

(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  
2013年9月6日 (06.09.2013)

WIPO | PCT

(10) 国际公布号  
WO 2013/127309 A1

- (51) 国际专利分类号: H03M 7/30 (2006.01)
(21) 国际申请号: PCT/CN2013/071725
(22) 国际申请日: 2013年2月21日 (21.02.2013)
(25) 申请语言: 中文
(26) 公布语言: 中文
(30) 优先权: 201210053609.6 2012年3月2日 (02.03.2012) CN
(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN).
(72) 发明人: 滕新东 (TENG, Xindong); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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, ML, MR, NE, SN, TD, TG).

[见续页]

(54) Title: DATA PROCESSING METHOD AND DATA PROCESSING DEVICE

(54) 发明名称: 数据处理方法及数据处理设备

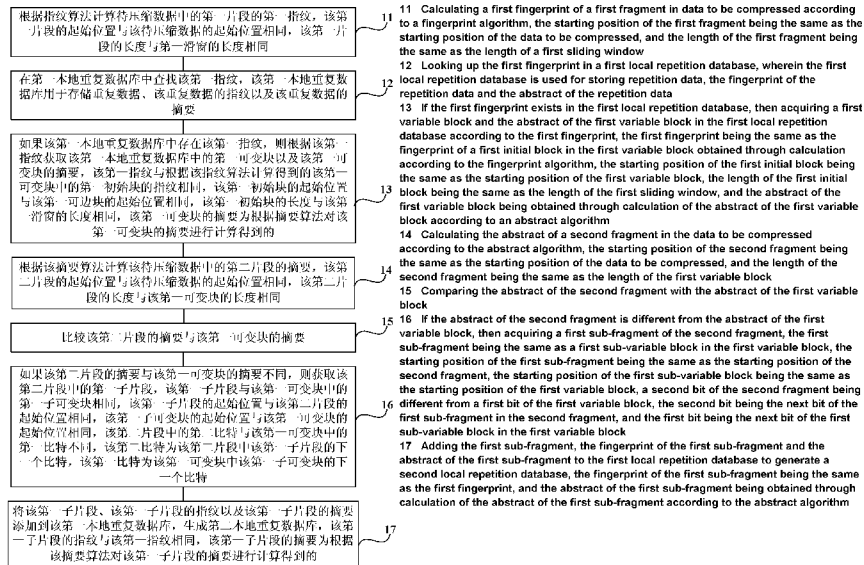


图 1 / FIG. 1

(57) Abstract: Provided are a data processing method and a data processing device. If data to be compressed includes a data fragment which is the same as the former part of a variable block in a repetition database while different from the latter part of the variable block, a new variable block of which the granularity is smaller than that of the matched variable block can be generated, and the new variable block is added to the repetition database. The granularity of the new variable block is smaller, improving the probability that subsequent data to be compressed is matched with the updated repetition database, thus improving the compression efficiency.

(57) 摘要: 本发明实施例提供了数据处理方法及数据处理设备, 如果待压缩的数据中包含了与重复数据库中的可变块的前半部分相同并且与可变块的后半部分不同的数据片段, 则能够生成粒度小于发生匹配的可变块的新的可变块, 并将新的可变块添加到重复数据库。新的可变块粒度较小, 提高了后续的特压缩数据与更新后的重复数据库发生匹配的概率, 进而提高了压缩的效率。

WO 2013/127309 A1

**本国际公布:**

- 包括国际检索报告(条约第 21 条(3))。

## 数据处理方法及数据处理设备

## 技术领域

本发明涉及报文处理技术，尤其涉及一种数据处理方法及数据处理设备。

## 5 背景技术

数据压缩是对报文内容进行算法处理，减小数据量，但不影响信息传递的过程。数据压缩是为了达到节约网络传送带宽、实现应用加速的目的。

根据 CDC (Content-Defined Chunking) 可变块算法生成包含可变块、可变块的指纹以及可变块的摘要的重复数据库。生成重复数据库后，如果待压缩的数据中包含了与可变块的前半部分相同并且与可变块的后半部分不同的数据片段，则数据片段的指纹与可变块的指纹匹配，则数据片段的摘要与可变块的摘要不匹配。CDC 算法认为重复数据库需要进行更新，并使用定界滑窗机制生成新的可变块。新的可变块粒度可能比较大，降低了后续待压缩数据与更新后的重复数据库发生匹配的概率。以上可能导致压缩效率下降。

## 发明内容

本发明实施例提供一种数据处理方法及数据处理设备，用于提高数据压缩效率。

一方面，本发明实施例提供了一种数据处理方法，包括：

20 根据指纹算法计算待压缩数据中的第一片段的第一指纹，所述第一片段的起始位置与所述待压缩数据的起始位置相同，所述第一片段的长度与第一滑窗的长度相同；

在第一本地重复数据库中查找所述第一指纹，所述第一本地重复数据库用于存储重复数据、所述重复数据的指纹以及所述重复数据的摘要；

如果所述第一本地重复数据库中存在所述第一指纹，则根据所述第一指纹获取所述第一本地重复数据库中的第一可变块以及所述第一可变块的摘要，所述第一指纹与根据所述指纹算法计算得到的所述第一可变块中的第一初始块的指纹相同，所述第一初始块的起始位置与所述第一可边块的起始位置相同，所述第一初始块的长度与所述第一滑窗的长度相同，所述  
5 第一可变块的摘要为根据摘要算法对所述第一可变块的摘要进行计算得到的；

根据所述摘要算法计算所述待压缩数据中的第二片段的摘要，所述第二片段的起始位置与所述待压缩数据的起始位置相同，所述第二片段的长  
10 度与所述第一可变块的长度相同；

比较所述第二片段的摘要与所述第一可变块的摘要；

如果所述第二片段的摘要与所述第一可变块的摘要不同，则获取所述第二片段中的第一子片段，所述第一子片段与所述第一可变块中的第一子可变块相同，所述第一子片段的起始位置与所述第二片段的起始位置相同，  
15 所述第一子可变块的起始位置与所述第一可变块的起始位置相同，所述第二片段中的第二比特与所述第一可变块中的第一比特不同，所述第二比特为所述第二片段中所述第一子片段的下一个比特，所述第一比特为所述第一可变块中所述第一子可变块的下一个比特；

将所述第一子片段、所述第一子片段的指纹以及所述第一子片段的摘要添加到所述第一本地重复数据库，生成第二本地重复数据库，所述第一子片段的指纹与所述第一指纹相同，所述第一子片段的摘要为根据所述摘要算法对所述第一子片段的摘要进行计算得到的。  
20

另一方面，本发明实施例还提供了一种数据压缩设备，包括：

第一指纹计算单元，用于根据指纹算法计算待压缩数据中的第一片段的第一指纹，所述第一片段的起始位置与所述待压缩数据的起始位置相同，所  
25

述第一片段的长度与第一滑窗的长度相同；

查找单元，用于在第一本地重复数据库中查找所述第一指纹，所述第一本地重复数据库用于存储重复数据、所述重复数据的指纹以及所述重复数据的摘要；

- 5 重复内容获取单元，用于如果所述第一本地重复数据库中不存在所述第一指纹，则根据所述第一指纹获取所述第一本地重复数据库中的第一可变块以及所述第一可变块的摘要，所述第一指纹与根据所述指纹算法计算得到的所述第一可变块中的第一初始块的指纹相同，所述第一初始块的起始位置与
- 10 所述第一可边块的起始位置相同，所述第一初始块的长度与所述第一滑窗的长度相同，所述第一可变块的摘要为根据摘要算法对所述第一可变块的摘要进行计算得到的；

摘要计算单元，用于根据所述摘要算法计算所述待压缩数据中的第二片段的摘要，所述第二片段的起始位置与所述待压缩数据的起始位置相同，所述

15 第二片段的长度与所述第一可变块的长度相同；

- 摘要比较单元，用于比较所述第二片段的摘要与所述第一可变块的摘要；

- 第一子片段获取单元，用于如果所述第二片段的摘要与所述第一可变块的摘要不同，则获取所述第二片段中的第一子片段，所述第一子片段与所述
- 20 第一可变块中的第一子可变块相同，所述第一子片段的起始位置与所述第二片段的起始位置相同，所述第一子可变块的起始位置与所述第一可变块的起始位置相同，所述
- 20 第二片段中的第二比特与所述第一可变块中的第一比特不同，所述第二比特为所述第二片段中所述第一子片段的下一个比特，所述第一比特为所述第一可变块中所述第一子可变块的下一个比特；

- 第一添加单元，用于将所述第一子片段、所述第一子片段的指纹以及所述
- 25 第一子片段的摘要添加到所述第一本地重复数据库，生成第二本地重复数据库，所述第一子片段的指纹与所述第一指纹相同，所述第一子片段的摘

要为根据所述摘要算法对所述第一子片段的摘要进行计算得到的。

又一方面，本发明实施例还提供了一种数据解压缩设备，包括：

第一指纹计算单元，用于根据指纹算法计算待压缩数据中的第一片段的第一指纹，所述第一片段的起始位置与所述待压缩数据的起始位置相同，所述第一片段的长度与第一滑窗的长度相同；

查找单元，用于在第一本地重复数据库中查找所述第一指纹，所述第一本地重复数据库用于存储重复数据、所述重复数据的指纹以及所述重复数据的摘要；

重复内容获取单元，用于如果所述第一本地重复数据库中不存在所述第一指纹，则根据所述第一指纹获取所述第一本地重复数据库中的第一可变块以及所述第一可变块的摘要，所述第一指纹与根据所述指纹算法计算得到的所述第一可变块中的第一初始块的指纹相同，所述第一初始块的起始位置与所述第一可边块的起始位置相同，所述第一初始块的长度与所述第一滑窗的长度相同，所述第一可变块的摘要为根据摘要算法对所述第一可变块的摘要进行计算得到的；

摘要计算单元，用于根据所述摘要算法计算所述待压缩数据中的第二片段的摘要，所述第二片段的起始位置与所述待压缩数据的起始位置相同，所述第二片段的长度与所述第一可变块的长度相同；

摘要比较单元，用于比较所述第二片段的摘要与所述第一可变块的摘要；

第一子片段获取单元，用于如果所述第二片段的摘要与所述第一可变块的摘要不同，则获取所述第二片段中的第一子片段，所述第一子片段与所述第一可变块中的第一子可变块相同，所述第一子片段的起始位置与所述第二片段的起始位置相同，所述第一子可变块的起始位置与所述第一可变块的起始位置相同，所述第二片段中的第二比特与所述第一可变块中的第一

比特不同，所述第二比特为所述第二片段中所述第一子片段的下一个比特，所述第一比特为所述第一可变块中所述第一子可变块的下一个比特；

第一添加单元，用于将所述第一子片段、所述第一子片段的指纹以及所述  
5 所述第一子片段的摘要添加到所述第一本地重复数据库，生成第二本地重复数据库，所述第一子片段的指纹与所述第一指纹相同，所述第一子片段的摘要为根据所述摘要算法对所述第一子片段的摘要进行计算得到的。

根据本发明实施例提供的技术方案，如果待压缩的数据中包含了与重复数据库中的可变块的前半部分相同并且与可变块的后半部分不同的数据片段，则能够生成粒度小于发生匹配的可变块的新的可变块，并将新的可  
10 变块添加到重复数据库。新的可变块粒度较小，提高了后续的待压缩数据与更新后的重复数据库发生匹配的概率，进而提高了压缩的效率。

#### 附图说明

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一简单地介绍，显而易见地，  
15 下面描述中的附图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。

图 1 为本发明实施例提供的一种数据处理方法的流程图；

图 2 为本发明实施例提供的数据处理方法的一种应用场景的示意图；

图 3 为本发明实施例提供的数据处理方法的另一种应用场景的示意图；

20 图 4 为本发明实施例提供的一种数据处理方法的一个具体实现方式的示意图；

图 5 为本发明实施例提供的另一种数据处理方法的流程图；

图 6 为本发明实施例提供的一种数据处理方法中压缩后报文的封装格式示意图；

25 图 7 为本发明实施例提供的一种数据处理方法应用到数据压缩侧设备

的示意图；

图 8 为本发明实施例提供的数据压缩设备的结构示意图；

图 9 为本发明实施例提供的数据压缩设备的一种应用场景的组网结构图；

5 图 10 为本发明实施例提供的数据压缩设备的另一种应用场景的组网结构图。

### 具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本  
10 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作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。

数据压缩技术包括两种：一种是通过无损压缩算法在发送端压缩数据，  
15 在接收端进行数据解压；另一种是基于 DRE (Data Redundancy Elimination, 数据冗余删除) 技术，也称为重复数据删除 (De-duplication)，将需要传送的数据中的重复内容消除后用特殊的 ID 代替，只传递增量信息，从而减小数据量，实现数据压缩目的。

DRE 技术一般应用于 WOC (WAN Optimization Controller, 广域网优化控制器)，以减少需要 WAN 传送的数据量，相当于增加 WAN 带宽，节约了宝贵的 WAN 资源。WOC 是一种应用加速设备。在 WOC 设备的 DRE 实现中，处理的对象可以是基于 IP 报文，也可以是基于一条会话 (即同一  
20 会话的多个连续 IP 报文)。基于 IP 报文的处理实现简单，不需要缓存报文，性能较高，但是由于重复数据可能包含在多个报文中，导致不容易将重复  
25 数据进行识别和缓存。另外，将多个报文组合起来进行算法压缩时，由于

不同包的类型不同（如不同协议，或不同格式），难以得到较高的压缩比。基于会话的处理则需要将一个会话的多个包进行缓存，在线处理时性能存在一定限制，但是可以最大限度地识别出重复数据，并且同一会话一般属于同一类型，进行算法压缩时能得到更高的压缩比。

- 5 DRE 实现的过程包括：压缩设备的本地 DRE 模块分析报文，判断和确定相应数据块；将确定的数据块与重复数据库中的已存数据块比较，如果查找到同样的块存在，则表示之前传输过该数据块（即为重复数据），此时在报文中用指纹代替该数据块；没有找到的数据块被加入重复数据库中；
- 10 可选地，对去冗余后的报文进一步进行压缩；远端设备的 DRE 模块解压（可选）后将指纹替换为原数据块，之后传送报文给用户；本地 DRE 模块与远端 DRE 模块需要进行数据块及指纹的同步。

- DRE 技术基于二进制数据进行删重处理，不需要感知上层的具体协议类型。其关键点是进行重复数据的识别和替换，即对内容完全相同的重复数据进行识别，并建立重复数据库（包括缓存重复数据和建立查找指纹），
- 15 当后续传送的数据存在已缓存的内容（即重复数据）时，则用指纹代替。由于需要建立重复数据库缓存（cache）已有数据，因此 DRE 技术也称为字节缓存（Byte cache）技术。

- 而重复数据的识别和替换是基于数据块实现的，数据块的识别算法包括 FSP（Fixed-Sized Partition，固定块）、CDC 以及 SB（Sliding Block，滑
- 20 动块）。

- 其中，CDC 算法采用一个滑动窗口（以下简称滑窗）对待压缩的数据进行块定界，滑窗从数据开始按字节向后移动，并采用特定的哈希（HASH）算法（如 rabin HASH、ELF HASH 等）计算出滑窗的指纹信息。当计算出的指纹满足一定条件（如对特定值取模结果为某设定值），则认为找到数据
- 25 块边界，然后向后滑动滑窗，再次计算滑窗的指纹信息，当计算条件成立，

则找到下一数据块的边界，由此可以将整个数据划分成大小可变的多个数据块。

当确定数据块之后，采用特定的算法（SHA-1、MD5 等，这些算法能对大量数据进行信息提取，形成固定长度的字段，由于算法的特定，不同的原始数据计算结果相同的概率非常低，可以忽略不计）来对数据块的内容进行计算，并将计算结果作为指纹查找本地的重复数据库，如果存在则表明已有相同内容的数据块存在，当前数据块为重复数据块，如果指纹不存在，则将当前数据块的指纹及块内容加入缓存（重复数据库），以备下次检查。

10 CDC 算法的块定界算法可能因为计算条件总是不满足导致块过大，具体实现上可以对块的大小设定上下限，当满足上下限条件时强制分块。CDC 算法的优点是对数据内容的变化不敏感，当插入或删除数据时只会影响到该变化数据相关的少量的数据块，而其它块不受影响。

图 1 为本发明实施例提供的一种数据处理方法的流程图。如图 1 所示，  
15 数据处理方法包括：

11、根据指纹算法计算待压缩数据中的第一片段的第一指纹，该第一片段的起始位置与该待压缩数据的起始位置相同，该第一片段的长度与第一滑窗的长度相同。

其中，第一滑窗为现有的滑窗，用于对数据块进行定界。

20 12、在第一本地重复数据库中查找该第一指纹，该第一本地重复数据库用于存储重复数据、该重复数据的指纹以及该重复数据的摘要。

13、如果该第一本地重复数据库中不存在该第一指纹，则根据该第一指纹获取该第一本地重复数据库中的第一可变块以及该第一可变块的摘要，该第一指纹与根据该指纹算法计算得到的该第一可变块中的第一初始块的

指纹相同，该第一初始块的起始位置与该第一可边块的起始位置相同，该第一初始块的长度与该第一滑窗的长度相同，该第一可变块的摘要为根据摘要算法对该第一可变块的摘要进行计算得到的。

14、根据该摘要算法计算该待压缩数据中的第二片段的摘要，该第二片段的起始位置与该待压缩数据的起始位置相同，该第二片段的长度与该第一可变块的长度相同。

其中，第二片段为现有 CDC 算法中进行匹配压缩处理的基本单元--数据块。

15、比较该第二片段的摘要与该第一可变块的摘要。

16、如果该第二片段的摘要与该第一可变块的摘要不同，则获取该第二片段中的第一子片段，该第一子片段与该第一可变块中的第一子可变块相同，该第一子片段的起始位置与该第二片段的起始位置相同，该第一子可变块的起始位置与该第一可变块的起始位置相同，该第二片段中的第二比特与该第一可变块中的第一比特不同，该第二比特为该第二片段中该第一子片段的下一个比特，该第一比特为该第一可变块中该第一子可变块的下一个比特。

17、将该第一子片段、该第一子片段的指纹以及该第一子片段的摘要添加到该第一本地重复数据库，生成第二本地重复数据库，该第一子片段的指纹与该第一指纹相同，该第一子片段的摘要为根据该摘要算法对该第一子片段的摘要进行计算得到的。

根据本发明实施例提供的技术方案，如果待压缩的数据中包含了与重复数据库中的可变块的前半部分相同并且与可变块的后半部分不同的数据片段，则能够生成粒度小于发生匹配的可变块的新的可变块，并将新的可变块添加到重复数据库。新的可变块粒度较小，提高了后续的待压缩数据与更新后的重复数据库发生匹配的概率，进而提高了压缩的效率。

上述 11 至 17 可由压缩侧设备执行，也可由解压缩设备执行。如压缩侧设备接收到一待去冗余的报文，该报文包括报文头及净荷，净荷为待压缩数据，压缩侧设备按照上述 11 至 17 对净荷进行处理。或者如解压缩侧设备接收到一去冗余后的报文，同样的，该去冗余后的报文包含报文头及净荷，只不过该净荷为压缩侧设备压缩后的数据，即部分数据已被替换为指纹，此时，解压缩侧设备处理的待压缩数据即净荷中其余未被替换的原始数据。

当上述方法由压缩侧设备实现时，本发明实施例提供的数据处理方法还可以包括：

10 接收第一报文，该第一报文包括第一报文头与第一净荷，该第一净荷包含第一净荷片段，该第一净荷片段的长度与该第一子片段的长度相同；

根据该指纹算法计算该第一净荷片段中第二初始块的指纹，该第二初始块的起始位置与该第一净荷片段的起始位置相同，该第二初始块的长度与该第一滑窗的长度相同；

15 根据该摘要算法计算该第一净荷片段的摘要；

如果该第二初始块的指纹与该第二本地重复数据库中的该第一子片段的指纹相等，并且该第一净荷片段的摘要与该第一子片段的摘要相等，则删除该第一报文中的该第一净荷片段，生成第二报文，该第二报文中包括第二报文头与第二净荷，该第二报文头与该第一报文头相同，该第二净荷包括该第一子片段的指纹。这样，解压缩侧设备可以根据指纹从重复数据库中找到对应的第一子片段，将指纹替换为第一子片段，将报文中的数据恢复为原始数据，实现数据的解压缩。

可选地，该第二净荷还包括该第一子片段在该第一报文中的位置信息。

当替换第一子片段的指纹未放置在第一子片段在第一报文中的原始位

置时，解压缩侧设备对第二报文执行解压缩操作时，解压缩侧设备可以根据该第二净荷中包括的该第一子片段在该第一报文中的位置信息，确定第一子片段在第一报文中的原始位置，进而还原出第一报文。第一子片段在第一报文中的位置信息可以是第一子片的最高比特相对于第一报文的报  
5 文头的最高比特的偏移量。

该获取该第二片段中的第一子片段之后，本发明实施例提供的数据处理方法还可以包括：

根据该指纹算法计算该待压缩数据中的第二子片段的第三初始块的指纹，该第二子片段的起始位置为该第二比特，该第二子片段的结束位置与  
10 该待压缩数据的结束位置相同；该第三初始块的起始位置为该第二比特，该第三初始块的长度等于该第一滑窗的长度；

根据第二滑窗获取第二子可变块中的第一检测块，该第二滑窗的起始位置与该第二子可变块中的第三比特对应，该第三比特为介于第二子可变块的起始位置与该第二子可变块的结束位置之间的比特，该第二滑窗的长  
15 度等于该第一滑窗的长度，该第二子可变块的起始位置为该第一比特，该第二子可变块的结束位置与该第一可变块的结束位置相同；第二滑窗可定义为内容滑窗，专用于判断重复数据库中是否存在对应的数据块。其计算指纹的方法可与现有的滑窗相同，只是现有技术中滑窗仅用于定界。

根据该指纹算法计算该第一检测块的指纹；

20 比较该第三初始块的指纹与该第一检测块的指纹；

如果该第三初始块的指纹与该第一检测块的指纹相同，则比较该第二子可变块中的第三子可变块的摘要与该第二子片段中的第三子片段的摘要，该第三子可变块的起始位置与该第一检测块的起始位置相同，该第三子可变块的结束位置与该第一可变块的结束位置相同，该第三子片段的起  
25 始位置与该第二子片段的起始位置相同，该第三子片段的长度与该第三子

可变块的长度相等，该第三子可变块的摘要为根据该摘要算法对该第三子可变块的摘要进行计算得到的；

如果该第三子片段的摘要与该第三子可变块的摘要相等，则将该第三子片段、该第三子片段的指纹以及该第三子片段的摘要添加到该第二本地重复数据库，生成第三本地重复数据库，该第三子片段的指纹等于该第三初始块的指纹，该第三子片段的摘要为根据该摘要算法对该第三子片段的摘要进行计算得到的。

在该第三子片段的摘要与该第三子可变块的摘要相等，并将该第三子片段、该第三子片段的指纹以及该第三子片段的摘要添加到该第二本地重复数据库的场景下，本发明实施例提供的数据处理方法还可以包括：

接收第三报文，该第三报文包括第三报文头与第三净荷，该第三净荷包含第二净荷片段以及第三净荷片段，该第二净荷片段的长度与该第一子片段的长度相同，该第三净荷片段的长度与该第三子片段的长度相同；

根据该指纹算法计算该第二净荷片段中第四初始块的指纹，根据该指纹算法计算该第三净荷片段中第五初始块的指纹，该第四初始块的起始位置与该第二净荷片段的起始位置相同，该第四初始块的长度与该第一滑窗的长度相同，该第五初始块的起始位置与该第三净荷片段的起始位置相同，该第五初始块的长度与该第一滑窗的长度相同；

根据该摘要算法计算该第二净荷片段的摘要，根据该摘要算法计算该第三净荷片段的摘要；

如果该第四初始块的指纹与该第三本地重复数据库中的该第一子片段的指纹相等，并且该第二净荷片段的摘要与该第一子片段的摘要相等，则删除该第三报文中的该第二净荷片段，如果该第五初始块的指纹与该第三本地重复数据库中的该第三子片段的指纹相等，并且该第三净荷片段的摘要与该第三子片段的摘要相等，则删除该第三报文中的该第三净荷片段，

生成第四报文，该第四报文中包括第四报文头与第四净荷，该第四报文头与该第三报文头相同，该第四净荷包括该第一子片段的指纹以及该第三子片段的指纹。

5 可选地，该第四净荷还包括该第一子片段在该第三报文中的位置信息、该第三子片段在该第三报文中的位置信息。

其中，第三子片段与第一子片段类似，为小于第二片段粒度的数据块，这样可以进一步提高数据压缩效率。第一子片段在第三报文中的位置信息可以是第一子片的最高比特相对于第三报文的报文头的最高比特的偏移量。

10 如图 2 所示，待压缩数据块 A' 与重复数据库中对应的数据块 A 的摘要不同，可以理解为上次发送的数据块 A 经过修改后变为数据块 A'，并进行再次发送。其中，待压缩数据块 A' 即上述第二片段，重复数据库中对应的数据块 A 即上述第一可变块。这样，上次发送数据时在重复数据库中保存的数据块 A 与当前发送的数据中的数据块 A' 相对应，具体地，当  
15 滑窗滑动到数据块 A' 时，计算指纹，若重复数据库中有相同的指纹，则表示重复数据库中存在有对应的数据块，这里对应的数据块为 A。由于此时不知道数据块 A' 的长度，假定数据块 A' 的长度与数据块 A 相同，计算数据块 A' 的摘要，并与数据块 A 的摘要比较。显然，数据块 A' 的摘要与数据块 A 的摘要不同，因为数据块 A 与数据块 A' 的内容不同。

20 之后再找到数据块 A' 中与数据块 A 相同的部分，以及不同的部分，进行进一步地压缩处理。

具体地，逐字节比较数据块 A' 与数据块 A，当重复数据库中对应的数据块 A 到第二拆分点，待压缩数据块 A' 到第一拆分点时，二者比较结果是不一致，此时，将滑窗滑动到第一拆分点处计算指纹，重复数据库中，  
25 将滑窗滑动到第二拆分点处计算指纹，与第一拆分点处的指纹进行比较，

不一致时，继续向后滑动滑窗，计算第二拆分点后的指纹，并继续与第一拆分点处的指纹进行比较，不一致时，继续向后滑动滑窗，直至计算出的指纹与第一拆分点处的指纹一致，或直至滑窗滑至数据块 A 的结束边界。

本实施例中，数据块 A 的第三拆分点处的指纹与待压缩数据块 A' 的第一拆分点处的指纹一致。当数据块 A 的第三拆分点处的指纹与待压缩数据块 A' 的第一拆分点处的指纹一致时，利用 HASH 表项中的块长度及第三拆分点到边界的长度计算出待压缩数据块 A' 中第一拆分点到边界的长度，然后计算第一拆分点至边界的摘要，并计算重复数据库中数据块 A 的第三拆分点至边界的摘要，将计算出的两个摘要进行比较，当不一致时，按照

10 前述操作比较第三拆分点后字节与第一拆分点后的字节，直至计算出的指纹与第一拆分点处的指纹一致，或直至滑窗滑至数据块 A 的结束边界；当一致时，待压缩数据块 A' 用第一拆分点拆分为两个子数据块 A1、A2，重复数据库中的数据块 A 被第二拆分点、第三拆分点拆分为三个子数据块 A1、A2、A3，可以看出，本次发送的数据块 A' 与上次发送的数据块 A 相比删

15 除了子数据块 A2，其余两个子数据块 A1、A3 不变，这样，可以用指纹替换 A1 或者 A3，实现对数据的精确压缩，提高数据压缩效率。其中，待压缩数据块 A' 中的子数据块 A1 即上述第一子片段，待压缩数据块 A' 中的子数据块 A3 即上述第三子片段。重复数据库中的子数据块 A1 即上述第一子可变块，重复数据库中的子数据块 A3 的初始块，也就是从重复数据库中的子数据块 A3 的第一个比特开始长度为第一滑窗长度的数据块，即上述第

20 二子可变块中的第一检测块。

替换 A1 或者 A3 也可以用其他方法实现。举例来说，可以用指纹、块长度及偏移值替换 A1 或者 A3。

可选地，本发明实施例提供的数据处理方法还可以包括：

25 根据该指纹算法计算该第一可变块中的第四子可变块的第六初始块的

指纹，该第四子可变块的起始位置为该第一比特，该第四子可变块的结束位置与该第一可变块的结束位置相同，该第六初始块的起始位置为该第一比特，该第六初始块的长度等于该第一滑窗的长度；

获取第四子片段中的第二检测块，该第二检测块的起始位置与该第四子片段中的第四比特对应，该第四比特为介于第三片段的起始位置与该第三片段的结束位置之间的比特，该第二检测块的长度等于该第一滑窗的长度，该第三片段的起始位置为该第二比特，该第三片段的结束位置通过定界算法确定，该第三片段为该待压缩数据中的片段，该第四子片段为该第三片段中的子片段，该第四子片段的起始位置与该第二检测块的起始位置  
10 相同，该第四子片段的长度与该第四子可变块的长度相同；

根据该指纹算法计算该第二检测块的指纹；

比较该第六初始块的指纹与该第二检测块的指纹；

如果该第六初始块的指纹与该第二检测块的指纹相同，则比较该第四子可变块的摘要与该第四子片段的摘要，该第四子可变块的摘要为根据该  
15 摘要算法对该第四子可变块的摘要进行计算得到的，该第四子片段的摘要为根据该摘要算法对该第四子片段的摘要进行计算得到的；

如果该第四子片段的摘要与该第四子可变块的摘要相同，则将该第四子片段、该第四子片段的指纹以及该第四子片段的摘要添加到该第二本地重复数据库，生成第四本地重复数据库，该第四子片段的指纹与该第六初  
20 始块的指纹相同。

可选地，在该第四子片段的摘要与该第四子可变块的摘要相同，将该第四子片段、该第四子片段的指纹以及该第四子片段的摘要添加到该第二本地重复数据库的场景下，本发明实施例提供的数据处理方法还可以包括：

接收第五报文，该第五报文包括第五报文头与第五净荷，该第五净荷

包含第四净荷片段以及第五净荷片段，该第四净荷片段的长度与该第一子片段相同，该第五净荷片段的长度与该第四子片段的长度相同；

根据该指纹算法计算该第四净荷片段中第七初始块的指纹，根据该指纹算法计算该第五净荷片段中第八初始块的指纹，该第七初始块的起始位置与该第四净荷片段的起始位置相同，该第七初始块的长度与该第一滑窗的长度相同，该第八初始块的起始位置与该第五净荷片段的起始位置相同，该第八初始块的长度与该第一滑窗的长度相同；

根据该摘要算法计算该第四净荷片段的摘要，根据该摘要算法计算该第五净荷片段的摘要；

如果该第七初始块的指纹与该第四本地重复数据库中的该第一子片段的指纹相等，并且该第四净荷片段的摘要与该第一子片段的摘要相等，则删除该第五报文中的该第四净荷片段，如果该第八初始块的指纹与该第四本地重复数据库中的该第四子片段的指纹相等，并且该第五净荷片段的摘要与该第四子片段的摘要相等，则删除该第五报文中的该第五净荷片段，生成第六报文，该第六报文中包括第六报文头与第六净荷，该第六报文头与该第五报文头相同，该第六净荷包括该第一子片段的指纹以及该第四子片段的指纹。

可选地，该第六净荷还包括该第一子片段在该第五报文中的位置信息、该第四子片段在该第五报文中的位置信息。

其中，第四子片段与第一子片段类似，为小于第二片段粒度的数据块，可以进一步提高数据压缩效率。第一子片段在第五报文中的位置信息可以是第一子片的最高比特相对于第五报文的报文头的最高比特的偏移量。

可选地，本发明实施例提供的数据处理方法还可以包括：

该第三片段的结束位置通过定界算法确定，包括：

从该第二比特开始向后滑动该定界算法中的第三滑窗，并判断该第三滑窗对应的数据是否符合该定界算法中的定界条件，当首次出现该第三滑窗对应的数据符合该定界条件时，在第一距离内继续向后滑动该第三滑窗，并判断该第三滑窗对应的数据是否符合该定界条件，如果出现该第三滑窗对应的第一数据符合该定界条件的情况时，则确定该第一数据的结束位置为该第三片段的结束位置，该第三滑窗的长度与该第一滑窗的长度相同，该第一距离的长度与该第一滑窗的长度相同。

如图 3 所示，待压缩数据块所属的上级数据块 A' 与重复数据块中对应的数据块 A 的摘要内容不同，可以理解为上次发送的数据块 A 经过修改后变为数据块 A'，并进行再次发送，这样，上次发送数据时在重复数据库中保存的数据块 A 与当前发送的数据中的数据块 A' 相对应，具体地，当滑窗滑动到数据块 A' 时，计算指纹，若重复数据库中有相同的指纹，则表示重复数据库中存在有对应的数据块，这里对应的数据块为 A。

其中，待压缩数据块所属的上级数据块 A' 可通过滑窗定界来确定。

计算数据块 A' 的摘要，并与数据块 A 的摘要比较。显然，数据块 A' 的摘要与数据块 A 的摘要不同，因为数据块 A 与数据块 A' 的内容不同。

之后再找到数据块 A' 中与数据块 A 相同的部分，以及不同的部分，进行进一步地压缩处理。

具体地，逐字节比较数据块 A' 与数据块 A，当重复数据库中对应的数据块 A 到第二拆分点，待压缩数据块 A' 到第一拆分点时，二者比较结果是不一致。此时，将滑窗滑动到第二拆分点处计算指纹，上级数据块 A' 中，将滑窗滑动到第一拆分点处计算指纹，与第二拆分点处的指纹进行比较，不一致时，继续向后滑动滑窗，计算第一拆分点后的指纹，并继续与第二拆分点处的指纹进行比较，不一致时，继续向后滑动滑窗，直至计算出的指纹与第二拆分点处的指纹一致，或直至滑窗滑至数据块 A' 的结束

边界。

本实施例中，数据块 A' 的第四拆分点处的指纹与重复数据库中数据块 A 的第二拆分点处的指纹一致。

当数据块 A' 的第四拆分点处的指纹与数据块 A 的第二拆分点处的指纹一致时，计算第四拆分点至边界的摘要，并计算重复数据库中数据块 A 的第二拆分点至边界的摘要，将计算出的两个摘要进行比较，当不一致时，按照前述操作比较第四拆分点后的字节与第一拆分点后的字节，直至计算出的指纹与第二拆分点处的指纹一致，或直至滑窗滑至数据块 A' 的结束边界；当一致时，上级数据块 A' 被第一拆分点、第四拆分点拆分为三个子数据块 A1、A3、A2，重复数据库中的数据块 A 被第二拆分点拆分为两个子数据块 A1、A2，可以看出，本次发送的数据块 A' 与上次发送的数据块 A 相比增加了子数据块 A3，其余两个子数据块 A1、A2 不变，这样，可以用指纹替换 A1 或者 A2，实现对数据的精确压缩，提高数据压缩效率。其中，待压缩数据块 A' 中的子数据块 A2 即上述第四子片段。待压缩数据块 A' 中的子数据块 A2 的初始块，也就是从待压缩数据块 A' 中的子数据块 A2 的第一比特开始长度为第一滑窗长度的数据块，即上述第四子片段中的第一检测块。

替换 A1 或者 A2 可以通过其他方法实现。举例来说，可以用指纹、块长度及偏移值替换 A1 或者 A2。

当图 1 所示实施例的方法由解压缩侧设备实现时，本发明实施例提供的数据处理方法还可包括：将该第一子片段的指纹以及该第一子片段的摘要同步到压缩侧的重复数据库。压缩侧的重复数据库被同步后，压缩侧设备可使用已有的 CDC 算法对报文进行删重处理即对报文中的数据进行压缩处理，这样，当报文中包含第一子片段时，就会被删除，从而进一步提高了数据压缩效率。

当图 1 所示实施例的方法由压缩侧设备实现时，本发明实施例提供的  
数据处理方法还可包括：将该第一子片段、该第一子片段的指纹以及该第  
一子片段的摘要同步到解压缩侧的重复数据库。这样，当后续报文中的第  
一子片段被删除后，解压缩侧设备就能够利用重复数据库中保存的第一子  
5 片段对压缩数据进行恢复，实现解压缩。

可选地，当压缩侧和解压缩侧的重复数据库均包含有重复数据的内容，  
且数据处理方法由压缩侧设备实现时，本发明实施例提供的数据处理方法  
还可以包括：

删除该待压缩数据中的该第一子片段，将该第一子片段的指纹、该第  
10 一子片段的长度及该第一子片段在该待压缩数据中的位置信息添加到第七  
报文的净荷中，该待压缩数据为该第七报文的净荷。

第一子片段在待压缩数据中的位置信息可以是第一子片的最高比特  
相对于待压缩数据的最高比特的偏移量。即，压缩侧设备在接收到该待压  
缩数据所在报文后，不仅能够匹配出第一子片段，在本地重复数据库中创  
15 建第一子片段，还能够对该待压缩数据进行压缩处理，删除第一子片段。  
其中，在去冗余报文中增加第一子片段的长度，是由于解压缩侧的重复数  
据库中已有第二片段，当去冗余报文发送到解压缩侧设备时，解压缩侧设  
备根据其中的第一子片段的指纹找到对应的第二片段，利用第一子片段的  
长度及第一子片段在该待压缩数据中的位置即第一子片段在压缩前报文中  
20 的原始位置，在第二片段中找到第一子片段，并对去冗余报文进行恢复，  
实现解压缩。

上述第三片段的结束位置通过定界算法确定，可包括：

从该第二比特开始向后滑动该定界算法中的第三滑窗，并判断该第三  
滑窗对应的数据是否符合该定界算法中的定界条件，当首次出现该第三滑  
25 窗对应的数据符合该定界条件时，在第一距离内继续向后滑动该第三滑窗，

并判断该第三滑窗对应的数据是否符合该定界条件，如果出现该第三滑窗对应的第一数据符合该定界条件的情况时，则确定该第一数据的结束位置为该第三片段的结束位置，该第三滑窗的长度与该第一滑窗的长度相同，该第一距离的长度与该第一滑窗的长度相同。

5 关于定界算法，可以参考 CDC 算法中的定界算法，此处不再赘述。

具体地，从上述第二比特开始向后滑动第三滑窗，判断第三滑窗对应的数据的指纹是否满足定界算法中的定界条件。

当首次出现第三滑窗对应的数据符合定界条件时，继续向后滑动第三滑窗。首次出现是指第一次出现第三滑窗对应的数据符合定界条件。第三滑窗的长度等于第一滑窗的长度。举例来说，第三滑窗的长度可以是 64 字节。定界算法中，第三滑窗向后滑动时，每次滑动的距离可以是 1 字节。

在第一距离内继续向后滑动第三滑窗，并判断第三滑窗对应的数据是否符合定界条件。如果在第一距离内多次出现第三滑窗对应的第一数据符合定界条件，则可以确定最后一次出现的第一数据为真正的边界。也就是说，可以确定最后一次出现的第一数据的结束位置为第三片段的结束位置。

图 4 为本发明实施例提供的一种数据压缩方法的一个具体实现方式的示意图。

如图 4 所示，根据 CDC 算法的定界原理分成三块的报文数据，其中，中间大块表示根据优化算法分成了三个子块。

20 假设该报文数据发送了三次，第一次发送时，子块 3356B、4505B、4520B 及 2988B 作为一个数据块发送，并被保存在重复数据库中。

第二次发送时，将该数据块拆分第一子块 3356B 及第二子块，第二子块由子块 4505B、4520B 与 2988B 组成。并且重复数据库中分别保存了第一子块 3356B 及第二子块。

第三次发送时，第一子块 3356B 成为数据块 3356B，第二子块中删除了子块 4520B，子块 4505B 及 2988B 作为一个数据块假设为数据块 C。则在本次发送进行压缩处理时，需要将数据块 C 拆分为子块 4505B 及子块 2988B。

5 具体地，当第三次发送数据时，以压缩数据块 6788B、数据块 3356B、数据块 C 及数据块 4800B 为例进行说明。

对图 4 中的数据块 6788B、数据块 3356B、数据块 C 及数据块 4800B 执行压缩操作，可以采用图 5 所示的数据处理方法。图 5 为本发明实施例提供的另一种数据处理方法的流程图。该数据处理方法包括：

10 51、滑动滑窗，计算指纹；这里的滑窗即内容滑窗，实现上内容滑动窗口可以与定界滑窗（即现有技术中的滑窗）合一，即对同样长度的数据进行 HASH 计算得到指纹，只是内容滑窗将指纹应用于判别数据块的内容是否发生变化，而定界滑窗将指纹应用于数据块定界。

15 52、判断是否存在 HASH 表项。具体地，用内容滑窗计算得到的指纹作为查找关键字（KEY）查找重复数据库，判断重复数据库是否存在 HASH 表项，该 HASH 表项包含有上述 51 计算得到的指纹。

若不存在 HASH 表项，说明指纹对应的数据块为本次新增加的内容，执行 531，将新增加内容添加到重复数据库中。

20 若存在 HASH 表项，说明指纹对应的数据块曾经被保存到重复数据库中，则执行 541，取 HASH 表项中的块信息，根据 HASH 表项中所保存的块信息（关键信息是块长度、SHA-1/MD5 算法得到的数据即摘要（checksum））来与当前数据进行比较。当然，块信息还可以包含该数据块的原始数据内容。

531、判断计算出的指纹是否满足定界条件。采用 CDC 算法中类似的

滑动窗口（称为定界滑窗）进行块定界，定界方法相同，即计算出指纹，并根据指纹判断是否满足定界条件，如果满足则确定找到数据的边界。

此处对 CDC 算法的定界过程有一点优化，因为有可能本数据块的内容发生了变化，该内容导致定界条件提前成立，而变化内容被计算到下一数据块中，结果可能导致下一块不能被判定为重复数据（实际上这部分确实为重复数据），因引考虑第一次满足条件的界为伪界，此时再向后滑动一部分距离（可称为数据扰动），如果再次发现定界条件满足，则认为后一次为真正的界（可以称为确界，前一次称为伪界），否则前一次就是真正的界（确界）。

10 532、最大界检查。通常数据块的长度有范围限制，可预先设定一个上限，如最长不能超过 64Kbyte（字节）。这里判断经过上述 531 确定的数据块的长度是否超过上限，若超过该上限，则以该上限作为数据块的长度，强制距离上述 51 滑窗字节为该上限的字节处定一个界，重新划定数据块，然后执行 534；若 531 确定的数据块的长度未超过该上限，则执行 532。

15 533、最小界检查。参见上述 532，通常不仅预设一个上限来限定数据块可具有的最长的长度，还会预先设定一个下限来限定数据块可具有的最短的长度，如最短不能短于 64 字节。小于下限的情况作为零碎数据不压缩，这是为了防止数据库的数据块太零碎，占用系统资源，使得压缩过程没有意义，假如一个字节就是一个块，这是没有意义的，反而更加浪费系统资源。  
20 源。

若 531 确定的数据块的长度大于该下限，则执行 534；若 531 确定的数据块的长度小于该下限，执行 57；

534、确定可变块，计算摘要，然后执行 56；

541、计算当前摘要是否一致。

具体的，计算当前数据块的摘要，并与重复数据库中的摘要进行比较。判断是否一致；若是，执行 58；否则，执行 542。

具体地，利用重复数据库中 HASH 表项记录的块长度直接确定上述步骤 51 滑窗对应的字节所属的数据块，然后计算该数据块的摘要。具体地，  
5 按照 HASH 表项中保存的块长度值定位当前数据的当前块即指纹对应的字节所属的数据块，通过 SHA-1/MD5 等算法计算当前块的 checksum。计算得到的 checksum 与表项中保存的 checksum 进行比较，如果相同，则认为当前块为重复数据，将定界滑窗和内容滑窗都直接滑过本次找到的数据块，进行下一块的定界、内容比对。对于图 4 中的数据块 3356B，按照上述方  
10 法，这次找到的实际上是一个重复块，此后先将内容滑窗滑到下一数据块 C，检查其对应的指纹是否在重复数据库中存在，即利用内容滑窗指纹作为查找 KEY 在重复数据库中查找 HASH 表项。查找到 HASH 表项后，利用表项中的块长度计算数据块 C 的 checksum，并与表项中的 checksum 进行比较，当检查到当前计算的 checksum 与表项中保存的 checksum 不一致，说  
15 明数据块 C 与重复数据库中对应的数据块的内容不同，执行 542 拆分数据块 C 及重复数据库中对应的数据块。

542、逐字节匹配拆分，对于每一个拆分出的子数据块，执行 55；完成拆分后，对于拆分出的所有满足数据块的长度限制的子数据块执行 543。在拆分过程中形成块的都要检查，不管是匹配部分还是不匹配部分，小于最  
20 小界的都作为零碎数据。在拆分过程中产生了新块都涉及到更新/删除表项，这是拆分过程中完成的。

具体地，从当前块的初始部分往后逐字节进行匹配，直到找到与重复数据库中不相同的字节，将当前块从不相同的字节处进行拆分，另外利用滑窗，检查从拆分点向后，是否有指纹与当前块的后续指纹相同，如果存  
25 在，检查两者完全匹配的部分，将其作为一个新块保存到重复数据库中，

此过程主要是识别原来的数据中被删除掉一小部分的情况。

543、更新或删除 HASH 表项；

544、计算拆分摘要及块长度，然后执行 56；

55、对子数据块进行最小界检查，若子数据块的长度大于上述下限，  
5 则对该子数据块执行 543；若子数据块的长度小于上述下限，则对该子数据  
块执行 57。55 可在执行 542 的过程中执行，即 542 每拆分出一部分数据，  
就执行 55，进行最小界检查，拆分完成后，也就对所有拆分出的子数据块  
完成了最小界检查，然后对所有拆分出的长度大于下限的数据块执行 543。  
或者，55 也可在 542 与 543 之间执行，即 542 先完成数据块的拆分，然后  
10 执行 55，对于拆分出的所有子数据块执行 55，再对 55 的判断结果执行 57  
或 543。

56、添加该可变块对应的 HASH 表项。

具体地，在重复数据库中，增加该数据块的内容，并添加 HASH 表项，  
即更新重复数据库，结束当前数据块的压缩处理。

15 57、非重复内容处理。具体地，单独处理非块内容即上述长度小于下  
限的零碎数据，这部分数据保留在报文中，不压缩，完成当前滑窗内字节  
的处理，继续向后滑动滑窗。

58、内容全部重复，重复内容替换成指纹，完成当前数据块的删除处  
理。当前块拆分完成后，继续滑动内容滑窗，检查后续数据块的内容，对  
20 当前块的后续数据进行处理。

上述流程中，定界滑窗和内容滑窗可为同一个滑窗，只是根据计算的  
结果（指纹）进行数据块定界、重复数据库中块内容的查找比较，据此，  
压缩方法流程分为定界滑窗流程和内容滑窗流程，最后的处理结果存在三  
种情况：一种是对原来的数据块进行了拆分，更新了重复数据库；另一种

是未能匹配到块内容并且不符合加入重复数据库条件（如长度太短），这部分数据作为非重复内容直接放到报文中；还有一种结果是找到重复数据，此时将表示该重复数据的指纹及相关信息以一定格式放到报文中，发送给对方设备。

5 可选地，压缩后的报文的封装格式如图 6 所示，方式 1 中非重复数据的位置不变，只是非重复数据的前后的数据块作为重复数据而被替换了，方式 2 中替换格式在压缩后的报文中的位置被指定。其中，替换格式包括指纹、块长度及被替换的数据块在本报文中的偏移。

具体应用到数据块 6788B 时，首先滑窗滑动到数据块 6788B 的初始部分时，执行上述 51 操作，计算出初始部分的指纹，然后执行 52，判断重复数据库是否存在包含该指纹的 HASH 表项，本实例中数据块 6788B 为重复数据块，因此，执行 541，计算出数据块 6788B 的摘要，并与重复数据库中 HASH 表项保存的摘要相比较，判断是否一致，本实例中判断结果一致，即内容部分重复，执行 58，删除数据中数据块 6788B 的内容，替换为指纹，  
15 完成对数据块 6788B 的处理。

之后滑窗滑动到数据块 3356B 的初始部分，执行 51，计算出初始部分的指纹；执行 52，从重复数据库中找到对应 HASH 表项，获知数据块 3356B 的块长度等块信息。执行 541，计算出数据块 3356B 的摘要，与重复数据块的 HASH 表项中的摘要一致，说明数据块 3356B 的内容与重复数据库中  
20 对应 HASH 表项的数据块内容全部重复，执行 58，删除数据中数据块 3356B 的内容，替换为指纹，完成对数据块 3356B 的处理。

之后滑窗滑动到数据块 C 的初始部分，执行 51，计算出初始部分的指纹；执行 52，查找重复数据库，判断是否存在包含该指纹的 HASH 表项。本实例中，重复数据库中存在数据块 C 的 HASH 表项。然后执行 541，根据查找到的 HASH 表项中的块长度界定数据块 C，计算出数据块 C 的摘要，  
25

并判断是否与 HASH 表项中的摘要一致。本实例中，数据块 C 的摘要与 HASH 表项中的摘要不一致，然后执行 542。从数据块 C 的初始部分开始向后逐字节与 HASH 表项对应的数据块内容进行比较，直到找到不相同的字节，将数据块 C 及重复数据库中 HASH 表项对应的数据块从此处即不相同的字节处进行拆分，本实例中，将数据块 C 拆分为子块 4505B 及子块 2988B，其中，子块 4505B 与重复数据库中的数据重复，但是由于本次发送的数据块 C 中删除了子块 4505B 与子块 2988B 之间的子块 4520B，导致数据块 C 的内容与重复数据库中的内容不匹配。然后执行 543，将拆分出的部分 4505B、2988B 均作为一个新数据块保存到重复数据库中，并生成新的 HASH 表项替换之前查找到的 HASH 表项。相应地，重复数据库中的 4520B 满足数据块长度要求时，作为新的数据块保存，否则，删除。再执行 544，计算子块 4505B、子块 2988B 的块长度及摘要，添加到新的 HASH 表项中，以用于下次发送数据删除作为数据块的 4505B、2988B 的数据内容，并进行替换，提高数据的压缩效率。

对于待压缩的数据，继续向后滑动滑窗至数据块 4800B 的初始部分，执行 51，计算初始部分的指纹。执行 52，从重复数据库中找到对应 HASH 表项，获知数据块 4800B 的块长度等块信息。执行 541，计算出数据块 4800B 的摘要，与重复数据块的 HASH 表项中的摘要一致，说明数据块 4800B 的内容与重复数据库中对应 HASH 表项的数据块内容全部重复，执行 58，删除数据中数据块 4800B 的内容，替换为指纹，完成对数据块 4800B 的处理。

按照上述方法完成对数据的压缩后，压缩侧设备如 WOC 将报文发送到解压缩侧设备。

如图 7 所示，压缩侧设备接收报文后，对接收到的报文进行解析，得到报文的净荷。根据本发明实施例提供的数据处理方法对净荷中的数据进行压缩处理。将压缩处理后的数据封装为新报文。将新报文发送到解压缩

侧设备。

解压缩侧设备接收到新报文后，利用重复数据库对新报文中的净荷进行解压缩处理。具体可以是将报文中的替换格式或指纹删除，并将替换格式或指纹替换为与替换格式或指纹对应的数据块，将解压缩的数据封装为报文。

其中，重复数据库的生成（即前面的算法过程）可以只在一端设备完成，然后将数据（重复数据块指纹和/或数据内容）同步到对端。当数据流量主要为下行，即服务器侧流向用户侧（如用户下载文件），算法可以在靠近用户侧的 WOC 设备实现，并且保存重复数据块内容，而靠近服务器一侧的 WOC 设备只需要同步指纹信息即可，不需要缓存所有内容，反之则在服务器侧的 WOC 设备处理算法和保存内容。当然两侧可以都计算和保存内容，只是会增加资源需求。

图 8 为本发明实施例提供的数据压缩设备的结构示意图。本实施例提供的数据压缩设备用于实施上述图 1、图 5 所示实施例的方法，如图 8 所示，数据处理设备包括：第一指纹计算单元 81、查找单元 82、重复内容获取单元 83、摘要计算单元 84、摘要比较单元 85、第一子片段获取单元 86 及第一添加单元 87。

第一指纹计算单元 81 用于根据指纹算法计算待压缩数据中的第一片段的第一指纹，该第一片段的起始位置与该待压缩数据的起始位置相同，该第一片段的长度与第一滑窗的长度相同；

查找单元 82 用于在第一本地重复数据库中查找该第一指纹，该第一本地重复数据库用于存储重复数据、该重复数据的指纹以及该重复数据的摘要；

重复内容获取单元 83 用于如果该第一本地重复数据库中不存在该第一指纹，则根据该第一指纹获取该第一本地重复数据库中的第一可变块以及该第一可变块的摘要，该第一指纹与根据该指纹算法计算得到的该第一可变

块中的第一初始块的指纹相同，该第一初始块的起始位置与该第一可边块的起始位置相同，该第一初始块的长度与该第一滑窗的长度相同，该第一可变块的摘要为根据摘要算法对该第一可变块的摘要进行计算得到的；

摘要计算单元 84 用于根据该摘要算法计算该待压缩数据中的第二片段的摘要，该第二片段的起始位置与该待压缩数据的起始位置相同，该第二片段的长度与该第一可变块的长度相同；

摘要比较单元 85 用于比较该第二片段的摘要与该第一可变块的摘要；

第一子片段获取单元 86 用于如果该第二片段的摘要与该第一可变块的摘要不同，则获取该第二片段中的第一子片段，该第一子片段与该第一可变块中的第一子可变块相同，该第一子片段的起始位置与该第二片段的起始位置相同，该第一子可变块的起始位置与该第一可变块的起始位置相同，该第二片段中的第二比特与该第一可变块中的第一比特不同，该第二比特为该第二片段中该第一子片段的下一个比特，该第一比特为该第一可变块中该第一子可变块的下一个比特；

第一添加单元 87 用于将该第一子片段、该第一子片段的指纹以及该第一子片段的摘要添加到该第一本地重复数据库，生成第二本地重复数据库，该第一子片段的指纹与该第一指纹相同，该第一子片段的摘要为根据该摘要算法对该第一子片段的摘要进行计算得到的。

本发明实施例提供的数据压缩设备还可包括：

第一报文接收单元，用于接收第一报文，该第一报文包括第一报文头与第一净荷，该第一净荷包含第一净荷片段，该第一净荷片段的长度与该第一子片段的长度相同；

第二初始指纹计算单元，用于根据该指纹算法计算该第一净荷片段中第二初始块的指纹，该第二初始块的起始位置与该第一净荷片段的起始位

置相同，该第二初始块的长度与该第一滑窗的长度相同；

第一净荷摘要计算单元，用于根据该摘要算法计算该第一净荷片段的摘要；

第二报文生成单元，用于如果该第二初始块的指纹与该第二本地重复  
5 数据库中的该第一子片段的指纹相等，并且该第一净荷片段的摘要与该第一子片段的摘要相等，则删除该第一报文中的该第一净荷片段，生成第二报文，该第二报文中包括第二报文头与第二净荷，该第二报文头与该第一报文头相同，该第二净荷包括该第一子片段的指纹。

该第二净荷还可包括该第一子片段在该第一报文中的位置信息。第一  
10 子片段在第一报文中的位置信息可以是第一子片的最高比特相对于第一报文的报文头的最高比特的偏移量。

本发明实施例提供的数据压缩设备还可包括：

第三初始指纹计算单元，用于根据该指纹算法计算该待压缩数据中的  
第二子片段的第三初始块的指纹，该第二子片段的起始位置为该第二比特，  
15 该第二子片段的结束位置与该待压缩数据的结束位置相同；该第三初始块的起始位置为该第二比特，该第三初始块的长度等于该第一滑窗的长度；

检测块获取单元，用于根据第二滑窗获取第二子可变块中的第一检测  
块，该第二滑窗的起始位置与该第二子可变块中的第三比特对应，该第三  
比特为介于第二子可变块的起始位置与该第二子可变块的结束位置之间的  
20 比特，该第二滑窗的长度等于该第一滑窗的长度，该第二子可变块的起始位置为该第一比特，该第二子可变块的结束位置与该第一可变块的结束位置相同；

检测块指纹计算单元，根据该指纹算法计算该第一检测块的指纹；

指纹比较单元，用于比较该第三初始块的指纹与该第一检测块的指纹；

第三摘要比较单元，用于如果该第三初始块的指纹与该第一检测块的指纹相同，则比较该第二子可变块中的第三子可变块的摘要与该第二子片段中的第三子片段的摘要，该第三子可变块的起始位置与该第一检测块的起始位置相同，该第三子可变块的结束位置与该第一可变块的结束位置相同，该第三子片段的起始位置与该第二子片段的起始位置相同，该第三子片段的长度与该第三子可变块的长度相等，该第三子可变块的摘要为根据该摘要算法对该第三子可变块的摘要进行计算得到的；

第三添加单元，用于如果该第三子片段的摘要与该第三子可变块的摘要相等，则将该第三子片段、该第三子片段的指纹以及该第三子片段的摘要添加到该第二本地重复数据库，生成第三本地重复数据库，该第三子片段的指纹等于该第三初始块的指纹，该第三子片段的摘要为根据该摘要算法对该第三子片段的摘要进行计算得到的。

本发明实施例提供的数据压缩设备还可包括：

第三报文接收单元，用于接收第三报文，该第三报文包括第三报文头与第三净荷，该第三净荷包含第二净荷片段以及第三净荷片段，该第二净荷片段的长度与该第一子片段的长度相同，该第三净荷片段的长度与该第三子片段的长度相同；

第四初始指纹计算单元，用于根据该指纹算法计算该第二净荷片段中第四初始块的指纹，根据该指纹算法计算该第三净荷片段中第五初始块的指纹，该第四初始块的起始位置与该第二净荷片段的起始位置相同，该第四初始块的长度与该第一滑窗的长度相同，该第五初始块的起始位置与该第三净荷片段的起始位置相同，该第五初始块的长度与该第一滑窗的长度相同；

第三净荷摘要计算单元，用于根据该摘要算法计算该第二净荷片段的摘要，根据该摘要算法计算该第三净荷片段的摘要；

第四报文生成单元，用于如果该第四初始块的指纹与该第三本地重复数据库中的该第一子片段的指纹相等，并且该第二净荷片段的摘要与该第一子片段的摘要相等，则删除该第三报文中的该第二净荷片段，如果该第五初始块的指纹与该第三本地重复数据库中的该第三子片段的指纹相等，并且该第三净荷片段的摘要与该第三子片段的摘要相等，则删除该第三报文中的该第三净荷片段，生成第四报文，该第四报文中包括第四报文头与第四净荷，该第四报文头与该第三报文头相同，该第四净荷包括该第一子片段的指纹以及该第三子片段的指纹。

该第四净荷还包括该第一子片段在该第三报文中的位置信息、该第三子片段在该第三报文中的位置信息。第一子片段在第三报文中的位置信息可以是第一子片的最高比特相对于第三报文的报文头的最高比特的偏移量。

本发明实施例提供的数据压缩设备还可包括：

初始块指纹计算单元，用于根据该指纹算法计算该第一可变块中的第四子可变块的第六初始块的指纹，该第四子可变块的起始位置为该第一比特，该第四子可变块的结束位置与该第一可变块的结束位置相同，该第六初始块的起始位置为该第一比特，该第六初始块的长度等于该第一滑窗的长度；

检测块获取单元，用于获取第四子片段中的第二检测块，该第二检测块的起始位置与该第四子片段中的第四比特对应，该第四比特为介于第三片段的起始位置与该第三片段的结束位置之间的比特，该第二检测块的长度等于该第一滑窗的长度，该第三片段的起始位置为该第二比特，该第三片段的结束位置通过定界算法确定，该第三片段为该待压缩数据中的片段，该第四子片段为该第三片段中的子片段，该第四子片段的起始位置与该第二检测块的起始位置相同，该第四子片段的长度与该第四子可变块的长度

相同；

检测块指纹计算单元，用于根据该指纹算法计算该第二检测块的指纹；

指纹比较单元，用于比较该第六初始块的指纹与该第二检测块的指纹；

5 第四摘要比较单元，用于如果该第六初始块的指纹与该第二检测块的指纹相同，则比较该第四子可变块的摘要与该第四子片段的摘要，该第四子可变块的摘要为根据该摘要算法对该第四子可变块的摘要进行计算得到的，该第四子片段的摘要为根据该摘要算法对该第四子片段的摘要进行计算得到的；

10 第四添加单元，用于如果该第四子片段的摘要与该第四子可变块的摘要相同，则将该第四子片段、该第四子片段的指纹以及该第四子片段的摘要添加到该第二本地重复数据库，生成第四本地重复数据库，该第四子片段的指纹与该第六初始块的指纹相同。

本发明实施例提供的数据压缩设备还可包括：

15 第五报文接收单元，用于接收第五报文，该第五报文包括第五报文头与第五净荷，该第五净荷包含第四净荷片度以及第五净荷片段，该第四净荷片度的长度与该第一子片段相同，该第五净荷片段的长度与该第四子片段的长度相同；

20 第七初始指纹计算单元，用于根据该指纹算法计算该第四净荷片段中第七初始块的指纹，根据该指纹算法计算该第五净荷片段中第八初始块的指纹，该第七初始块的起始位置与该第四净荷片段的起始位置相同，该第七初始块的长度与该第一滑窗的长度相同，该第八初始块的起始位置与该第五净荷片段的起始位置相同，该第八初始块的长度与该第一滑窗的长度相同；

第五净荷摘要计算单元，用于根据该摘要算法计算该第四净荷片段的

摘要，根据该摘要算法计算该第五净荷片段的摘要；

第六报文生成单元，用于如果该第七初始块的指纹与该第四本地重复数据库中的该第一子片段的指纹相等，并且该第四净荷片段的摘要与该第一子片段的摘要相等，则删除该第五报文中的该第四净荷片段，如果该第八初始块的指纹与该第四本地重复数据库中的该第四子片段的指纹相等，并且该第五净荷片段的摘要与该第四子片段的摘要相等，则删除该第五报文中的该第五净荷片段，生成第六报文，该第六报文中包括第六报文头与第六净荷，该第六报文头与该第五报文头相同，该第六净荷包括该第一子片段的指纹以及该第四子片段的指纹。

该第六净荷还可包括该第一子片段在该第五报文中的位置信息、该第四子片段在该第五报文中的位置信息。第一子片段在第五报文中的位置信息可以是第一子片的最高比特相对于第五报文的报文头的最高比特的偏移量。

可选地，从该第二比特开始向后滑动该定界算法中的第三滑窗，并判断该第三滑窗对应的数据是否符合该定界算法中的定界条件，当首次出现该第三滑窗对应的数据符合该定界条件时，在第一距离内继续向后滑动该第三滑窗，并判断该第三滑窗对应的数据是否符合该定界条件，如果出现该第三滑窗对应的第一数据符合该定界条件的情况时，则确定该第一数据的结束位置为该第三片段的结束位置，该第三滑窗的长度与该第一滑窗的长度相同，该第一距离的长度与该第一滑窗的长度相同。

本发明实施例提供的数据压缩设备还可包括：

第二同步单元，用于将该第一子片段、该第一子片段的指纹以及该第一子片段的摘要同步到解压缩侧的重复数据库。

本发明实施例提供的数据压缩设备还可包括：

压缩处理单元, 用于删除该待压缩数据中的该第一子片段, 将该第一子片段的指纹、该第一子片段的长度及该第一子片段在该待压缩数据中的位置信息添加到第七报文的净荷中, 该待压缩数据为该第七报文的净荷。

5 第一子片段在待压缩数据中的位置信息可以是第一子片段的最高比特相对于待压缩数据的最高比特的偏移量。

本发明实施例提供的数据压缩设备可为 WOC。本领域技术人员应理解, WOC 除了上述功能单元, 还应具有报文接收、网络连接等基本功能单元, 这里不再赘述。

本发明实施例提供的数据解压缩设备包括:

10 第一指纹计算单元, 用于根据指纹算法计算待压缩数据中的第一片段的第一指纹, 该第一片段的起始位置与该待压缩数据的起始位置相同, 该第一片段的长度与第一滑窗的长度相同;

查找单元, 用于在第一本地重复数据库中查找该第一指纹, 该第一本地重复数据库用于存储重复数据、该重复数据的指纹以及该重复数据的摘要;

15 重复内容获取单元, 用于如果该第一本地重复数据库中不存在该第一指纹, 则根据该第一指纹获取该第一本地重复数据库中的第一可变块以及该第一可变块的摘要, 该第一指纹与根据该指纹算法计算得到的该第一可变块中的第一初始块的指纹相同, 该第一初始块的起始位置与该第一可变块的起始位置相同, 该第一初始块的长度与该第一滑窗的长度相同, 该第一可变块  
20 的摘要为根据摘要算法对该第一可变块的摘要进行计算得到的;

摘要计算单元, 用于根据该摘要算法计算该待压缩数据中的第二片段的摘要, 该第二片段的起始位置与该待压缩数据的起始位置相同, 该第二片段的长度与该第一可变块的长度相同;

摘要比较单元, 用于比较该第二片段的摘要与该第一可变块的摘要;

第一子片段获取单元,用于如果该第二片段的摘要与该第一可变块的摘要不同,则获取该第二片段中的第一子片段,该第一子片段与该第一可变块中的第一子可变块相同,该第一子片段的起始位置与该第二片段的起始位置相同,该第一子可变块的起始位置与该第一可变块的起始位置相同,该第二片段中的第二比特与该第一可变块中的第一比特不同,该第二比特为该第二片段中该第一子片段的下一个比特,该第一比特为该第一可变块中该第一子可变块的下一个比特;

第一添加单元,用于将该第一子片段、该第一子片段的指纹以及该第一子片段的摘要添加到该第一本地重复数据库,生成第二本地重复数据库,该第一子片段的指纹与该第一指纹相同,该第一子片段的摘要为根据该摘要算法对该第一子片段的摘要进行计算得到的。

本领域技术人员应理解为:本发明实施例提供的数据解压缩设备除具备上述单元外,还具有解压缩等基本功能单元。

本发明实施例提供的数据解压缩设备还可包括:

第一同步单元,用于将该第一子片段的指纹以及该第一子片段的摘要同步到压缩侧的重复数据库。

图9为本发明实施例提供的一种数据压缩方法的一种应用场景的组网结构图。如图9所示,在WAN两端用来压缩数据及解压缩数据的设备均为WOC。即,WOC可同时具备压缩数据和解压缩数据的功能。

图10为本发明实施例提供的一种数据压缩方法的另一种应用场景的组网结构图。如图10所示,两端WOC设备的交互包括控制和数据两个层面。

控制层面是两端设备需要通过私有协议交互控制信息,自动发现对端设备,通过下游设备发送响应到上游设备,可以使上游设备感知下游设备的存在及地址信息(如IP地址)。并且,两端设备需要互相通知采用的算

法信息并达成一致（如都采用 MD5 进行内容 checksum 的计算），还需要同步重复数据库信息，包括指纹及指纹对应的信息表项（如块长度，checksum 值等）。同步过程分为定时同步和增量同步，定时同步指定时将所有的重复数据库内容进行一致性检查，而增量同步则只对新增或更新/删除的特定块信息通知对端。

数据层面是指对通过两端设备的业务数据进行算法处理和数据替换，从而减少报文的长度和个数，达到数据压缩的目的。

根据本发明实施例提供的技术方案，如果待压缩的数据中包含了与重复数据库中的可变块的前半部分相同并且与可变块的后半部分不同的数据片段，则能够生成粒度小于发生匹配的可变块的新的可变块，并将新的可变块添加到重复数据库。

相对于直接对定位到的块进行查找方法，新的可变块粒度较小。因此，本发明实施例提供的技术方案提高了后续的待压缩数据与更新后的重复数据库发生匹配的概率，进而提高了压缩的效率。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅

是示意性的，例如，所述单元的划分，可以仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置  
5 或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

10 另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用  
15 时，可以存储在一个计算机可读取存储介质中。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或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，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  
20 等）执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（英文缩写为 ROM，英文全称为 Read-Only Memory）、随机存取存储器（英文缩写为 RAM，英文全称为 Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可

轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应所述以权利要求的保护范围为准。

## 权利要求

1、一种数据数据处理方法，其特征在于，包括：

根据指纹算法计算待压缩数据中的第一片段的第一指纹，所述第一片段的起始位置与所述待压缩数据的起始位置相同，所述第一片段的长度与  
5 第一滑窗的长度相同；

在第一本地重复数据库中查找所述第一指纹，所述第一本地重复数据库用于存储重复数据、所述重复数据的指纹以及所述重复数据的摘要；

如果所述第一本地重复数据库中不存在所述第一指纹，则根据所述第一指纹获取所述第一本地重复数据库中的第一可变块以及所述第一可变块的  
10 摘要，所述第一指纹与根据所述指纹算法计算得到的所述第一可变块中的第一初始块的指纹相同，所述第一初始块的起始位置与所述第一可边块的起始位置相同，所述第一初始块的长度与所述第一滑窗的长度相同，所述第一可变块的摘要为根据摘要算法对所述第一可变块的摘要进行计算得到的；

15 根据所述摘要算法计算所述待压缩数据中的第二片段的摘要，所述第二片段的起始位置与所述待压缩数据的起始位置相同，所述第二片段的长度与所述第一可变块的长度相同；

比较所述第二片段的摘要与所述第一可变块的摘要；

如果所述第二片段的摘要与所述第一可变块的摘要不同，则获取所述  
20 第二片段中的第一子片段，所述第一子片段与所述第一可变块中的第一子可变块相同，所述第一子片段的起始位置与所述第二片段的起始位置相同，所述第一子可变块的起始位置与所述第一可变块的起始位置相同，所述第二片段中的第二比特与所述第一可变块中的第一比特不同，所述第二比特为所述第二片段中所述第一子片段的下一个比特，所述第一比特为所述第  
25 一可变块中所述第一子可变块的下一个比特；

将所述第一子片段、所述第一子片段的指纹以及所述第一子片段的摘要添加到所述第一本地重复数据库，生成第二本地重复数据库，所述第一子片段的指纹与所述第一指纹相同，所述第一子片段的摘要为根据所述摘要算法对所述第一子片段的摘要进行计算得到的。

5        2、根据权利要求 1 所述方法，其特征在于，还包括：

接收第一报文，所述第一报文包括第一报文头与第一净荷，所述第一净荷包含第一净荷片段，所述第一净荷片段的长度与所述第一子片段的长度相同；

10        根据所述指纹算法计算所述第一净荷片段中第二初始块的指纹，所述第二初始块的起始位置与所述第一净荷片段的起始位置相同，所述第二初始块的长度与所述第一滑窗的长度相同；

根据所述摘要算法计算所述第一净荷片段的摘要；

15        如果所述第二初始块的指纹与所述第二本地重复数据库中的所述第一子片段的指纹相等，并且所述第一净荷片段的摘要与所述第一子片段的摘要相等，则删除所述第一报文中的所述第一净荷片段，生成第二报文，所述第二报文中包括第二报文头与第二净荷，所述第二报文头与所述第一报文头相同，所述第二净荷包括所述第一子片段的指纹。

3、根据权利要求 2 所述方法，其特征在于，所述第二净荷还包括所述第一子片段在所述第一报文中的位置信息。

20        4、根据权利要求 1 所述方法，其特征在于，所述获取所述第二片段中的第一子片段之后，所述方法还包括：

根据所述指纹算法计算所述待压缩数据中的第二子片段的第三初始块的指纹，所述第二子片段的起始位置为所述第二比特，所述第二子片段的结束位置与所述待压缩数据的结束位置相同；所述第三初始块的起始位置

为所述第二比特，所述第三初始块的长度等于所述第一滑窗的长度；

根据第二滑窗获取第二子可变块中的第一检测块，所述第二滑窗的起始位置与所述第二子可变块中的第三比特对应，所述第三比特为介于第二子可变块的起始位置与所述第二子可变块的结束位置之间的比特，所述第二滑窗的长度等于所述第一滑窗的长度，所述第二子可变块的起始位置为  
5 所述第一比特，所述第二子可变块的结束位置与所述第一可变块的结束位置相同；

根据所述指纹算法计算所述第一检测块的指纹；

比较所述第三初始块的指纹与所述第一检测块的指纹；

10 如果所述第三初始块的指纹与所述第一检测块的指纹相同，则比较所述第二子可变块中的第三子可变块的摘要与所述第二子片段中的第三子片段的摘要，所述第三子可变块的起始位置与所述第一检测块的起始位置相同，所述第三子可变块的结束位置与所述第一可变块的结束位置相同，所述第三子片段的起始位置与所述第二子片段的起始位置相同，所述第三子  
15 片段的长度与所述第三子可变块的长度相等，所述第三子可变块的摘要为根据所述摘要算法对所述第三子可变块的摘要进行计算得到的；

如果所述第三子片段的摘要与所述第三子可变块的摘要相等，则将所述第三子片段、所述第三子片段的指纹以及所述第三子片段的摘要添加到所述第二本地重复数据库，生成第三本地重复数据库，所述第三子片段的  
20 指纹等于所述第三初始块的指纹，所述第三子片段的摘要为根据所述摘要算法对所述第三子片段的摘要进行计算得到的。

5、根据权利要求4所述方法，其特征在于，还包括：

接收第三报文，所述第三报文包括第三报文头与第三净荷，所述第三净荷包含第二净荷片段以及第三净荷片段，所述第二净荷片段的长度与所

述第一子片段的长度相同，所述第三净荷片段的长度与所述第三子片段的长度相同；

根据所述指纹算法计算所述第二净荷片段中第四初始块的指纹，根据所述指纹算法计算所述第三净荷片段中第五初始块的指纹，所述第四初始块的起始位置与所述第二净荷片段的起始位置相同，所述第四初始块的长度与所述第一滑窗的长度相同，所述第五初始块的起始位置与所述第三净荷片段的起始位置相同，所述第五初始块的长度与所述第一滑窗的长度相同；

根据所述摘要算法计算所述第二净荷片段的摘要，根据所述摘要算法计算所述第三净荷片段的摘要；

如果所述第四初始块的指纹与所述第三本地重复数据库中的所述第一子片段的指纹相等，并且所述第二净荷片段的摘要与所述第一子片段的摘要相等，则删除所述第三报文中的所述第二净荷片段，如果所述第五初始块的指纹与所述第三本地重复数据库中的所述第三子片段的指纹相等，并且所述第三净荷片段的摘要与所述第三子片段的摘要相等，则删除所述第三报文中的所述第三净荷片段，生成第四报文，所述第四报文中包括第四报文头与第四净荷，所述第四报文头与所述第三报文头相同，所述第四净荷包括所述第一子片段的指纹以及所述第三子片段的指纹。

6、根据权利要求 4 所述方法，其特征在于，所述第四净荷还包括所述第一子片段在所述第三报文中的位置信息、所述第三子片段在所述第三报文中的位置信息。

7、根据权利要求 1 所述方法，其特征在于，所述方法还包括：

根据所述指纹算法计算所述第一可变块中的第四子可变块的第六初始块的指纹，所述第四子可变块的起始位置为所述第一比特，所述第四子可变块的结束位置与所述第一可变块的结束位置相同，所述第六初始块的起

始位置为所述第一比特，所述第六初始块的长度等于所述第一滑窗的长度；

获取第四子片段中的第二检测块，所述第二检测块的起始位置与所述第四子片段中的第四比特对应，所述第四比特为介于第三片段的起始位置与  
5 所述第三片段的结束位置之间的比特，所述第二检测块的长度等于所述第一滑窗的长度，所述第三片段的起始位置为所述第二比特，所述第三片段的结束位置通过定界算法确定，所述第三片段为所述待压缩数据中的片段，所述第四子片段为所述第三片段中的子片段，所述第四子片段的起始位置与  
10 所述第二检测块的起始位置相同，所述第四子片段的长度与所述第四子可变块的长度相同；

10 根据所述指纹算法计算所述第二检测块的指纹；

比较所述第六初始块的指纹与所述第二检测块的指纹；

如果所述第六初始块的指纹与所述第二检测块的指纹相同，则比较所述第四子可变块的摘要与所述第四子片段的摘要，所述第四子可变块的摘要为根据所述摘要算法对所述第四子可变块的摘要进行计算得到的，所述  
15 第四子片段的摘要为根据所述摘要算法对所述第四子片段的摘要进行计算得到的；

如果所述第四子片段的摘要与所述第四子可变块的摘要相同，则将所述第四子片段、所述第四子片段的指纹以及所述第四子片段的摘要添加到  
20 所述第二本地重复数据库，生成第四本地重复数据库，所述第四子片段的指纹与所述第六初始块的指纹相同。

8、根据权利要求7所述方法，其特征在于，还包括：

接收第五报文，所述第五报文包括第五报文头与第五净荷，所述第五净荷包含第四净荷片度以及第五净荷片段，所述第四净荷片度的长度与所述  
25 第一子片段相同，所述第五净荷片段的长度与所述第四子片段的长度相

同；

根据所述指纹算法计算所述第四净荷片段中第七初始块的指纹，根据所述指纹算法计算所述第五净荷片段中第八初始块的指纹，所述第七初始块的起始位置与所述第四净荷片段的起始位置相同，所述第七初始块的长度与所述第一滑窗的长度相同，所述第八初始块的起始位置与所述第五净荷片段的起始位置相同，所述第八初始块的长度与所述第一滑窗的长度相同；

根据所述摘要算法计算所述第四净荷片段的摘要，根据所述摘要算法计算所述第五净荷片段的摘要；

10 如果所述第七初始块的指纹与所述第四本地重复数据库中的所述第一子片段的指纹相等，并且所述第四净荷片段的摘要与所述第一子片段的摘要相等，则删除所述第五报文中的所述第四净荷片段，如果所述第八初始块的指纹与所述第四本地重复数据库中的所述第四子片段的指纹相等，并且所述第五净荷片段的摘要与所述第四子片段的摘要相等，则删除所述第五报文中的所述第五净荷片段，生成第六报文，所述第六报文中包括第六报文头与第六净荷，所述第六报文头与所述第五报文头相同，所述第六净荷包括所述第一子片段的指纹以及所述第四子片段的指纹。

20 9、根据权利要求 8 所述方法，其特征在于，所述第六净荷还包括所述第一子片段在所述第五报文中的位置信息、所述第四子片段在所述第五报文中的位置信息。

10、根据权利要求 7 所述方法，其特征在于，所述第三片段的结束位置通过定界算法确定，包括：

25 从所述第二比特开始向后滑动所述定界算法中的第三滑窗，并判断所述第三滑窗对应的数据是否符合所述定界算法中的定界条件，当首次出现所述第三滑窗对应的数据符合所述定界条件时，在第一距离内继续向后滑

动所述第三滑窗，并判断所述第三滑窗对应的数据是否符合所述定界条件，如果出现所述第三滑窗对应的第一数据符合所述定界条件的情况时，则确定所述第一数据的结束位置为所述第三片段的结束位置，所述第三滑窗的长度与所述第一滑窗的长度相同，所述第一距离的长度与所述第一滑窗的  
5 长度相同。

11、根据权利要求 1 所述方法，其特征在于，还包括：将所述第一子片段的指纹以及所述第一子片段的摘要同步到压缩侧的重复数据库。

12、根据权利要求 1-10 任一项所述方法，其特征在于，还包括：将所述第一子片段、所述第一子片段的指纹以及所述第一子片段的摘要同步到  
10 解压缩侧的重复数据库。

13、根据权利要求 1 所述方法，其特征在于，还包括：

删除所述待压缩数据中的所述第一子片段，将所述第一子片段的指纹、所述第一子片段的长度及所述第一子片段在所述待压缩数据中的位置信息添加到第七报文的净荷中，所述待压缩数据为所述第七报文的净荷。

15 14、一种数据压缩设备，其特征在于，包括：

第一指纹计算单元，用于根据指纹算法计算待压缩数据中的第一片段的第一指纹，所述第一片段的起始位置与所述待压缩数据的起始位置相同，所述第一片段的长度与第一滑窗的长度相同；

查找单元，用于在第一本地重复数据库中查找所述第一指纹，所述第  
20 一本本地重复数据库用于存储重复数据、所述重复数据的指纹以及所述重复数据的摘要；

重复内容获取单元，用于如果所述第一本地重复数据库中不存在所述第一指纹，则根据所述第一指纹获取所述第一本地重复数据库中的第一可变块以及所述第一可变块的摘要，所述第一指纹与根据所述指纹算法计算得

到的所述第一可变块中的第一初始块的指纹相同，所述第一初始块的起始位置与所述第一可边块的起始位置相同，所述第一初始块的长度与所述第一滑窗的长度相同，所述第一可变块的摘要为根据摘要算法对所述第一可变块的摘要进行计算得到的；

- 5 摘要计算单元，用于根据所述摘要算法计算所述待压缩数据中的第二片段的摘要，所述第二片段的起始位置与所述待压缩数据的起始位置相同，所述第二片段的长度与所述第一可变块的长度相同；

摘要比较单元，用于比较所述第二片段的摘要与所述第一可变块的摘要；

- 10 第一子片段获取单元，用于如果所述第二片段的摘要与所述第一可变块的摘要不同，则获取所述第二片段中的第一子片段，所述第一子片段与所述第一可变块中的第一子可变块相同，所述第一子片段的起始位置与所述第二片段的起始位置相同，所述第一子可变块的起始位置与所述第一可变块的起始位置相同，所述第二片段中的第二比特与所述第一可变块中的  
15 第一比特不同，所述第二比特为所述第二片段中所述第一子片段的下一个比特，所述第一比特为所述第一可变块中所述第一子可变块的下一个比特；

- 第一添加单元，用于将所述第一子片段、所述第一子片段的指纹以及所述  
20 所述第一子片段的摘要添加到所述第一本地重复数据库，生成第二本地重复数据库，所述第一子片段的指纹与所述第一指纹相同，所述第一子片段的摘要为根据所述摘要算法对所述第一子片段的摘要进行计算得到的。

15、根据权利要求 14 所述设备，其特征在于，还包括：

第一报文接收单元，用于接收第一报文，所述第一报文包括第一报文头与第一净荷，所述第一净荷包含第一净荷片段，所述第一净荷片段的长度与所述第一子片段的长度相同；

第二初始指纹计算单元，用于根据所述指纹算法计算所述第一净荷片段中第二初始块的指纹，所述第二初始块的起始位置与所述第一净荷片段的起始位置相同，所述第二初始块的长度与所述第一滑窗的长度相同；

5 第一净荷摘要计算单元，用于根据所述摘要算法计算所述第一净荷片段的摘要；

第二报文生成单元，用于如果所述第二初始块的指纹与所述第二本地重复数据库中的所述第一子片段的指纹相等，并且所述第一净荷片段的摘要与所述第一子片段的摘要相等，则删除所述第一报文中的所述第一净荷片段，生成第二报文，所述第二报文中包括第二报文头与第二净荷，所述  
10 第二报文头与所述第一报文头相同，所述第二净荷包括所述第一子片段的指纹。

16、根据权利要求 15 所述设备，其特征在于，所述第二净荷还包括所述第一子片段在所述第一报文中的位置信息。

17、根据权利要求 14 所述设备，其特征在于，所述设备还包括：

15 第三初始指纹计算单元，用于根据所述指纹算法计算所述待压缩数据中的第二子片段的第三初始块的指纹，所述第二子片段的起始位置为所述第二比特，所述第二子片段的结束位置与所述待压缩数据的结束位置相同；所述第三初始块的起始位置为所述第二比特，所述第三初始块的长度等于所述  
20 第一滑窗的长度；

20 检测块获取单元，用于根据第二滑窗获取第二子可变块中的第一检测块，所述第二滑窗的起始位置与所述第二子可变块中的第三比特对应，所述第三比特为介于第二子可变块的起始位置与所述第二子可变块的结束位置之间的比特，所述第二滑窗的长度等于所述第一滑窗的长度，所述第二子可变块的起始位置为所述第一比特，所述第二子可变块的结束位置与所  
25 述第一可变块的结束位置相同；

检测块指纹计算单元，根据所述指纹算法计算所述第一检测块的指纹；

指纹比较单元，用于比较所述第三初始块的指纹与所述第一检测块的指纹；

第三摘要比较单元，用于如果所述第三初始块的指纹与所述第一检测  
5 块的指纹相同，则比较所述第二子可变块中的第三子可变块的摘要与所述  
第二子片段中的第三子片段的摘要，所述第三子可变块的起始位置与所述  
第一检测块的起始位置相同，所述第三子可变块的结束位置与所述第一可  
变块的结束位置相同，所述第三子片段的起始位置与所述第二子片段的起  
始位置相同，所述第三子片段的长度与所述第三子可变块的长度相等，所  
10 述第三子可变块的摘要为根据所述摘要算法对所述第三子可变块的摘要进  
行计算得到的；

第三添加单元，用于如果所述第三子片段的摘要与所述第三子可变块  
的摘要相等，则将所述第三子片段、所述第三子片段的指纹以及所述第三  
子片段的摘要添加到所述第二本地重复数据库，生成第三本地重复数据库，  
15 所述第三子片段的指纹等于所述第三初始块的指纹，所述第三子片段的摘  
要为根据所述摘要算法对所述第三子片段的摘要进行计算得到的。

18、根据权利要求 17 所述设备，其特征在于，还包括：

第三报文接收单元，用于接收第三报文，所述第三报文包括第三报文  
头与第三净荷，所述第三净荷包含第二净荷片段以及第三净荷片段，所述  
20 第二净荷片段的长度与所述第一子片段的长度相同，所述第三净荷片段  
的长度与所述第三子片段的长度相同；

第四初始指纹计算单元，用于根据所述指纹算法计算所述第二净荷片  
段中第四初始块的指纹，根据所述指纹算法计算所述第三净荷片段中第五  
初始块的指纹，所述第四初始块的起始位置与所述第二净荷片段的起始位  
置相同，所述第四初始块的长度与所述第一滑窗的长度相同，所述第五初  
25 置相同；

始块的起始位置与所述第三净荷片段的起始位置相同，所述第五初始块的长度与所述第一滑窗的长度相同；

第三净荷摘要计算单元，用于根据所述摘要算法计算所述第二净荷片段的摘要，根据所述摘要算法计算所述第三净荷片段的摘要；

- 5 第四报文生成单元，用于如果所述第四初始块的指纹与所述第三本地重复数据库中的所述第一子片段的指纹相等，并且所述第二净荷片段的摘要与所述第一子片段的摘要相等，则删除所述第三报文中的所述第二净荷片段，如果所述第五初始块的指纹与所述第三本地重复数据库中的所述第三子片段的指纹相等，并且所述第三净荷片段的摘要与所述第三子片段的摘要相等，则删除所述第三报文中的所述第三净荷片段，生成第四报文，  
10 所述第四报文中包括第四报文头与第四净荷，所述第四报文头与所述第三报文头相同，所述第四净荷包括所述第一子片段的指纹以及所述第三子片段的指纹。

- 19、根据权利要求 17 所述设备，其特征在于，所述第四净荷还包括所述  
15 第一子片段在所述第三报文中的位置信息、所述第三子片段在所述第三报文中的位置信息。

20、根据权利要求 14 所述设备，其特征在于，所述设备还包括：

- 初始块指纹计算单元，用于根据所述指纹算法计算所述第一可变块中的  
20 第四子可变块的第六初始块的指纹，所述第四子可变块的起始位置为所述第一比特，所述第四子可变块的结束位置与所述第一可变块的结束位置相同，所述第六初始块的起始位置为所述第一比特，所述第六初始块的长度等于所述第一滑窗的长度；

- 检测块获取单元，用于获取第四子片段中的第二检测块，所述第二检测  
25 块的起始位置与所述第四子片段中的第四比特对应，所述第四比特为介于第三片段的起始位置与所述第三片段的结束位置之间的比特，所述第二

检测块的长度等于所述第一滑窗的长度，所述第三片段的起始位置为所述第二比特，所述第三片段的结束位置通过定界算法确定，所述第三片段为所述待压缩数据中的片段，所述第四子片段为所述第三片段中的子片段，所述第四子片段的起始位置与所述第二检测块的起始位置相同，所述第四子片段的长度与所述第四子可变块的长度相同；

检测块指纹计算单元，用于根据所述指纹算法计算所述第二检测块的指纹；

指纹比较单元，用于比较所述第六初始块的指纹与所述第二检测块的指纹；

10 第四摘要比较单元，用于如果所述第六初始块的指纹与所述第二检测块的指纹相同，则比较所述第四子可变块的摘要与所述第四子片段的摘要，所述第四子可变块的摘要为根据所述摘要算法对所述第四子可变块的摘要进行计算得到的，所述第四子片段的摘要为根据所述摘要算法对所述第四子片段的摘要进行计算得到的；

15 第四添加单元，用于如果所述第四子片段的摘要与所述第四子可变块的摘要相同，则将所述第四子片段、所述第四子片段的指纹以及所述第四子片段的摘要添加到所述第二本地重复数据库，生成第四本地重复数据库，所述第四子片段的指纹与所述第六初始块的指纹相同。

21、根据权利要求 20 所述设备，其特征在于，还包括：

20 第五报文接收单元，用于接收第五报文，所述第五报文包括第五报文头与第五净荷，所述第五净荷包含第四净荷片度以及第五净荷片段，所述第四净荷片度的长度与所述第一子片段相同，所述第五净荷片段的长度与所述第四子片段的长度相同；

第七初始指纹计算单元，用于根据所述指纹算法计算所述第四净荷片

段中第七初始块的指纹，根据所述指纹算法计算所述第五净荷片段中第八初始块的指纹，所述第七初始块的起始位置与所述第四净荷片段的起始位置相同，所述第七初始块的长度与所述第一滑窗的长度相同，所述第八初始块的起始位置与所述第五净荷片段的起始位置相同，所述第八初始块的长度与所述第一滑窗的长度相同；

第五净荷摘要计算单元，用于根据所述摘要算法计算所述第四净荷片段的摘要，根据所述摘要算法计算所述第五净荷片段的摘要；

第六报文生成单元，用于如果所述第七初始块的指纹与所述第四本地重复数据库中的所述第一子片段的指纹相等，并且所述第四净荷片段的摘要与所述第一子片段的摘要相等，则删除所述第五报文中的所述第四净荷片段，如果所述第八初始块的指纹与所述第四本地重复数据库中的所述第四子片段的指纹相等，并且所述第五净荷片段的摘要与所述第四子片段的摘要相等，则删除所述第五报文中的所述第五净荷片段，生成第六报文，所述第六报文中包括第六报文头与第六净荷，所述第六报文头与所述第五报文头相同，所述第六净荷包括所述第一子片段的指纹以及所述第四子片段的指纹。

22、根据权利要求 21 所述设备，其特征在于，所述第六净荷还包括所述第一子片段在所述第五报文中的位置信息、所述第四子片段在所述第五报文中的位置信息。

23、根据权利要求 20 所述设备，其特征在于，所述检测块获取单元具体用于从所述第二比特开始向后滑动所述定界算法中的第三滑窗，并判断所述第三滑窗对应的数据是否符合所述定界算法中的定界条件，当首次出现所述第三滑窗对应的数据符合所述定界条件时，在第一距离内继续向后滑动所述第三滑窗，并判断所述第三滑窗对应的数据是否符合所述定界条件，如果出现所述第三滑窗对应的第一数据符合所述定界条件的情况时，

则确定所述第一数据的结束位置为所述第三片段的结束位置，所述第三滑窗的长度与所述第一滑窗的长度相同，所述第一距离的长度与所述第一滑窗的长度相同。

24、根据权利要求 14-23 任一项所述设备，其特征在于，还包括：

5 第二同步单元，用于将所述第一子片段、所述第一子片段的指纹以及所述第一子片段的摘要同步到解压缩侧的重复数据库。

25、根据权利要求 14 所述设备，其特征在于，还包括：

10 压缩处理单元，用于删除所述待压缩数据中的所述第一子片段，将所述第一子片段的指纹、所述第一子片段的长度及所述第一子片段在所述待压缩数据中的位置信息添加到第七报文的净荷中，所述待压缩数据为所述第七报文的净荷。

26、一种数据解压缩设备，其特征在于，包括：

15 第一指纹计算单元，用于根据指纹算法计算待压缩数据中的第一片段的第一指纹，所述第一片段的起始位置与所述待压缩数据的起始位置相同，所述第一片段的长度与第一滑窗的长度相同；

查找单元，用于在第一本地重复数据库中查找所述第一指纹，所述第一本地重复数据库用于存储重复数据、所述重复数据的指纹以及所述重复数据的摘要；

20 重复内容获取单元，用于如果所述第一本地重复数据库中不存在所述第一指纹，则根据所述第一指纹获取所述第一本地重复数据库中的第一可变块以及所述第一可变块的摘要，所述第一指纹与根据所述指纹算法计算得到的所述第一可变块中的第一初始块的指纹相同，所述第一初始块的起始位置与所述第一可边块的起始位置相同，所述第一初始块的长度与所述第一滑窗的长度相同，所述第一可变块的摘要为根据摘要算法对所述第一可

变块的摘要进行计算得到的；

摘要计算单元，用于根据所述摘要算法计算所述待压缩数据中的第二片段的摘要，所述第二片段的起始位置与所述待压缩数据的起始位置相同，所述第二片段的长度与所述第一可变块的长度相同；

5 摘要比较单元，用于比较所述第二片段的摘要与所述第一可变块的摘要；

第一子片段获取单元，用于如果所述第二片段的摘要与所述第一可变块的摘要不同，则获取所述第二片段中的第一子片段，所述第一子片段与所述第一可变块中的第一子可变块相同，所述第一子片段的起始位置与所述第二片段的起始位置相同，所述第一子可变块的起始位置与所述第一可变块的起始位置相同，所述第二片段中的第二比特与所述第一可变块中的第一比特不同，所述第二比特为所述第二片段中所述第一子片段的下一个比特，所述第一比特为所述第一可变块中所述第一子可变块的下一个比特；

15 第一添加单元，用于将所述第一子片段、所述第一子片段的指纹以及所述第一子片段的摘要添加到所述第一本地重复数据库，生成第二本地重复数据库，所述第一子片段的指纹与所述第一指纹相同，所述第一子片段的摘要为根据所述摘要算法对所述第一子片段的摘要进行计算得到的。

27、根据权利要求 26 所述设备，其特征在于，还包括：

20 第一同步单元，用于将所述第一子片段的指纹以及所述第一子片段的摘要同步到压缩侧的重复数据库。

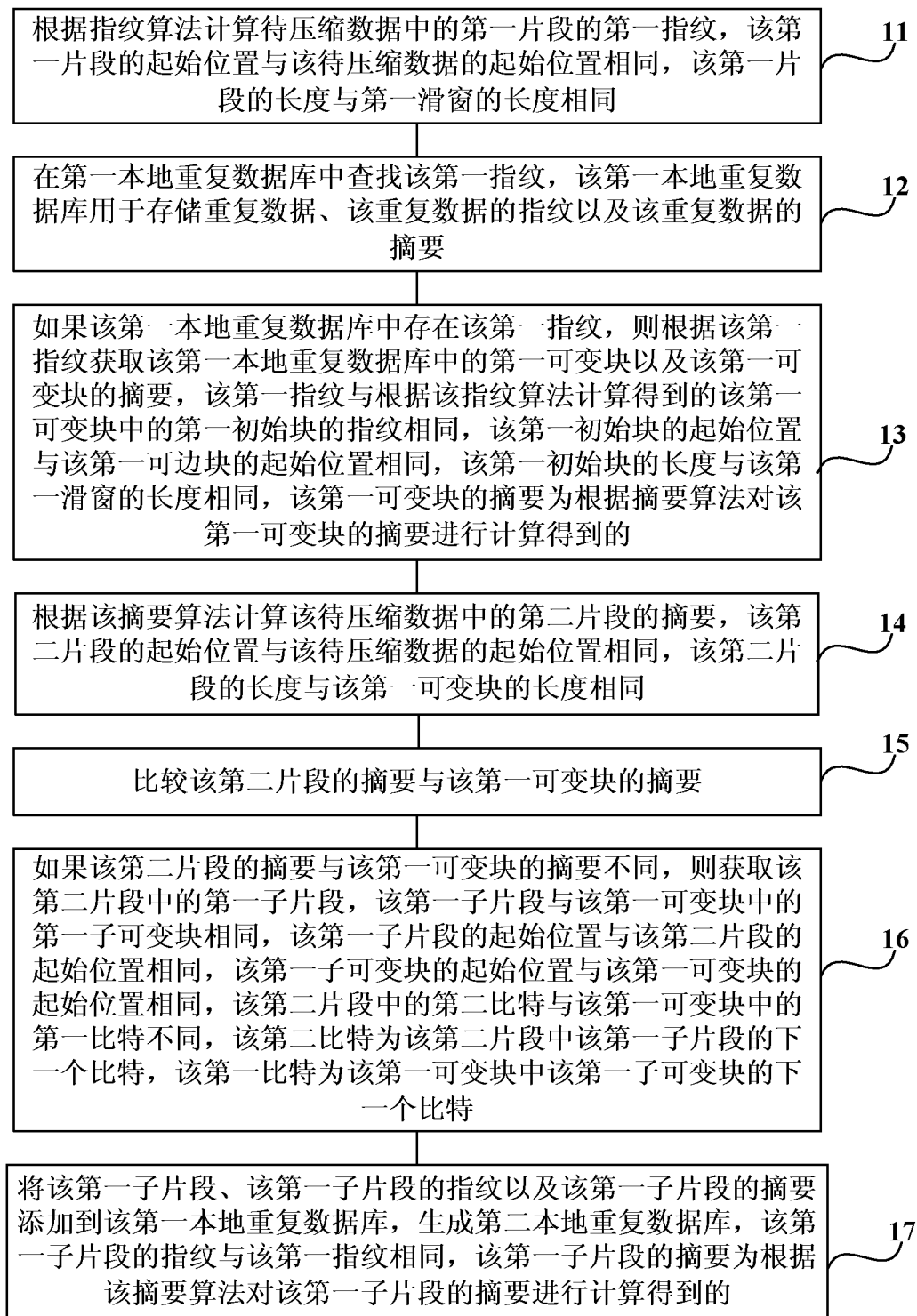


图 1

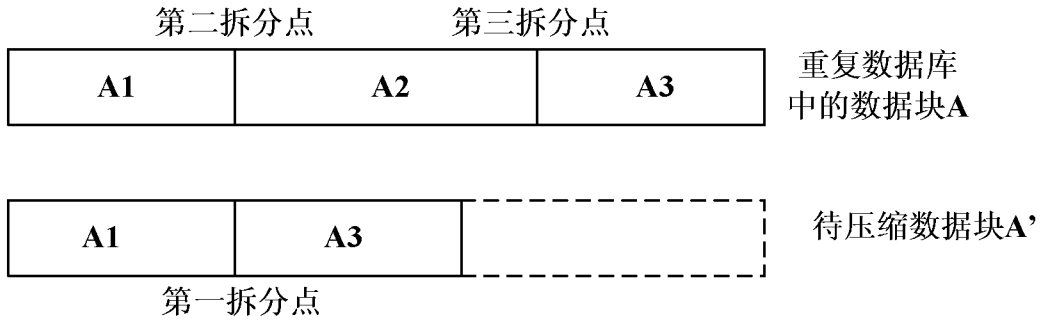


图 2

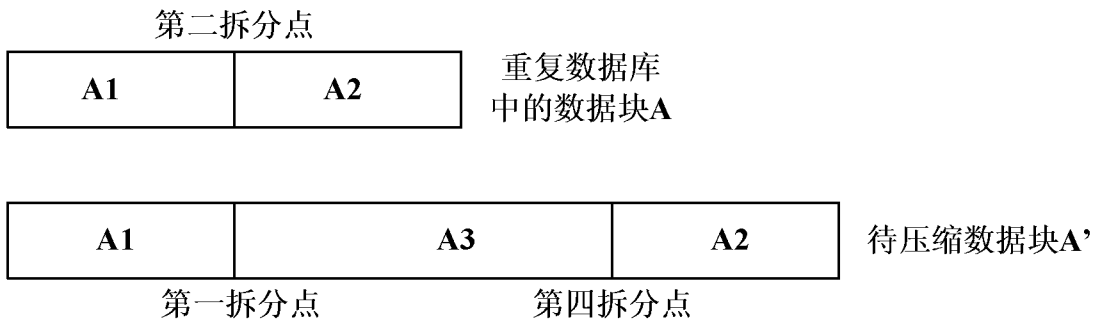


图 3

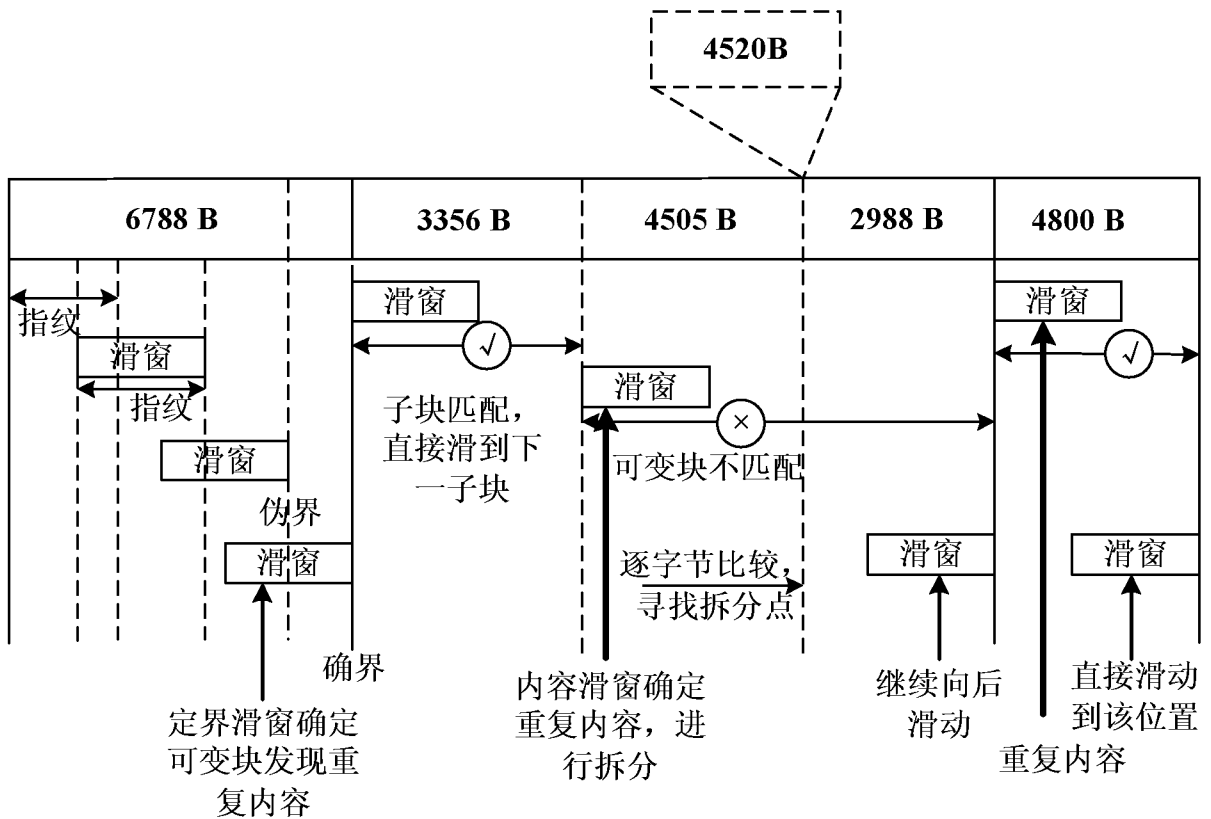


图 4

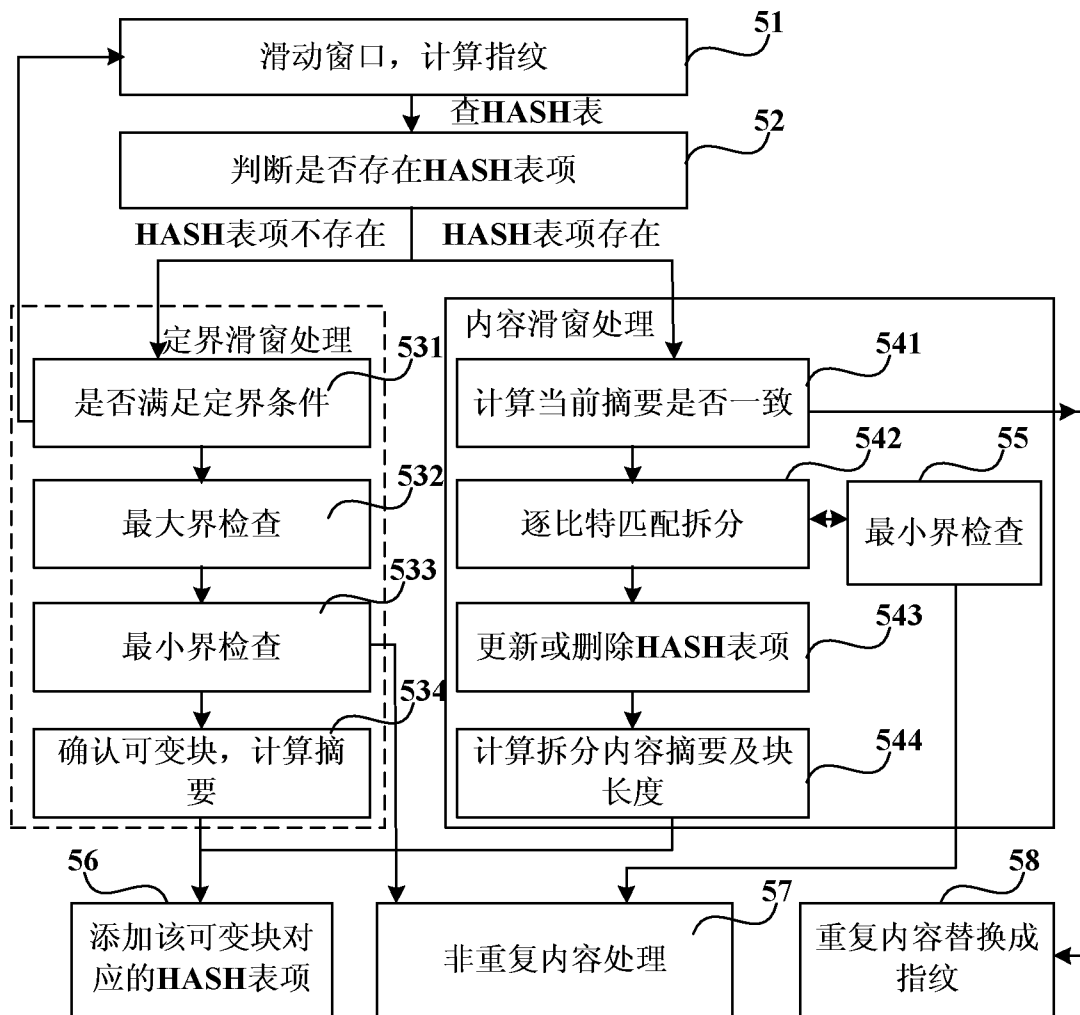


图 5



图 6

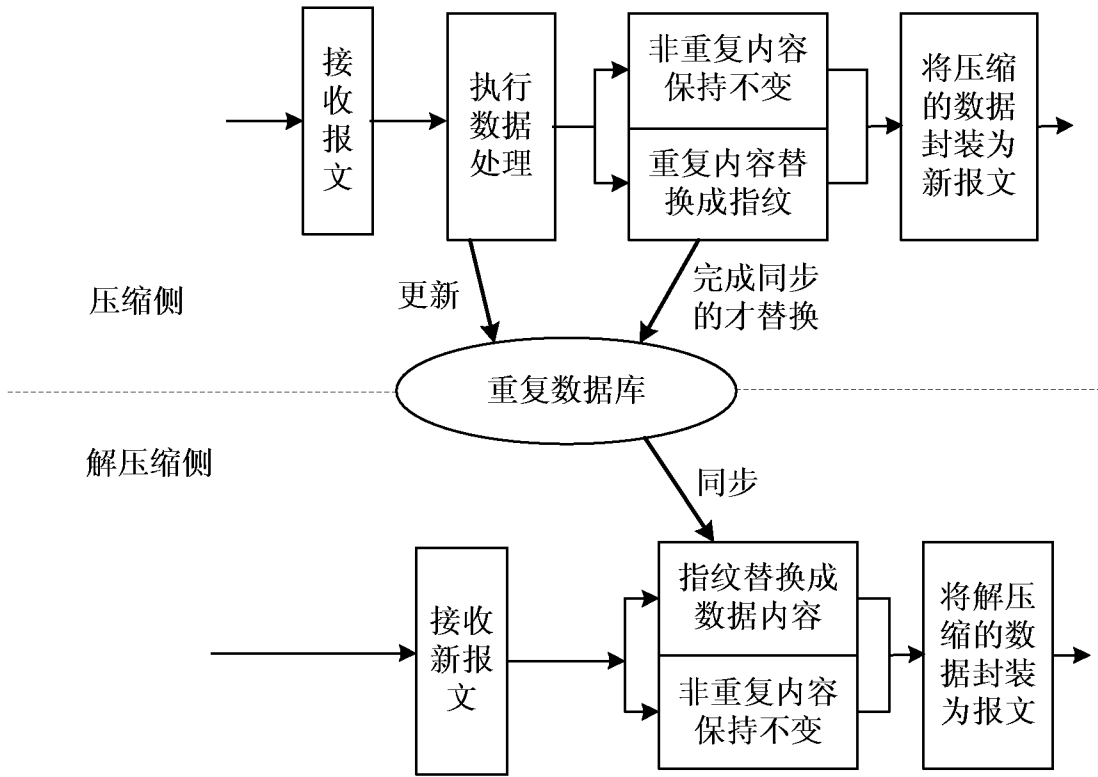


图 7

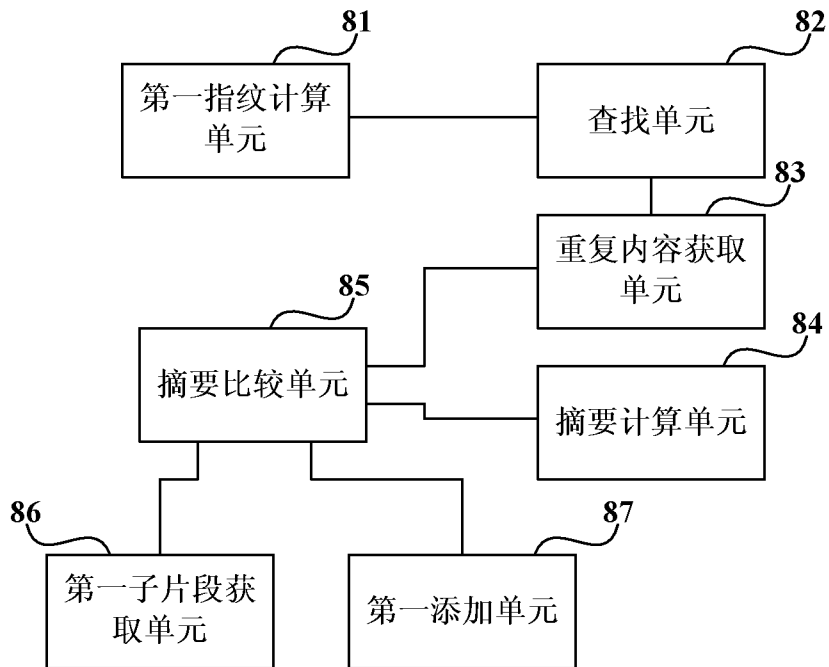


图 8

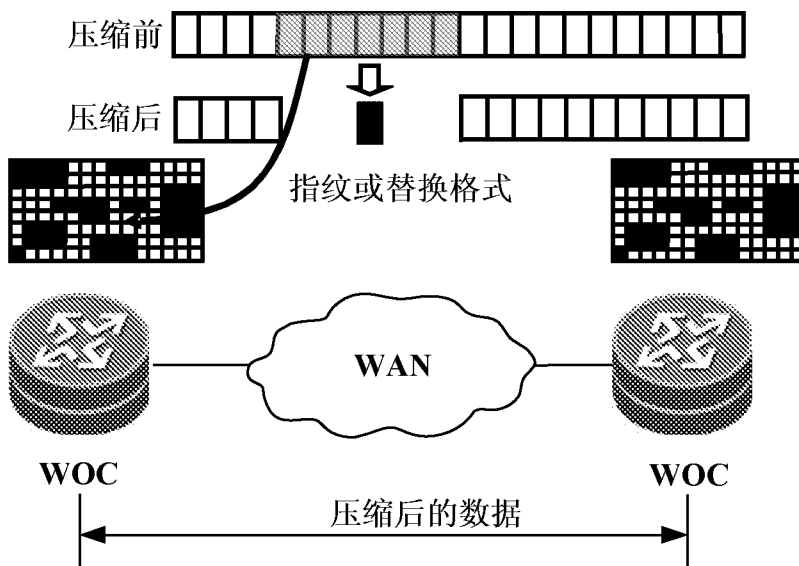


图 9



图 10

# INTERNATIONAL SEARCH REPORT

International application No.  
**PCT/CN2013/071725**

## A. CLASSIFICATION OF SUBJECT MATTER

H03M 7/30 (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)

IPC: H03M, H04L

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)

WPI, EPODOC, IEEE, CNKI, CNPAT: compress, variable block, bandwidth, data compress, repeat, fingerprint, window, slide, sliding window

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 1901549 A (BAI, Jie et al.) 24 January 2007 (24.01.2007) the whole document	1-27
A	CN 101599091 A (RIVERBED TECHNOLOGY INC.) 09 December 2009 (09.12.2009) the whole document	1-27
A	CN 1630984 A (PERIBIT NETWORKS INC.) 22 June 2005 (22.06.2005) the whole document	1-27
PX	CN 102684827 A (HUAWEI TECHNOLOGY CO., LTD.) 19 September 2012 (19.09.2012) claims 1-27	1-27

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p>
---	---

Date of the actual completion of the international search  
03 April 2013 (03.04.2013)

Date of mailing of the international search report  
**23 May 2013 (23.05.2013)**

Name and mailing address of the ISA  
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  
  
GONG, Yu  
Telephone No. (86-10) **62413682**

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
**PCT/CN2013/071725**

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date		
CN 1901549 A	24.01.2007	None			
CN 101599091 A	09.12.2009	US 6667700 B1	23.12.2003		
		WO 2004042930 A2	21.05.2004		
		AU 2003288962 A1	07.06.2004		
		EP 1584139 A2	12.10.2005		
		JP 2006505217 A	09.02.2006		
		CN 1736030 A	15.02.2006		
		INCHENP 200501053 E	10.08.2007		
		JP 2009171635 A	30.07.2009		
		EP2093885 A1	26.08.2009		
		DE 60328528 E	03.09.2009		
		CN 1630984 A	22.06.2004	WO 0209339 A2	31.01.2002
				US 2002037035 A1	28.03.2002
				AU 7791401 A	05.02.2002
EP 1307967 A2	07.05.2003				
JP 2004511928 A	15.04.2004				
NZ 523657 A	26.11.2004				
US 2005169364 A1	04.08.2005				
INCHENP 200300124 E	08.04.2005				
US 2007115151 A1	24.05.2007				
IL 153957 A	31.10.2007				
US 2008016246 A1	17.01.2008				
EP 1895663 A2	05.03.2008				
EP 1895664 A2	05.03.2008				
EP 1895665 A2	05.03.2008				
CN 101277117 A	01.10.2008				
US 2009022413 A1	22.01.2009				
DE 60140256 E	03.12.2009				
CA 2418314 C	13.10.2009				
CN 102684827 A	19.09.2012	None			

<b>A. 主题的分类</b>  H03M7/30(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>B. 检索领域</b> 检索的最低限度文献(标明分类系统和分类号) IPC: H03M, H04L 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)) WPI,EPODOC,IEEE,CNKI,CNPAT: 压缩, 指纹, 可变块, 重复, 带宽, 窗口, 滑窗, data compress, repeat, fingerprint, window, slide		
<b>C. 相关文件</b>		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN1901549A(白杰 等) 24.1 月 2007(24.01.2007) 全文	1-27
A	CN101599091A(河床技术股份有限公司) 09.12 月 2009(09.12.2009) 全文	1-27
A	CN1630984A(派里比特网络股份有限公司) 22.6 月 2005(22.06.2004) 全文	1-27
PX	CN102684827A(华为技术有限公司) 19.9 月 2012(19.09.2012) 权利要求 1-27	1-27
<input type="checkbox"/> 其余文件在 C 栏的续页中列出。 <input checked="" type="checkbox"/> 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 03.4 月 2013(03.04.2013)	国际检索报告邮寄日期 <b>23.5 月 2013 (23.05.2013)</b>	
ISA/CN 的名称和邮寄地址: 中华人民共和国国家知识产权局 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  <p style="text-align: center;">巩瑜</p> 电话号码: (86-10) <b>62413682</b>	

国际检索报告  
关于同族专利的信息

国际申请号  
**PCT/CN2013/071725**

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN1901549A	24.01.2007	无	
CN101599091A	09.12.2009	US6667700B1	23.12.2003
		WO2004042930A2	21.05.2004
		AU2003288962A1	07.06.2004
		EP1584139A2	12.10.2005
		JP2006505217A	09.02.2006
		CN1736030A	15.02.2006
		INCHENP200501053E	10.08.2007
		JP2009171635A	30.07.2009
		EP2093885A1	26.08.2009
		DE60328528E	03.09.2009
CN1630984A	22.06.2004	WO0209339A2	31.01.2002
		US2002037035A1	28.03.2002
		AU7791401A	05.02.2002
		EP1307967A2	07.05.2003
		JP2004511928A	15.04.2004
		NZ523657A	26.11.2004
		US2005169364A1	04.08.2005
		INCHENP200300124E	08.04.2005
		US2007115151A1	24.05.2007
		IL153957A	31.10.2007
		US2008016246A1	17.01.2008
		EP1895663A2	05.03.2008
		EP1895664A2	05.03.2008
		EP1895665A2	05.03.2008
		CN101277117A	01.10.2008
		US2009022413A1	22.01.2009
		DE60140256E	03.12.2009
		CA2418314C	13.10.2009
CN102684827A	19.09.2012	无	