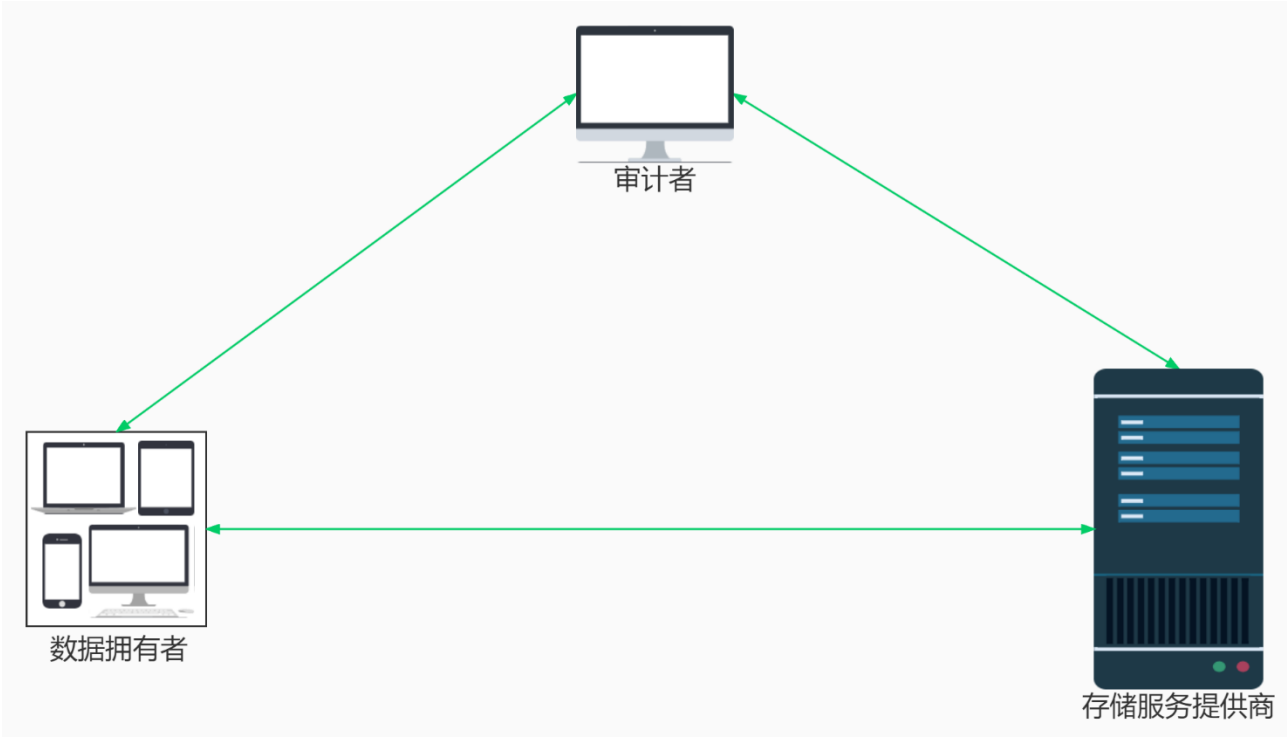
说 明 书 摘 要

本发明涉及一种轻量级可去重的密文完整性审计方法，上述系统包括数据拥有者，审计者，存储服务提供商；数据拥有者拥有原始数据，数据拥有者会将原始数据存储在存储服务提供商处，同时，释放存储在自身的原始数据，对于存储在存储服务提供商处的数据有保密性和完整性的需求；审计者为可信第三方，审计者执行数据拥有者委托的完整性审计任务；存储服务提供商会存储来自数据拥有者的数据和完整性标签，并进行数据去重。通过实施本发明实施例，在可去重情况下，数据拥有者的完整性标签计算负担和通信成本降为常量级别；同时还能降低在存储服务提供商的标签存储空间成本。

摘 要 附 图



权 利 要 求 书

1、轻量级可去重的密文完整性审计方案，其特征在于轻量级可去重的密文完整性审计方案包括数据拥有者、审计者以及存储服务提供商三种角色；方法大致有 3 个阶段：初始化阶段，文件上传阶段，审计阶段；在初始化阶段，数据拥有者初始化后，将各自对应的数据发送至存储服务提供商和审计者；在文件上传阶段，用户将原始数据进行预处理后，将预处理后的数据发送至存储服务提供商，存储服务提供商做出响应，数据拥有者根据响应结果判断是否应计算标签，如需计算则产生标签，否则跳过标签产生的步骤，直接产生标签转化辅助材料以及审计材料发送给存储服务提供商；在审计阶段，审计者发送对目标文件的挑战给存储服务提供商，存储服务提供商对挑战进行响应，审计者对响应结果进行验证。构建方法包括以下步骤：

S1、可去重的密文完整性审计方案初始化阶段，包括用户密钥的初始化，用于审计的公私钥对初始化等；

S2、可去重的密文完整性审计方案文件上传阶段，包括产生去重密钥，对原文件的加密，对加密后的文件进行切块并编码，对去重密钥的加密，上传加密、切块并编码后的文件和加密后的去重密钥，对块数据计算并上传完整性标签（如果需要）、标签转化辅助材料以及审计材料至存储服务提供商等；

S3、可去重的密文完整性审计方案审计阶段，包括审计者发送挑战以及验证对挑战的响应结果的正确性，存储服务提供者利用自身持有的数据对挑战计算响应结果；

2、根据权利要求1所述的轻量级可去重的密文完整性审计方案的构建方法，其特征在于，步骤S1过程为：

首先，用户随机选择数ku作为用户密钥；对于审计私钥，用户随机选择数x作为审计私钥，从循环乘法群$\mathbb{G}$中选择生成元g并计算g^x作为审计公钥。

3、根据权利要求1所述的轻量级可去重的密文完整性审计方案的构建方法，其特征在于，步骤S2过程为：

第一步，计算消息明文的哈希值$h(m)$作为去重密钥记为k_{dedup}。

第二步，利用去重密钥k_{dedup}对原文件数据进行加密、切块和编码后，获得密态文件数据C_{data}。

第三步，以用户密钥ku作为加密密钥对去重密钥k_{dedup}进行加密获得密态去重密钥C_{key}。

第四步，上传文件目录dir、文件名$filename$、密态文件数据C_{data}和密态密钥C_{key}至服务器，根据dir创建文件目录，将C_{key}存储到相应目录中，并以$filename.key$命名；如果服务器中未存在密态文件数据C_{data}，则进入第五步，否则进入第六步。

第五步，告知用户文件不存在重复。用户收到不存在重复的通知后，为C_{data}中的块数据依次计算标签σ_i，产生文件摘要ma、利用审计私钥x和文件摘要ma计算用户标签转化辅助变量$trans$，然后将所有块标签σ_i、对关键参数r加密后的辅助变量C_{trans}和审计者的审计材料TAM发送至存储服务提供商。

第六步，告知用户文件可以去重，并返回C_{trans}至用户，用户通过去重密钥k_{dedup}解密出C_{trans}中被加密的变量得到$trans$，通过$trans$计算出自己的标签转化辅助材料$trans'$，随后发送标签转化辅助材料$trans'$和审计者的审计材料TAM'至存储服务提供商。

4、根据权利要求 1 所述的轻量级可去重的密文完整性审计方案的构建方法，其特征在于，步骤 S3 中：

当审计者想要审计数据时，需要告知存储服务提供商被审计目标文件的文件名$filename$、存储服务提供商响应总数据块数n，审计者利用总数据块数随机产生的块下标集合和随机数集合$Q = \{i, v_i\}_{i\in[1,n]}$；存储服务提供商根据块下标集合与随机数集合$Q$以及所存储的相应的用户密态文件数据$C_{data}$产生数据完整性证明$(\bar{\rho}, \bar{\sigma})$，随后将审计者的审计材料$TAM$与完整性证明$(\bar{\rho}, \bar{\sigma})$返回给用户。最后审计者利用审计公钥$g^x$与审计材料$TAM$对完整性证明进行检验，检验通过则说

明被抽查的数据是完整的，反之则说明被抽查的数据遭到了破坏。

轻量级可去重的密文完整性审计方案构建方法

技术领域

本发明属于网络空间安全技术领域，尤其涉及一种轻量级可去重的密文完整性审计方案的构建方法。

背景技术

由于云计算技术的进步，大数据已经融入了人们的日常生活，用户数据迅速增长。

一方面，为了降低运维成本，个人或者企业倾向于将数据外包存储到云存储服务提供商的平台上且往往不会继续保留原始数据，一旦删除原始数据，这也意味着数据拥有者失去了对原有数据的控制。由于云存储服务提供商具有强烈的动机来破坏用户数据的完整性（例如通过压缩用户存储在云上的图片或者视频来达到节省实际存储空间获取额外利润的目的），因此上述场景对于用户的数据完整性来说构成了巨大的威胁。消息验证码和数字签名是以往密码学用来解决数据完整性的有效手段，然而这两种传统方案并不能直接用于海量数据的场景下。这两种方案需要用户拥有所有原始数据，由于目前用于归档存储的云服务器的单盘数据吞吐量（吞吐量为读写速度的总和）是 30~40MB/s，即便为理想情况，单盘读速度为 40MB/s，读取 1TB 的文件约需 7 小时，性能消耗代价极大，因此不现实。2007 年，Juels 和 Ateniese 提出了 PoR 和 PDP 方案，他们都在纠删码的辅助下，利用随机抽样形式的概率性算法来规避了待检验数据需要被全文读取的问题，初步解决了大文件数据完整性审计的问题。随后国内外许多研究者在他们的思想之上对方案进行了迭代，实现安全加固、效率优化和安全模型的扩展等。

另一方面，云存储服务提供商们发现用户上传的文件具有相似性，存在大量的重复文件，大数据平台对未经加密的文件进行重复数据删除已经拥有较为成

熟的技术和相关实践。然而，近年来国内相关法律法规对用户数据的保密性加大了保护力度，要求企业对用户数据进行加密存储，因此传统的明文重复数据删除技术已经长期难以适应互联网的发展，密态数据的重复数据删除技术将被逐渐应用于商业应用中。密文去重的基本思想是通过文件数据产生用于可去重加密的密钥，随后，利用该密钥对文件数据进行加密，上述操作可将相同的文件加密成获得相同的密文文件，从而实现重复数据删除。

随着完整性审计研究和密文去重研究的深入，同时支持密文去重和完整性审计的系统被越来越多的学者们考虑，要同时支持两种功能比较直接的方法是将去重和审计方案进行简单叠加使用，但用户每一次上传即便此文件能够被去重，都需要为文件重新计算一套标签，标签的计算时间与文件大小呈线性关系，效率极低；近年来，有学者提出以去重密钥产生审计私钥，即完整性标签的产生与去重密钥直接相关，当文件存在重复时，便无需重新计算标签，因为相同的文件其完整性标签是完全相同的，这种方法虽然提高了文件上传的效率，但是由于采用刻板的标签生成方式也使审计公钥的管理变得困难，更具体地说，因为审计密钥是根据文件内容衍生的，所以每个文件的审计公钥都不同。如果一个用户拥有 100 个文件，那么用户就需要管理 100 个审计公钥，这是不切实际的。此外，为用户间的重复文件产生个性化的标签也代表了用户对文件的拥有权。

为此，我们设计了一套既能高效实现文件上传，又能保持标签个性化的方案，此方案可以适用于多种密文去重系统。此方案在已存在重复数据的情况下，重新产生该文件的一组用户个性化完整性标签仅仅需要常量级别的计算代价、通信代价也降低至常量级别，且存储服务提供商仅需存储常量级别的标签辅助变量，而不是像传统方案需要存储一组用户标签，近似地实现重复标签删除。

发明内容

为了解决现有技术所存在的问题，本发明提供轻量级可去重的密文完整性审计方案构建方法。

本发明是这样实现的：轻量级可去重的密文完整性审计方案构建方法，其特征在于轻量级可去重的密文完整性审计方案包括数据拥有者、审计者以及存储服务提供商三种角色；方法大致有 3 个阶段：初始化阶段，文件上传阶段，审计阶段；在初始化阶段，数据拥有者初始化后，将各自对应的数据发送至存储服务提供商和审计者；在文件上传阶段，用户将原始数据进行预处理后，将预处理后的数据发送至存储服务提供商，存储服务提供商做出响应，数据拥有者根据响应结果判断是否应计算标签，如需计算则产生标签，否则跳过标签产生的步骤，直接产生标签转化辅助材料以及审计材料发送给存储服务提供商；在审计阶段，审计者发送对目标文件的挑战给存储服务提供商，存储服务提供商对挑战进行响应，审计者对响应结果进行验证。构建方法包括以下步骤：

S1、可去重的密文完整性审计方案初始化阶段，包括用户密钥的初始化，用于审计的公私钥对初始化等；

S2、可去重的密文完整性审计方案文件上传阶段，包括产生去重密钥，对原文件的加密，对加密后的文件进行切块并编码，对去重密钥的加密，上传加密、切块并编码后的文件和加密后的去重密钥，对块数据计算并上传完整性标签（如果需要）、标签转化辅助材料以及审计材料至存储服务提供商等；

S3、可去重的密文完整性审计方案审计阶段，包括审计者发送挑战以及验证对质询的响应结果的正确性，存储服务提供者利用自身持有的数据对质询计算响应结果；

优选地，步骤 S1 过程为：首先，用户随机选择数ku作为用户密钥；对于审计私钥，用户随机选择数x作为审计私钥，从循环乘法群$\mathbb{G}$中选择生成元g并计算g^x作为审计公钥。

优选地，步骤 S2 过程为：第一步，计算消息明文的哈希值$h(m)$作为去重密钥记为k_{dedup}。第二步，利用去重密钥k_{dedup}对原文件数据进行加密、切块和编码后，获得密态文件数据C_{data}。第三步，以用户密钥ku作为加密密钥对去重密钥k_{dedup}进行加密获得密态去重密钥C_{key}。第四步，上传文件目录dir、文件名

$filename$、密态文件数据C_{data}和密态密钥C_{key}至服务器，根据dir创建文件目录，将C_{key}存储到相应目录中，并以$filename.key$命名；如果服务器中未存在密态文件数据C_{data}，则进入第五步，否则进入第六步。第五步，告知用户文件为首次存储。用户收到首次存储的通知后，为C_{data}中的块数据依次计算标签σ_i，产生文件摘要ma、利用审计私钥x和文件摘要ma计算用户标签转化辅助变量$trans$，然后将所有块标签σ_i、对关键参数r加密后的辅助变量C_{trans}和审计者的审计材料TAM发送至存储服务提供商。第六步，告知用户文件可以去重，并返回C_{trans}至用户，用户通过去重密钥k_{dedup}解密出C_{trans}中被加密的变量得到$trans$，通过$trans$计算出自己的标签转化辅助材料$trans'$，随后发送标签转化辅助材料$trans'$和审计者的审计材料TAM'至存储服务提供商。

优选地，步骤 S3 过程为：当审计者想要审计数据时，需要告知存储服务提供商被审计目标文件的文件名$filename$、存储服务提供商响应总数据块数n，审计者利用总数据块数随机产生的块下标集合和随机数集合$Q = \{i, v_i\}_{i\in[1,n]}$；存储服务提供商根据块下标集合与随机数集合$Q$以及所存储的相应的用户密态文件数据$C_{data}$产生数据完整性证明$(\bar{\rho}, \bar{\sigma})$，随后将审计者的审计材料$TAM$与完整性证明$(\bar{\rho}, \bar{\sigma})$返回给用户。最后审计者利用审计公钥$g^x$与审计材料$TAM$对完整性证明进行检验，检验通过则说明被抽查的数据是完整的，反之则说明被抽查的数据遭到了破坏。

从上述技术方案可知，本发明将完整性审计技术和密文去重技术应用于构建轻量级可去重的密文完整性审计系统，该系统包括文件的完整性审计和密态文件的重复数据删除。首先，用户先对明文文件进行可去重加密，随后将密态文件上传至存储服务提供商，存储服务提供商判断是否已存在系统文件，如果不存在则要求用户上传一组用户标签、标签转化辅助材料和审计材料。如果存在则返回标签转化辅助材料，用户通过标签转化辅助材料产生该用户自己的标签转化辅助材料和对其标签进行审计的审计者的审计材料。接着将自己的标签转化辅助材料和审计材料上传至存储服务提供商。

与现有技术相比，本发明具有如下有益效果：

1、用于审计的公钥不再同以往的方案一样，根据文件数量来决定，而是根据用户的数量来决定，每个用户的文件拥有自己个性化的完整性标签，具有极高的灵活性，解决了文件数量多的情况下，用户的审计公钥多、管理困难的问题。

2、在存在重复文件的情况下，用户无需从头开始计算一组用户个性化完整性标签，而是进行常量级别的计算就可以实现标签的转化，大大降低了文件上传阶段的用时，通信代价也因此降低，值得一提的是，存储服务提供商也可以选择仅存储标签转化辅助材料，这实现了类似于完整性标签去重的效果，降低了存储成本。

3、 系统对密文去重技术的具体方案是解耦的，可以适用于大部分密文去重方案。

附图说明

图 1 是本发明实施例提供的一种轻量级可去重的密文完整性审计方案框架图。

图 2 是本发明实施例提供的一种轻量级可去重的密文完整性审计方案文件

上传阶段流程图。

图 3 是本发明实施例提供的一种轻量级可去重的密文完整性审计方案审计阶段流程图。

具体实施方式

下面结合实施例和附图对本发明做进一步阐述，但本发明的实施方式不限于此。

实施例

本发明轻量级可去重的密文完整性审计方案框架如图 1 所示。在本实施例中，我们假设审计者是诚实的，且审计者、数据拥有者和存储服务提供商两两之间不会合谋。

首先对本发明中所涉及的一些字母及公式的定义进行说明：

$\mathbb{G}$：阶为p的循环群，其中p为素数，群定义为乘法群。

g：乘法群G的一个生成元，是全系统共用的公开安全参数。

x, y：秘密参数，分别作为用户 A、B 的审计私钥，选自Z_p。

pk_A, pk_B, sk_A, sk_B：分别为用于完整性审计的公钥和私钥，其中pk_A为g^x，sk_A为x，pk_B为g^y，sk_B为y。

$h(\cdot)$：为任意哈希函数，其作用为将数据m映射到去重密钥的群上。

k_{dedup}：对文件进行可去重加密时使用的密钥。

ku：用户密钥，作为加密用户去重密钥k_{dedup}时的使用的加密算法的密钥，其可选值域与使用的加密算法有关。

C_{data}：对文件内容进行可去重加密后得到的密态数据。

C_{key}：已用户密钥ku利用加密算法对k_{dedup}加密后获得的密文。

dir、$filename$、ma：文件存储目录、文件名、文件的摘要。

σ_i：文件中第i块数据的完整性标签。

ω、ω'、W：分别为$H_2(x||ma)$、$H_2(y||ma)$、$\omega' - \omega$，其中$H_2(\cdot)$的作用是将

任意数据映射到Z_p。

$H_1(\cdot)$：作用是将任意数据映射到$\mathbb{G}$。

r、r'：临时秘密参数，为随机数，选自Z_p。

v、v'、h、h'、h_1、$h_1{}'$、V：分别为u^r、$u^{r'}$、g^{ω}、$g^{\omega\prime}$、$g^{x\omega}$、$g^{y\omega\prime}$、vv'的简写。

$trans$：$-(\frac{1}{x+H_2(ma||h)}+r)$，用户标签转化辅助变量，具有两个作用，第一，用户可以利用其他用户的此变量派生出用户自身的标签转化辅助变量；第二，服务器能够利用此变量将某特定用户的完整性标砖转化为此变量产生者的完整性标签。（例如用户 B 经由用户 A 的标签转化辅助变量产生了用户 B 的标签转化辅助变量，则存储服务提供商能够借由用户 A 的完整性标签和用户 B 的用户标签转化辅助变量计算出用户 B 的完整性标签，即此处用户 A 被视为某特定用户）。

$trans'$：$\frac{1}{y+H_2(ma||h')}-r'+trans$。

C_{trans}：$E_{k_{dedup}}(\omega)||trans||v$，对部分秘密参数加密后的用户标签转化辅助变量。

TAM、TAM'：分别为$(ma||n||h||h_1)$、$(ma||n||h'||h_1{}')$，审计者的审计材料。

v_i：为审计者产生的挑战中的第i个随机数。

$\bar{\rho}$：$\sum m_i v_i$。

$\bar{\sigma}$：聚合标签，用于验证完整性。

本发明密文完整性审计方案的构建依赖于构建在完整性审计方案上的密文去重技术，利用重复文件的完整性标签可以在辅助变量的帮助下进行转化的特性，从而实现用户上传重复文件时的低计算代价的标签产生算法和标签上传时的低通信代价。如图 2、3 所示，主要包括以下步骤：

步骤一、无重复文件上传阶段，包括上传可去重加密后的文件，计算用户完整性标签和标签转化辅助材料等。

假设数据拥有者 A 拥有文件m，此文件在服务器中不存在重复数据，首先

数据拥有者对文件m进行 hash 计算获得$h(m)$，此外，产生文件m的摘要ma。令去重密钥k_{dedup}为$h(m)$，利用去重密钥k_{dedup}加密文件m获得C_{data}；选取用户密钥ku，具体取值需根据使用的具体加密算法而定，利用用户密钥ku加密去重密钥k_{dedup}获得C_{key}；将文件存储目录dir，文件名$filename$，密态文件数据C_{data}，C_{key}一并发送至存储服务提供商，如果存储服务提供商发现没有重复数据C_{data}，则会将C_{data}，C_{key}分别以$filename.data$和$filename.key$命名，存储于目录dir下，并通知数据拥有者 A 文件m不含有重复文件，并在数据库中记录文件路径。

数据拥有者 A 收到无重复文件的通知后，将C_{data}中的块数据编码到群Z_p上并进一步计算块数据的完整性标签$\sigma_i = H_1(ma||i)^{\omega} \cdot (u^{m_i})^{\frac{1}{x+H_2(ma||h)}}$、计算$trans$进而得到$C_{trans}$；最后，向存储服务提供商发送$\{\sigma_1, \dots, \sigma_n\}$、$C_{trans}$、$TAM$。

步骤二、重复文件上传阶段，包括上传可去重加密后的文件，计算用户标签转化辅助材料等。

数据拥有者 B 向存储服务提供商上传可去重加密后的文件此部分过程与步骤一中数据拥有者 A 向存储服务提供商上传可去重加密后的文件部分过程相同，但在存储服务提供商检测到文件存在重复时，仅在目录dir中存储以$filename.key$命名存储C_{key}，并在数据库中记录重复文件实际存储路径$dir'/filename'.data$，当下次检索$dir/filename.data$时，重定向到$dir'/filename'.data$。

在存储服务提供商通知数据拥有者 B 存在重复文件的同时，一并返回数据拥有者 A 的用户标签转化辅助材料C_{trans}。当数据拥有者 B 获取到C_{trans}时，利用去重密钥k_{dedup}解析并解密出ω、$trans$、v，计算$Trans' = \omega'||trans'||V$，最后将审计材料$TAM'$、$Trans'$一并发送给存储服务提供商，存储服务提供商将$TAM'$、$Trans'$进行存储。

步骤三、密文去重审计系统的审计阶段，包括审计者生成挑战、存储服务提供商为挑战生成完整性证明、审计者验证证明等。

首先，审计者向存储服务提供商发送需要审计的文件名$filename$，存储服

务提供商根据文件名$filename$返回该文件的块数n，审计者根据n产生挑战$Q \in \{(i_1, v_{i_1}), \dots, (i_c, v_{i_c})\}_{i_j \in [1,n]}$，其中$c$是被挑战的数据块总个数，随后发送挑战$Q$、文件名$filename$至存储服务提供商。

然后，存储服务提供商根据$filename$检索出C_{data}，根据Q中下标集合索引出对应数据块，此时有两种情况：第一种情况是审计者帮助数据拥有者 A（即文件的首次存储者）审计，则计算证明$\bar{\rho}$，$\bar{\sigma} = \prod_{(i,v_i)\in Q} {\sigma_i}^{v_i}$；第二种情况是审计者帮助数据审计者 B 审计，则计算证明$\bar{\rho}$，$\bar{\sigma} = \prod_{(i,v_i)\in Q} \left(\sigma_i \cdot H_1(ma||i)^W \cdot (u^{m_i})^{trans'} \cdot V^{m_i}\right)^{v_i}$。最后将$\bar{\rho}$，$\bar{\sigma}$及审计该用户标签的审计材料$TAM$或$TAM'$返回给审计者。

最后，如果审计者接收到审计材料为TAM，则利用对应的审计公钥验证如下式子：

$$e\left(\bar{\sigma}, pk_A \cdot g^{H_2(ma||h)}\right) = e\left(\prod_{(i,v_i)\in Q} [H_1(ma||i)]^{v_i},\ h_1 \cdot h^{H_2(ma||h)}\right) \cdot e(u^{\bar{\rho}}, g)$$

是否成立即可，成立则本次挑战通过，不成立则挑战失败，说明文件的完整性已经受到破坏。如果接收到的审计材料为TAM'，则验证

$e\left(\bar{\sigma}, pk_B \cdot g^{H_2(ma||h')}\right) = e\left(\prod_{(i,v_i)\in Q} [H_1(ma||i)]^{v_i},\ {h_1}' \cdot (h')^{H_2(ma||h')}\right) \cdot e(u^{\bar{\rho}}, g)$

是否成立。

本发明另一实施例提供了一种通信代价更低的方法，将步骤一的上传可去重加密后的文件操作替换成上传文件摘要，随后，存储服务提供商对用户的文件所有权进行验证，从而告知用户是否存在重复文件。

需要说明的是这一方法项实施例是与本发明上述系统项实施例相对应的，其具体的实现原理与上述系统项实施所公开的原理相同，在此不再赘述。

上述实施例为本发明较佳的实施方式，但本发明的实施方式并不受上述实施例的限制，其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化，均应为等效的置换方式，都包含在本发明的保护范围之内。

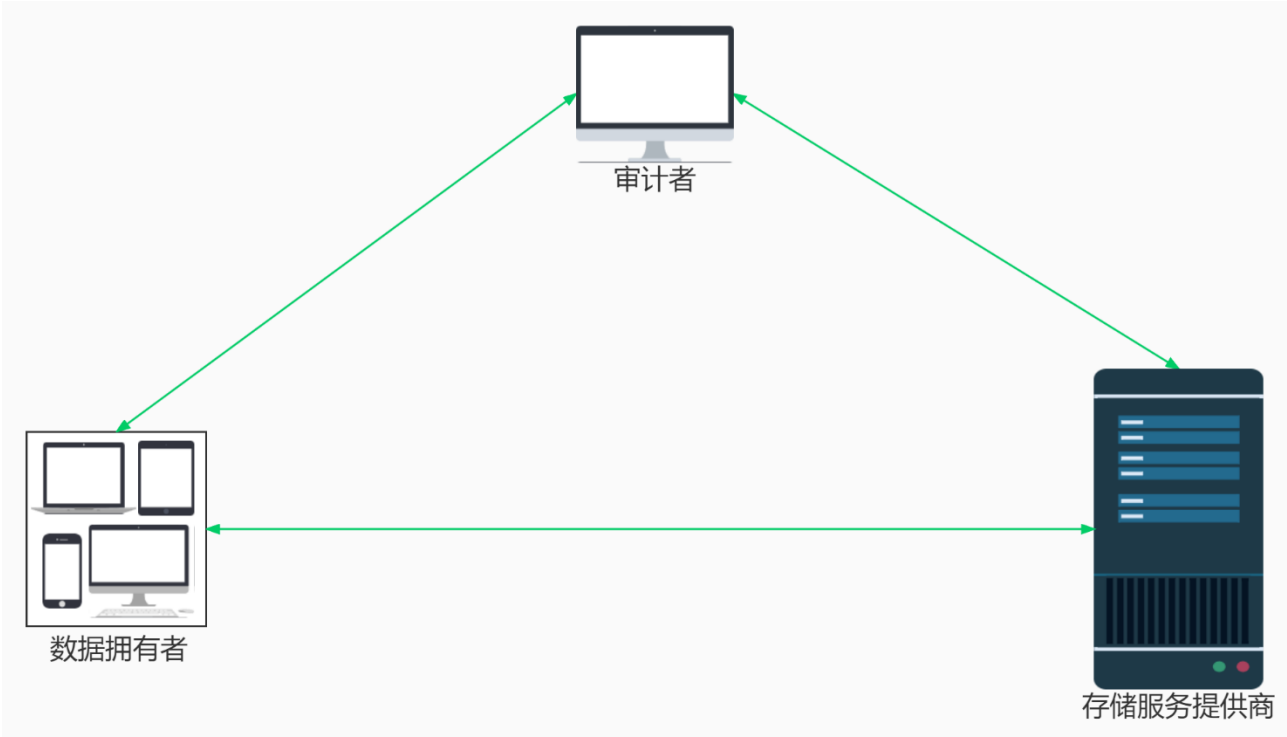


图 1

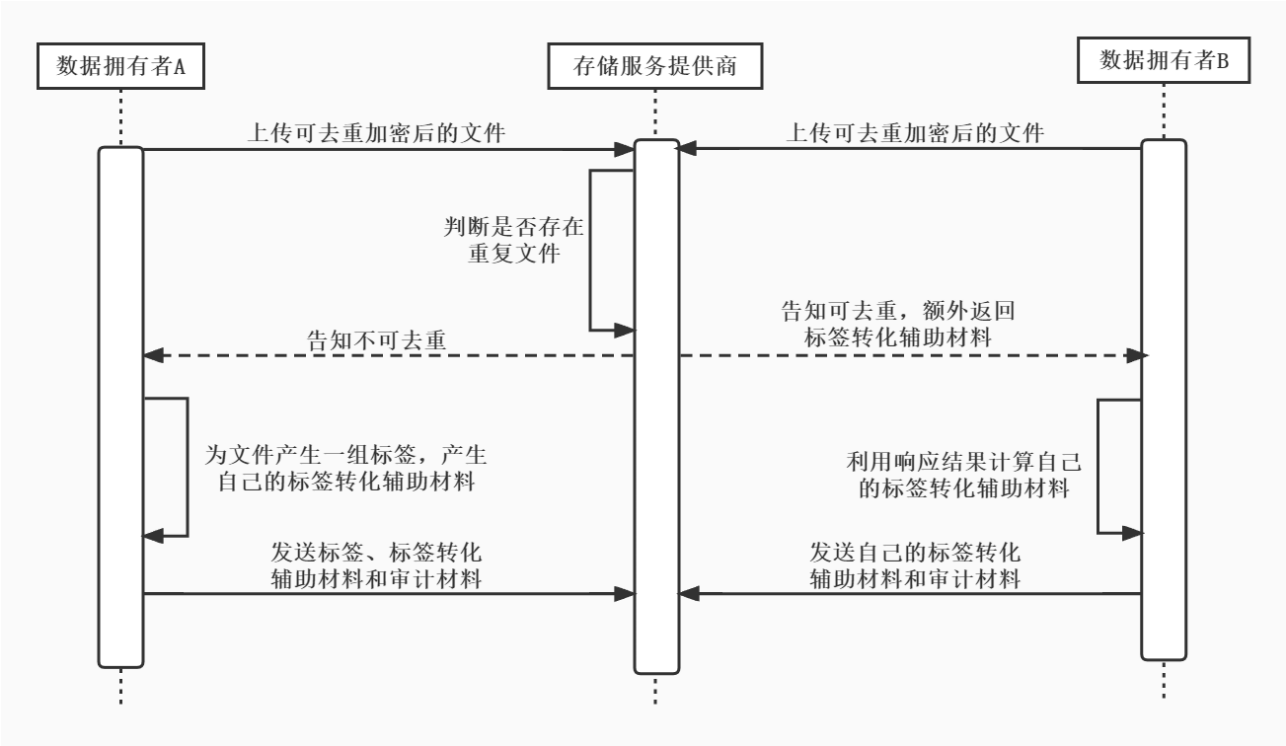

数据拥有者A
存储服务提供商
数据拥有者B
上传可去重加密后的文件
上传可去重加密后的文件
判断是否存在
重复文件
告知可去重，额外返回
标签转化辅助材料
告知不可去重
为文件产生一组标签，产生
自己的标签转化辅助材料
利用响应结果计算自己
的标签转化辅助材料
发送标签、标签转化
辅助材料和审计材料
发送自己的标签转化
辅助材料和审计材料

图 2

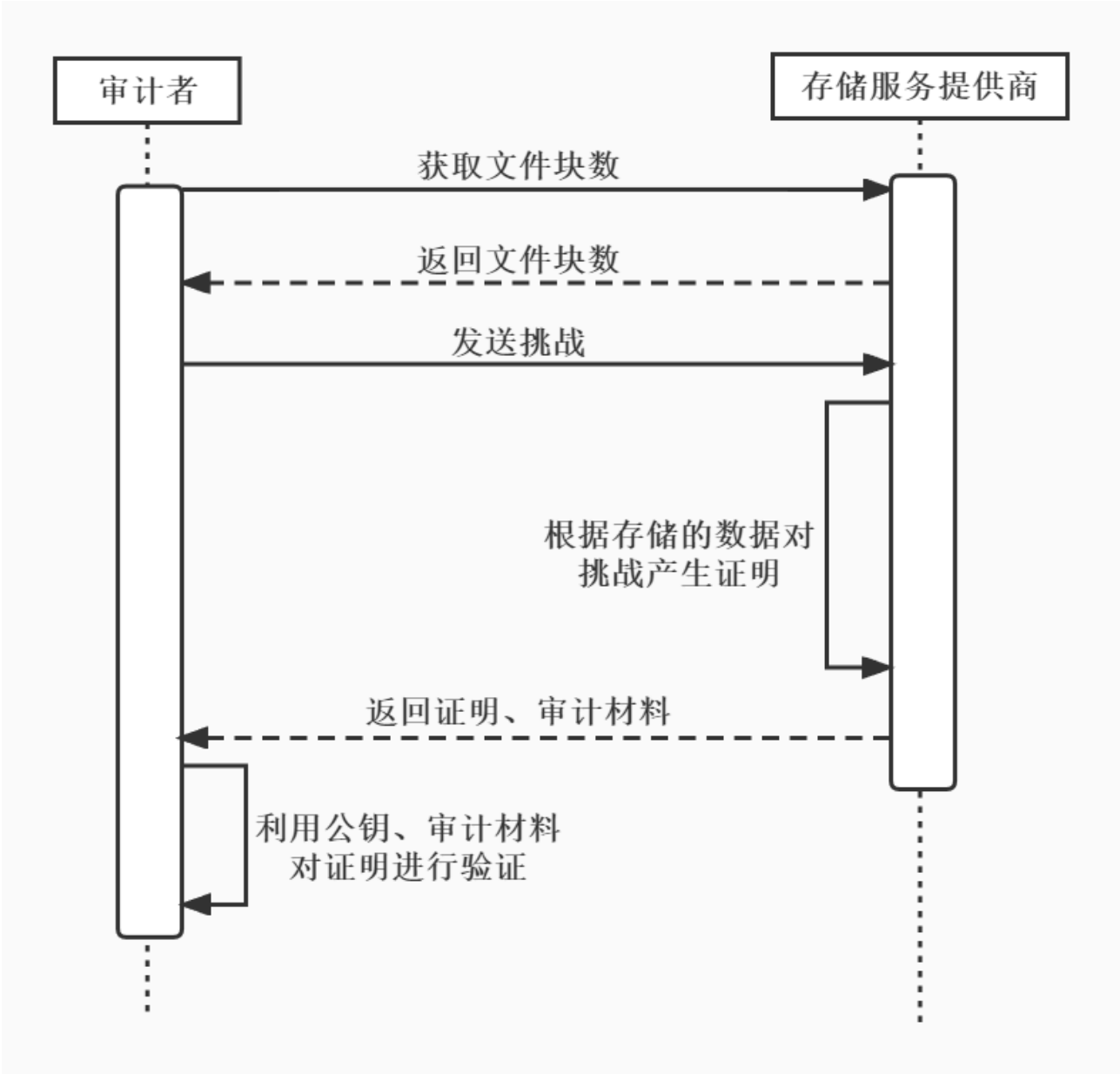

审计者
存储服务提供商
获取文件块数
返回文件块数
发送挑战
根据存储的数据对
挑战产生证明
返回证明、审计材料
利用公钥、审计材料
对证明进行验证

图 3