(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103138926 A
(43)申请公布日 2013.06.05

(21)申请号 201110388841.0

(22)申请日 2011.11.30

(71)申请人 中国电信股份有限公司
地址 100032 北京市西城区金融大街 31 号

(72)发明人 章军 唐维 李文宇 田朝文
贾海燕 冯晓东 张鉴 常力元
赵洪波 赵敬谦 俞韶桢

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038
代理人 毛丽琴

(51)Int.Cl.
H04L 9/32(2006.01)
G06F 21/64(2013.01)
G06F 21/16(2013.01)

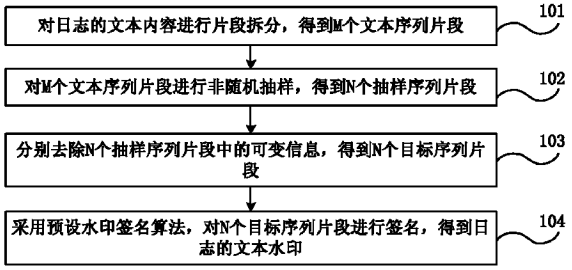
权利要求书3页 说明书8页 附图3页

(54)发明名称

水印签名方法与装置

(57)摘要

本发明实施例公开了一种水印签名方法与装置，其中，方法包括：对日志的文本内容进行片段拆分，得到M个文本序列片段，其中，M为大于1的整数；对M个文本序列片段进行非随机抽样，得到N个抽样序列片段，其中，N为大于0且不大于M的整数；分别去除N个抽样序列片段中的可变信息，得到N个目标序列片段；采用预设水印签名算法对N个目标序列片段进行签名，得到所述日志的文本水印。本发明实施例可以提高文本水印的生成效率，从而提高对日志的处理性能。



CN 103138926 A

1. 一种水印签名方法，其特征在于，包括：

对日志的文本内容进行片段拆分，得到 M 个文本序列片段，其中，M 为大于 1 的整数；

对 M 个文本序列片段进行非随机抽样，得到 N 个抽样序列片段，其中，N 为大于 0 且不大于 M 的整数；

分别去除 N 个抽样序列片段中的可变信息，得到 N 个目标序列片段；

采用预设水印签名算法对 N 个目标序列片段进行签名，得到所述日志的文本水印。

2. 根据权利要求 1 所述的方法，其特征在于，对日志的文本内容进行片段拆分包括：

采用定长的拆分方法，将日志的文本内容拆分为 M 个长度相同的文本序列片段；或者

采用固定片段数量的拆分方法，将日志的文本内容拆分为 M 个文本序列片段，M 为预先设定的固定片段数量。

3. 根据权利要求 1 所述的方法，其特征在于，分别去除 N 个抽样序列片段中的可变信息包括：

分别从 N 个抽样序列片段中去除预先设定的可变信息。

4. 根据权利要求 3 所述的方法，其特征在于，分别从 N 个抽样序列片段中去除预先设定的可变信息包括：

分别从 N 个抽样序列片段中去除阿拉伯数字信息；

分别从 N 个抽样序列片段中去除成对的符号中间的内容；

分别从 N 个抽样序列片段中去除等号之后直到分割边界的所有内容，以及冒号之后直到分割边界的所有内容。

5. 根据权利要求 4 所述的方法，其特征在于，所述分割边界包括空格、TAB 制表符、行尾、小于号、前中括号、前大括号、前小括号、引号与单引号。

6. 根据权利要求 1 所述的方法，其特征在于，所述预设水印签名算法包括消息摘要算法第五版 MD5 或安全哈希算法 SHA1。

7. 根据权利要求 1 至 6 任意一项所述的方法，其特征在于，N 为预先设定的固定整数。

8. 根据权利要求 7 所述的方法，其特征在于，采用预设水印签名算法对 N 个目标序列片段进行签名，得到所述日志的文本水印包括：

将 N 个目标序列片段依次拼接，得到一个完整序列；

采用预设水印签名算法对所述完整序列进行签名，得到所述日志的文本水印。

9. 根据权利要求 8 所述的方法，其特征在于，还包括：

将所述日志的文本水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否有与所述日志的文本水印一致的水印；

若历史记录中未识别日志的水印中有与所述日志的文本水印一致的水印，确认所述日志无法被解析识别；

若历史记录中未识别日志的水印中没有与所述日志的文本水印一致的水印，采用预先设置的解析识别规则列表，逐条对所述日志进行匹配处理；

若解析识别规则列表中的全部解析识别规则对所述日志均匹配失败，将所述日志的文本水印添加为历史记录中未识别日志的水印。

10. 根据权利要求 7 所述的方法，其特征在于，采用预设水印签名算法对 N 个目标序列片段进行签名，得到所述日志的文本水印包括：

分别采用预设水印签名算法对 N 个目标序列片段进行签名，得到 N 个目标序列片段的水印，所述日志的文本水印包括 N 个目标序列片段的水印。

11. 根据权利要求 10 所述的方法，其特征在于，还包括：

分别将 N 个目标序列片段的水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否包括 N 个目标序列片段的水印；

若历史记录中未识别日志的水印中包括 N 个目标序列片段中一个或多个片段的水印，确认所述日志无法被解析识别；

若历史记录中未识别日志的水印中不包括 N 个目标序列片段中任意一个片段的水印，采用预先设置的解析识别规则列表，逐条对所述日志进行匹配处理；

若解析识别规则列表中的全部解析识别规则对所述日志均匹配失败，将 N 个目标序列片段的水印作为所述日志的文本水印，添加为历史记录中未识别日志的水印。

12. 一种水印签名装置，其特征在于，包括：

拆分单元，用于对日志的文本内容进行片段拆分，得到 M 个文本序列片段，其中，M 为大于 1 的整数；

抽样单元，用于对 M 个文本序列片段进行非随机抽样，得到 N 个抽样序列片段，其中，N 为大于 0 且不大于 M 的整数；

信息去除单元，用于分别去除 N 个抽样序列片段中的可变信息，得到 N 个目标序列片段；

签名单元，用于采用预设水印签名算法对 N 个目标序列片段进行签名，得到所述日志的文本水印。

13. 根据权利要求 12 所述的装置，其特征在于，所述拆分单元对日志的文本内容进行片段拆分时，具体采用定长的拆分方法，将日志的文本内容拆分为 M 个长度相同的文本序列片段；或者采用固定片段数量的拆分方法，将日志的文本内容拆分为 M 个文本序列片段，M 为预先设定的固定片段数量。

14. 根据权利要求 12 所述的装置，其特征在于，所述信息去除单元具体用于分别从 N 个抽样序列片段中去除预先设定的可变信息，得到 N 个目标序列片段。

15. 根据权利要求 14 所述的装置，其特征在于，所述信息去除单元分别从 N 个抽样序列片段中去除预先设定的可变信息时，具体用于分别从 N 个抽样序列片段中去除阿拉伯数字信息；分别从 N 个抽样序列片段中去除成对的符号中间的内容；以及分别从 N 个抽样序列片段中去除等号之后直到分割边界的所有内容，以及冒号之后直到分割边界的所有内容。

16. 根据权利要求 15 所述的装置，其特征在于，所述分割边界包括空格、TAB 制表符、行尾、小于号、前中括号、前大括号、前小括号、引号与单引号。

17. 根据权利要求 12 所述的装置，其特征在于，所述预设水印签名算法包括 MD5 或 SHA1。

18. 根据权利要求 12 至 17 任意一项所述的装置，其特征在于，N 为预先设定的固定整数。

19. 根据权利要求 18 所述的装置，其特征在于，所述签名单元具体将 N 个目标序列片段依次拼接，得到一个完整序列；并采用预设水印签名算法对所述完整序列进行签名，得到所述日志的文本水印。

20. 根据权利要求 19 所述的装置，其特征在于，还包括：

存储单元，用于存储历史记录，所述历史记录中包括未识别日志的水印；

第一判断单元，用于将所述日志的文本水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否有与所述日志的文本水印一致的水印；若历史记录中未识别日志的水印中有与所述日志的文本水印一致的水印，确认所述日志无法被解析识别；

第一日志处理单元，用于根据第一判断单元的判断结果，在历史记录中未识别日志的水印中没有与所述日志的文本水印一致的水印时，采用预先设置的解析识别规则列表，逐条对所述日志进行匹配处理；并在解析识别规则列表中的全部解析识别规则对所述日志均匹配失败时，将所述日志的文本水印添加为历史记录中未识别日志的水印。

21. 根据权利要求 18 所述的装置，其特征在于，所述签名单元具体分别采用预设水印签名算法对 N 个目标序列片段进行签名，得到 N 个目标序列片段的水印，所述日志的文本水印包括 N 个目标序列片段的水印。

22. 根据权利要求 21 所述的装置，其特征在于，还包括：

存储单元，用于存储历史记录，所述历史记录中包括未识别日志的水印；

第二判断单元，用于分别将 N 个目标序列片段的水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否包括 N 个目标序列片段的水印；若历史记录中未识别日志的水印中包括 N 个目标序列片段中一个或多个片段的水印，确认所述日志无法被解析识别；

第二日志处理单元，用于根据第二判断单元的判断结果，在历史记录中未识别日志的水印中不包括 N 个目标序列片段中任意一个片段的水印时，采用预先设置的解析识别规则列表，逐条对所述日志进行匹配处理；并在解析识别规则列表中的全部解析识别规则对所述日志均匹配失败时，将 N 个目标序列片段的水印作为所述日志的文本水印，添加为历史记录中未识别日志的水印。

水印签名方法与装置

技术领域

[0001] 本发明涉及通信技术，尤其是一种水印签名方法与装置。

背景技术

[0002] 在通信等各种业务中，经常需要对记录本次业务事件的日志进行解析识别处理。在日志处理领域，对日志的解析识别处理通常包括以下两个阶段：第一个阶段是，判断一条日志能否被解析识别；第二个阶段是，将可以解析识别的日志与解析识别规则进行匹配，并根据匹配上的解析识别规则对该日志进行相应处理。目前，第二个阶段中，通常通过将可以解析识别的日志与类似于访问控制链表（Access Control Link，以下简称：ACL）结构的链式匹配规则列表逐一进行匹配，其中的链式匹配规则列表中的解析识别规则数量较高，可能高达1000条以上。因此，如果一种文本日志在第一个阶段无法被有效识别是否能被解析识别，每次接收到这种文本日志，在第二个阶段中，都要遍历全部的解析识别规则列表对其进行匹配处理，需要消耗大量的计算资源。

[0003] 现有技术在第一个阶段中，采用摘要算法第五版（Message DigestAlgorithm 5，以下简称：MD5）或安全哈希算法（Secure HashAlgorithm，以下简称：SHA1）对日志的全部文本内容进行签名，对日志的原始文本内容抽取特征值并采用水印签名算法进行处理，从而得到该日志的文本水印，来识别该文本水印能否被解析识别，从而判断该文本水印对应的日志能否被解析识别。

[0004] 在实现本发明的过程中，发明人发现，上述现有技术判断日志能否被解析识别的方法至少存在以下问题：

[0005] 由于需要对日志的全部文本内容进行签名，当日志的文本内容较长时，由于水印签名算法本身性能的限制，导致文本水印的生成效率较低，从而影响对日志的处理性能，使得对日志的处理性能较差；

[0006] 在类似日志处理等领域，日志的文本内容变化较大，例如，网络设备的同一种类型的日志，文本的部分内容，例如来源IP地址、来源端口等可变信息，会发生变化，而且变化的值域会非常大，例如会在整个IP地址范围、TCP/UDP端口范围内变化，再考虑到文本内容中多部分可变内容的交叉组合，值域范围无法穷尽。由于需要对全部文本内容进行抽样，当文本内容出现部分变化时，最终生成的文本水印会不同，从而增加了判断日志能否被解析识别的工作量，影响了对日志的处理性能。

发明内容

[0007] 本发明实施例所要解决的技术问题是：提供一种水印签名方法与装置，以提高文本水印的生成效率，从而提高对日志的处理性能。

[0008] 本发明实施例提供的一种水印签名方法，包括：

[0009] 对日志的文本内容进行片段拆分，得到M个文本序列片段，其中，M为大于1的整数；

[0010] 对M个文本序列片段进行非随机抽样，得到N个抽样序列片段，其中，N为大于0且不大于M的整数；

[0011] 分别去除N个抽样序列片段中的可变信息，得到N个目标序列片段；

[0012] 采用预设水印签名算法对N个目标序列片段进行签名，得到所述日志的文本水印。

[0013] 本发明实施例提供的一种水印签名装置，包括：

[0014] 拆分单元，用于对日志的文本内容进行片段拆分，得到M个文本序列片段，其中，M为大于1的整数；

[0015] 抽样单元，用于对M个文本序列片段进行非随机抽样，得到N个抽样序列片段，其中，N为大于0且不大于M的整数；

[0016] 信息去除单元，用于分别去除N个抽样序列片段中的可变信息，得到N个目标序列片段；

[0017] 签名单元，用于采用预设水印签名算法对N个目标序列片段进行签名，得到所述日志的文本水印。

[0018] 基于本发明上述实施例提供的水印签名方法与装置，对日志的文本内容进行片段拆分，对拆分得到的M个文本序列片段进行非随机抽样，并去除N个抽样序列片段中的可变信息，再采用预设水印签名算法对N个目标序列片段进行签名，得到日志的文本水印。与现有技术相比，本发明实施例仅对日志的若干片段进行签名，而无需对日志的全部文本内容进行签名，从而不会由于日志的文本内容长度影响文本水印的生成效率，有效提高了文本水印的生成效率与对日志的处理性能；并且，去除了签名片段中的可变信息部分，避免了可变信息对最终生成的文本水印的影响，减少了判断日志能否被解析识别的工作量，提高了对日志的处理性能。由此，本发明实施例提高了第一个阶段中日志能否被解析识别的识别效率与准确率，使无法匹配解析识别规则的日志在第一个阶段尽可能的被发现出来，避免其进入第二个阶段对解析识别规则列表的遍历，从而极大的提高日志处理的性能。

[0019] 下面通过附图和实施例，对本发明的技术方案做进一步的详细描述。

附图说明

[0020] 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。

[0021] 图1为本发明水印签名方法一个实施例的流程图；

[0022] 图2为本发明水印签名方法另一个实施例的流程图；

[0023] 图3为本发明水印签名装置一个实施例的结构示意图；

[0024] 图4为本发明水印签名装置另一个实施例的结构示意图；

[0025] 图5为本发明水印签名装置又一个实施例的结构示意图。

具体实施方式

[0026] 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。

[0027] 图 1 为本发明水印签名方法一个实施例的流程图。如图 1 所示，该实施例的水印签名方法包括 ：

[0028] 101，对日志的文本内容进行片段拆分，得到 M 个文本序列片段。

[0029] 其中，M 为大于 1 的整数。

[0030] 示例性地，本发明实施例中可以采用定长的拆分方法，将日志的文本内容拆分为 M 个长度相同的文本序列片段 ；或者，也可以采用固定片段数量的拆分方法，将日志的文本内容拆分为 M 个文本序列片段，其中，M 为预先设定的固定片段数量。

[0031] 102，对 M 个文本序列片段进行非随机抽样，得到 N 个抽样序列片段。

[0032] 其中，N 为大于 0 且不大于 M 的整数。

[0033] 示例性地，本发明实施例中，采用相同的方式对对 M 个文本序列片段进行多次非随机抽样，可以保证抽样结果的一致，即 ：保证抽样得到的 N 个抽样序列片段的正确性。

[0034] 其中，非随机抽样，是指按照预设规定有规律的抽样，以确保相同或类似的内容，在多次抽样处理后，得到的抽样结果是一样的 ；并且，避免采用随机抽样的方式导致不同或不相似的内容在多次抽样后，得到的抽样结果反而是相同的，从而引起结果冲撞，这是必须要避免的。示例性地，在本发明实施例中，可以采用包括但不限于如下所示的两种非随机抽样方式 ：一是采用固定抽取方式，例如，抽取 M 个文本序列片段中的前 N 个 ；二是奇偶位抽取，例如，抽取 M 个文本序列片段中的奇数位片段。

[0035] 103，分别去除 N 个抽样序列片段中的可变信息，得到 N 个目标序列片段。

[0036] 示例性地，可以根据预先设定的可变信息定义，分别从 N 个抽样序列片段中去除预先设定的可变信息。

[0037] 104，采用预设水印签名算法，例如，包括但不限于 MD5 或 SHA1，对 N 个目标序列片段进行签名，得到日志的文本水印。

[0038] 本发明上述实施例提供的水印签名方法，对日志的文本内容进行片段拆分，对拆分得到的 M 个文本序列片段进行非随机抽样，并去除 N 个抽样序列片段中的可变信息，再采用预设水印签名算法对 N 个目标序列片段进行签名，得到日志的文本水印。由于仅对日志的若干片段进行签名，而无需对日志的全部文本内容进行签名，从而不会由于日志的文本内容长度影响文本水印的生成效率，有效提高了文本水印的生成效率与对日志的处理性能 ；并且，去除了签名片段中的可变信息部分，避免了可变信息对最终生成的文本水印的影响，实现了基于模式的水印，其中的模式是指部分内容变化而大部分内容不变的形式，减少了判断日志能否被解析识别的工作量，提高了对日志的处理性能。由此，本发明实施例提高了第一个阶段中日志能否被解析识别的识别效率与准确率，使无法匹配解析识别规则的日志在第一个阶段尽可能的被发现出来，避免其进入第二个阶段对解析识别规则列表的遍历，从而极大的提高日志处理的性能。

[0039] 另外，采用 MD5、SHA1 等方式对 N 个目标序列片段进行签名时，由于采用单向不可逆转换方法，从得到的签名中，无法还原签名前的内容，有效提高了日志的安全性 ；并且，两个不同的内容信息，经过签名算法处理后，得到的水印不相同。因此，可以以及极低的冲撞

几率，应用于各种各样的系统。

[0040] 根据本发明的一个示例而非限制，图 1 所示的上述各实施例中，具体可以通过以下方式实现 103 的操作 ：

[0041] 分别从 N 个抽样序列片段中去除阿拉伯数字信息 ；

[0042] 分别从 N 个抽样序列片段中去除成对的符号或其它预设符号中间的内容 ；

[0043] 分别从 N 个抽样序列片段中去除等号之后直到分割边界的所有内容，以及冒号之后直到分割边界的所有内容。其中的分割边界包括但不限于空格、TAB 制表符、行尾、小于号、前中括号、前大括号、前小括号、引号与单引号等。

[0044] 根据本发明的另一个示例而非限制，在本发明上述各实施例的水印签名方法中，N 为预先设定的固定整数，即 ：抽样序列片段的数量是固定的。

[0045] 由于抽样序列片段的数量是固定的，即 N 的取值一定，该数值不会由于日志的文本内容的长度不同而变化，由于进行签名的目标序列片段数量一定，可以保证签名得到水印的过程中系统开销上限是收敛的，不会由于文本长度增加而导致性能的过度下降，当文本内容较长时，也不会出现性能下降，确保性能在可控范围之内。

[0046] 图 2 为本发明水印签名方法另一个实施例的流程图。如图 2 所示，该实施例的水印签名方法包括 ：

[0047] 201，对日志的文本内容进行片段拆分，得到 M 个文本序列片段。

[0048] 示例性地，本发明实施例中可以采用定长的拆分方法，将日志的文本内容拆分为 M 个长度相同的文本序列片段 ；或者，也可以采用固定片段数量的拆分方法，将日志的文本内容拆分为 M 个文本序列片段，其中，M 预先设定的固定片段数量，M 的取值为大于 1 的整数。

[0049] 202，对 M 个文本序列片段进行非随机抽样，得到 N 个抽样序列片段。

[0050] 其中，N 为预先设定的固定整数，且 N 的取值为大于 0 且不大于 M 的整数。

[0051] 203，别从 N 个抽样序列片段中去除阿拉伯数字信息。

[0052] 204，分别从 N 个抽样序列片段中去除成对的符号中间的内容。

[0053] 其中，成对的符号例如尖括号 <>、中括号 []、大括号 {}、小括号 ()、双引号“”、单引号‘’等。

[0054] 205，分别从 N 个抽样序列片段中去除等号之后直到分割边界的所有内容，以及冒号之后直到分割边界的所有内容，得到 N 个目标序列片段。

[0055] 其中的分割边界例如，空格、TAB 制表符、行尾、小于号、前中括号、前大括号、前小括号、引号与单引号。

[0056] 206，采用预设水印签名算法，例如，包括但不限于 MD5 或 SHA1，对 N 个目标序列片段进行签名，得到日志的文本水印。

[0057] 根据本发明的一个具体实施例，在图 2 所示实施例的 206 中，具体可以将 N 个目标序列片段依次拼接，得到一个完整序列 ；并采用预设水印签名算法对该完整序列进行签名，得到日志的文本水印。相应的，本发明实施例判断该日志能否被解析识别以及将第一个阶段无法判定为无法解析识别的日志与解析识别规则进行匹配时，具体可以通过如下方式实现 ：将得到的日志的文本水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否有与日志的文本水印一致的水印 ；若历史记录中未识别日志的水印中有与该日志的文本水印一致的水印，确认该日志无法被解析识别 ；若历史记录中未

识别日志的水印中没有与该日志的文本水印一致的水印，采用预先设置的解析识别规则列表，逐条对该日志进行匹配处理；若解析识别规则列表中的全部解析识别规则对该日志均匹配失败，将该日志的文本水印添加为历史记录中未识别日志的水印，以便据此判断后续日志的水印能否被解析识别。

[0058] 根据本发明的另一个具体实施例，在图 2 所示实施例的 206 中，具体可以分别采用预设水印签名算法对 N 个目标序列片段进行签名，得到 N 个目标序列片段的水印，其中，日志的文本水印包括该 N 个目标序列片段的水印。相应的，本发明实施例判断该日志能否被解析识别以及将第一个阶段无法判定为无法解析识别的日志与解析识别规则进行匹配时，具体可以通过如下方式实现：分别将 N 个目标序列片段的水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否包括 N 个目标序列片段的水印；若历史记录中未识别日志的水印中包括该 N 个目标序列片段中一个或多个片段的水印，确认日志无法被解析识别；若历史记录中未识别日志的水印中不包括该 N 个目标序列片段中任意一个片段的水印，采用预先设置的解析识别规则列表，逐条对日志进行匹配处理；若解析识别规则列表中的全部解析识别规则对日志均匹配失败，将该 N 个目标序列片段的水印作为日志的文本水印，添加为历史记录中未识别日志的水印，以便据此判断后续日志的水印能否被解析识别。

[0059] 相对于上述一个具体实施例而言，由于该另一个具体实施例中无需将 N 个目标序列片段依次拼接，可以进一步减少由此带来的运算开销，进一步提高了日志处理性能。无论采用上述两个具体实施例中的哪种方式，得到的日志的文本水印结果是一致的，此文本水印不会随着文本内容中可变信息部分的改变而改变，从而满足了同一种类日志处理系统的需求，解决了传统的 MD5、SHA1 等方法在此类系统中的不适用性的问题。

[0060] 以下以对一个具体日志的处理为例，来进一步说明本发明水印签名方法的具体应用。如下所示为通信系统中的一条普通日志内容实例：

[0061] <189>gzgt-nsg2 :NetScreen device_id = gzgt-nsg2 [Root] system-notification-00015 :Infranet Enforcer could not connect to the InfranetController because the Controller could not be reached on the network. (2010-10-09 11:05:42)

[0062] 在上述格式的日志中，“189”、“gzgt-nsg2”、“Root”、“00015”、“2010-10-09 11:05:42”这几部分为可变信息，其余部分为不变信息。依据本发明上述实施例的水印签名方法，采用定长的拆分方法或者固定片段数量的拆分方法，对日志的文本内容进行片段拆分，得到 M 个文本序列片段。然后对 M 个文本序列片段进行非随即抽样，得到 N 个抽样序列片段，记为：P1，P2，...，PN，N 为预先设定的抽样序列片段的最大数量值，例如取值为 10，其取值不大于 M，N 的取值可以任意选取但是需要在日志开始处理前就确定并在日志处理的过程中始终保持不变。M 个文本序列片段中未被抽样的其它文本序列片段丢弃。接下来，从这 N 个抽样序列片段中，依次除去可变信息。去掉可变信息的操作如下：首先，去掉 N 个抽样序列片段中的阿拉伯数字信息，即去掉该实例中的“189”数字信息；去掉成对的符号中间的内容，成对的符号例如尖括号 <>、中括号 []、大括号 {}、小括号 ()、双引号“”、单引号‘’等，即去掉该实例中的“Root”、“2010-10-09 11:05:42”信息；然后去掉等号＝、冒号：后面直到分割边界的所有内容，其中的分割边界例如，空格、TAB 制表符、行尾、小于号、

前中括号、前大括号、前小括号、引号、单引号等，即去掉该实例中的“gzgt-nsg2”信息。经过上述操作，可以得到全部为不可变内容的 N 个目标序列片段，记为 :C1，C2，...，CN。

[0063] 图 3 为本发明水印签名装置一个实施例的结构示意图。该实施例的水印签名装置可用于实现本发明上述各水印签名方法实施例的相应流程。如图 3 所示，其包括拆分单元 301、抽样单元 302、信息去除单元 303 与签名单元 304。

[0064] 其中，拆分单元 301，用于对日志的文本内容进行片段拆分，得到 M 个文本序列片段，其中，M 为大于 1 的整数。示例性地，拆分单元 301 对日志的文本内容进行片段拆分时，具体可以采用定长的拆分方法，将日志的文本内容拆分为 M 个长度相同的文本序列片段 ；或者，也可以采用固定片段数量的拆分方法，将日志的文本内容拆分为 M 个文本序列片段，M 为预先设定的固定片段数量。

[0065] 抽样单元 302，用于对拆分单元 301 得到的 M 个文本序列片段进行非随机抽样，得到 N 个抽样序列片段，其中，N 为大于 0 且不大于 M 的整数。

[0066] 信息去除单元 303，用于分别去除抽样单元 302 得到的 N 个抽样序列片段中的可变信息，得到 N 个目标序列片段。

[0067] 签名单元 304，用于采用预设水印签名算法，例如，MD5、SHA1 等，对信息去除单元 303 得到的 N 个目标序列片段进行签名，得到日志的文本水印。

[0068] 本发明上述实施例提供的水印签名装置，对日志的文本内容进行片段拆分，对拆分得到的 M 个文本序列片段进行非随机抽样，并去除 N 个抽样序列片段中的可变信息，再采用预设水印签名算法对 N 个目标序列片段进行签名，得到日志的文本水印。本发明实施例仅对日志的若干片段进行签名，而无需对日志的全部文本内容进行签名，从而不会由于日志的文本内容长度影响文本水印的生成效率，有效提高了文本水印的生成效率与对日志的处理性能 ；并且，去除了签名片段中的可变信息部分，避免了可变信息对最终生成的文本水印的影响，减少了判断日志能否被解析识别的工作量，提高了对日志的处理性能。由此，本发明实施例提高了第一个阶段中日志能否被解析识别的识别效率与准确率，使无法匹配解析识别规则的日志在第一个阶段尽可能的被发现出来，避免其进入第二个阶段对解析识别规则列表的遍历，从而极大的提高日志处理的性能。

[0069] 根据本发明的一个示例而非限制，与本发明上述水印签名方法实施例相应的，信息去除单元 303 具体可以根据预先设定的可变信息定义，分别从 N 个抽样序列片段中去除预先设定的可变信息，得到 N 个目标序列片段。进一步示例性地，信息去除单元 303 可以分别从 N 个抽样序列片段中去除阿拉伯数字信息 ；分别从 N 个抽样序列片段中去除成对的符号中间的内容 ；以及分别从 N 个抽样序列片段中去除等号之后直到分割边界的所有内容，以及冒号之后直到分割边界的所有内容，得到 N 个目标序列片段。其中的分割边界可以包括但不限于空格、TAB 制表符、行尾、小于号、前中括号、前大括号、前小括号、引号与单引号等。

[0070] 根据本发明的另一个示例而非限制，与本发明上述水印签名方法实施例相应的，N 为预先设定的固定整数，即 :抽样序列片段的数量是固定的。

[0071] 图 4 为本发明水印签名装置另一个实施例的结构示意图。与图 3 所示实施例的水印签名装置相比，该实施例中，签名单元 304 具体将 N 个目标序列片段依次拼接，得到一个完整序列，并采用预设水印签名算法对完整序列进行签名，得到日志的文本水印。相应的，

如图4所示，该实施例中，水印签名装置还包括存储单元305、第一判断单元306与第一日志处理单元307。

[0072] 其中，存储单元305，用于存储历史记录，该历史记录中包括未识别日志的水印。示例性地，该未识别日志的水印可以预先设置并可以在后续更新。

[0073] 第一判断单元306，用于将签名单元304得到的日志的文本水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否有与签名单元304得到的日志的文本水印一致的水印；若历史记录中未识别日志的水印中有与签名单元304得到的日志的文本水印一致的水印，确认签名单元304得到的日志无法被解析识别。

[0074] 第一日志处理单元307，用于根据第一判断单元306的判断结果，在历史记录中未识别日志的水印中没有与签名单元304得到的日志的文本水印一致的水印时，采用预先设置的解析识别规则列表，逐条对签名单元304得到的日志进行匹配处理；并在解析识别规则列表中的全部解析识别规则对签名单元304得到的日志均匹配失败时，将签名单元304得到的日志的文本水印添加倒存储单元305存储的历史记录中未识别日志的水印。

[0075] 图5为本发明水印签名装置又一个实施例的结构示意图。与图3所示实施例的水印签名装置相比，该实施例中，签名单元304分别采用预设水印签名算法对N个目标序列片段进行签名，得到N个目标序列片段的水印，其中，日志的文本水印包括该N个目标序列片段的水印。相应的，如图5所示，该实施例中，水印签名装置还包括存储单元305、第二判断单元308与第二日志处理单元309。

[0076] 其中，存储单元305，用于存储历史记录，该历史记录中包括未识别日志的水印。示例性地，该未识别日志的水印可以预先设置并可以在后续更新。

[0077] 第二判断单元308，用于分别将签名单元304得到的N个目标序列片段的水印依次与历史记录中未识别日志的水印比较，识别历史记录中未识别日志的水印中是否包括该N个目标序列片段的水印；若历史记录中未识别日志的水印中包括该N个目标序列片段中一个或多个片段的水印，确认日志无法被解析识别。

[0078] 第二日志处理单元309，用于根据第二判断单元308的判断结果，在历史记录中未识别日志的水印中不包括该N个目标序列片段中任意一个片段的水印时，采用预先设置的解析识别规则列表，逐条对日志进行匹配处理；并在解析识别规则列表中的全部解析识别规则对日志均匹配失败时，将该N个目标序列片段的水印作为对应日志的文本水印，添加在存储单元305存储的历史记录中未识别日志的水印。

[0079] 本说明书中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它实施例的不同之处，各个实施例之间相同或相似的部分相互参见即可。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。

[0080] 本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0081] 本发明实施例无需对日志的全部文本内容进行签名，从而不会由于日志的文本内容长度影响文本水印的生成效率，有效提高了文本水印的生成效率与对日志的处理性能；

并且，去除了签名片段中的可变信息部分，当文本内容出现部分变化时，生成的文本水印会忽略内容变化部分，避免了可变信息对最终生成的文本水印的影响，实现了对文本内容模式的水印，保证用有限的、可枚举的模式水印，支持对无穷日志解析识别前的处理，减少了判断日志能否被解析识别的工作量，提高了对日志的处理性能。由此，本发明实施例提高了第一个阶段中日志能否被解析识别的识别效率与准确率，使无法匹配解析识别规则的日志在第一个阶段尽可能的被发现出来，避免其进入第二个阶段对解析识别规则列表的遍历，从而极大的提高日志处理的性能。

[0082] 本发明的描述是为了示例和描述起见而给出的，而并不是无遗漏的或者将本发明限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本发明的原理和实际应用，并且使本领域的普通技术人员能够理解本发明从而设计适于特定用途的带有各种修改的各种实施例。

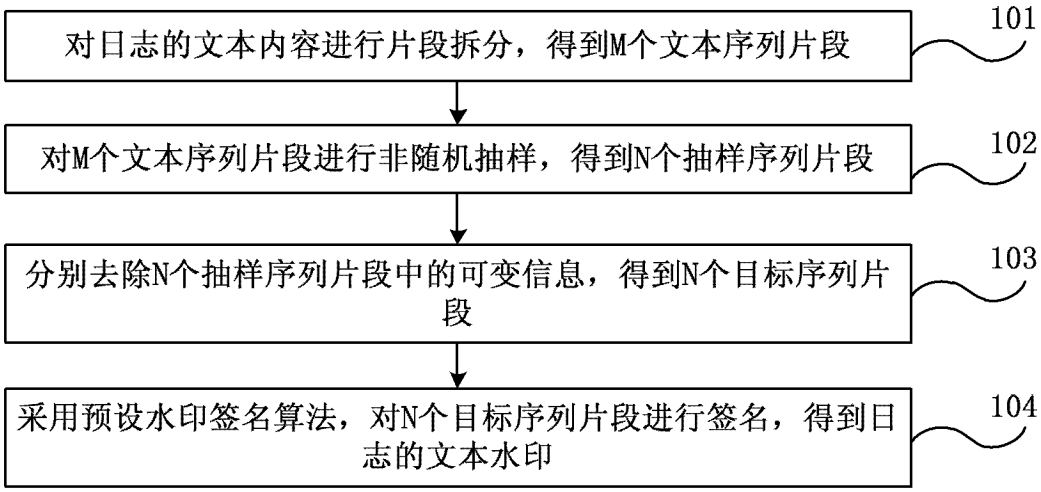

对日志的文本内容进行片段拆分，得到M个文本序列片段
101
对M个文本序列片段进行非随机抽样，得到N个抽样序列片段
102
分别去除N个抽样序列片段中的可变信息，得到N个目标序列片段
103
采用预设水印签名算法，对N个目标序列片段进行签名，得到日志的文本水印
104

图 1

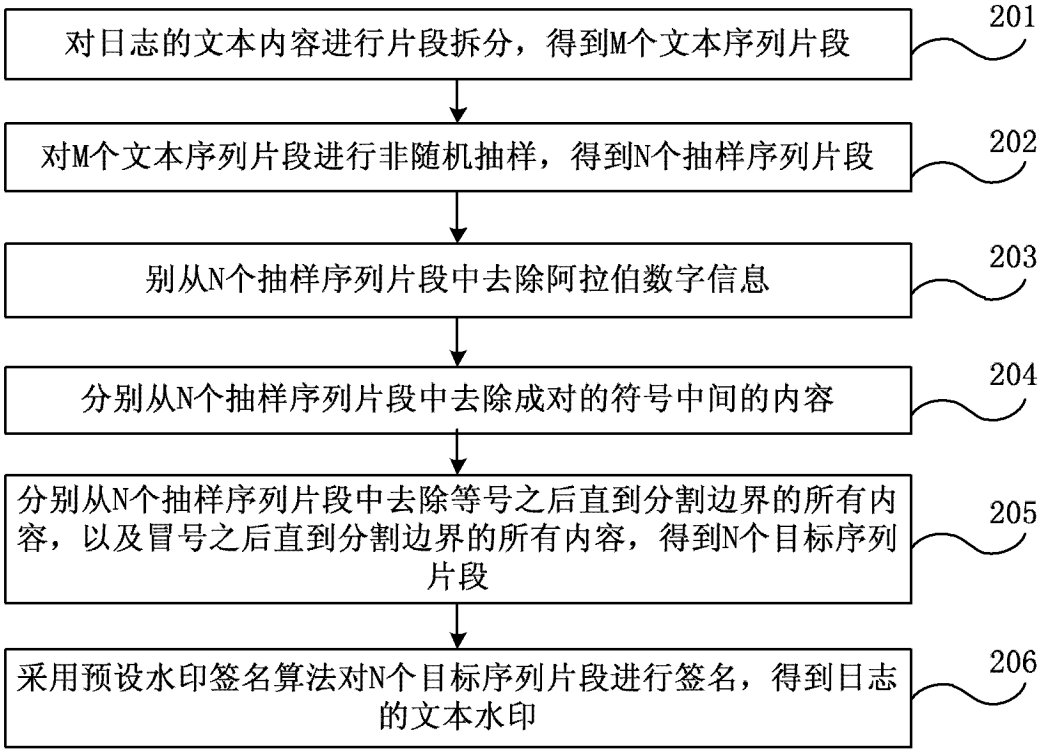

对日志的文本内容进行片段拆分，得到M个文本序列片段
201
对M个文本序列片段进行非随机抽样，得到N个抽样序列片段
202
别从N个抽样序列片段中去除阿拉伯数字信息
203
分别从N个抽样序列片段中去除成对的符号中间的内容
204
分别从N个抽样序列片段中去除等号之后直到分割边界的所有内容，以及冒号之后直到分割边界的所有内容，得到N个目标序列片段
205
采用预设水印签名算法对N个目标序列片段进行签名，得到日志的文本水印
206

图 2

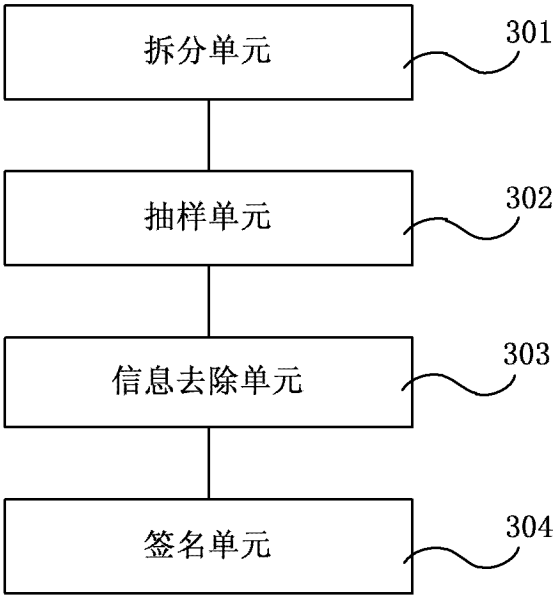

拆分单元
301
抽样单元
302
信息去除单元
303
签名单元
304

图 3

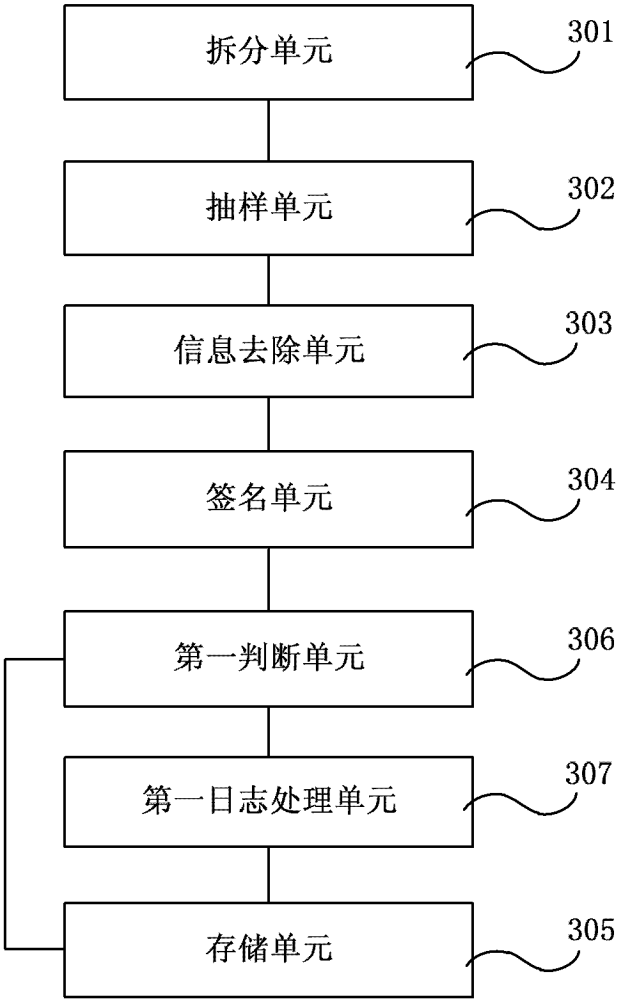

拆分单元
301
抽样单元
302
信息去除单元
303
签名单元
304
第一判断单元
306
第一日志处理单元
307
存储单元
305

图 4

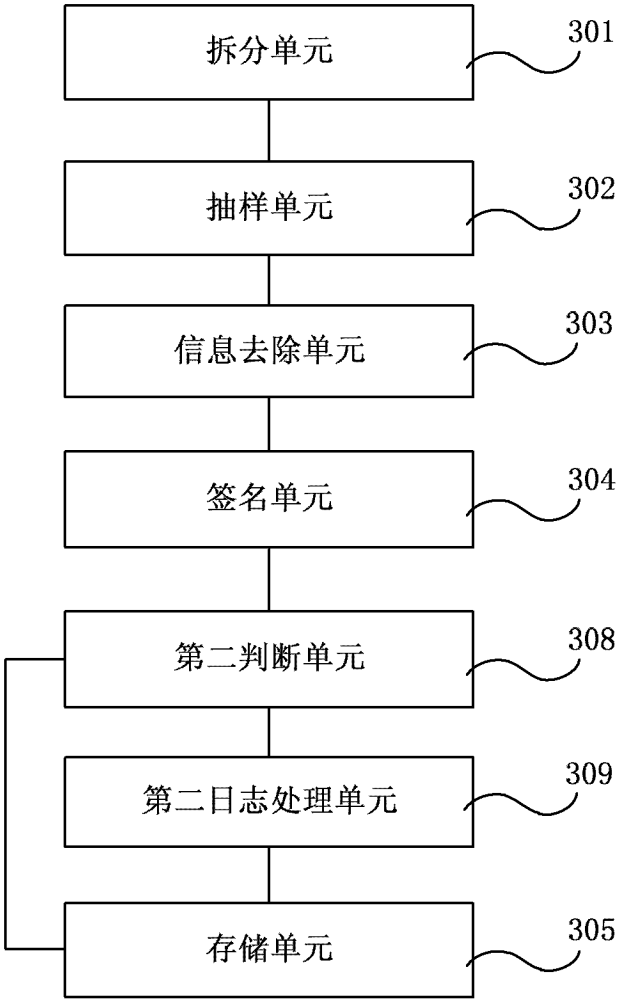

拆分单元
301
抽样单元
302
信息去除单元
303
签名单元
304
第二判断单元
308
第二日志处理单元
309
存储单元
305

图 5