计算机科学
COMPUTER SCIENCE

对一个基于身份远程数据完整性验证方案的分析与改进

王少辉，赵铮宇，王化群，肖甫

引用本文

王少辉，赵铮宇，王化群，肖甫.对一个基于身份远程数据完整性验证方案的分析与改进[J]．计算机科学，2023，50(7)：302-307.

WANG Shaohui，ZHAO Zhengyu，WANG Huaqun，XIAO Fu. Analysis and Improvement on Identity-based Remote Data Integrity Verification Scheme [J]. Computer Science，2023，50(7)：302-307.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220600067

对一个基于身份远程数据完整性验证方案的分析与改进

王少辉　赵铮宇　王化群　肖　甫
南京邮电大学计算机学院、软件学院、网络空间安全学院　南京 210003
江苏省无线传感网高技术研究重点实验室　南京 210003

摘　要　云存储服务能够让个人或者企业以更低的成本轻松地维护和管理大量数据，但其在为人们带来便利的同时却无法保证其外包数据的完整性和隐私性。远程数据完整性验证方案可以使用户在不下载全部数据的情况下对外包数据的完整性进行验证，即云服务器能够向验证者证明它实际上是在诚实地存储用户的数据。对 Li 等提出的基于身份云存储远程数据完整性验证方案的安全性进行了分析，结果表明该方案容易受到伪造攻击，即云服务器仅需保存少量的数据就能够生成合法的数据完整性证明。在 Li 等方案的基础上，提出了一个新的基于身份远程数据完整性验证方案。分析表明，所提方案能够满足健壮性和隐私性的安全需求，且与 Li 等方案的计算开销也能保持基本一致。

关键词：云存储；数据完整性；隐私保护；基于身份的密码体制；数据安全
中图法分类号　TP309

Analysis and Improvement on Identity-based Remote Data Integrity Verification Scheme

WANG Shaohui，ZHAO Zhengyu，WANG Huaqun and XIAO Fu
1 School of Computer Science，Nanjing University of Posts & Telecommunications，Nanjing 210003，China
2 Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks，Nanjing 210003，China

Abstract　Cloud storage services enable individuals or enterprises to easily maintain and manage large amounts of data at a low cost，but they cannot guarantee the integrity and privacy of outsourced data at the same time. The remote data integrity verification schemes allow users to verify the integrity of outsourced data without downloading all the data，that is，the cloud server can prove to the verifier that it is actually store the user's data honestly. The security of an identity-based privacy preserving remote data integrity verification scheme proposed by Li et al. is analyzed. The analysis shows that the scheme is subjected to forgery attack，that is，the cloud server only needs to store a small amount of data to generate legitimate data integrity proof. Based on Li et al.'s scheme，a new identity-base remote data integrity verification scheme is proposed. The analysis shows that the new scheme can meet the security requirements of privacy and soundness，and the computational cost is basically comparable to that of Li et al.'s scheme.

Keywords　Cloud storage，Data integrity，Privacy preserving，Identity-based cryptography，Data security

1　引言

云计算是一种在存储、服务、应用和处理能力等方面共享虚拟化计算资源的分布式计算模型，这一新型模型受到了学术界和工业界的广泛关注。通过虚拟化技术，云计算能够汇聚和整合巨大的计算资源和强大的计算能力，为客户提供所需的服务。在云计算环境中，云用户可以根据自己的需要来分配和释放资源，目前越来越多的客户和公司租用云存储服务器来维护它们的海量数据。一方面，与部署和维护昂贵的 IT 基础设施相比，租用云服务器需要的投资更少，大大降低了用户在硬件、软件和服务上的成本支出；另一方面，用户在任何有网络的地方都能访问数据，摆脱了地域的限制。

由于云服务器并非完全可信，当数据被外包给云服务器时，必然会存在严重的安全问题，其中最重要的问题之一是云服务器是否能保持用户数据的完整性。首先，云服务器硬件或软件异常可能导致数据损坏或丢失。其次，云服务提供商可能会为了自身的利益而没有正确地保存外包数据，例如，云服务器可能会为了金钱利益去删除一些很少被访问或没有被访问的数据，也可能会为了保持自身良好的信誉而不透露数据丢失的事件。因此，用户需要一种有效的方法来定期验证云服务器中数据的完整性。

为了解决这个问题，2003 年，Deswarte 等[1] 首次提出了利用消息认证码验证远程云服务器上存储数据的完整性的方案，该方案的计算开销和通信开销都较大。2007 年，Ateniese

到稿日期：2022-06-07　返修日期：2022-12-12
基金项目：国家自然科学基金(61872192)
This work was supported by the National Natural Science Foundation of China(61872192).
通信作者：王少辉(wangshaohui@njupt.edu.cn)

等[2]首次提出可证明数据持有（Provable Data Possession，PDP）的模型和方案来验证云存储环境下的数据完整性。在该方案中，用户不需要在本地备份数据，每次验证时，服务器不需要对外包数据的所有数据块进行验证，而是由用户随机指定需要进行完整性验证的数据块。该方案的设计基于RSA算法，方案的计算效率较低。同年，Juels等[3]提出数据可恢复证明（Proof of Retrievability，POR）的模型和方案。该方案使用纠错码对数据进行编码，不仅可以提供存储在云服务器中的数据完整性验证，同时保证了数据丢失后可使用纠错码对数据进行恢复。此后，远程数据完整性验证（Remote Data Integrity Checking，RDIC）方案得到了深入的研究[4-7]。目前多数方案的设计通常基于公钥基础设施（Public Key Infrastructure，PKI）技术，但是基于PKI技术的方案必须要处理复杂的证书管理等问题，其中包括证书的生成、存储、更新和验证，这些操作都比较耗时，且开销较大。Shamir于1984年提出了基于身份的密码体制（Identity-based Cryptography，IBC），在一定程度上解决了基于PKI技术的方案必须要处理复杂的证书管理等问题。在基于身份的密码体制方案中，用户的身份，如姓名、身份证号码、邮箱地址等均被视为公钥，而用户的私钥则由密钥生成中心（Private Key Generation，PKG）为用户生成。此外，在RDIC方案中，通常会选择第三方审计机构来代替用户进行完整性验证，以降低用户的成本开销。由第三方审计机构向云服务器发送验证挑战，而云服务器则依据挑战生成数据完整性证明，进而由第三方审计机构判定数据是否存储完整。

2014年，Wang等[8]首次提出了基于身份的云存储数据完整性验证方案，以避免RDIC方案中复杂的PKI证书管理问题。2016年，Yu等[9]提出了一个基于身份的云存储数据完整性验证方案，该方案的设计基于双线性配对，为了保障验证过程中数据的隐私性，完整性证明阶段需要不同群下两个变量离散对数相等的零知识证明，故效率较低。Wang等[10]提出了面向代理的基于身份的云存储数据完整性验证方案。在文献[11]中，Wang等提出基于身份数据的外包方案，允许用户委托代理人上传数据给云服务器。Zhang等[12]基于身份密码系统提出了对云存储外包数据进行高效公共验证方案，但其无法抵御恶意服务器的攻击。Wang等[13]研究了在多云环境下的分布式完整性验证方案的设计问题。2017年，Li等[14]提出了一个模糊的基于身份的云存储完整性验证方案，该方案进一步简化了密钥证书管理问题。

最近，Li等在《IEEE Systems Journal》期刊上发表的文章中提出了一个新的基于身份的远程数据完整性验证方案[15]，其采用与文献[9]不同的同态可验证标签来降低验证的复杂度，并在证明过程中加入随机数以保护用户数据的隐私性不受TPA侵害。本文首先对文献[15]中方案的安全性进行了分析研究，研究指明该方案不能满足健壮性的安全需求，即云服务器即便没有完整的存储用户数据，也能够通过已有的数据伪造证明来通过TPA的验证。进而，我们在文献[15]的基础上，采用零知识证明提出了一个改进的基于身份远程数据完整性验证方案。本文方案在提供正确性、健壮性和隐私性安全需求的同时，其存储成本和文献[15]中的方案一致，而计算和传输成本略高于文献[15]中的方案。

2 预备知识

2.1 双线性映射和计算性难题

设G_1和G_2是阶均为大素数q的两个乘法循环群，g是群G_1的一个生成元。G_1和G_2间的双线性映射$e:G_1\times G_2\rightarrow G_2$，具有以下3个性质。

（1）可计算性：对于任意$g_1,g_2\in G_1$，存在有效的算法计算$e(g_1,g_2)$。

（2）双线性：对于任意$a,b\in Z_q^*$和任意$g_1,g_2\in G_1$，有$e(g_1^a,g_2^b)=e(g_1,g_2)^{ab}$。

（3）非退化性：存在$g_1,g_2\in G_1$，使得$e(g_1,g_2)\neq 1$。

定义1（Computational Diffie-Hellman（CDH）problem）假设g为乘法循环群G的生成元，a,b为群Z_p^*中的两个随机数，给定值g^a和g^b，那么求解值g^{ab}在计算是上困难的。

2.2 系统模型与安全定义

基于身份云存储远程数据完整性验证方案主要由4方实体组成，分别是密钥生成中心（KGC）、云存储服务器（CSP）、用户和第三方审计机构（TPA），其系统模型如图1所示。

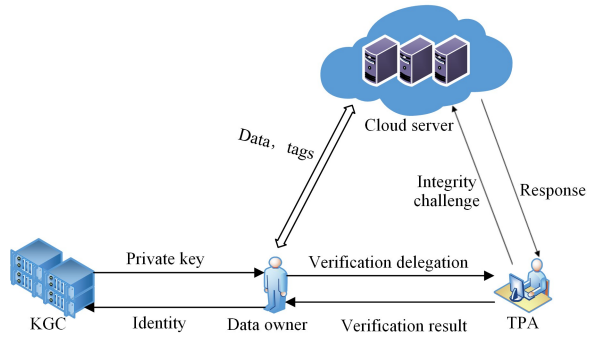

图1 基于身份的数据完整性验证方案系统模型

Fig. 1 System model of identity-based remote data integrity verification scheme

（1）KGC是完全可信任的机构，它利用系统用户的身份，计算出对应的用户私钥，并通过安全通道将私钥传输给用户。

（2）CSP具有庞大的存储容量和强大的计算能力，它向用户提供云存储服务，CSP收到TPA的数据完整性挑战后，根据挑战生成并返回相应的完整性证明。

（3）用户租用CSP的云存储服务，将自身大量的隐私数据外包给CSP，用户也可以委托TPA对存储在CSP上的外包数据进行数据完整性审计工作。

（4）作为第三方审计机构，TPA根据用户的委托，发送数据完整性挑战给CSP。当收到由CSP生成的证明后，TPA验证所收到证明的有效性，进而验证数据是否完整，并将结果返回给用户。

基于身份云存储数据完整性验证方案一般由6个算法组成，分别是Setup，Extract，TagGen，Challenge，ProofGen和ProofVerify，具体描述如下。

$Setup(1^k)\rightarrow(params,msk)$：该算法由KGC执行，它以安全参数$k$作为输入，输出系统的公共系统$params$和主密钥$msk$。

$Extract(ID, msk) \rightarrow sk_{ID}$：该算法是由 KGC 和用户交互执行，它以主密钥 msk 和用户的身份信息 $ID \in \{0,1\}^*$ 作为输入，输出用户私钥 sk_{ID}。

$TagGen(params, sk_{ID}, F) \rightarrow T$：该算法由身份信息为 ID 的用户为数据 F 生成认证标签。通常用户会将外包数据 F 分成 n 个数据块 m_i，利用系统公共参数 $params$ 和用户私钥 sk_{ID}，为每个数据块 m_i 计算得到认证标签 T_i。最后用户将外包数据 F 和认证标签 $T=(T_1, \cdots, T_n)$ 一并存储至云服务器中。

$Challenge(params, F_{id}, ID) \rightarrow chal$：该随机算法由 TPA 执行，它以系统公共参数 $params$、文件数据名 F_{id} 和用户身份 ID 作为输入，输出挑战信息 $chal$。

$ProofGen(params, ID, chal, F, T) \rightarrow P$：云服务器接收到由 TPA 产生的挑战信息 $chal$，根据存储在服务器中的数据文件 F 和认证标签 T，生成并返回对应挑战的完整性证明 P。

$ProofVerify(params, ID, chal, P, F_{id}) \rightarrow \{1,0\}$：TPA 执行该算法，它对云服务器的完整性证明 P 进行验证，如果通过验证则输出 1，否则输出 0。

一个安全的基于身份远程数据完整性验证方案通常需要满足正确性、健壮性和隐私性等安全需求。其中，正确性指如果用户的外包数据完整地存储于 CSP，且 CSP 和 TPA 都诚实而正确地执行数据完整性验证方案的操作，则 CSP 生成的证明 P 一定能够通过 TPA 的认证；而隐私性指 TPA 在数据完整性验证过程中没有获得关于原始数据的任何信息，即方案对 TPA 提供存储数据的机密性保护。

健壮性指当用户的外包数据部分丢失或损坏，方案可以抵抗 CSP 通过伪造证明来欺骗 TPA 的行为。也就是说，对于没有被完整存储的用户数据块，CSP 不能生成并返回正确的数据完整性证明。下面我们给出健壮性的定义。

定义 1（健壮性） 方案通过如下挑战者和敌手之间的安全游戏对健壮性安全需求进行定义，安全游戏中敌手扮演 CSP 的角色，试图通过伪造数据完整性证明欺骗 TPA。

(1)初始化：挑战者运行 Setup 算法，获取系统参数 $param$ 和主私钥 msk，挑战者将系统参数 $param$ 发送给敌手，并秘密保存主私钥 msk。

(2)质询：敌手向挑战者发起一系列质询，其中包括 Extract 质询和 TagGen 质询。

1)Extract 质询：敌手可以向挑战者质询任意身份 ID_i 所对应的私钥，挑战者执行 Extract 算法计算得到相应的私钥 sk_{ID_i}，并返回给敌手。

2)TagGen 质询：敌手可以请求身份为 ID_i 的用户为文件 F 生成验证标签。此时挑战者先执行 Extract 算法获取私钥 sk_{ID_i}，再执行 TagGen 算法来生成文件 F 的标签，最后将验证标签返回给敌手。

(3)ProofCheck：敌手执行 ProofGen 算法为文件 F 的数据块生成完整性证明，并将证明返回给挑战者，挑战者再执行 ProofVerify 算法验证证明的合法性。

(4)Output：最后，敌手给出文件 $F_{id'}$ 中一组数据块的完整性证明 P，并且这一组数据块至少有一个数据块没有执行过 TagGen 质询。如果证明 P 通过了完整性验证，那么敌手就赢得了这场安全游戏。

如果任意多项式时间的敌手在上述安全游戏中获胜的概率是可忽略的，则称方案满足健壮性的安全需求。

3 对 Li 等方案的安全性分析

本节给出对 Li 等[15]提出的基于身份的云存储远程数据完整性验证方案的安全分析。通过分析可知，文献[15]中的方案不能满足健壮性的安全需求，即云服务器能够在未完整保存数据的条件下伪造合法的证明。

下面我们给出方案的简要介绍和必要说明，方案的具体细节可参考文献[15]。

Setup：针对安全参数 k，KGC 随机选取阶为素数 q 的循环群 G_1 和 G_2。g 是群 G_1 的生成元，e 表示双线性配对运算。方案涉及两个抗碰撞哈希函数 $H_1: \{0,1\}^* \rightarrow G_1$ 和 $H_2: \{0,1\}^* \rightarrow G_1$。$\phi: Z_q^* \times \{1, \cdots, n\} \rightarrow Z_q^*$ 和 $\pi: Z_q^* \times \{1, \cdots, n\} \rightarrow \{1, \cdots, n\}$ 分别表示伪随机函数运算和伪随机置换运算。KGC 选取随机值 $s \in Z_q^*$ 作为主私钥，计算 $P_0 = g^s$ 作为主公钥，系统公共参数为 $params=(q, g, G_1, G_2, e, P_0, H_1, H_2, \phi, \pi)$。

Extract：当收到用户身份信息 ID，KGC 输出用户私钥为 $sk_{ID} = H_1(ID)^s$。

TagGen：设数据文件 F 文件名为 F_{id}，用户选取两个随机值 $\chi \in G_1$，$\lambda \in Z_q^*$。数据块 $m_i \in Z_q$ 的认证标签计算式如式(1)所示：

$$T_i = sk_{ID} \cdot (H_2(F_{id} \| i) \cdot \chi^{m_i})^{\lambda} \tag{1}$$

然后，用户计算 $R=g^{\lambda}$，并选择签名算法 Sig 来计算文件标签 $T_{Fid} = Sig(R \| \chi \| Fid)$。最后用户将 $(F, R, \chi, \{T_i\}_{i=1}^n, T_{Fid})$ 上传给云服务器。云服务器利用式(2)来检查每个数据块标签的正确性。

$$e(T_i, g) = e(H_1(ID), P_0) \cdot e(H_2(F_{id} \| i) \cdot \chi^{m_i}, R) \tag{2}$$

Challenge：TPA 随机选取 $k_1, k_2 \in Z_q^*$ 以及挑战块的数量 $c(1 \leqslant c \leqslant n)$，将挑战 $chal=(c, k_1, k_2)$ 和文件名 F_{id} 传输给 CSP。

ProofGen：CSP 收到挑战后，首先计算挑战集 $C=\{(v_i, a_i)\}_{i=1}^c$，其中 $v_i = \pi(k_1, i)$ 表示第 i 个挑战块的索引，$a_i = \phi(k_2, i)$ 是随机值。然后 CSP 选取 $r \in Z_q^*$，并计算 $W=\chi^{-r}$，$\sigma = \prod_{(v_i, a_i) \in C} T_{v_i}^{a_i}$ 和 $M = \sum_{(v_i, a_i) \in C} a_i (r + m_{v_i})$。最后，CSP 将证明 $P=(W, \sigma, M)$ 返回给 TPA。

ProofVerify：TPA 收到证明 P 后，验证以下等式是否成立。

$$e(\sigma, g) = e(H_1(ID)^{\sum_{(v_i, a_i) \in C} a_i}, P_0) \cdot e(\prod_{(v_i, a_i) \in C} H_2(F_{id} \| v_i)^{a_i} \cdot \chi^M \cdot W^{\sum_{(v_i, a_i) \in C} a_i}, R) \tag{3}$$

安全性分析：假设 CSP 每次需验证 k 块数据的完整性，即 c 设定为 k。下面我们说明 CSP 只需要存储 k 块数据及其标签(不妨设为 $\{m_i, T_i\}_{i=1}^k$)，即可伪造合法的证明通过 TPA 的挑战。当 CSP 接收到 TPA 的挑战 $chal=(c, k_1, k_2)$ 后，

计算 $C=\{(v_i,a_i)\}_{i=1}^{k}$，其中 $v_i=\pi(k_1,i)$，$a_i=\phi(k_2,i)$，设 $\beta=(\sum_{(v_i,a_i)\in C}a_i)\bmod q$，然后计算并返回如下的$(W^*,\sigma^*,M^*)$：

$$W^*=(\prod_{j=1}^{k}H_2(F_{id}\|j)^{a_j})^{\beta^{-1}}\cdot(\prod_{v_j\in C}H_2(F_{id}\|v_j)^{-a_i})^{\beta^{-1}}\cdot\chi^{-r}$$

$$\sigma^*=\prod_{i=1}^{k}T_i^{a_i},M^*=\sum_{j=1}^{k}a_j(r+m_j)$$

下面我们证明(W^*,σ^*,M^*)可以满足式(3)的验证，即云服务器可以伪造合法的证明以通过 TPA 的验证。

$$\begin{aligned}\text{等式右边}&=e(H_1(ID)^{\beta},P_0)\cdot e(\prod_{v_i\in C}H_2(F_{id}\|v_i)^{a_i}\cdot\chi^{M^*}\cdot W^{*\beta},R)\\&=e(H_1(ID)^{\beta},g^s)\cdot e(\prod_{j=1}^{k}H_2(F_{id}\|j)^{a_j}\cdot\chi^{M^*}\cdot\chi^{-r\beta},g^{\lambda})\\&=e((sk_{ID})^{\beta},g)\cdot e(\prod_{j=1}^{k}H_2(F_{id}\|j)^{a_j}\cdot\chi^{\sum_{j=1}^{k}a_jm_j},g^{\lambda})\\&=e((sk_{ID})^{\beta},g)\cdot e((\prod_{j=1}^{k}H_2(F_{id}\|j)^{a_j}\cdot\chi^{\sum_{j=1}^{k}a_jm_j})^{\lambda},g)\\&=e(\sigma^*,g)\end{aligned}$$

4 新的改进方案

在文献[9]中，为了获得健壮性和隐私性的安全需求，完整性证明阶段需要 G_1，G_2 群中两个变量离散对数相等的零知识证明。可以看出，针对文献[15]中的方案，CSP 之所以能够伪造成功，在于发送的证明消息 $W=\chi^{-r}$ 中缺少了 CSP 知道 W 关于 χ 的离散对数的证明。与文献[9]不同，这里 CSP 只需要零知识证明其知道 W 关于 χ 的离散对数，而基于 Schnorr 签名方案的零知识证明方案可高效地解决这一问题，其证明效率远优于文献[9]所需要的零知识证明方案。下面我们在文献[15]的基础上，提出基于身份云存储远程数据完整性验证改进方案。改进方案的 Setup，Extract，TagGen，Challenge 算法与文献[15]的方案一致，这里不再赘述。下面给出改进后 ProofGen 和 ProofVerify 算法的具体介绍：

ProofGen：CSP 收到 $chal=(c,k_1,k_2)$后，计算挑战集 $C=\{(v_i,a_i)\mid i\in[1,c]\}$，其中，$v_i=\pi(k_1,i)$表示第 i 个挑战块的位置索引，$a_i=\phi(k_2,i)$是随机数。然后 CSP 选取随机值 $k,r\in Z_q^*$，并计算 $W=\chi^{-r}$，$\sigma=\prod_{(v_i,a_i)\in C}T_{v_i}^{a_i}$，$M=\sum_{(v_i,a_i)\in C}a_i(r+m_{v_i})$。为了证明 CSP 知道 W 关于 χ 的离散对数，CSP 还需要计算 $V=\chi^k$，$e=H_1(W\|V)$，$s=k+r\cdot e$。

最后，CSP 将证明 $P=(W,\sigma,M,e,s)$返回给 TPA。

ProofVerify：TPA 收到证明 P 后，首先计算 $V'=\chi^s\cdot W^e$，再验证 $e'=H_1(W\|V')$是否与 e 相等。如果不相等，则 TPA 验证不通过，否则，继续验证式(4)是否成立。

$$e(\sigma,g)=e(H_1(ID)^{\sum_{(v_i,a_i)\in C}a_i},P_0)\cdot e(\prod_{(v_i,a_i)\in C}H_2(F_{id}\|v_i)^{a_i}\cdot\chi^{M}\cdot W^{\sum_{(v_i,a_i)\in C}a_i},R)\tag{4}$$

5 安全性分析和性能分析

5.1 安全性分析

本节分别对方案的正确性、健壮性和隐私性进行分析。

正确性：相较于文献[15]，我们利用 Schnorr 签名增加了对 $W=\chi^{-r}$ 关于 χ 的零知识证明。如果用户和云服务器都是诚实的，参考文献[15]，验证式(4)和式(3)一致，显然成立。而由 $s=k+r\cdot e$ 可知：$\chi^s=\chi^k+\chi^{r\cdot e}$，即 $\chi^k=\chi^s\chi^{-r\cdot e}=\chi^sW^e$，从而 $e=H_1(W\|\chi^k)$成立。

健壮性：根据文献[15]，我们通过定理 1 证明任何数据块的标签都不能被云服务器伪造，并且云服务器不能伪造对应挑战的数据完整性证明。

定理 1[15] 如果一个多项式时间的敌手 A 在最多进行了 qH_1,q_k,qH_2,q_T 次的 H_1 质询，$Extract$ 质询，H_2 质询和 Tag 质询，并且在时间 t 内以 ε 的优势成功伪造了数据块的标签，那么存在算法 S 在 $t'\leqslant t+O(qH_1+q_k+qH_2+q_T)$的时间内以优势 $\varepsilon'\geqslant\varepsilon/((q_k+q_T)\cdot 2e)$解决了 CDH 问题。

下面说明新方案中 CSP 不能伪造数据完整性证明。假设挑战信息为 $chal=(c,k_1,k_2)$，通过计算 $v_i=\pi(k_1,i)$和 $a_i=\phi(k_2,i)$来分别表示挑战块索引和随机参数。数据完整性证明 $P=(W,\sigma,M,e,s)$包含 5 个参数，其中 e,s 采用了标准的 Schnorr 签名，如果通过验证，则证明 CSP 知道 W 关于 χ 的离散对数。我们不妨假设此时 $W=\chi^{-z}$。设 $\beta=(\sum_{(v_i,a_i)\in C}a_i)\bmod q$，由验证式(4)的右端，可知：

$$\begin{aligned}&e(H_1(ID)^{\beta},P_0)\cdot e(\prod_{(v_i,a_i)\in C}H_2(F_{id}\|v_i)^{a_i}\cdot\chi^{M}\cdot W^{\beta},R)\\&=e(H_1(ID)^{\beta},g^s)\cdot e(\prod_{(v_i,a_i)\in C}H_2(F_{id}\|v_i)^{a_i}\cdot\chi^{M}\cdot\chi^{-z\beta},g^{\lambda})\\&=e(sk_{ID}^{\beta},g)\cdot e((\prod_{(v_i,a_i)\in C}H_2(F_{id}\|v_i)^{a_i}\cdot\chi^{M}\cdot\chi^{-z\beta})^{\lambda},g)\\&=e(sk_{ID}^{\beta}(\prod_{(v_i,a_i)\in C}H_2(F_{id}\|v_i)^{a_i}\cdot\chi^{M}\cdot\chi^{-z\beta})^{\lambda},g)\end{aligned}$$

此时必有：

$$\sigma=sk_{ID}^{\beta}(\prod_{(v_i,a_i)\in C}H_2(F_{id}\|v_i)^{a_i}\cdot\chi^{M}\cdot\chi^{-z\beta})^{\lambda}$$

由 $\beta=(\sum_{(v_i,a_i)\in C}a_i)\bmod q$ 和定理 1 有：

$$\sigma=\prod_{(v_i,a_i)\in C}(sk_{ID}(H_2(F_{id}\|v_i)\chi^{m_{v_i}})^{\lambda})^{a_i}\cdot(\chi^{M}\cdot\chi^{-z\beta}\cdot\chi^{-\sum_{(v_i,a_i)\in C}a_im_{v_i}})^{\lambda}$$

从而有：

$$M=z(\sum_{(v_i,a_i)\in C}a_i)+\sum_{(v_i,a_i)\in C}a_im_{v_i}=\sum_{(v_i,a_i)\in C}a_i(z+m_{v_i})$$

即 CSP 只能以协议要求的方式生成合法的安全证明，而此时 CSP 必须保存所有的外包数据。

隐私性：由定理 2 可以看出，TPA 不能从数据完整性证明的过程中获得关于用户数据的任何信息。

定理 2 在新方案中，TPA 在收到数据完整性证明 $P=(W,\sigma,M,e,s)$后，不能通过计算获取用户存储在云服务器的数据。

证明：TPA 在接收到证明 $P=(W,\sigma,M,e,s)$之后，其中 W,e,s 都不包含任何数据块的信息。由于任意单个数据块的

标签都不能被云服务器伪造，因此 TPA 不能从 σ 中计算并推导出数据块的信息。对于 $M=\sum_{(v_i,a_i)\in C} a_i(r+m_{v_i})$，因为 r 的值是由 CSP 随机选取，对于 TPA 而言，M 与随机值不可区分，所以 TPA 也无法从 M 中获得数据块 m_{v_i} 的值。因此，TPA 不能从数据完整性证明中获得关于被挑战数据块的信息，故新方案满足隐私性的安全需求。

5.2 效率分析

可以看出，新方案的存储成本和文献[15]中的方案一致。而相较于文献[15]，新方案云服务器需要额外传送两个消息 $\{e,s\}$，传输成本略有增加。下面我们从计算成本的角度对新方案和文献[15]中的方案进行比较。假设在这些方案中存储相同的数据文件 F，而 TPA 产生的挑战信息 $chal$ 也一样。这里我们仅考虑成本较大的双线性配对和群 G_1 中的模乘、模幂运算，而忽略哈希、模加等运算。我们分别用 T_{mul}，T_p 和 T_{exp} 表示在群 G_1 中执行一次模乘、一次配对和一次模幂运算所需要的开销。假设数据文件一共被分为 n 块，其中有 c 块被挑战。

KGC 运行 *Extract* 算法为用户生成私钥，这里需要进行一次 T_{exp} 操作。

TagGen 算法需要运行 n 次才能生成所有数据块的标签，计算总开销为 $2nT_{exp}+2nT_{mul}$。

为了生成完整性证明，云服务器需要 $(c+2)\cdot T_{exp}+(c-1)\cdot T_{mul}$ 的计算开销，即 $c+2$ 次模幂运算、$c-1$ 次模乘运算。

在 ProofVerify 算法中，TPA 需要执行 $(c+5)T_{exp}+(c+2)T_{mul}+3T_p$ 的开销，即 $c+5$ 次模幂运算、$c+2$ 次模乘运算和 3 次配对运算。

本文方案与文献[15]中方案的计算开销对比如表 1 所列。从表中可以看出，在 ProofGen 算法中，本文方案比文献[15]多 1 次模幂运算，在 ProofVerify 算法中，本文方案比文献[15]多 2 次模幂和 1 次模乘。虽然本文方案的计算成本略高于文献[15]的方案，但是在一定程度上，本文方案与文献[15]中方案的计算成本相差不大。

表 1 本文方案与文献[15]方案的计算成本比较

Table 1 Comparison of computational cost between our scheme and literature [15] 's scheme

方案	Tag generation	Proof generation	Proof verify
文献[15]	$2T_{exp}+2T_{mul}$	$(c+1)T_{exp}+(c-1)T_{mul}$	$(c+3)T_{exp}+(c+1)T_{mul}+3T_p$
本文方案	$2T_{exp}+2T_{mul}$	$(c+2)T_{exp}+(c-1)T_{mul}$	$(c+5)T_{exp}+(c+2)T_{mul}+3T_p$

5.3 实验仿真

本节对本文方案进行了实验仿真，并给出了本文方案和文献[15]中方案在 ProofGen 和 ProofVerify 算法中的性能比较。实验基于 Java 语言和 JPBC 库，采用 AMD Ryzen 54600 H with Radeon Graphics 3.00 GHz 处理器，16 GB 运行内存，在 windows 1064 位系统的主机环境下完成。我们选择参数 a.properties 作为 JPBC 库的参数，对比实验分为两组。第一组实验中，系统生成的文件大小为 1 MB，分块大小为 164 bit，数据块共计 50000 块；第二组实验中，系统生成的文件大小为 2 MB，分块大小为 328 bit，数据块共计 50000 块。为了得到更精确的结果，各阶段都进行 50 次实验。

当系统生成的文件大小为 1MB 时，在分别选择不同的挑战块数量时，两个方案 ProofGen 算法和 ProofVerify 算法的计算时间比较结果如图 2 和图 3 所示。

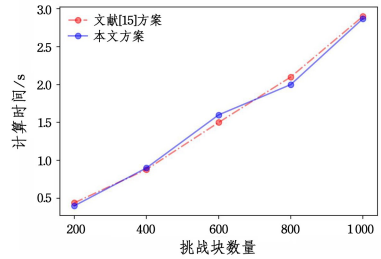

图 2 实验 1 Proof generation 计算时间对比

Fig. 2 Computing time comparison of Proof generation in experiement 1

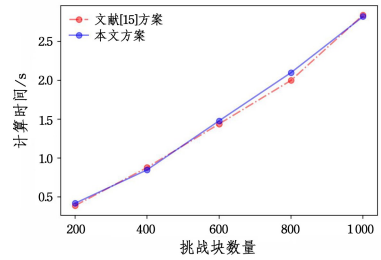

图 3 实验 1 Proof verify 计算时间对比

Fig. 3 Computing time comparison of Proof verify in experiement 1

当系统生成的文件大小为 2MB 时，在分别选择不同的挑战块数量时，两个方案 ProofGen 算法和 ProofVerify 算法的计算时间比较结果如图 4 和图 5 所示。

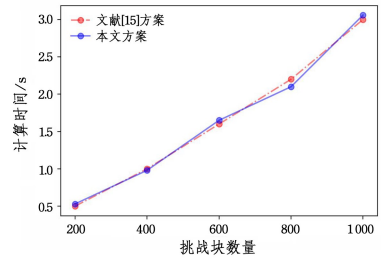

图 4 实验 2 Proof generation 计算时间对比

Fig. 4 Compuing time comparison of Proof generation in experiement 2

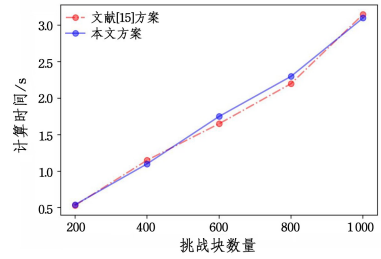

图 5 实验 2 Proof verify 计算时间对比

Fig. 5 Computing time comparison of Proof verify in experiement 2

由图 2 和图 4 可以得出，数据块由 164 bit 增大到 328 bit

时,计算时间略有增加。在选择不同数量的挑战块时,在ProofGen算法阶段本文方案和文献[15]方案的计算时间相差并不大;且随着挑战块数量的增多,计算时间呈线性增加。

同样地,由图3和图5可以得出,在ProofVerify算法阶段,选择不同数量的挑战块以及不同大小的数据块时,本文方案的计算开销与所选择数据块的数量是线性相关的,且与文献[15]中方案的计算时间相差不大。可以看出,本文所提改进方案在与文献[15]中方案的计算时间保持基本一致的同时,还满足了健壮性的安全需求。

结束语 本文对文献[15]所提基于身份的远程数据完整性验证方案进行了安全性分析,分析结果指出该方案易遭受伪造攻击,即云服务器即便没有完整地保存全部用户数据,也能够利用保存的少量外包数据来伪造合法的证明,从而通过TPA的挑战。因此,文献[15]方案不能满足方案健壮性的安全需求。针对文献[15]方案存在的安全问题,我们提出了一个新的改进方案。在新方案中,TPA可以正确地验证数据的完整性;并且效率分析和实验仿真表明,新方案的计算效率和文献[15]方案可以保持基本一致,且新方案能够满足健壮性和隐私性的安全需求。

在目前基于身份的密码体制下,云存储数据完整性验证方案存在支持用户撤销、外包数据的动态更新等扩展性问题,这些问题将是我们下一步研究的重点问题。

参 考 文 献

[1] DESWARTE Y,QUISQUATER J J,SAIDANE A. Remote Integrity Checking[C]// Working Conference on Integrity & Internal Control in Information Systems. Springer,2003:1-11.

[2] ATENIESE G,BURNS R,CURTMOLA R,et al. Provable Data Possession at Untrusted Stores[C]// Proceedings of the 14th ACM Conference on Computer & Communications Security. 2007:598-609.

[3] JUELS A,KALISKI JR B S. PORs:Proofs of Retrievability for Large Files[C]// Proceedings of the 14th ACM Conference on Computer & Communications Security. 2007:584-597.

[4] SHACHAM H,WATERS B. Compact Proofs of Retrievability [C]// International Conference on the Theory & Application of Cryptology & Information Security. Springer,2008:90-107.

[5] ATENIESE G,DI PIETRO R,MANCINI L V,et al. Scalable and Efficient Provable Data Possession[C]// Proceedings of the 4th International Conference on Security & Privacy in Communication Networks. ACM,2008:1-10.

[6] WANG Q,WANG C,REN K,et al. Enabling public auditability and data dynamics for storage security in cloud computing[J]. IEEE Transactions on Parallel and Distributed Systems,2010,22(5):847-859.

[7] SHEN W,YU J,XIA H,et al. Light-weight and privacy-preserving secure cloud auditing scheme for group users via the third party medium[J]. Journal of Network & Computer Applications,2017,82:56-64.

[8] WANG H,WU Q,QIN B,et al. Identity-based remote data possession checking in public clouds[J]. IET Information Security,2014,8(2):114-121.

[9] YU Y,MAN H A A,ATENIESE G,et al. Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage[J]. IEEE Transactions on Information Forensics and Security,2017,12(4):767-778.

[10] WANG H,HE D,TANG S. Identity-based proxy-oriented data uploading and remote data integrity checking in public cloud[J]. IEEE Transactions on Information Forensics and Security,2016,11(6):1165-1176.

[11] WANG Y,WU Q,QIN B,et. al. Identity-based data outsourcing with comprehensive auditing in clouds[J]. IEEE Transactions on Information Forensics and Security,2016,12(4):940-952.

[12] ZHANG J,DONG Q. Efficient ID-based public auditing for the outsourced data in cloud storage[J]. Information Sciences,2016,343:1-14.

[13] WANG H Q. Identity-based distributed provable data possession in multicloud storage[J]. IEEE Transactions on Services Computing,2014,8(2):328-340.

[14] LI Y,YU Y,MIN G,et al. Fuzzy identity-based data integrity auditing for reliable cloud storage systems[J]. IEEE Transactions on Dependable & Secure Computing,2017,16(1):72-83.

[15] LI J,YAN H,ZHANG Y. Identity-based privacy preserving remote data integrity checking for cloud storage[J]. IEEE Systems Journal,2020,15(1):577-585.

WANG Shaohui, born in 1977, Ph. D, vice-professor. His main research interests include information security and applied cryptography.

(责任编辑:杨雪敏)