值可以是各个标准分片校验值,或者是标准文件校验值,标准校验值记录于预设的配置文件中。

[0070] S106:在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对。

[0071] 在本实施例中,可以将待核验文件与待核验文件对应的标准文件进行文本强一致性比对。

[0072] S107:在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。

[0073] 在本实施例中,在待核验文件与标准文件比对不通过的情况下,可以利用标准文件替换待核验文件,具体可以将标准文件覆盖待核验文件。

[0074] S108:确定所述待核验文件核验通过。

[0075] S109:判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致,若是,则执行S108;若否,则执行S110。

[0076] S110:利用待核验文件对应的标准文件替换所述待核验文件。

[0077] 在本实施例中,利用标准文件替换待核验文件之后,可以在配置文件中更新该待核验文件的文件修改时间。

[0078] 应用本发明实施例提供的方法,能够快速准确地处理发生不一致异常的文件。

[0079] 在本发明提供的一实施例中,基于上述的实施过程,可选的,S105中提及的判断待核验文件的校验值与待核验文件对应的标准校验值是否一致的过程,包括:

[0080] 在所述待核验文件满足预设的分片条件的情况下,对所述待核验文件进行切分,获得各个分片;计算出每一所述分片的分片校验值,并判断每一所述分片的校验值与预先记录的所述分片对应的标准分片校验值是否一致;

[0081] 在所述待核验文件不满足预设的分片条件的情况下,则计算所述待核验文件的文件校验值,并判断所述待核验文件的文件校验值与预先记录的所述待核验文件对应的标准文件校验值是否一致。

[0082] 在本实施例中,分片条件可以是待核验文件所占的存储空间容量大于预设的阈值,在待核验文件满足分片条件的情况下,可以按预设的分片方式对待核验文件进行切分,得到多个分片。

[0083] 可选的,在至少一个分片的校验值与该分片对应的标准分片校验值不一致的情况下,确定待核验文件的校验值与待核验文件对应的标准校验值一致,在所有分片校验值均与其对应的标准分片校验值一致的情况下,确定待核验文件的校验值与待核验文件对应的标准校验值不一致。

[0084] 可选的,在待核验文件的文件校验值与待核验文件对应的标准文件校验值的情况下,确定待核验文件的校验值与标准文件的校验值一致。

[0085] 在本发明提供的一实施例中,基于上述的实施过程,可选的,所述将所述待核验文件与所述待核验文件对应的标准文件进行比对的过程,如图2所示,可以包括:

[0086] S201:分别对所述待核验文件以及所述标准文件进行预处理,以剔除所述待核验文件以及所述标准文件中的无效信息。

[0087] 在本实施例中,无效信息可以无效换行和无效空格等其中至少一种。

[0088] S202:计算出预处理后的所述待核验文件与预处理后的所述标准文件之间的匹配度。

[0089] 在本实施例中,可以采用余弦距离算法或编辑距离算法计算出待核验文件与标准文件之间的匹配度,匹配度可以表示待核验文件与标准文件之间的距离或相似程度,匹配度可以范围可以为0到1之间,越接近1,则两个文件越相似。

[0090] S203:在所述匹配度满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对通过。

[0091] 在本实施例中,该匹配度阈值可以为1,在匹配度为1的情况下,确定待核验文件与标准文件比对通过。

[0092] S204:在所述匹配度不满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对不通过。

[0093] 在本发明提供的一实施例中,基于上述的实施过程,可选的,还包括:

[0094] 在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息不一致的情况下,确定所述文件目录信息异常,并输出所述文件目录信息中与所述标准文件目标信息不一致的部分。

[0095] 本发明实施例提供的方法,可以应用于各种类型的分布式程序的程序文件的校验。在实际应用过程中,参见图3,为本发明实施例提供的一种文件处理过程的流程图,首先收集标准信息,然后初始化信息校验,再进行日常信息校验。

[0096] 在本实施例中,收集标准信息的过程如图4所示,可以在多套分布式程序中选取一套标准的程序目录,该目录为已明确的标准程序目录,将该程序目录作为基准,递归遍历该目录及该目录下所有对象,如遍历选取的对象为链接或者目录则直接跳过,如对象为文件则先判断其是否为文本文件,如果为非文本文件则直接计算其MD5散列值,如果该程序文件为文本文件则获取文件大小,当文件超过指定大小时,对文件进行切片,并记录每个分片的MD5值;标准信息可以包括文件的MD5值及文件名称及路径信息。

[0097] 在收集好标准信息后,初始化信息校验,具体过程如图5所示,可以将每一套待比较的分布式程序,与标准程序目录进行比对,链接和文件仅比对名称,文件比对其MD5的值或分片MD5值,如MD5值或分片MD5值不一致则使用标准文件覆盖掉该目标文件。当文件MD5值相同时,则将该文件的修改信息及MD5值或分片MD5值记录到本地配置文件中,供下次日常比较使用。

[0098] 进一步的,初始化信息校验后,进行日常信息校验,具体过程如图6所示,日常信息校验采用一种快速的分布式程序一致性校验方法进行分布式程序校验,按照运维人员需求的比对频率对分布式程序进行一致性校验,比较程序文件的内容。比对时,先比对目录结构,确保文件目录结构一致性,接着再比对文件的修改时间,如果文件的修改时间为该套分布式程序本地配置文件的初始化的修改时间,则直接判定认为该文件未改变,否则开始比较文件的MD5值或分片MD5值,如MD5值或者分片MD5值与本地配置文件上记录的值一致则判断文件未改变,此时仅修改本地配置文件中该文件的修改时间,如不一致,则进行文件强一致性校验,强一致性校验采用格式塔模式匹配及自动垃圾启发式算法,将标准文本文件每一行与待比较程序每一行进行处理,找到不包含垃圾元素(空行、空格)的最长连续匹配序列,进行匹配,并对整个文本计算匹配度,如匹配度为1,则判断该文件正常,仅修改本地配置文件中该文件的修改时间,否则判定文本文件自动到不包含将该文件识别成异常文件,选取标准文件进行覆盖,并输出异常信息。

[0099] 在本实施例中,为解决MD5算法Hash碰撞问题及大文本文件MD5校验速度缓慢问题,本发明提供了一种快速MD5散列算法,将超过一定大小的文本文件每N行切分成一个子文件,并提供多线程技术,使用多个线程计算并比对分片MD5文件的散列值,此种算法充分利用了多核服务器,有效的提升了MD5计算的速度,同时由于采用了切片技术,缩小了关键字集合,进而缩小了MD5校验的信息的地址集合,降低了MD5算法Hash碰撞的概率。

[0100] 在本实施例中,采用了基于格式塔模式的强一致性文本匹配算法,该算法是本实施例中使用的一种相似度计算方法,作为MD5分片算法的补充方法,主要针对文本中增加无效的换行,空格等进行垃圾进行去除比对,使用可理解的方式比对文件。

[0101] 应用本实施例提供的方法,通过基于文件修改时间、分片MD5算法、强一致性文本匹配算法的分布式程序一致性校验方法,比传统单纯的文件的MD5算法速度更快,实测使用四核服务器校验大文件时校验所消耗的时间缩小将近50%左右,且分片算法从一定程度上缩小了HASH算法输入的地址集合范围,减少了Hash碰撞的概率,并且当所有分片校验全部成功时,该算法才判定文件为相同文件,从某种程度上增加了MD5算法的可靠性,除上述优点外,还提供文本的强一致性比对,去除文本中无效换行,空格等无意义信息后,将文本文件进行比对,并计算相似度,从而确认文本是否一致。

[0102] 与图1所述的方法相对应,本发明实施例还提供了一种文件处理装置,用于对图1中方法的具体实现,本发明实施例提供的文件处理装置可以应用于电子设备中,其结构示意图如图7所示,具体包括:

[0103] 获取单元701,用于响应于校验指令,获取待处理分布式程序的文件目录信息;

[0104] 确定单元702,用于在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件;

[0105] 第一判断单元703,用于判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;

[0106] 第二判断单元704,用于在所述待核验文件的修改时间与所述标准修改时间不一致的情况下,判断所述待核验文件是否为文本类型文件;

[0107] 第三判断单元705,用于在所述待核验文件为文本类型文件的情况下,判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;

[0108] 比对单元706,用于在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文件与所述待核验文件对应的标准文件进行比对;

[0109] 替换单元707,用于在所述待核验文件与所述标准文件比对不通过的情况下,利用所述标准文件替换所述待核验文件。

[0110] 在本发明实施例提供的一实施例中,基于上述的方案,具体的,所述第三判断单元705,包括:

[0111] 第一执行子单元,用于在所述待核验文件满足预设的分片条件的情况下,对所述待核验文件进行切分,获得各个分片;计算出每一所述分片的分片校验值,并判断每一所述分片的分片校验值与预先记录的所述分片对应的标准分片校验值是否一致;

[0112] 第二执行子单元,用于在所述待核验文件不满足预设的分片条件的情况下,则计算所述待核验文件的文件校验值,并判断所述待核验文件的文件校验值与预先记录的所述待核验文件对应的标准文件校验值是否一致。

[0113] 在本发明实施例提供的一实施例中,基于上述的方案,具体的,所述比对单元706,包括:

[0114] 预处理子单元,用于分别对所述待核验文件以及所述标准文件进行预处理,以剔除所述待核验文件以及所述标准文件中的无效信息;

[0115] 计算子单元,用于计算出预处理后的所述待核验文件与预处理后的所述标准文件之间的匹配度;

[0116] 第一确定子单元,用于在所述匹配度满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对通过;

[0117] 第二确定子单元,用于在所述匹配度不满足预设的匹配度阈值的情况下,确定所述待核验文件与所述标准文件比对不通过。

[0118] 在本发明实施例提供的一实施例中,基于上述的方案,具体的,所述文件处理装置还包括:

[0119] 执行单元,用于在所述待核验文件不为文本类型文件的情况下,若待核验文件的校验值与所述待核验文件对应的标准校验值不一致,则利用所述待核验文件对应的标准文件替换所述待核验文件。

[0120] 在本发明实施例提供的一实施例中,基于上述的方案,具体的,所述文件处理装置还包括:

[0121] 输出单元,用于在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息不一致的情况下,确定所述文件目录信息异常,并输出所述文件目录信息中与所述标准文件目标信息不一致的部分。

[0122] 上述本发明实施例公开的文件处理装置中的各个单元和模块具体的原理和执行过程,与上述本发明实施例公开的文件处理方法相同,可参见上述本发明实施例提供的文件处理方法中相应的部分,这里不再赘述。

[0123] 本发明实施例还提供了一种存储介质,所述存储介质包括存储的指令,其中,在所述指令运行时控制所述存储介质所在的设备执行上述文件处理方法。

[0124] 本发明实施例还提供了一种电子设备,其结构示意图如图8所示,具体包括存储器801,以及一个或者一个以上的指令802,其中一个或者一个以上指令802存储于存储器801中,且经配置以由一个或者一个以上处理器803执行所述一个或者一个以上指令802进行以下操作:

[0125] 响应于校验指令,获取待处理分布式程序的文件目录信息;

[0126] 在所述文件目录信息与所述待处理分布式程序对应的标准文件目录信息一致的情况下,确定所述文件目录信息下的待核验文件;

[0127] 判断所述待核验文件的修改时间与所述待核验文件对应的标准修改时间是否一致;

[0128] 在所述待核验文件的修改时间与所述标准修改时间不一致的情况下,判断所述待核验文件是否为文本类型文件;

[0129] 在所述待核验文件为文本类型文件的情况下,判断所述待核验文件的校验值与所述待核验文件对应的标准校验值是否一致;

[0130] 在所述待核验文件的校验值与所述标准校验值不一致的情况下,将所述待核验文

件与所述待核验文件对应的标准文件进行比对；

[0131] 在所述待核验文件与所述标准文件比对不通过的情况下，利用所述标准文件替换所述待核验文件。

[0132] 需要说明的是，本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。对于装置类实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0133] 最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0134] 为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本发明时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0135] 通过以上的实施方式的描述可知，本领域的技术人员可以清楚地了解到本发明可借助软件加必需的通用硬件平台的方式来实现。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可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如ROM/RAM、磁碟、光盘等，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或者实施例的某些部分所述的方法。

[0136] 以上对本发明所提供的一种文件处理方法进行了详细介绍，本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

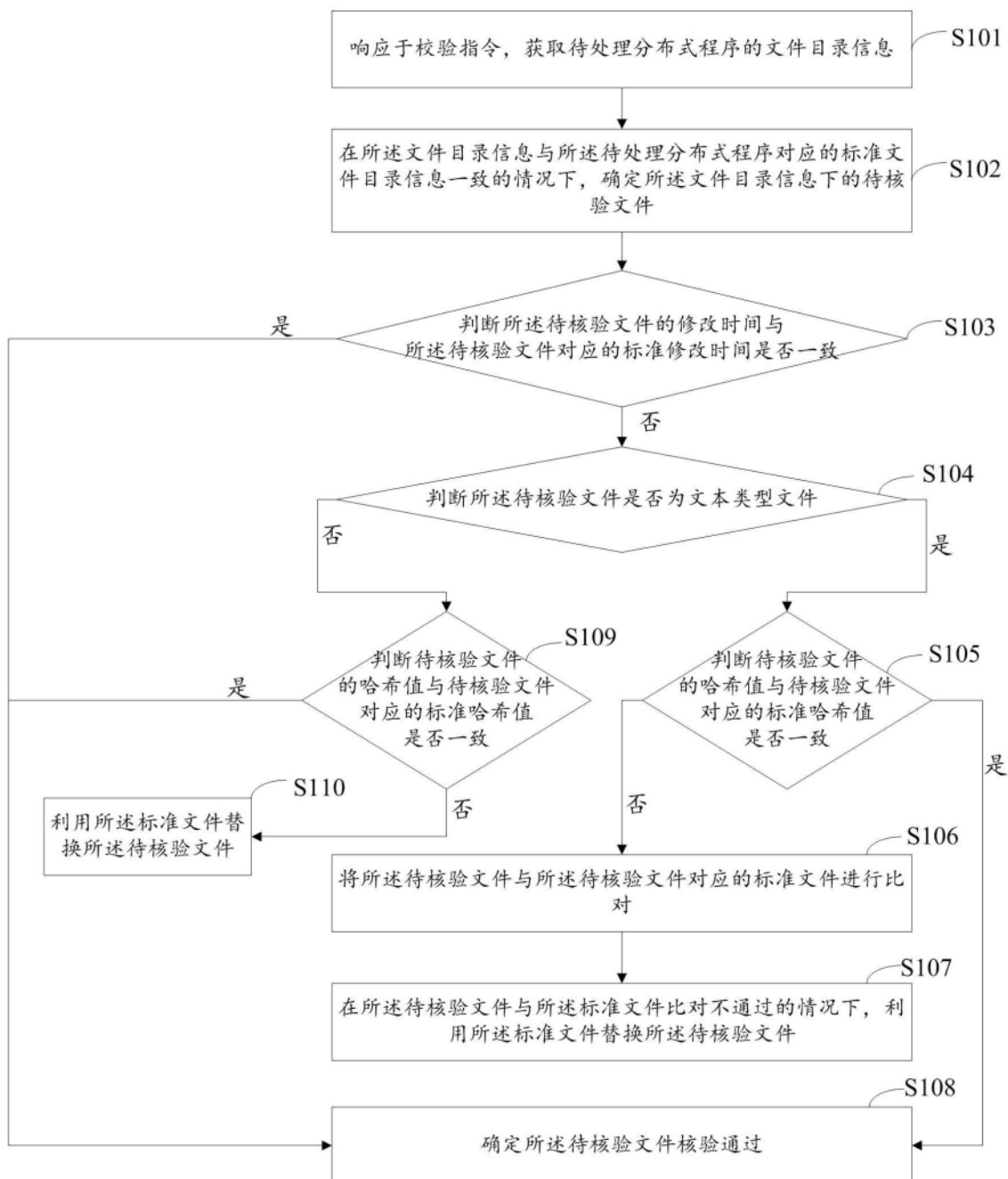


图1

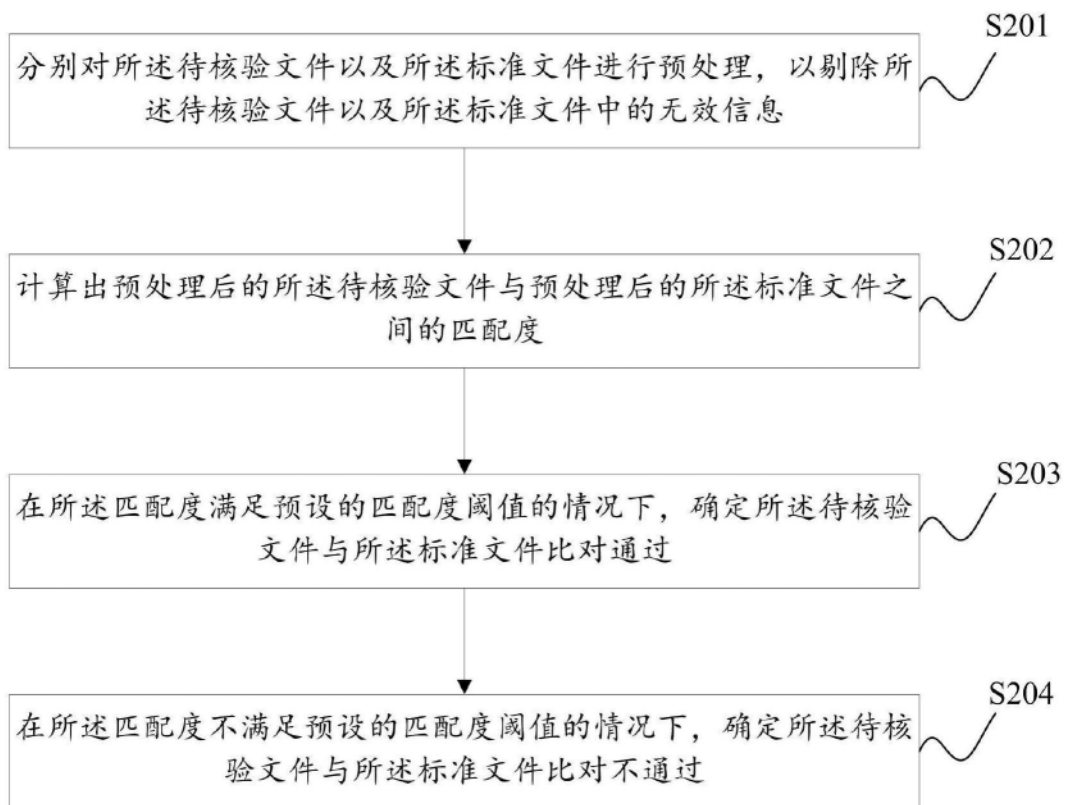


图2

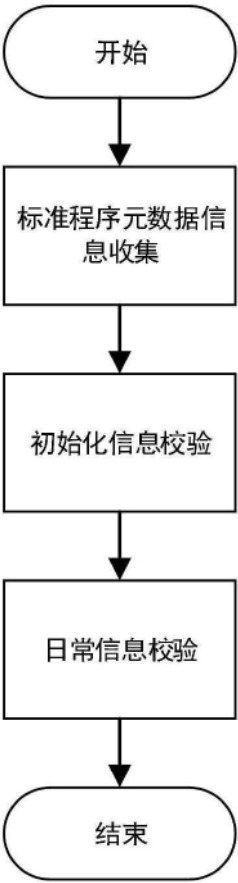


图3

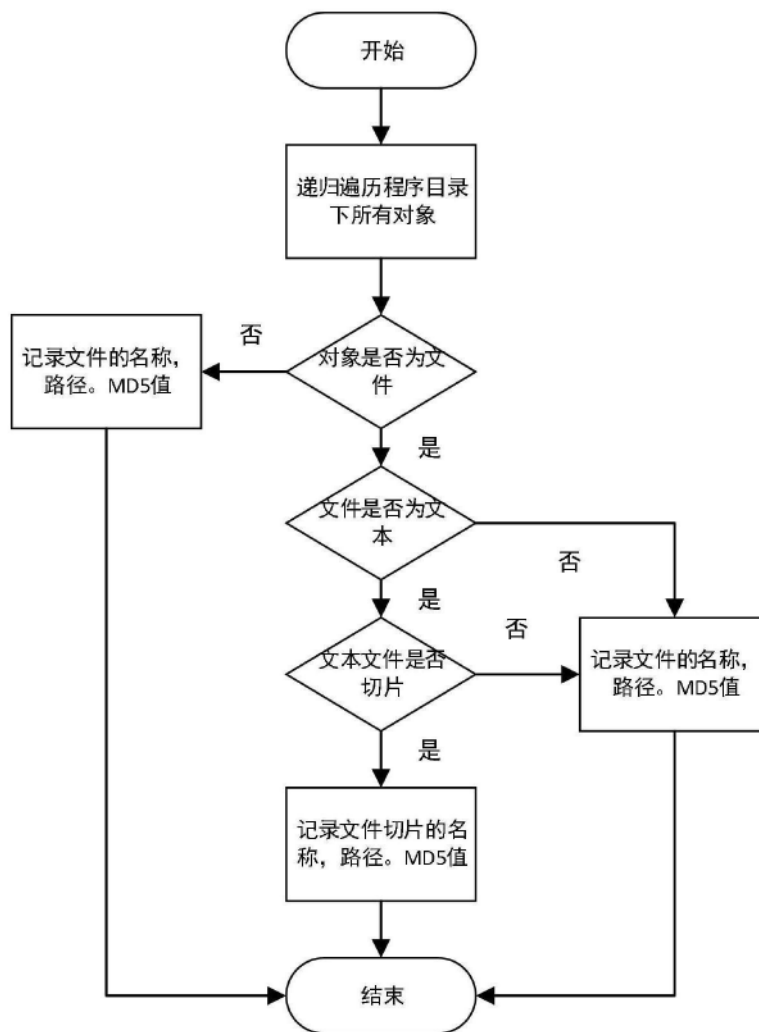


图4

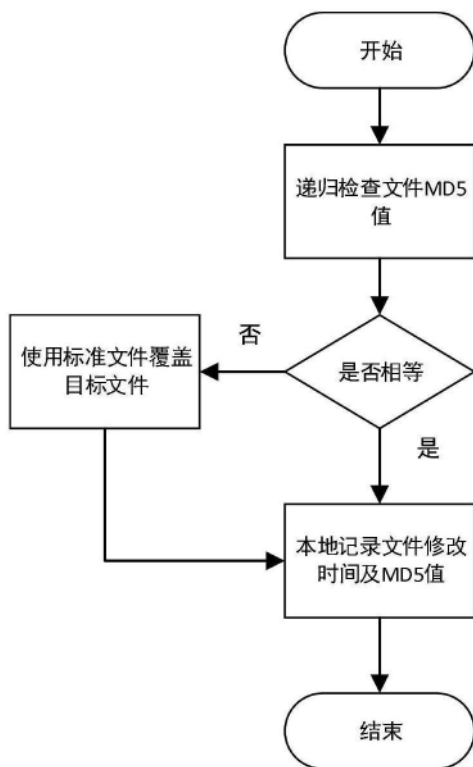


图5

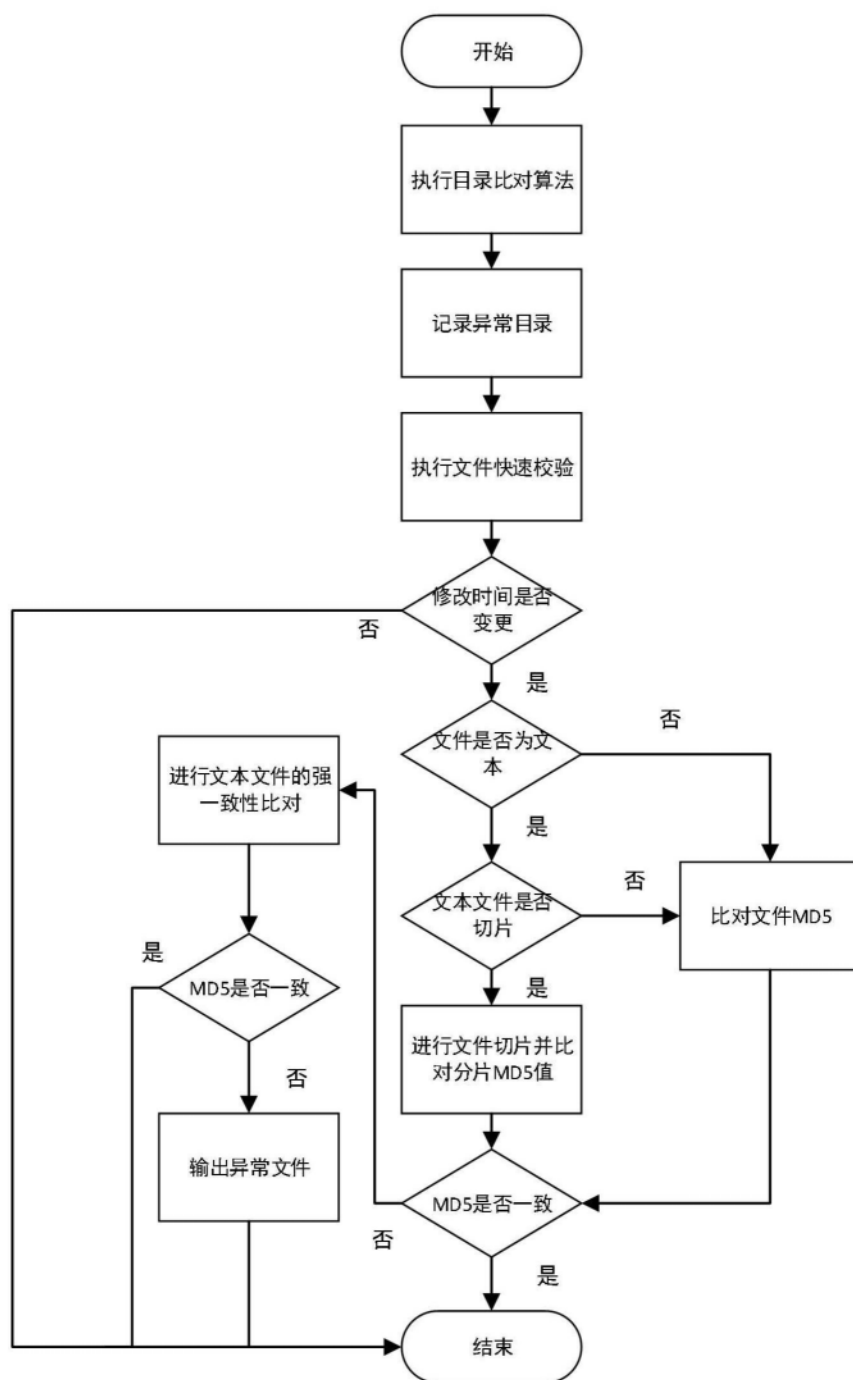


图6

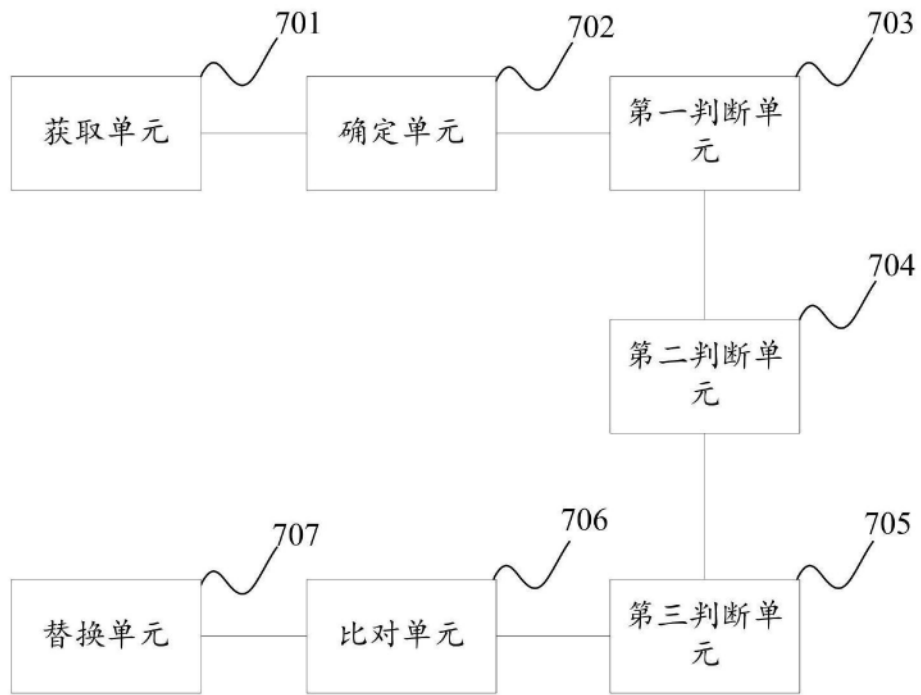


图7

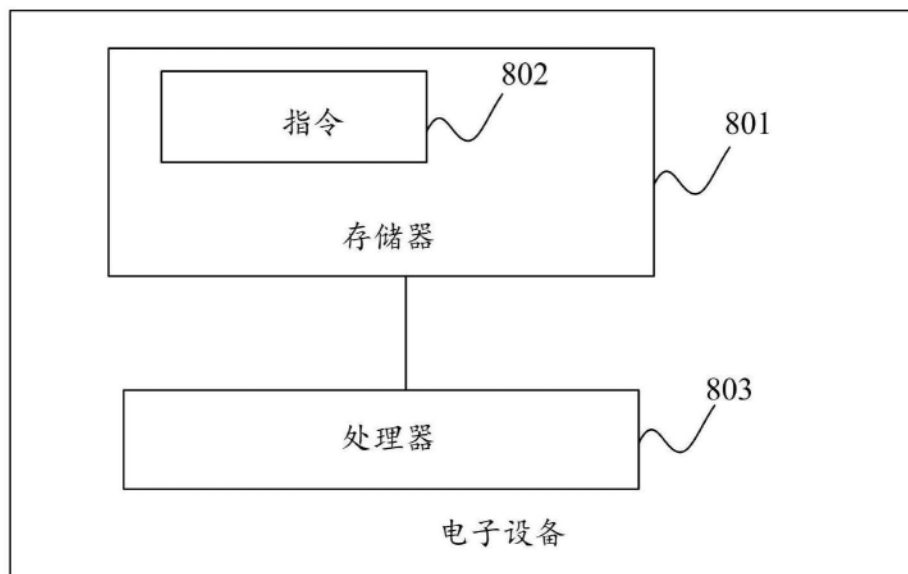


图8