适用于再生编码分布式存储的轻量型隐私保护审计方案

刘光军[1]，郭网媚[2]，熊金波[3]，刘西蒙[4]，董长宇[5]

（1. 西安文理学院信息工程学院，陕西 西安 710065； 2. 西安电子科技大学综合业务网理论及关键技术国家重点实验室，陕西 西安 710071；
3. 福建师范大学数学与信息学院，福建 福州 350117； 4. 福州大学数学与计算机科学学院，福建 福州 350108；
5. 英国纽卡斯尔大学计算机学院，纽卡斯尔 NE4 5TG）

摘 要：为了降低面向再生编码分布式存储系统的外包数据审计机制的安全实现开销，提出了一种正交化代数编码方法，以此构造一类基于线性同态认证的轻量型隐私保护审计方案。利用文件编码数据与私有密钥向量的正交化构造外包存储向量的同态认证标签，并提出利用密钥特定分量的正交基向量组的随机化掩码来完成审计响应消息的隐私保护，实现代数编码、隐私保护和安全审计的高效融合。理论分析表明，所提方案在再生编码分布式存储应用中可实现信息理论意义下的安全性。与现有同类工作相比，该方案计算复杂度低，通信开销小，具有更好的性能优势。

关键词：数据审计；隐私保护；再生码；网络编码；分布式存储

中图分类号：TP309

文献标识码：A

DOI: 10.11959/j.issn.1000−436x.2021116

Lightweight privacy protection data auditing scheme for regenerating-coding-based distributed storage

LIU Guangjun[1], GUO Wangmei[2], XIONG Jinbo[3], LIU Ximeng[4], DONG Changyu[5]

1. School of Information Engineering, Xi'an University, Xi'an 710065, China
2. State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China
3. College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350117, China
4. College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China
5. School of Computing, Newcastle University, Newcastle NE4 5TG, UK

Abstract: To reduce the security implementation cost of the outsourcing data audit mechanism for the regenerating-coding-based distributed storage systems, an orthogonal algebraic coding method was put forward to construct a lightweight privacy-preserving audit scheme based on linear homomorphic authentication. The homomorphic authentication tags were generated with the orthogonalization between the file encoded data and the private secret key vector, and the privacy protection of the auditing response message was achieved by using the random masking that was constructed by randomizing the orthogonal basis vectors of the specific sub-vector of the user's secret key. The work realized the effective integration of algebraic coding, privacy protection, and security auditing. Theoretical analysis shows that the proposed scheme can realize the information-theoretic security in the regenerating-coding-based storage applications. Compared with the existing works, the proposed scheme is of low computational complexity and communication overhead, and better performance advantages.

Keywords: data auditing, privacy protection, regenerating code, network coding, distributed storage

---

收稿日期：2020−12−03；**修回日期**：2021−03−01

基金项目：国家自然科学基金资助项目（No.61872088, No.U1905211, No.62072109, No.U1804263）；福建省自然科学基金资助项目（No.2019J01276）；陕西省自然科学基金资助项目（No.2021JQ-196）；中国博士后基金资助项目（No.2019M663629）；西安市科技计划资助项目（No.2020KJWL02）

Foundation Items: The National Natural Science Foundation of China (No.61872088, No.U1905211, No.62072109, No.U1804263), The Natural Science Foundation of Fujian Province (No.2019J01276), The Natural Science Foundation of Shaanxi Province (No.2021JQ-196), The China Postdoctoral Science Foundation (No.2019M663629), The Xi'an Science and Technology Project (No.2020KJWL02)

1　引言

当前，云计算和大数据已成为推动信息技术发展和促进应用创新的热点技术，具有非常广阔的应用前景[1]。同时，基于网络编码的再生编码技术，已经得到了学术界的广泛关注[2]。研究表明，将再生编码技术应用在分布式云存储系统中，可以实现节点存储量和失效修复带宽方面的最优均衡。然而，如何保证再生编码分布式云存储数据的安全可靠是该类系统应用中面临的一个重要挑战[3-5]。

在外包数据安全可靠性方面，学术界已经取得了大量研究成果。根据使用的系统场景，这些成果大致可以分为数据持有性证明（PDP, provable data posesion）和数据可恢复证明（PoR, proof of retrievability）[6-10]、容错编码方案[2,4,11-14]等。前者主要适用于单服务器或单云系统，使用聚合审计方式提高审计的效率；后者主要适用于多服务器或多云系统，大多采用多备份和纠错编码等技术实现数据的有效恢复。虽然这些方案在传统云计算应用中表现尚好，但是如果把它们直接应用到再生编码分布式存储应用中，其技术实现付出的开销可能已经抵消了再生码本应具备的性能优势。

当前，学者在再生编码云存储的外包数据审计检验领域也做出了大量的工作。根据审计密钥共享的方式，这些工作大体可分成 2 类，即公开审计策略[10,15-17]和私有审计策略[11-12,18-19]。前者只能用公钥密码技术实现，任何公开实体都可以对用户的云端存储数据进行完整性检测，计算代价较高；后者中审计者在检测时通常需要知晓用户的私钥信息，常用相对高效的对称密码技术实现，多适用于用户能力受限的应用场景。虽然公开审计策略具有更强的安全性，但运行过程中引入了高昂的计算和通信开销，已经严重影响了再生编码技术的可用性。因此，出于对存储系统综合效能的考虑，私有审计策略是当前基于再生编码大数据云存储系统较实用合理的选择。

考虑到离线用户计算资源的有限性，与传统云审计方法类似，现有再生编码云存储系统大都将审计工作委托给第三方实体来进行。相应地，用户数据审计交互过程的隐私保护也成为用户主要考量的因素。为解决这个问题，常见的做法是用户在数据外包上传之前对其进行离线静态预加密（例如使用对称加密[10]或同态加密[20-21]等手段）。但是，这种方法一旦应用于实时大数据分布式处理系统，不仅严重限制了系统为用户提供数据的在线分析和处理服务，而且大幅增加了数据应用中的计算负担和通信代价。此外，失效节点数据修复过程中需要进行频繁的加解密操作，将降低严重拖累再生码存储系统的性能。

值得注意的是，如果云处理中心或相应的数据编码机制能保证数据的隐私安全，用户只需要将数据编码信息及其认证标签外包存储到云上即可，这将为大数据的实时分析提供极大的便利，是与传统预加密存储完全不同的情形，而在此情形下实现审计的高效性、可靠性和隐私保护等问题并未得到有效解决。采用云服务器直接对数据进行简单的在线实时加密并不是一种安全可行的策略，因为用户无法对这些云端加密结果实时生成合法有效的认证标签。

迄今为止，现有研究还没有给出一种上述可在再生编码存储系统中在线实施的高效隐私保护安全审计方法。文献[17]提出了利用公钥掩码技术来解决这个问题，但在大数据外包处理的应用场景将会产生很高的计算开销，而且已被证明存在一些审计安全问题[22]。文献[11-12]分别提出了适用于再生码分布式存储系统的远程认证机制，但前者涉及大量对编码向量的对称加解密操作，后者完全不支持隐私保护功能。文献[23]利用拟态技术解决了存储数据的可靠性和安全性问题，但没有考虑服务器不可信场景下的审计认证问题。Le 等[18]利用网络编码认证的思想提出了一种性能高效的分布式审计方案 NC-Audit，可以高效地实现节点修复功能。但是，该方案要求云存储中心必须知道用户的主密钥信息，并不满足隐私保护实际应用需求。文献[19]利用数据预加密、基于纠错码的同态加密和 Toeplitz 哈希函数[24]构造了存储数据的完整性检测方案，但该方案的安全性还有待论证。同时，该方案涉及复杂的纠错译码操作，故计算开销较大。此外，文献[25]将再生码存储技术引入了区块链应用场景，构造了区块链网络中的数据安全存储和恢复方案，但没有涉及数据审计问题。总体来看，现有适用于再生码系统的数据审计研究虽取得了一定的研究成果，但在执行效率或安全性能上还很不理想，仍然没有打破分布式存储系统整体效能提升的性能瓶颈。

与现有离线预加密实现隐私保护的云存储审

计方式不同，本文主要针对基于再生码的大数据存储系统，提出了一种高效的具备隐私保护功能的外包数据审计方案。该方案具有以下特点。

1) 动态加密和在线审计的融合。传统离线预加密策略不能有效适用于再生编码系统的在线审计场景，简单在线加密方法无法在外包审计场景实现阻止审计端获取用户隐私信息的目的。为解决此问题，本文在存储服务器端设计了一种可以快速实施的动态随机掩码机制，不仅保证了与审计服务的有效兼容，而且可以阻止审计端对审计响应消息的隐私提取。

2) 轻量型代数审计。利用网络编码代数子空间认证思想，所提方案对数据向量和线性掩码向量进行了一些特殊的正交化同态认证编码，构造了一种基于代数正交判定的审计验证技术。本文采用这种正交代数编码手段完成了在线隐私加密与数据审计功能的无缝融合，实现了传统预加密存储机制无法达成的隐私认证功能。

3) 低开销的实现效能。所提方案有效融合了密码技术和信息理论安全技术，尽可能避免了使用耗时的公钥密码操作（例如，大整数模指数或双线性对运算）。方案的核心操作主要涉及有限域上的乘法和加法运算，具有较低的计算开销。方案采用向量聚合认证模式，审计过程通信开销也较小。

2 系统与安全模型

2.1 系统和功能模型

本文考虑基于云计算的分布式聚集存储和处理的应用场景。数据处理中心采用基于再生码的分布式云架构，系统架构和功能模型如图 1 所示。分布式云存储审计系统包含 3 类实体，即云服务提供商（CSP, cloud service provider）、用户和第三方审计者（TPA, third party auditor）。CSP 由若干分布式存储节点组成，这些节点协作运行网络编码分布式存储协议；用户即云数据服务的订阅使用者；审计者可以实时对云端外包数据进行完整性检测，一旦 TPA 检测到云平台中某个存储节点出现存储错误，系统将会启动该存储节点的数据修复过程。

为方便后文描述，表 1 列举了本文中常用的参数符号表示。

用户 U 利用网络编码技术对文件数据进行编码，然后将其分布式地存放在云存储系统各节点 $\mathcal{N}_1,\mathcal{N}_2,\cdots,\mathcal{N}_{\tilde{n}}$ 上。不失一般性，接下来描述 U 对节点 $\mathcal{N}_s$ $(\forall s\in[\tilde{n}])$ 上存储数据的编码过程。

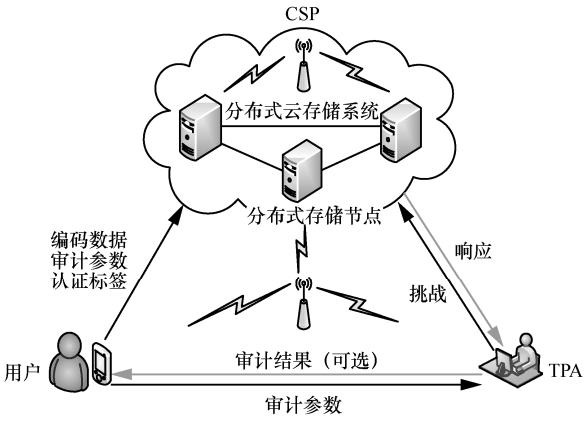

图 1 基于再生码的云存储系统架构和功能模型

表 1 参数符号表示

参数	含义
m	用户文件数据向量的个数
n	用户文件数据向量的长度
$\tilde{n}$	系统中存储节点的个数
σ	用户文件在每个存储节点存储的消息向量个数
k	成功恢复文件时需要协作的存储节点最少个数
$\boldsymbol{y}_{sj}$	$\mathcal{N}_s$ 中存储的第 j 个消息向量
$\boldsymbol{g}_{\boldsymbol{y}_{sj}}$	$\boldsymbol{y}_{sj}$ 的随机编码向量
$\boldsymbol{\kappa}_e$	用户与 CSP 的共享临时密钥
id, wid	外包存储文件和当前审计任务的唯一标识符
$\tilde{\boldsymbol{r}}$	存储节点持有的密钥向量（长度为 $m+n+1$）
$\bar{\boldsymbol{r}}$	$\tilde{\boldsymbol{r}}$ 的前 n 个元素组成的向量
$\boldsymbol{r}$	$\tilde{\boldsymbol{r}}$ 的前 $n+m$ 个元素组成的向量
$\bar{\boldsymbol{p}}_i$	存储节点持有的第 i 个掩码向量（长度为 n）

1) 将待存储文件的数据解析为有限域 $\mathbb{F}_q$ 上的 n 维向量组成的序列 $\{\bar{\boldsymbol{v}}_i\}_{i=1}^m$，其中，$\bar{\boldsymbol{v}}_i\in\mathbb{F}_q^n$ 且 $n\gg m$。

2) 对 $\{\bar{\boldsymbol{v}}_i\}_{i=1}^m$ 进行扩展，生成文件的扩展编码向量序列 $\{\boldsymbol{v}_i\}_{i=1}^m$，其中，$i\in[m]$，$\boldsymbol{v}_i=(\bar{\boldsymbol{v}}_i,\boldsymbol{e}_i)\in\mathbb{F}_q^{n+m}$，$\boldsymbol{e}_i$ 为第 i 个元素为 1 的 m 维单位向量。

3) 对序列 $\{\boldsymbol{v}_i\}_{i=1}^m$ 进行参数为 $(\tilde{n},k)$ 的最大距离可分（MDS, maximum distance separable）编码，生成 σ 个外包存储向量序列 $\{\boldsymbol{y}_{sj}\}_{j=1}^{\sigma}$ 并将其存储至节点 $\mathcal{N}_s$ 上。$\boldsymbol{y}_{sj}$ 计算式为

$$\boldsymbol{y}_{sj}=\sum_{i=1}^{m}\alpha_{sji}\boldsymbol{v}_i=\left(\bar{\boldsymbol{y}}_{sj},\boldsymbol{g}_{\boldsymbol{y}_{sj}}\right) \tag{1}$$

其中，$\alpha_{sji}\in\mathbb{F}_q$ 为编码系数，$\boldsymbol{g}_{\boldsymbol{y}_{sj}}=(\alpha_{sj1},\alpha_{sj2},\cdots,\alpha_{sjm})$ 为向量 $\boldsymbol{y}_{sj}$ 的最后 m 个元素组成的向量。

图 2 给出了用户将一个由 m 个数据块组成的文件经过再生编码后外包存储到包含 $\tilde{n}$ 个存储节点的分布式系统的过程。用户通过任意 k 个存储节点就可以完全恢复文件数据。当某个存储节点发生故障时，系统可以借助任意 $d(k \leqslant d \leqslant \tilde{n}-1)$ 个正常节点获取 $\gamma = d\beta(\beta \leqslant \sigma)$ 个数据块来恢复故障节点中存储的数据。根据节点恢复后的数据与之前存储的数据的异同，再生码修复机制包括功能性修复和精确性修复。根据 σ 和 γ 的最优折中曲线，常用的 2 种再生编码是最小存储再生（MSR, minimum-storage regenerating）码和最小带宽再生（MBR, minimum-bandwidth regenerating）码[2,4]，其对应的 (σ,γ) 值分别为

$$(\sigma_{\mathrm{MSR}},\ \gamma_{\mathrm{MSR}}) = \left(\frac{m}{k},\ \frac{md}{k(d-k+1)}\right)$$

$$(\sigma_{\mathrm{MBR}},\ \gamma_{\mathrm{MBR}}) = \left(\frac{2md}{2kd-k^2+k},\ \frac{2md}{2kd-k^2+k}\right)$$

实际上，图 2 即为 MSR 码的典型构造过程。

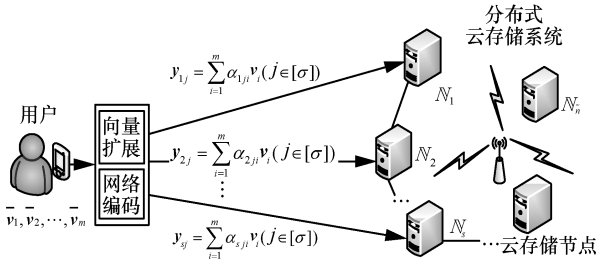

图 2　再生码存储系统中用户文件的编码和外包存储过程

由于节点数据修复问题并不会对本文审计机制造成实质性的影响，因此后文中将不再考虑此问题，仅专注讨论系统审计过程的安全可靠性。

2.2　安全模型

本文假设 CSP 中各存储节点之间可以安全通信，始终忠实地执行云存储审计相关协议。各节点之间不存在共谋，每个存储节点都有充分可靠的安全机制来保护用户的外包信息数据，但各个节点总是试图获取用户私有密钥的信息。系统中各节点可能为了节省存储开销而删除用户极少访问的部分数据，也可能为了自身商业信誉或利益而向用户隐瞒因各种原因造成的存储数据损坏。系统中各存储节点与 TPA 之间的信道是不安全的，因此需要对传输中的消息进行保护处理。为了节省计算资源和降低审计开销，用户将云存储数据的审计任务授权委托给 TPA。系统中 TPA 是一个独立可靠的实体，可以保证密钥的安全性和知晓必要的审计元数据，虽无法与 CSP 中各存储节点进行共谋，但有窥探和泄露用户文件隐私数据的强烈动机和可能性。这些是远程数据审计中依赖 TPA 进行完整性检查时的合理假设[10]。在系统协议执行期间，CSP、TP 和 U 三者之间的交互或响应都能正确合法地执行。

用户除了对数据进行必要的预编码操作之外，不对存储在云端的数据进行静态预加密操作，其目的是确保 CSP 能有效实施大规模数据的实时分析和处理。这对再生编码存储数据的外包应用具有非常重要的实用价值。

3　面向再生编码云存储的隐私保护审计协议

3.1　隐私保护审计模型

根据系统和安全模型，本文设计的隐私保护审计协议包括 Setup 阶段和 Audit 阶段。其中，Setup 阶段包含 2 个子阶段 KeyGen 和 SigUpload，Audit 阶段包含 2 个子阶段 ChalResp 和 VeriProof。

在 Setup 运行阶段，系统首先利用 KeyGen 设置安全密钥和协议执行参数，然后在 SigUpload 运行阶段生成用户编码数据的审计认证标签，并将用户编码数据及其认证标签信息统一上传到分布式云存储系统中，同时生成（针对敌手 TPA 的）隐私加密辅助信息。

TPA 和 CSP 在 Audit 阶段执行“挑战–应答”模式的审计交互操作。首先，TPA 向 CSP 发起审计挑战（Challenge）；其次，CSP 执行响应操作（Response），实时构造挑战数据向量（和认证标签）的聚合消息，并将该消息的即时加密结果（挑战应答）发送给 TPA；最后 TPA 可以根据 CSP 的应答，执行 VeriProof 算法完成审计验证。

与现有同类方案（例如文献[10]）相比，本文方案的不同之处在于 ChalResp 操作。该操作不仅完成了数据审计操作，而且实现了挑战信息（针对敌手 TPA）的实时隐私保护功能。

3.2　隐私保护审计协议描述

为区分各次审计检测任务，协议为每一次审计检测过程设定了唯一的任务标签 wid，该标签与存储服务器索引和文件标签关联。同时，本文引入 2 个伪随机函数（PRF, pseudo-random function），即 $F_1 : K_{\mathrm{ID}} \times \mathbb{Z}^+ \to \mathbb{F}_q$，$F_2 : K_{\mathrm{WID}} \times \mathbb{Z}^+ \to \mathbb{F}_q$。其中，$\mathbb{Z}^+$ 为正整数集合，K 为伪随机函数的密钥集，

ID 为文件标识符集合， WID 为审计任务的标识符集合。

由于 U 对各存储节点的审计方式相同，因此，本文仅描述 U 与 CSP 中单个存储节点 $\mathcal{N}_s(\forall s\in[\tilde{n}])$ 的协议交互执行过程。再生编码云存储隐私保护审计协议交互过程如图 3 所示，该过程包括 Setup 阶段和 Audit 阶段。

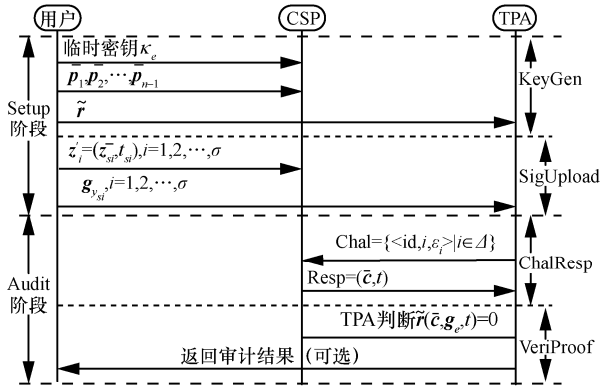

图 3 再生编码云存储隐私保护审计协议交互过程

3.2.1 Setup 阶段

1) KeyGen 子阶段

此阶段主要完成审计系统相关密钥生成和服务器端掩码向量空间的配置工作。

① 系统选定安全参数 λ，确定 2 个 PRF，即 F_1 和 F_2，生成用户的主密钥 κ_0 以及用户和 $\mathcal{N}_s$ 的共享临时密钥 κ_e。

② U 利用 id 和 F_1 计算私有密钥向量 $\tilde{\boldsymbol{r}}\in\mathbb{F}_q^{n+m+1}$、$\bar{\boldsymbol{r}}\in\mathbb{F}_q^{n}$ 和 $\boldsymbol{r}\in\mathbb{F}_q^{n+m}$。

$\tilde{\boldsymbol{r}}=(r_1,r_2,\cdots,r_{n+m+1})\triangleq(\bar{\boldsymbol{r}}\,\|\,r_{n+1},\cdots,r_{n+m+1})\triangleq(\boldsymbol{r}\,\|\,r_{n+m+1})$

其中，$r_i=F_1(\kappa_0,\mathrm{id},i),i\in[n+m+1]$，$\|$表示向量连接，$\bar{\boldsymbol{r}}$ 和 $\boldsymbol{r}$ 分别是由 $\tilde{\boldsymbol{r}}$ 的前 n 个和前 $n+m$ 个元素构成的向量。

③ U 构造线性方程组

$$\bar{\boldsymbol{r}}\boldsymbol{x}^{\mathrm{T}}=0 \tag{2}$$

并随机选择 $n-1$ 个线性无关且不包含 0 元素的解向量 $\bar{\boldsymbol{p}}_1,\bar{\boldsymbol{p}}_2,\cdots,\bar{\boldsymbol{p}}_{n-1}\in\boldsymbol{W}$。这里，$\boldsymbol{W}$ 为式(2)的解空间，也是私有密钥向量 $\bar{\boldsymbol{r}}$ 的正交补空间。显然，$\boldsymbol{W}=\mathrm{span}\{\bar{\boldsymbol{p}}_1,\bar{\boldsymbol{p}}_2,\cdots,\bar{\boldsymbol{p}}_{n-1}\}$。

④ U 将上述基向量 $\bar{\boldsymbol{p}}_1,\bar{\boldsymbol{p}}_2,\cdots,\bar{\boldsymbol{p}}_{n-1}$ 发送给 $\mathcal{N}_s$，同时将 $\tilde{\boldsymbol{r}}$ 发送给 TPA。

⑤ $\mathcal{N}_s$ 对 $\bar{\boldsymbol{p}}_1,\bar{\boldsymbol{p}}_2,\cdots,\bar{\boldsymbol{p}}_{n-1}$ 进行随机线性组合，生成 2 个不含 0 元素的向量 $\bar{\boldsymbol{p}}_n$ 和 $\bar{\boldsymbol{p}}_{n+1}$。

2) SigUpload 子阶段

本节描述采用前文系统模型中的符号表示。

用户 U 首先需要生成文件编码向量 $\{\boldsymbol{v}_i\}_{i=1}^m$ 的认证标签，然后生成外包存储向量及其认证标签，并将外包数据上传至云存储中心。

① U 计算数据文件的扩展编码序列向量 $\boldsymbol{v}_i=(\bar{\boldsymbol{v}}_i,\boldsymbol{e}_i)(i\in[m])$ 的认证标签 t_i，即

$$t_i=-\left(r_{n+m+1}\right)^{-1}(\boldsymbol{r}\boldsymbol{v}_i) \tag{3}$$

② 针对存储节点 $\mathcal{N}_s$，用户对向量序列 $\{\tilde{\boldsymbol{v}}_i\}_{i=1}^m=(\boldsymbol{v}_i,t_i)_{i=1}^m$ 进行 MDS 编码，生成向量序列 $\{\tilde{\boldsymbol{z}}_{sj}\}_{j=1}^{\sigma}$，编码计算过程为

$$\begin{aligned}&\tilde{\boldsymbol{z}}_{sj}=\left(\boldsymbol{z}_{sj},t_{sj}\right)=\left(\bar{\boldsymbol{z}}_{sj},\boldsymbol{g}_{z_{sj}},t_{sj}\right)=\\&\left(\sum_{i=1}^m\alpha_{sji}\bar{\boldsymbol{v}}_i,\sum_{i=1}^m\alpha_{sji}\boldsymbol{e}_i,\sum_{i=1}^m\alpha_{sji}t_i\right)=\\&\left(\sum_{i=1}^m\alpha_{sji}\left(\bar{\boldsymbol{v}}_i,\boldsymbol{e}_i\right),\sum_{i=1}^m\alpha_{sji}t_i\right)=\left(\sum_{i=1}^m\alpha_{sji}\boldsymbol{v}_i,\sum_{i=1}^m\alpha_{sji}t_i\right)=\\&\sum_{i=1}^m\alpha_{sji}\left(\boldsymbol{v}_i,t_i\right)=\sum_{i=1}^m\alpha_{sji}\hat{\boldsymbol{v}}_i\end{aligned} \tag{4}$$

其中，式(4)中第二个等号取自式(1)。

③ 用户将向量组 $\left\{\boldsymbol{z}'_{sj}\right\}_{j=1}^{\sigma}=\left\{\left(\bar{\boldsymbol{z}}_{sj},t_{sj}\right)\right\}_{j=1}^{\sigma}$ 上传到 $\mathcal{N}_s$，同时将 $\boldsymbol{g}_{z_{sj}}$ 发送给 TPA。

④ 用户可以删除本地文件数据，仅保留 $\boldsymbol{g}_{z_{sj}}$ 和相关密钥即可。

在上述过程中，用户首先利用式(3)，即通过文件编码向量 $\boldsymbol{v}_i$ 与私有密钥向量 $\tilde{\boldsymbol{r}}$ 的正交化来计算生成 $\boldsymbol{v}_i$ 的同态认证标签 t_i，然后利用式(4)的同态计算生成外包存储向量及其认证标签，使其满足 $\tilde{\boldsymbol{z}}_{sj}\tilde{\boldsymbol{r}}=0$。

3.2.2 Audit 阶段

此阶段是 TPA 和存储节点 $\mathcal{N}_s$ 之间的数据审计检测的交互过程。

1) ChalResp 子阶段

① TPA 任选集合 $\Delta\subseteq\{1,2,\cdots,M\}$，随机选取 $\varepsilon_i\in\mathbb{F}_q(i\in\Delta)$，生成针对节点 $\mathcal{N}_s$ 的审计挑战消息 $\mathrm{Chal}=\{<\mathrm{id},i,\varepsilon_i>|\,i\in\Delta\}$，并将其发送至 $\mathcal{N}_s$。

② $\mathcal{N}_s$ 利用 Chal 计算聚合向量

$$\boldsymbol{e}=\sum_{i\in\Delta}\varepsilon_i\boldsymbol{z}'_{si}=\left(\bar{\boldsymbol{e}},t\right)\in\mathbb{F}_q^{n+m+1} \tag{5}$$

③ $\mathcal{N}_s$ 对 $\bar{\boldsymbol{e}}$ 进行随机加密，生成密文 $\bar{\boldsymbol{c}}$，生成

响应消息 Resp，具体如下。

计算随机掩码系数 $\beta_z = F_2(k_e, \mathrm{wid}, z)$, $z = 1,2,\cdots,n+1$ 并生成掩码向量

$$\bar{\boldsymbol{m}} = \sum_{i=1}^{n+1} \beta_i \bar{\boldsymbol{p}}_i \in \mathbb{F}_q^n \tag{6}$$

生成密文向量

$$\bar{\boldsymbol{c}} = \bar{\boldsymbol{e}} + \bar{\boldsymbol{m}} \tag{7}$$

发送 $\mathrm{Resp} = (\bar{\boldsymbol{c}}, t)$ 至 TPA。

2) VeriProof 子阶段

TPA 收到 Resp 后，利用 Chal 计算编码系数向量 $\boldsymbol{g}_e$，解析出待检测消息 $\boldsymbol{c} = (\bar{\boldsymbol{c}}, \boldsymbol{g}_e, t)$，判断式(8)是否成立。如果成立，TPA 返回 1；否则，判断 $\mathbb{N}_s$ 的数据审计响应错误。

$$\tilde{\boldsymbol{r}}\boldsymbol{c} = 0 \tag{8}$$

在 Audit 阶段中，式(6)和式(7)的主要作用是实现对响应消息 $\bar{\boldsymbol{e}}$ 的隐私加密基本功能，阻止了 TPA 通过收集响应消息解出存储在 $\mathbb{N}_s$ 上的外包编码数据，同时确保了 TPA 对响应消息的可验证性。这种方法特有的高效性和有效性是常规的预加密外包存储技术难以达到的。

4　方案理论分析

4.1　正确性验证

如果存储节点正确存储了待审计文件，它返回的响应消息必然可以通过式(8)的验证。

定理 1　如果上述方案能得到正确执行，那么对于用户存储在分布式云处理系统中的任意一个文件，TPA 都可以正确地验证存储在系统各存储节点上属于该文件的编码数据。

证明　根据方案描述，只需要说明 TPA 能正确验证存储节点 $\mathbb{N}_s$ 上的文件编码数据即可。根据方案中 $\tilde{\boldsymbol{r}}$ 和认证标签的生成方法，可得

$$\tilde{\boldsymbol{r}}\boldsymbol{c} = \tilde{\boldsymbol{r}}(\bar{\boldsymbol{c}}, \boldsymbol{g}_e, t) = \tilde{\boldsymbol{r}}(\bar{\boldsymbol{e}} + \bar{\boldsymbol{m}}, \boldsymbol{g}_e, t) =$$
$$\tilde{\boldsymbol{r}}(\bar{\boldsymbol{e}}, \boldsymbol{g}_e, t) + \tilde{\boldsymbol{r}}(\bar{\boldsymbol{m}}, \boldsymbol{0}_{m+1}) = \tilde{\boldsymbol{r}}(\bar{\boldsymbol{e}}, \boldsymbol{g}_e, t) + \bar{\boldsymbol{r}}\bar{\boldsymbol{m}} = 0 \tag{9}$$

其中，$\boldsymbol{0}_{m+1}$ 表示长为 $m+1$ 的零向量。

证毕。

实际上，式(9)说明任意一个合法的响应消息向量 $\boldsymbol{c}$ 总是在向量 $\tilde{\boldsymbol{r}}$ 的正交子空间中，而合法的掩码向量 $\bar{\boldsymbol{m}}$ 总是在向量 $\bar{\boldsymbol{r}}$ 的正交子空间里。这意味着式(8)同时满足 2 种正交关系。这种双重正交的特性保证了式(9)和数据审计验证的正确性和合理性。

4.2　安全性证明

所提方案使用随机掩码和正交编码结合的技术实现了适用于再生编码分布式存储系统的隐私保护审计方法。根据系统安全模型和方案关键技术特征，该方法的安全性取决于用户私有向量 $\bar{\boldsymbol{r}}$（或 $\tilde{\boldsymbol{r}}$）、针对响应消息 $\bar{\boldsymbol{e}}$ 的随机加密和审计认证机制三方面的安全性。相应地，这三方面的安全性分别由定理 2~定理 4 来保证。

定理 2　对于使用本文审计方案的再生编码分布式云存储系统，如果 F_1 是理想安全的 PRF，则对于任意一个具有无限计算能力的敌手（CSP），解出用户私有向量 $\bar{\boldsymbol{r}}$（或 $\tilde{\boldsymbol{r}}$）的最大概率仅为 q^{-1}。

证明　根据式(2)和 $\bar{\boldsymbol{p}}_1, \bar{\boldsymbol{p}}_2, \cdots, \bar{\boldsymbol{p}}_{n+1}$ 的线性相关性可知，除了猜测之外，CSP 得到向量 $\bar{\boldsymbol{r}}$ 的最有利的方法是利用已知信息构造关于向量 $\bar{\boldsymbol{r}}$（含 n 个未知量）的线性方程组，即

$$\bar{\boldsymbol{r}}\left(\bar{\boldsymbol{p}}_1^{\mathrm{T}}, \bar{\boldsymbol{p}}_2^{\mathrm{T}}, \cdots, \bar{\boldsymbol{p}}_{n-1}^{\mathrm{T}}\right) = \bar{\boldsymbol{r}}\boldsymbol{P}_{n\times(n-1)} = \boldsymbol{0}_{n-1} \tag{10}$$

由于 $\bar{\boldsymbol{r}}$ 是由安全的 PRF 生成，可知 CSP 能解出 $\bar{\boldsymbol{r}}$ 的概率是 $q^{-(n-(n-1))} = q^{-1}$。

证毕。

利用定理 2 可以很容易地发现文献[18]中的 NC-Audit 方案无法保证审计参数的安全性。这个结论是显然的。事实上，CSP 完全可以构造类似式(10)的可解方程组。根据 NC-Audit 的方案设计，该方程组的未知量个数不会多于关于未知量的方程个数，致使 CSP 能以很大的概率解出用户的隐私密钥向量，使该方案的审计功能失效[26]。

定理 3　对于使用本文审计方案的再生编码分布式云存储系统，如果 F_1 和 F_2 是理想安全的 PRF，则对于任意一个具有无限计算能力的敌手（TPA），协议中的随机掩码技术实现了完善的隐私保密性。

证明　记 $\langle\beta\rangle = \{\beta_1, \beta_2, \cdots, \beta_{n+1}\}$，$\langle\boldsymbol{p}\rangle = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_{n+1}\}$。根据 KenGen 阶段的过程描述，可得

$$I(\bar{\boldsymbol{e}};\bar{\boldsymbol{c}}) \leqslant I(\bar{\boldsymbol{e}};\bar{\boldsymbol{c}},\bar{\boldsymbol{e}}) = I(\bar{\boldsymbol{e}};\bar{\boldsymbol{m}}) \leqslant I(\bar{\boldsymbol{e}};\bar{\boldsymbol{m}},\langle\beta\rangle,\langle\boldsymbol{p}\rangle) =$$
$$I(\bar{\boldsymbol{e}};\langle\boldsymbol{p}\rangle) + I(\bar{\boldsymbol{e}};\langle\beta\rangle|\langle\boldsymbol{p}\rangle) + I(\bar{\boldsymbol{e}};\bar{\boldsymbol{m}}|\langle\beta\rangle,\langle\boldsymbol{p}\rangle) =$$
$$I(\bar{\boldsymbol{e}};\langle\beta\rangle|\langle\boldsymbol{p}\rangle) + I(\bar{\boldsymbol{e}};\bar{\boldsymbol{m}}|\langle\beta\rangle,\langle\boldsymbol{p}\rangle) =$$
$$I(\bar{\boldsymbol{e}};\langle\beta\rangle|\langle\boldsymbol{p}\rangle) + H(\bar{\boldsymbol{e}}|\langle\beta\rangle,\langle\boldsymbol{p}\rangle) - H(\bar{\boldsymbol{e}}|\bar{\boldsymbol{m}},\langle\beta\rangle,\langle\boldsymbol{p}\rangle) =$$
$$I(\bar{\boldsymbol{e}};\langle\beta\rangle|\langle\boldsymbol{p}\rangle) + H(\bar{\boldsymbol{e}}|\langle\beta\rangle,\langle\boldsymbol{p}\rangle) - H(\bar{\boldsymbol{e}}|\langle\beta\rangle,\langle\boldsymbol{p}\rangle) =$$
$$I(\bar{\boldsymbol{e}};\langle\beta\rangle|\langle\boldsymbol{p}\rangle) = I(\bar{\boldsymbol{m}};\langle\beta\rangle|\langle\boldsymbol{p}\rangle) \tag{11}$$

下一步，只需证明式(11)中 $I\left(\bar{\boldsymbol{m}};\langle\beta\rangle|\langle\boldsymbol{p}\rangle\right)=0$ 即可。记式(2)中方程组选取的解向量为 $\bar{\boldsymbol{p}}_i=(p_{i1},p_{i2},\cdots,p_{i,n})$，$i\in[n+1]$，审计聚合形成的掩码向量为 $\bar{\boldsymbol{m}}=(m_1,m_2,\cdots,m_n)$。假设敌手已经获得向量组 $\langle\boldsymbol{p}\rangle$ 和掩码向量 $\bar{\boldsymbol{m}}$，则敌手可以根据 式(6)和式(7)构造关于 $\beta_i(i\in[n+1])$ 的线性方程组。

$$\begin{cases}\beta_1 p_{11}+\beta_2 p_{21}+\cdots+\beta_{n+1}p_{n+1,1}=m_1\\ \beta_1 p_{12}+\beta_2 p_{22}+\cdots+\beta_{n+1}p_{n+1,2}=m_2\\ \qquad\cdots\\ \beta_1 p_{1n}+\beta_2 p_{2n}+\cdots+\beta_{n+1}p_{n+1,n}=m_n\end{cases}\tag{12}$$

在假设条件下，该方程组中所有的 β_i 和 p_{ij} 都是 $\mathbb{F}_q$ 中的非 0 值。令该方程组的系数矩阵的秩为 ρ，显然有 $\rho<n+1$，即可能解个数始终为 $q^{n+1-\rho}$。也就是说，除了猜测，敌手得不到 $\beta_i(i\in[n+1])$ 的任何信息。

证毕。

根据定理 3，对于一次审计来说，只要能保证 $\beta_i(i=1,2,\cdots,n+1)$ 的随机性和保密性，向量集 $\langle\boldsymbol{p}\rangle$ 完全可以是公开的，不用对其进行专门的保密设计。根据式(12)，如果 CSP 能保证集合 $\langle\boldsymbol{p}\rangle$ 中若干（至少一个）向量是保密的，而其他向量是公开的，那么 CSP 在每次审计任务中只需要更新任意一个保密向量和所有公开向量的相应掩码系数即可。这将显著降低 CSP 在协议运行期间的计算开销。

值得注意的是，定理 3 也保证了 TPA 无法通过自身拥有的向量 $\boldsymbol{r}$ 反推关于明文消息向量的有用信息。因为 TPA 无法得到 $\langle\beta\rangle$ 的具体值，仅能利用式(8)来推导密文 $\boldsymbol{c}$ 的信息，显然是不可能的。

定理 4 对于使用本文审计方案的再生编码分布式云存储系统，如果 F_1 和 F_2 是理想安全的 PRF，则对于任意一个具有无限计算能力的敌手（CSP），伪造了一个非法的认证三元组 $\langle\mathrm{id}',\boldsymbol{y}',t'\rangle$ 并成功通过审计检测的概率最高为 q^{-d}，其中，$d=n+m-\rho+1$，$\rho\leqslant n+m$。

证明 对于敌手伪造的认证三元组，可能包含 2 种情况。

情况 1 $\langle\mathrm{id}',\boldsymbol{y}',t'\rangle=\langle\mathrm{id},\boldsymbol{y}',t'\rangle$，即敌手选用某个合法的 id，但 $\boldsymbol{y}'$ 不属于文件 id。

情况 2 $\langle\mathrm{id}',\boldsymbol{y}',t'\rangle=\langle\mathrm{id}',\boldsymbol{y},t\rangle$，即敌手选用某个伪造的 id′，但 $\boldsymbol{y}$ 和 t 是文件 id（$\mathrm{id}\neq\mathrm{id}'$）的合法认证数据。

根据方案构造，在外包存储不同的文件（具有不同的 id）时，由于用户选用了不同的密钥，则 TPA 对这 2 个文件的存储数据进行审计检测时也相应地需要使用完全独立的审计向量（即用户私有向量）。对于敌手来说，上述 2 种伪造形式对审计机制的破解难度是相同的。基于此，这里只需证明上述情况 1 下审计检测的安全性即可。

假设对于文件 id，正如方案 Setup 阶段所示，用户选用的私有审计向量为 $\tilde{\boldsymbol{r}}$。如果敌手可以输出一个非法的认证三元组 $\langle\mathrm{id},\boldsymbol{y}',t'\rangle$ 且能通过审计机制的检测，则 $\boldsymbol{y}'$ 和 t' 必然满足以下 2 个条件

$$t'=-\left(r_{n+m+1}\right)^{-1}\boldsymbol{y}'\boldsymbol{r}\tag{13}$$

$$\boldsymbol{y}'\notin\mathrm{span}\left(\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_m\right)\tag{14}$$

这时，敌手将有能力构造一个关于 $\tilde{\boldsymbol{r}}$ 中 $n+m+1$ 个元素 $r_i(i\in[n+m+1])$ 的线性方程组

$$\begin{cases}\bar{\boldsymbol{p}}_1\bar{\boldsymbol{r}}=0\\ \quad\vdots\\ \bar{\boldsymbol{p}}_{n-1}\bar{\boldsymbol{r}}=0\\ \boldsymbol{y}_1\boldsymbol{r}=t_1\\ \quad\vdots\\ \boldsymbol{y}_m\boldsymbol{r}=t_m\\ \boldsymbol{y}'\boldsymbol{r}=t'r_{n+m+1}\end{cases}\tag{15}$$

令式(15)中方程组系数矩阵的秩为 ρ，则其解空间的维数为 $d=n+m-\rho+1\geqslant 1$。此时无论 t' 取何值，该方程组可能解的个数都为 q^d。从敌手的视角来看，t' 是 $\mathbb{F}_q$ 中按均匀分布选取的随机值，因而敌手成功猜测到 t' 的真实值的最大概率为 q^{-d}。

证毕。

4.3 扩展功能分析

4.3.1 外包存储隐私保护

为保证用户外包数据的隐私性，常见的做法是在数据外包存储之前用户对数据实施预加密处理操作，使云服务器无法获取用户数据的内容。但在再生编码存储环境中，这种做法是影响再生编码存储系统性能的关键瓶颈。所以，本文没有讨论这种做法。实际上，再生编码云存储系统中对外包服务器的隐私保护是很容易实现的。一种最简单的方法是规定每个文件在每个存储节点存储的外包编码向量的个数少于 m。通过合理设置参数和预编码操作，常用的再生编码基本上都能满足这种需求。根

据系统和安全模型描述，用户文件数据向量组成了一个 m 维的向量空间，并且各个存储节点之间不能共谋，那么各存储节点就无法利用自身存储的编码向量恢复出用户的数据向量。已有研究表明，通过灵活运用代数编码方法，再生编码存储系统可以很容易地部署信息理论意义下更实用的弱安全隐私保护机制[27-28]。需要强调的是，针对云服务器的隐私保护机制设计超出了本文的研究范围，但可以作为提升再生编码存储系统数据审计安全性的支撑技术。

4.3.2 存储更新动态审计

基于再生码的分布式存储机制在数据很少被读取或修改的应用场景（例如长期归档、数据托管和监管存储等）具有最佳的性能优势[14]。但在实际应用中，用户可能会主动更新存储的文件数据（包括插入、修改或删除等操作）。从原理上说，本文的审计策略也可应用在此种情形。根据前文的系统描述，数据的动态更新（尤其是修改和删除操作）可能会涉及目标文件和认证元数据的重构，从而会产生文件级规模的传输量，这对系统性能造成了一定的负面影响。但是，这个问题是由再生码本身的编码特性所决定的，除非引入公钥密码技术，单凭私有审计机制本身是无法解决的。因此，本文暂不考虑用户主动更新外包数据时的动态审计过程。

5 方案性能分析

针对单文件的存储过程，本节对审计协议中 TPA 与 CSP（仅考虑单个存储节点）间的交互过程的计算量、通信量和存储开销进行具体分析。在计算开销方面，本节主要考虑认证元数据（认证标签和全局编码向量）生成、在线隐私保护机制和完整性检测中的乘法计算量；在通信开销方面，本节主要考虑用户数据上传量、审计交互消息传输及安全存储开销；在存储开销方面，本节主要考虑协议运行过程中的实际存储需求量。

5.1 计算开销

在系统初始化阶段，各种认证参数可以通过离线预计算完成，所以本节忽略该部分的计算开销。

1) 认证标签计算

由于每个文件包含 m 个数据向量，为了计算这 m 个数据向量的认证标签，用户为此需要计算 $\sigma(m+n)$ 次乘法运算。

2) 在线隐私保护机制

假设 $|\Delta|=\xi$，被审计的存储节点需要生成挑战响应消息。首先，它要对 ξ 个消息向量进行向量聚合，需要进行 $\xi(n+1)$ 次乘法；然后，存储节点为了生成 $\bar{\boldsymbol{m}}$，可以最少只需要进行一次向量的标量乘运算。因此，在线加密的乘法计算量是 $n+\xi(m+n+1)$。

3) 完整性检测

TPA 首先需要恢复 $\boldsymbol{g}_e$ 和计算 $\tilde{\boldsymbol{r}}\boldsymbol{c}$，乘法计算量分别为 $m\xi$ 和 $n+m+1$。因此，该部分总的计算量是 $n+m(\xi+1)+1$。

5.2 通信开销

同计算开销分析，这里只考虑 TPA 和 CSP 间审计交互过程中的在线通信开销。

1) 用户数据上传

为了存储单个文件，用户不仅需要向每个存储节点上传存储数据，同时也要向 TPA 上传全局编码向量及认证标签，通信开销分别为 $\sigma(n+m+1)$ 和 $m(\sigma+1)+n$。

2) 审计挑战及响应

此过程中，TPA 先向存储节点发送挑战消息 Chal，然后存储节点向 TPA 发送响应消息 Resp。此阶段总通信开销为 $n+3\xi+1$。

5.3 存储开销

与同类方案 NC-Audit 类似，每个用户文件在各存储节点上需要的存储空间为 $O(\sigma n)$，与其相对应的认证标签产生的存储开销仅为 $O(\sigma)$。TPA 付出的存储开销为 $O(m\sigma\tilde{n})$，因为它需要存储所有存储节点的编码系数。由于 $n \gg m$，可知编码系数的存储空间量级远小于外包数据的存储量级，完全可以忽略。实际上，在系统实现时，编码系数向量的存储开销完全可以保持在 160 B 以内[14]。

5.4 性能比较与实验分析

表 2 将本文方案与文献[10,11,17-19]方案在功能特征、计算复杂度和通信开销等方面进行了综合比较。结合各方案的技术实现特征，本文得出以下的性能比较结论。

1) 文献[10,17]方案使用公钥签名和双线性对技术，计算和通信开销都明显高于其他方案。但在基于公钥的审计机制中，用户和 TPA 仅需要存储必要的私钥和少量的认证元数据，不需要保留过多的预编码信息，与私有审计机制相比，用户和 TPA 的存储开销相对较少。

2) 在文献[11]方案中，CSP 需要根据 TPA 的挑

表 2 本文方案与代表性方案性能比较

方案名称	实现机制	审计安全	数据修复	隐私保护	审计通信量	TPA 存储量	TPA 计算量	CSP 计算量
文献[10]方案	Public Key	Yes	No	Yes	$m+n+2\xi+1$	$O(1)$	Heavy	Heavy
文献[11]方案	Private Key	Yes	Yes	No	$m+3\xi+m\sigma$	$O(1)$	$O(\xi\sigma)$	$O(\sigma(\xi+m))$
文献[17]方案	Public Key	Yes	Yes	No	$m+3\xi$	$O(1)$	Heavy	Heavy
文献[18]方案	Private Key	No	Yes	Yes	$m+n+3\xi+3$	$O(m\sigma)$	$O(\xi\sigma)$	$O(\xi n+n^2)$
文献[19]方案	Private Key	Yes	Yes	Yes	$m+n+3\xi+1$	$O(m\sigma)$	$O(n\sigma)$	$O(\xi n)$
本文方案	Private Key	Yes	Yes	Yes	$m+n+3\xi+1$	$O(m\sigma)$	$O(\xi\sigma)$	$O(\xi n)$

战指令对存储的数据向量进行对应的采样计算，计算复杂度为 $O(\xi\sigma)$；CSP 把数据文件的所有编码向量的密文发送给 TPA，产生额外通信开销 $O(m\sigma)$；TPA 还需要对这些密文进行解密，解密复杂度也为 $O(m\sigma)$。总体来看，文献[11]方案在 TPA 端的处理复杂度比本文方案和 NC-Audit 都要大，但 CSP 端的计算复杂度要低很多。

3) 虽然文献[18]方案与本文方案具有相同的计算和通信开销量级，但文献[18]方案在认证签名时额外需要在每个外包向量中填充一定量的随机字符，这使其在存储和通信带宽资源利用上略逊于本文方案。

4) 在文献[19]方案中，CSP 端的开销较小，但由于在数据认证签名阶段需要进行大量的向量−矩阵乘法计算，且在用户审计检测时涉及 LDPC 译码纠错操作，因而总体复杂度要比本文方案高。

5) 在外包数据认证标签生成阶段，现有方案普遍采用“先编码后签名”的模式。与此不同，本文方案采用了“先签名后编码”的方式，从而大幅降低了用户端的乘法计算开销，代价仅为现有同类方案的 m/M。

综合而言，本文方案的计算复杂度和通信开销具有较好的比较优势，在理论分析上具有更好的系统实现效率，因而对再生码技术性能优势的负面影响也最小。为了比较方案执行性能，本文选取文献[17]中的公开审计方案，以及文献[11,18-19]中的私有审计方案，在相同安全强度（80 bit）下，与本文方案进行计算性能的测试比较。实验使用了 4 台配置为 Intel(R) Core(TM) i5-2520M CPU@ 2.50 GHz（6 GB RAM）的 64 bit 同型主机，其中一台为控制节点，用以测试 TPA（或用户端）的计算性能，另外 3 台组成并行处理集群实现 CSP 端的计算功能。实验系统环境为 Debian 10，利用 Mpich2、Torque+Maui 实现集群并行计算和任务调度功能。实验过程仅统计数据审计过程中的在线计算量。各方案中使用的伪随机函数 PRF 均采用 AES 的 CTR 模式来实现。同时，设定 $q=2^{160}$，$n=2^{12}$（即消息向量大小为 4 KB），$m=400$，$\sigma=500$。实验中将文献[19]方案中编码向量的分块个数设为 200。

根据方案描述可知，本文方案与文献[18]方案的计算开销主要来自向量乘法，文献[11]方案的主要操作包括对称加解密操作和向量乘法计算，而文献[19]方案的运行耗时来自矩阵乘法和纠错译码操作。在不考虑通信时延的情形下，表 3 给出了 ξ 取 250 和 350 时 4 种方案经过 1 000 次审计运行时间开销的均值（包括 CSP 平均响应时间和 TPA 审计验证时间）。由于需要进行双线性对运算和处理大量的乘幂运算，文献[17]方案付出的时间开销较大，明显高于其他方案。在 TPA 端，文献[11]方案的运行效率较低，文献[19]方案运行稍快，但性能不及文献[18]方案和本文方案。在 CSP 端，本文方案比文献[11]方案稍慢，但具有比文献[18-19]方案更好的性能优势，并没有因安全审计功能的完善而影响审计运算效率。

表 3 TPA 和 CSP 审计平均运行时间开销比较

方案	ξ=250		ξ=350	
	TPA/ms	CSP/ms	TPA/ms	CSP/ms
文献[11]方案	23.35	2.029	23.557	2.243
文献[17]方案	125.78	46.5	187.23	69.01
文献[18]方案	9.634	3.325	9.74	4.011
文献[19]方案	13.32	3.045	16.512	3.595
本文方案	9.631	2.99	9.735	3.107

为了验证用户端安全编码处理效率，表 4 给出本文方案与文献[17-19]方案中数据外包单轮安

全编码操作的时间开销比较（忽略了对数据进行再生编码的时间开销）。由于文献[19]方案需要代数预加密操作，文献[17]方案需要大量的公钥签名计算，因此这些方案需要更多的计算处理开销。从表 4 可以看出，本文方案可以大幅降低用户端的计算开销，在实现性能方面具有轻量型的特征。

表 4　　用户端计算时间开销比较

方案	实现机制	安全编码时间开销/ms
文献[17]方案	Public Key	302.116
文献[18]方案	Private Key	19.209
文献[19]方案	Private Key	26.157
本文方案	Private Key	13.828

6　结束语

外包数据的隐私保护安全审计是基于再生码的分布式云存储系统应用中的关键问题之一。虽然使用公钥密码技术能很容易地实现明文的实时隐私保护，但需要付出较大的在线计算开销。本文通过深入挖掘再生码存储系统线性编码存储特性，将线性随机掩码技术和代数正交验证技术进行有效融合，提出了一种适用于再生编码分布式存储系统的隐私保护审计方案。该方案可以实现轻量型的安全性认证，在计算资源受限的环境下具有一定的实用价值。从系统和安全模型来看，该方案不仅适用于基于编码的云存储系统，而且可以推广到一般性的云存储场景。

虽然再生编码分布式存储技术在大数据静态存储修复方面具有优异的性能，但在某些大数据应用场景下，用户可能仍会对系统存储的文件数据执行一些动态更新操作。由于用户本身不可能去备份所有数据，更新数据可能会涉及相关数据的解码和析取操作，随之产生较大的通信开销。因此，如何在审计过程中最大限度地降低数据更新处理开销仍是该领域中尚未解决的挑战性问题。为此，未来的工作是将代数正交编码审计思想与公钥密码技术进行深度结合，设计能高效地适应动态存储场景的轻量型审计机制。

参考文献：

[1] NACHIAPPAN R, JAVADI B, CALHEIROS R N, et al. Cloud storage reliability for big data applications: a state of the art survey[J]. Journal of Network and Computer Applications, 2017, 97: 35-47.

[2] BALAJI S B, KRISHNAN M N, VAJHA M, et al. Erasure coding for distributed storage: an overview[J]. Science China Information Sciences, 2018, 61(10): 1-45.

[3] THAKUR N, SINGH A, SANGAL A L. Data integrity authentication techniques in cloud computing: a survey[C]//Soft Computing: Theories and Applications. Berlin: Springer, 2020: 1255-1267.

[4] 王意洁, 许方亮, 裴晓强. 分布式存储中的纠删码容错技术研究[J]. 计算机学报, 2017(1): 236-255.
WANG Y J, XU F L, PEI X Q. Research on error code-based fault-tolerant technology for distributed storage[J]. Chinese Journal of Computers, 2017(1): 236-255.

[5] ZHOU L, FU A M, YU S, et al. Data integrity verification of the outsourced big data in the cloud environment: a survey[J]. Journal of Network and Computer Applications, 2018, 122: 1-15.

[6] ARMKNECHT F, BOHLI J M, KARAME G, et al. Outsourcing proofs of retrievability[J]. IEEE Transactions on Cloud Computing, 2021, 9(1): 286-301.

[7] WANG H Q, HE D B, FU A M, et al. Provable data possession with outsourced data transfer[J]. IEEE Transactions on Services Computing, 2019, PP(99): 1.

[8] TAN C B, HIJAZI M H A, LIM Y, et al. A survey on proof of retrievability for cloud data integrity and availability: cloud storage state-of-the-art, issues, solutions and future trends[J]. Journal of Network and Computer Applications, 2018, 110: 75-86.

[9] HAHN C, KWON H, KIM D, et al. Enabling fast public auditing and data dynamics in cloud services[J]. IEEE Transactions on Services Computing, 2020, PP(99): 1.

[10] WANG C, CHOW S S M, WANG Q, et al. Privacy-preserving public auditing for secure cloud storage[J]. IEEE Transactions on Computers, 2013, 62(2): 362-375.

[11] CHEN B, CURTMOLA R, ATENIESE G, et al. Remote data checking for network coding-based distributed storage systems[C]//Proceedings of the 2010 ACM Workshop on Cloud computing Security Workshop. New York: ACM Press, 2010: 31-42.

[12] CHEN H C H, LEE P P C. Enabling data integrity protection in regenerating-coding-based cloud storage: theory and implementation[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(2): 407-416.

[13] BOWERS K D, JUELS A, OPREA A. HAIL: a high-availability and integrity layer for cloud storage[C]//Proceedings of the 16th ACM Conference on Computer and Communications Security. New York: ACM Press, 2009: 187-198.

[14] CHEN H C H, HU Y C, LEE P P C, et al. NCCloud: a network-coding-based storage system in a cloud-of-clouds[J]. IEEE Transactions on Computers, 2014, 63(1): 31-44.

[15] HE K, HUANG C H, SHI J L, et al. Public integrity auditing for dynamic regenerating code based cloud storage[C]//2016 IEEE Symposium on Computers and Communication. Piscataway: IEEE Press, 2016: 581-588.

[16] REN Z W, WANG L N, WANG Q, et al. Dynamic proofs of retrievability for coded cloud storage systems[J]. IEEE Transactions on Services Computing, 2018, 11(4): 685-698.

[17] LIU J, HUANG K, RONG H, et al. Privacy-preserving public auditing for regenerating-code-based cloud storage[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(7): 1513-1528.
[18] LE A, MARKOPOULOU A, DIMAKIS A G. Auditing for distributed storage systems[J]. IEEE/ACM Transactions on Networking, 2016, 24(4): 2182-2195.
[19] VS L, PP D. A secure regenerating code-based cloud storage with efficient integrity verification[J]. International Journal of Communication Systems, 2019, 32(9): e3948.
[20] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes[C]//International Conference on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 1999: 223-238.
[21] CHILLOTTI I, GAMA N, GEORGIEVA M, et al. TFHE: fast fully homomorphic encryption over the torus[J]. Journal of Cryptology, 2020, 33(1): 34-91.
[22] LIU M P, JIANG R, KONG H F. Cryptanalysis and countermeasures on privacy-preserving public auditing for regenerating-code-based cloud storage[C]//International Conference on Communication and Electronic Information Engineering. Dordrecht: Atlantis Press, 2016: 275-283.
[23] 陈越, 王龙江, 严新成, 等. 基于再生码的拟态数据存储方案[J]. 通信学报, 2018, 39(4): 21-34.
CHEN Y, WANG L J, YAN X C, et al. Mimic storage scheme based on regenerated code[J]. Journal on Communications, 2018, 39(4): 21-34.
[24] KRAWCZYK H. New hash functions for message authentication[C]//International Conference on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 1995: 301-310.
[25] LIANG W, FAN Y K, LI K C, et al. Secure data storage and recovery in industrial blockchain network environments[J]. IEEE Transactions on Industrial Informatics, 2020, 16(10): 6543-6552.
[26] LIU G J, GUO W M, LIU X M, et al. Security analysis and improvements on a remote integrity checking scheme for regenerating-coding-based distributed storage[J]. Security and Communication Networks, 2021, 2021: 1-8.
[27] COHEN A, D'OLIVEIRA R G L, SALAMATIAN S, et al. Network coding-based post-quantum cryptography[J]. IEEE Journal on Selected Areas in Information Theory, 2021, 2(1): 49-64.
[28] KADHE S, SPRINTSON A. Weakly secure regenerating codes for distributed storage[C]//2014 International Symposium on Network Coding. Piscataway: IEEE Press, 2014: 1-6.

[作者简介]

刘光军（1980– ），男，安徽六安人，博士，西安文理学院副教授，主要研究方向为密码学与编码理论、网络编码、安全编码计算等。

郭网媚（1984– ），女，陕西周至人，博士，西安电子科技大学副教授、博士生导师，主要研究方向为网络编码、无线通信等。

熊金波（1981– ），男，湖南益阳人，博士，福建师范大学教授，主要研究方向为安全深度学习、移动群智感知、隐私保护技术等。

刘西蒙（1988– ），男，陕西西安人，博士，福州大学研究员，主要研究方向为隐私计算、密文数据挖掘、大数据隐私保护、可搜索加密等。

董长宇（1977– ），男，黑龙江齐齐哈尔人，博士，纽卡斯尔大学教授、博士生导师，主要研究方向为可搜索加密、隐私保护、人工智能安全等。