

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2017 年 1 月 19 日 (19.01.2017)



(10) 国际公布号
WO 2017/008677 A1

- (51) 国际专利分类号:
G06F 11/34 (2006.01)
- (21) 国际申请号: PCT/CN2016/089079
- (22) 国际申请日: 2016 年 7 月 7 日 (07.07.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510419821.3 2015 年 7 月 16 日 (16.07.2015) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 梁永锋 (LIANG, Yongfeng); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) Title: DATA DETECTION METHOD AND DEVICE

(54) 发明名称: 一种数据检测方法和装置

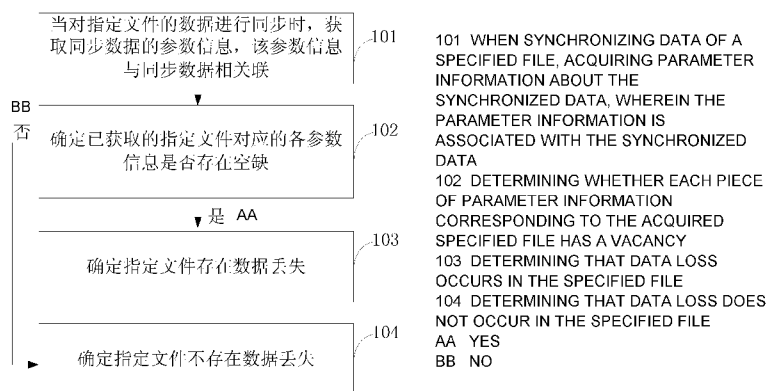


图 1

(57) Abstract: Disclosed are a data detection method and device. The data detection method comprises: when synchronizing data of a specified file, acquiring parameter information about the synchronized data, wherein the parameter information is associated with the synchronized data (101); determining whether each piece of parameter information corresponding to the acquired specified file has a vacancy (102); and if so, determining that data loss occurs in the specified file (103). The detection method not only improves the detection accuracy for lost data, but also has a simple and feasible detection process and relatively high detection efficiency.

(57) 摘要: 一种数据检测方法和装置, 该数据检测方法包括: 当对指定文件的数据进行同步时, 获取同步数据的参数信息, 所述参数信息与所述同步数据相关联 (101); 确定已获取的所述指定文件对应的各参数信息是否存在空缺 (102); 若是, 则确定所述指定文件存在数据丢失 (103)。该检测方法不仅提高了对丢失数据的检测准确度, 而且检测过程简单易行, 检测效率较高。



WO 2017/008677 A1

一种数据检测方法和装置

本申请要求 2015 年 07 月 16 日递交的申请号为 201510419821.3、发明名称为“一种数据检测方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请涉及数据处理技术领域，特别是涉及一种数据检测方法和一种数据检测装置。

背景技术

在各种 web 应用中，会有很多日志文件不断产生，比如页面上用于记录用户行为的埋点日志等。这些日志分布在很多台服务器上，可以多达 10W 台服务器，每台机器上又有多个日志文件。业务上需要将这些日志文件同步到数据仓库系统中进行分析，那么，在同步过程中需要保证数据不丢失，才能为业务提供真正的保障。

目前，在海量数据的同步过程中，有两种方案用于检测数据是否丢失。一种是抽样检查，在日志中定时输出有规律的日志，然后在目标地（数据仓库系统）检测这些规律日志是否存在。如果这些规律日志不存在，就认为数据丢失。另外一种是最最终一致性检查，在数据收集过程中将文件的最后记录数发送给目标地，然后在目标地统计出最终收到的记录数，如记录数不一致则认为数据丢失。

然而，对于抽样检查，由于存在一定的抽样率，极易出现数据丢失但无法检测到的情况。对于最终一致性检查，在数据传输的过程中，时常出现数据重复的情况，如果没有去重，那么在目标地统计出来的最终记录数很多情况下是不准确的。由此可见，上述检测方式对丢失数据的检测准确度较低。

因此，目前需要本领域技术人员迫切解决的一个技术问题就是：如何提高对丢失数据的检测准确度。

25 发明内容

本申请实施例所要解决的技术问题是提供一种数据检测方法，能够提高对丢失数据的检测准确度。

相应的，本申请实施例还提供了一种数据检测装置，用以保证上述方法的实现及应用。

30 为了解决上述问题，本申请公开了一种数据检测方法，包括：

当对指定文件的数据进行同步时，获取同步数据的参数信息，所述参数信息与所述同步数据相关联；

确定已获取的所述指定文件对应的各参数信息是否存在空缺；

若是，则确定所述指定文件存在数据丢失。

5 进一步，所述确定已获取的所述指定文件对应的各参数信息是否存在空缺，包括：

将获取的所述同步数据的参数信息插入所述指定文件的参数信息链表中，所述参数信息链表中包含有在先获取的所述指定文件对应的参数信息；

将所述参数信息链表中相邻的参数信息进行合并；

按照预置条件查找合并后的所述参数信息链表中是否存在空缺的参数信息。

10 进一步，按照预置条件查找合并后的所述参数信息链表中是否存在空缺的参数信息，包括：

按照预设的次数阈值重复查找合并后的所述参数信息链表中是否存在空缺的参数信息，当查找结果为存在空缺的参数信息且查找次数达到所述次数阈值时，判定所述已获取的所述指定文件对应的参数信息存在空缺。

15 进一步，所述方法还包括：

将所述合并后的所述参数信息链表中空缺的参数信息对应的数据确定为所述指定文件丢失的数据。

进一步，所述同步数据的参数信息包括偏移量和数据长度，且下一次同步数据的偏移量为本次同步数据的偏移量与数据长度的和。

20 本申请实施例还公开了一种数据检测装置，包括：

参数获取单元，被配置为当对指定文件的数据进行同步时，获取同步数据的参数信息，所述参数信息与所述同步数据相关联；

参数确认单元，被配置为确定已获取的所述指定文件对应的各参数信息是否存在空缺；

25 结果确认单元，被配置为当所述参数确认单元确定存在空缺时，确定所述指定文件存在数据丢失。

进一步，所述参数确认单元包括：

插入子单元，被配置为将获取的所述同步数据的参数信息插入所述指定文件的参数信息链表中，所述参数信息链表中包含有在先获取的所述指定文件对应的参数信息；

30 合并子单元，被配置为将所述参数信息链表中相邻的参数信息进行合并；

确认子单元，被配置为按照预置条件查找合并后的所述参数信息链表中是否存在空缺的参数信息。

进一步，所述确认子单元，被配置为按照预设的次数阈值重复查找合并后的所述参数信息链表中是否存在空缺的参数信息，当查找结果为存在空缺的参数信息且查找次数达到所述次数阈值时，判定所述已获取的所述指定文件对应的参数信息存在空缺。

进一步，所述装置还包括：

丢失数据确定单元，被配置为将所述合并后的所述参数信息链表中空缺的参数信息对应的数据确定为所述指定文件丢失的数据。

进一步，所述同步数据的参数信息包括偏移量和数据长度，且下一次同步数据的偏移量为本次同步数据的偏移量与数据长度的和。

与现有技术相比，本申请实施例包括以下优点：

本申请实施例通过利用指定文件的参数信息，以及参数信息与指定文件的同步数据之间的关联关系，只需确定已获取的参数信息是否存在空缺，即可确定出指定文件在数据同步时是否存在数据丢失。该方法中由于每次指定文件的数据同步都会携带参数信息，因此不会存在抽样检查中抽样率的问题，所以避免了没有抽样到的数据存在数据丢失时无法检测到的问题，而且，由于参数信息与指定文件的同步数据相关联，参数信息可以表征指定文件在不同同步次数的数据，所以相对于最终一致性检查，解决了数据重复传输时目标地统计记录数有误而导致的检测不准确的问题。本申请实施例公开的方法不仅提高了对丢失数据的检测准确度，而且检测过程简单易行，检测效率较高。

附图说明

图 1 是本申请的一种数据检测方法实施例的步骤流程图；

图 2 是本申请中的一种确定已获取的所述指定文件对应的各参数信息是否存在空缺的方法实施例的步骤流程图；

图 3 是本申请中的一种参数信息链表的示意图；

图 4 是本申请的一种数据检测装置实施例的结构框图；

图 5 是本申请中的一种参数确认单元的结构框图；

图 6 是本申请的另一种数据检测装置实施例的结构框图。

具体实施方式

为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

参照图 1，示出了本申请的一种数据检测方法实施例的步骤流程图，具体可以包括如下步骤：

- 5 步骤 101，当对指定文件的数据进行同步时，获取同步数据的参数信息，该参数信息与同步数据相关联。

本申请实施例中，该用于检测数据是否丢失的装置（以下简称该装置）可以是进行数据同步时的目标服务器本身，或者是设置在该目标服务器内的一模块，也可以是独立于目标服务器之外且可以与目标服务器进行交互的装置。

- 10 在对指定文件的数据进行同步时，例如对日志的更新数据进行同步，发送数据的源服务器可以将指定文件的数据连同本次数据的参数信息一并发送至目标服务器，该装置向目标服务器请求获得该同步数据的参数信息；或者，源服务器可以将指定文件的数据和本次数据的参数信息分别发送至目标服务器和该装置。

- 15 其中，该参数信息可以是新增的信息，由源服务器和该装置约定，源服务器每次对指定文件进行数据同步，就携带本次数据的参数信息。参数信息与同步数据相关联，不同数据的参数信息不同，参数信息之间可以遵循一定的规律，例如连续发送的同步数据的参数信息是连续的。

- 20 该参数信息也可以是现有文件自带的信息，例如，偏移量和数据长度。每一次文件的读取都会有偏移量(offset)和数据长度(length)两个参数信息，在源服务器读取该文件的更新数据进行数据同步时，同时传输该同步数据在本次读取的偏移量和数据长度。该参数信息与同步数据相关联，下一次同步数据的偏移量为本次同步数据的偏移量与数据长度的和，即：

$$\text{offset}_{\text{next}} = \text{offset}_{\text{now}} + \text{length}_{\text{now}}$$

其中，每次同步数据的 length 可能相同也可能不同。

- 25 该参数信息也可以是其它内容的信息，只要该参数信息与指定文件的同步数据相关联，可以用于区分指定文件在每次同步时的同步数据即可。

步骤 102，确定已获取的指定文件对应的各参数信息是否存在空缺。

- 30 在上步骤获得参数信息后，本步骤即可根据所有已获取的该指定文件对应的参数信息确定参数信息中是否存在空缺。具体的可以根据参数信息之间的规律来确定是否存在空缺或缺失的参数信息。

例如，上步骤中约定连续发送的同步数据的参数信息是连续的，则在本步骤中可以判断已获取的指定文件的参数信息之间是否是连续的，如果是连续的，说明不存在空缺，如果是不连续的，则说明参数信息存在缺失。

再例如，若上步骤采用偏移量和数据长度作为参数信息，则可以根据已获得的某一偏移量和数据长度，根据连续两次同步的偏移量之间的关系 $\text{offset}_{\text{next}} = \text{offset}_{\text{now}} + \text{length}_{\text{now}}$ 来确定出邻近的偏移量，然后在已获取的参数信息中查找是否存在该确定出的邻近的偏移量，以此类推，即可确定出已获取的指定文件对应的参数信息中是否存在空缺。

若采用其它内容的参数信息，由于参数信息与指定文件的同步数据相关联，可以利用不同的参数信息之间的规律来确定已获取的指定文件的参数信息中是否存在空缺。

若确定存在空缺，则执行步骤 103，若不存在空缺，则执行步骤 104。

步骤 103，确定指定文件存在数据丢失。

由于源服务器在进行数据同步时，是根据指定文件的数据更新按次序进行数据同步的，如果该装置所获得的参数信息存在空缺，则说明空缺的参数信息对应的数据丢失，数据同步存在问题。

步骤 104，确定指定文件不存在数据丢失。

本申请实施例通过利用指定文件的参数信息，以及参数信息与指定文件的同步数据之间的关联关系，只需确定已获取的参数信息是否存在空缺，即可确定出指定文件在数据同步时是否存在数据丢失。该方法中由于每次指定文件的数据同步都会携带参数信息，因此不会存在抽样检查中抽样率的问题，所以避免了没有抽样到的数据存在数据丢失时无法检测到的问题，而且，由于参数信息与指定文件的同步数据相关联，参数信息可以表征指定文件在不同同步次数的数据，所以相对于最终一致性检查，解决了数据重复传输时目标地统计记录数有误而导致的检测不准确的问题。本申请实施例公开的方法不仅提高了对丢失数据的检测准确度，而且检测过程简单易行，检测效率较高。

在本申请的另一实施例中，确定已获取的所述指定文件对应的各参数信息是否存在空缺的过程，如图 2 所示，可以进一步包括：

步骤 201，将获取的同步数据的参数信息插入指定文件的参数信息链表中，该参数信息链表中包含有在先获取的该指定文件对应的参数信息。

为了便于确定参数信息中是否存在空缺，本实施例中可以建立该指定文件的参数信息链表，该参数信息链表中包含有所有已获取的该指定文件对应的参数信息，也即在指

定文件的各次数据同步过程中获得的同步数据的参数信息。该参数信息链表可以按参数信息的大小顺序等排列参数信息，以便于查找。

当该装置在步骤 101 中获得该指定文件的新的同步数据的参数信息时，将该参数信息按照排列规律插入对应的参数信息链表中。

5 步骤 202，将参数信息链表中相邻的参数信息进行合并。

在插入参数信息链表后，将相邻的参数信息进行合并。其中，相邻的参数信息也即相邻的两次数据同步中同步数据的参数信息。具体可以依据参数信息的设置规律来推定两参数信息是否相邻，例如，如果连续的同步数据对应的参数信息是连续的，则将连续的参数信息是相邻的参数信息，可以合并；如果按照上述 $\text{offset}_{\text{next}} = \text{offset}_{\text{now}} + \text{length}_{\text{now}}$

10 设置参数信息，则根据数据长度来推定两参数信息是否相邻，再确定是否可以合并。

步骤 203，按照预置条件查找合并后的参数信息链表中是否存在空缺的参数信息。

在本步骤中，不能合并的参数信息之间可能就是存在空缺的参数信息，在确定是否存在空缺的参数信息时，还可以设置条件，查找结果满足预设条件时，再确定是否存在空缺的参数信息。

15 例如预先设置查找的次数阈值，按照预设的次数阈值重复查找合并后的参数信息链表中是否存在空缺的参数信息，当查找结果为存在空缺的参数信息且查找次数达到该次数阈值时，判定已获取的指定文件对应的参数信息存在空缺。

另外，在确定出空缺的参数信息后，该合并后的所述参数信息链表中空缺的参数信息对应的数据即可确定为指定文件在数据同步过程中丢失的数据。

20 下面以参数信息为 offset 和 length 为例进行说明。在每次对指定文件进行数据同步时，都会携带上这 2 个参数，并且在目标服务器保存起来。在目标服务器，参数信息和真实同步数据的保存，结果状态必须是一致的，同时成功或者失败。该用于检测数据是否丢失的装置可以扫描目标服务器保存的该指定文件的参数，进而判断是否存在数据丢失的情况。

25 该装置建立指定文件对应的参数信息链表（TailHead），不断从目标服务器读取新的同步数据的参数信息，并将参数信息插入该参数信息链表中。为便于理解，假设每次数据同步获得的参数信息中 length 一直为 1，已建立的参数信息链表如图 3 所示，以 TailHead 的 d1 对象为例，tail 为 1 表示它的 offset 为 1， length 为 1，head 为 5 表示 offset 为 5， length 为 1。d1 表示 1 到 5 之间的同步数据都已检测到，则如图 3 所示，图中的数据表示 1~5、8~10、12、m~n 中的同步数据都已经存在，未丢失。

30

当该装置执行前述步骤 101，从目标服务器中读取到该指定文件的新的同步数据的参数信息时，例如读取到同步数据的 offset 为 11 和 length 为 1，执行前述步骤 201，将读取的参数信息插入到链表 TailHead 中，那么根据参数信息 offset 为 11 和 length 为 1，将 d2 的 tail 值更新为 11 即可。然后执行前述步骤 202 合并链表中相邻的参数信息，比如 d2 的 tail 为 11 也即 offset 为 11，length 为 1，d3 的 head 为 12 也即 offset 为 12，length 为 1，那么由于 length 均为 1，可以确定 d2 与 d3 相邻，合并 d2 和 d3，其中 head 为 8，tail 为 13。然后执行前述步骤 203 检测参数信息是否存在空缺，以 d1、d2 为例，他们之间间隔的 offset 为 6、7，表示这 2 个参数信息对应的数据不存在。如果这 2 个参数经过多次（次数阈值可以调整）判断仍然不存在，那么就可以认为 offset6、7 对应的数据丢失。

需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本申请实施例所必须的。

参照图 4，示出了本申请一种数据检测装置实施例的结构框图，具体可以包括如下单元：

参数获取单元 401，被配置为当对指定文件的数据进行同步时，获取同步数据的参数信息，所述参数信息与所述同步数据相关联。

参数确认单元 402，被配置为确定已获取的所述指定文件对应的各参数信息是否存在空缺。

结果确认单元 403，被配置为当所述参数确认单元确定存在空缺时，确定所述指定文件存在数据丢失。

本申请实施例中，该装置通过利用指定文件的参数信息，以及参数信息与指定文件的同步数据之间的关联关系，只需确定已获取的参数信息是否存在空缺，即可确定出指定文件在数据同步时是否存在数据丢失。该装置不仅提高了对丢失数据的检测准确度，而且检测过程简单易行，检测效率较高。

在另一实例中，当参数确认单元确定不存在空缺时，结果确认单元 403 可以确定所述指定文件不存在数据丢失。

在另一实施例中，如图 5 所示，该参数确认单元 402 可以进一步包括：

插入子单元 501，被配置为将获取的所述同步数据的参数信息插入所述指定文件的参数信息链表中，所述参数信息链表中包含有在先获取的所述指定文件对应的参数信息。

合并子单元 502，被配置为将所述参数信息链表中相邻的参数信息进行合并。

5 确认子单元 503，被配置为按照预置条件查找合并后的所述参数信息链表中是否存在
在空缺的参数信息。

其中，确认子单元 503，可以具体被配置为按照预设的次数阈值重复查找合并后的所述参数信息链表中是否存在空缺的参数信息，当查找结果为存在空缺的参数信息且查找次数达到所述次数阈值时，判定所述已获取的所述指定文件的参数信息存在空缺。

在另一实施例中，如图 6 所示，该装置还可以包括：

10 丢失数据确定单元 601，被配置为将所述合并后的所述参数信息链表中空缺的参数
信息对应的数据确定为所述指定文件丢失的数据。

上述同步数据的参数信息可以包括偏移量和数据长度，且下一次同步数据的偏移量为本次同步数据的偏移量与数据长度的和。

本申请实施例还公开了一种电子设备，包括存储器和处理器。

15 处理器与存储器通过总线相互连接；总线可以是 ISA 总线、PCI 总线或 EISA 总线等。
所述总线可以分为地址总线、数据总线、控制总线等。

其中，存储器用于存储一段程序，具体地，程序可以包括程序代码，所述程序代码包括计算机操作指令。存储器可能包含高速 RAM 存储器，也可能还包括非易失性存储器（non-volatile memory），例如至少一个磁盘存储器。

20 处理器用于读取存储器中的程序代码，执行以下步骤：

当对指定文件的数据进行同步时，获取同步数据的参数信息，所述参数信息与所述同步数据相关联；

确定已获取的所述指定文件对应的各参数信息是否存在空缺；

若是，则确定所述指定文件存在数据丢失。

25 对于装置实施例而言，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。

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。

30 本领域内的技术人员应明白，本申请实施例的实施例可提供为方法、装置、或计算

机程序产品。因此，本申请实施例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，本申请实施例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。

- 5 在一个典型的配置中，所述计算机设备包括一个或多个处理器 (CPU)、输入/输出接口、网络接口和内存。内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM) 和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、
- 10 程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用
- 15 于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非持续性的电脑可读媒体(transitory media)，如调制的数据信号和载波。

本申请实施例是参照根据本申请实施例的方法、终端设备(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终端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终端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。

20 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终端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终端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。

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终端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。

25 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终端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。

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终端设备上，使得在计算机或其他可编程终端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终端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

30 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终端设备上，使得在计算机或其他可编程终端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终端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。

尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。

5 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、物品或者终端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
10 种过程、方法、物品或者终端设备所固有的要素。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、物品或者终端设备中还存在另外的相同要素。

 以上对本申请所提供的一种数据检测方法和一种数据检测装置，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是
15 用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，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申请的限制。

权利要求书

1、一种数据检测方法，其特征在于，包括：

当对指定文件的数据进行同步时，获取同步数据的参数信息，所述参数信息与所述同步数据相关联；

5 确定已获取的所述指定文件对应的各参数信息是否存在空缺；
若是，则确定所述指定文件存在数据丢失。

2、根据权利要求 1 所述的方法，其特征在于，所述确定已获取的所述指定文件对应的各参数信息是否存在空缺，包括：

10 将获取的所述同步数据的参数信息插入所述指定文件的参数信息链表中，所述参数信息链表中包含有在先获取的所述指定文件对应的参数信息；

将所述参数信息链表中相邻的参数信息进行合并；

按照预置条件查找合并后的所述参数信息链表中是否存在空缺的参数信息。

3、根据权利要求 2 所述的方法，其特征在于，按照预置条件查找合并后的所述参数信息链表中是否存在空缺的参数信息，包括：

15 按照预设的次数阈值重复查找合并后的所述参数信息链表中是否存在空缺的参数信息，当查找结果为存在空缺的参数信息且查找次数达到所述次数阈值时，判定所述已获取的所述指定文件对应的参数信息存在空缺。

4、根据权利要求 2 所述的方法，其特征在于，所述方法还包括：

20 将所述合并后的所述参数信息链表中空缺的参数信息对应的数据确定为所述指定文件丢失的数据。

5、根据权利要求 1 至 4 中任意一项所述的方法，其特征在于，所述同步数据的参数信息包括偏移量和数据长度，且下一次同步数据的偏移量为本次同步数据的偏移量与数据长度的和。

6、一种数据检测装置，其特征在于，包括：

25 参数获取单元，被配置为当对指定文件的数据进行同步时，获取同步数据的参数信息，所述参数信息与所述同步数据相关联；

参数确认单元，被配置为确定已获取的所述指定文件对应的各参数信息是否存在空缺；

30 结果确认单元，被配置为当所述参数确认单元确定存在空缺时，确定所述指定文件存在数据丢失。

7、根据权利要求 6 所述的装置，其特征在于，所述参数确认单元包括：

插入子单元，被配置为将获取的所述同步数据的参数信息插入所述指定文件的参数信息链表中，所述参数信息链表中包含有在先获取的所述指定文件对应的参数信息；

合并子单元，被配置为将所述参数信息链表中相邻的参数信息进行合并；

5 确认子单元，被配置为按照预置条件查找合并后的所述参数信息链表中是否存在空缺的参数信息。

8、根据权利要求 7 所述的装置，其特征在于，

所述确认子单元，被配置为按照预设的次数阈值重复查找合并后的所述参数信息链表中是否存在空缺的参数信息，当查找结果为存在空缺的参数信息且查找次数达到所述
10 次数阈值时，判定所述已获取的所述指定文件对应的参数信息存在空缺。

9、根据权利要求 7 所述的装置，其特征在于，所述装置还包括：

丢失数据确定单元，被配置为将所述合并后的所述参数信息链表中空缺的参数信息对应的数据确定为所述指定文件丢失的数据。

10、根据权利要求 6 至 9 中任意一项所述的装置，其特征在于，所述同步数据的参
15 数信息包括偏移量和数据长度，且下一次同步数据的偏移量为本次同步数据的偏移量与数据长度的和。

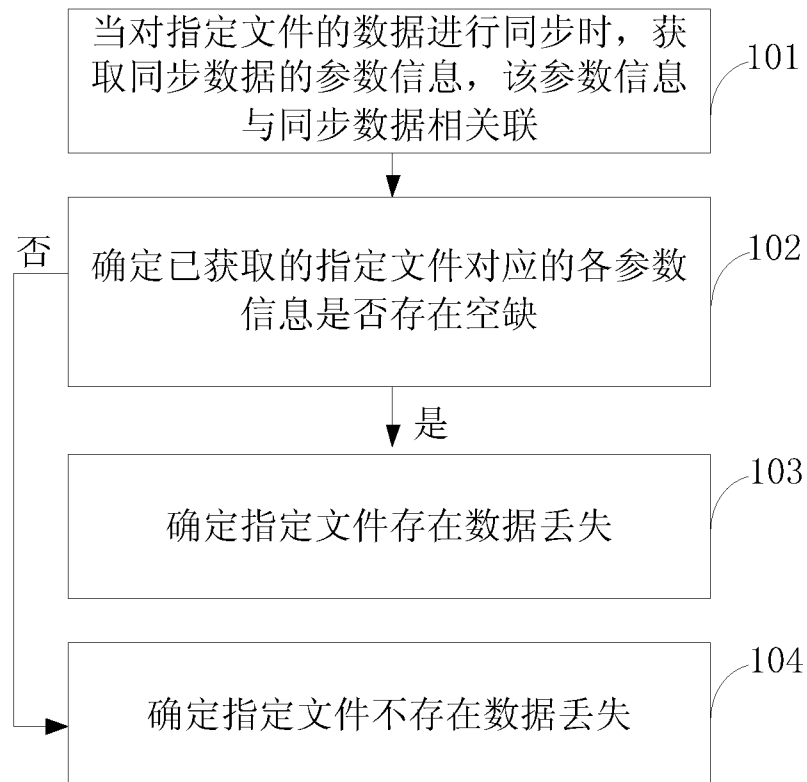


图 1

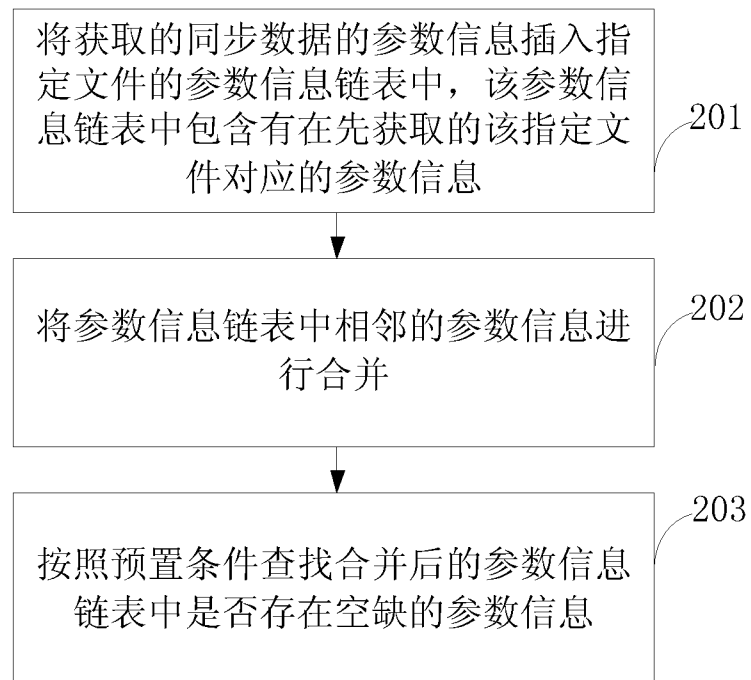


图 2

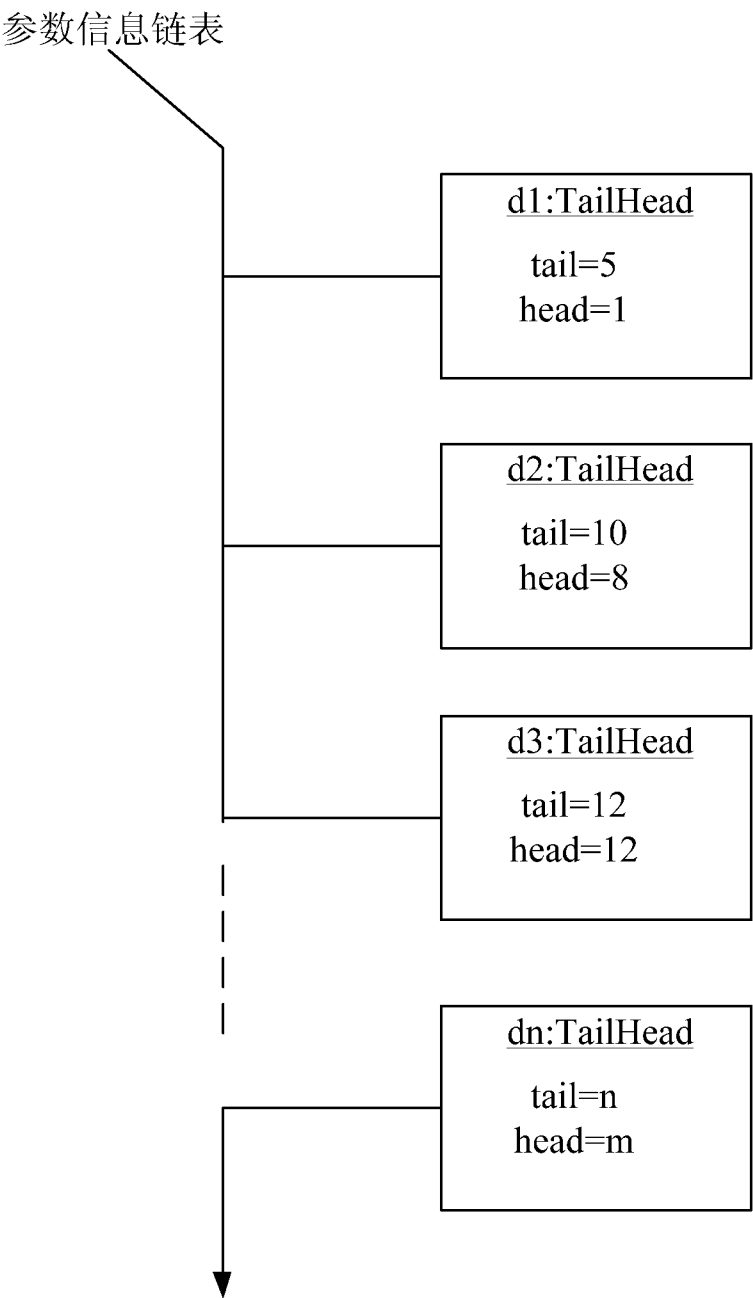


图 3

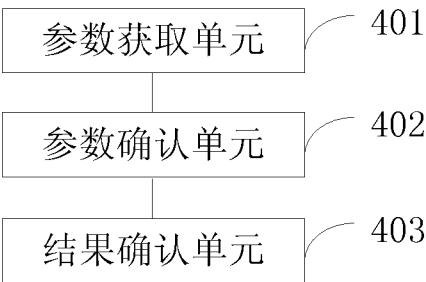


图 4

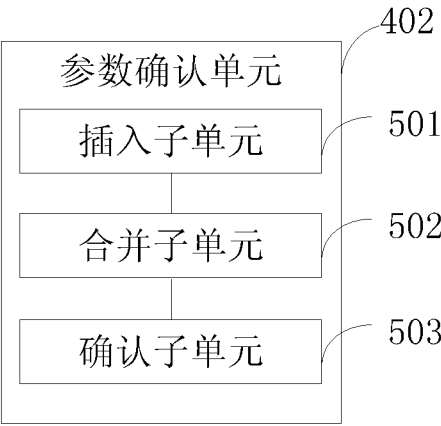


图 5

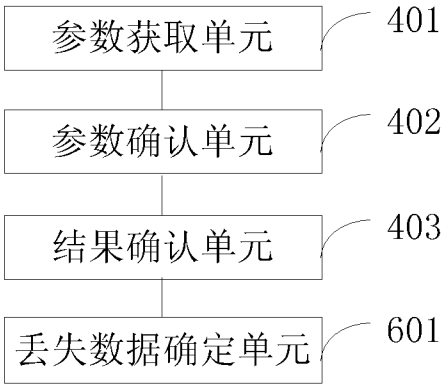


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/089079

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/34 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC, WPI, CNPAT, CNKI, IEEE, GOOGLE: synchronizat+, log, relat+, parameter, lose, loss, miss+, detect+, monitor+, judg+, bias, length

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101751394 A (HISENSE MEDIA NETWORKS CO., LTD.), 23 June 2010 (23.06.2010), description, paragraphs [0007]-[0020] and [0062], and figure 1	1-10
X	CN 103338131 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 02 October 2013 (02.10.2013), description, paragraphs [0003]-[0008] and [0032]-[0037], and figures 1-3	1-10
A	CN 103514223 A (ALIBABA GROUP HOLDING LIMITED), 15 January 2014 (15.01.2014), the whole document	1-10
A	CN 103988189 A (INTERNATIONAL BUSINESS MACHINES CORPORATION), 13 August 2014 (13.08.2014), the whole document	1-10
A	CN 101686184 A (ZTE CORP.), 31 March 2010 (31.03.2010), the whole document	1-10
A	US 6604205 B1 (HEWLETT-PACKARD DEVELOPMENT CO., L.P.), 05 August 2003 (05.08.2003), the whole document	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 09 September 2016 (09.09.2016)	Date of mailing of the international search report 27 September 2016 (27.09.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer WU, Qing Telephone No.: (86-10) 61648111

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2016/089079

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101751394 A	23 June 2010	None	
CN 103338131 A	02 October 2013	None	
CN 103514223 A	15 January 2014	None	
CN 103988189 A	13 August 2014	US 2014331091 A1	06 November 2014
		JP WO2013084571 A1	27 April 2015
		DE 112012005154 T5	19 March 2015
		WO 2013084571 A1	13 June 2013
		GB 2511969 A	17 September 2014
CN 101686184 A	31 March 2010	None	
US 6604205 B1	05 August 2003	None	

A. 主题的分类 G06F 11/34 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																							
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) EPODOC, WPI, CNPAT, CNKI, IEEE, GOOGLE: 同步, 日志, 关联, 参数, 丢失, 检测, 判断, 偏移, 长度, synchronizat+, log, relat+, parameter, lose, loss, miss+, detect+, monitor+, judg+, bias, length																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 101751394 A (青岛海信传媒网络技术有限公司) 2010年 6月 23日 (2010 - 06 - 23) 说明书第[0007]-[0020]、[0062]段, 附图1</td> <td>1-10</td> </tr> <tr> <td>X</td> <td>CN 103338131 A (百度在线网络技术北京有限公司) 2013年 10月 2日 (2013 - 10 - 02) 说明书第[0003]-[0008]、[0032]-[0037]段, 附图1-3</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 103514223 A (阿里巴巴集团控股有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 103988189 A (国际商业机器公司) 2014年 8月 13日 (2014 - 08 - 13) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 101686184 A (中兴通讯股份有限公司) 2010年 3月 31日 (2010 - 03 - 31) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 6604205 B1 (HEWLETT-PACKARD DEVELOPMENT CO., L.P.) 2003年 8月 5日 (2003 - 08 - 05) 全文</td> <td>1-10</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 101751394 A (青岛海信传媒网络技术有限公司) 2010年 6月 23日 (2010 - 06 - 23) 说明书第[0007]-[0020]、[0062]段, 附图1	1-10	X	CN 103338131 A (百度在线网络技术北京有限公司) 2013年 10月 2日 (2013 - 10 - 02) 说明书第[0003]-[0008]、[0032]-[0037]段, 附图1-3	1-10	A	CN 103514223 A (阿里巴巴集团控股有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文	1-10	A	CN 103988189 A (国际商业机器公司) 2014年 8月 13日 (2014 - 08 - 13) 全文	1-10	A	CN 101686184 A (中兴通讯股份有限公司) 2010年 3月 31日 (2010 - 03 - 31) 全文	1-10	A	US 6604205 B1 (HEWLETT-PACKARD DEVELOPMENT CO., L.P.) 2003年 8月 5日 (2003 - 08 - 05) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
X	CN 101751394 A (青岛海信传媒网络技术有限公司) 2010年 6月 23日 (2010 - 06 - 23) 说明书第[0007]-[0020]、[0062]段, 附图1	1-10																					
X	CN 103338131 A (百度在线网络技术北京有限公司) 2013年 10月 2日 (2013 - 10 - 02) 说明书第[0003]-[0008]、[0032]-[0037]段, 附图1-3	1-10																					
A	CN 103514223 A (阿里巴巴集团控股有限公司) 2014年 1月 15日 (2014 - 01 - 15) 全文	1-10																					
A	CN 103988189 A (国际商业机器公司) 2014年 8月 13日 (2014 - 08 - 13) 全文	1-10																					
A	CN 101686184 A (中兴通讯股份有限公司) 2010年 3月 31日 (2010 - 03 - 31) 全文	1-10																					
A	US 6604205 B1 (HEWLETT-PACKARD DEVELOPMENT CO., L.P.) 2003年 8月 5日 (2003 - 08 - 05) 全文	1-10																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																							
国际检索实际完成的日期 2016年 9月 9日		国际检索报告邮寄日期 2016年 9月 27日																					
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10) 62019451		受权官员 吴卿 电话号码 (86-10) 61648111																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/089079

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	101751394	A	2010年 6月 23日	无	
CN	103338131	A	2013年 10月 2日	无	
CN	103514223	A	2014年 1月 15日	无	
CN	103988189	A	2014年 8月 13日	US 2014331091 A1	2014年 11月 6日
				JP WO2013084571 A1	2015年 4月 27日
				DE 112012005154 T5	2015年 3月 19日
				WO 2013084571 A1	2013年 6月 13日
				GB 2511969 A	2014年 9月 17日
CN	101686184	A	2010年 3月 31日	无	
US	6604205	B1	2003年 8月 5日	无	