DOI：10.13718/j.cnki.xdzk.2020.10.005

面向再生编码云存储的高效隐私保护审计方案

刘光军[1]，　熊金波[2]，　张力宁[1]，
杨渭清[1]，　李少武[3]，　许迅雷[1]

1. 西安文理学院 信息工程学院，西安 710065；2. 福建师范大学 数学与信息学院，福州 350117；
3. 北方自动控制技术研究所，太原 030006

摘要：再生码技术在大数据分布式云存储中具有重要的应用价值，如何利用编码技术构造轻量型的隐私保护和审计机制仍然是基于再生码的分布式云存储系统尚未解决的问题. 采用质询响应计算安全外包的策略和动态随机线性掩码技术，提出了一种适用于再生码云存储系统隐私保护数据审计方案，方案不仅能实时完成服务器质询响应数据的在线动态加密，而且在云存储节点和审计者之间构造了一种隐私计算协同外包的审计策略. 理论分析和实验表明，该方案实现了完备的隐私保护机制和审计安全性，与现有工作相比，具有较快的实现效率.

关　键　词：数据审计；再生码；隐私保护；分布式存储；大数据

中图分类号：TP391　　**文献标志码**：A　　**文章编号**：1673－9868(2020)10－0037－09

再生码是近年来提出的一种适用于分布式数据存储的冗余编码机制，已被证明可以达到存储和修复带宽的最优权衡[1]. 基于再生码技术的分布式云存储系统，在数据修复时的带宽利用方面具有明显的性能优势. 然而，如何保证存储数据的安全可靠性是基于再生码的大数据云存储系统有待解决的关键问题之一[2].

迄今为止，学者们已经提出了众多云计算数据审计方案，但这些方案大多依赖于公钥密码技术，具有很高的计算复杂度，存在着验证效率低下和安全实现条件过于苛刻等问题[3-4]，难以适用于需要进行频繁代数编码操作的基于再生码的云存储系统.

当前，基于再生码技术的云存储数据审计也取得一些代表性的成果. Chen 等[5]首先利用网络编码和随机采样技术提出了一种远程数据检测方案，但该方案不仅需要将编码向量进行加密操作，而且也要对所有的外包存储向量进行采样编码，计算量和通信量都很高，严重影响了系统存储性能. Chen 等[6]采用最大距离可分码(MDS，Maximum Distance Separable)实现了抗拜占庭攻击的编码方案. Ren 等[7]利用网络编码(外码)和纠错编码(内码)技术对数据进行双重编码，有效提升了数据的可用性. Sengupta 等[8]将网络编码抗污染方案和公钥密码技术进行结合，实现了一种可公开审计的机制，但该方案和文献[6-7]都没有考虑数据的隐私保护. Liu 等[9]利用 BLS 签名设计了一种审计方案，预编码操作代价很大，但被发现存在着一些安全缺陷[10]. 考虑到存储系统效能，选择私有审计策略是当前基于再生码的云存储系统较为合理的选择. Le 等[11]利用消息认证码和线性加密的思想提出了一种性能高效的分布式隐私保护审计方案 NC-Audit，但

收稿日期：2020－09－29
基金项目：国家自然科学基金项目(61872088)；西安市科技计划项目(2020KJWL02，2017CGWL35，2016CXWL22).
作者简介：刘光军(1980－)，男，博士，副教授，主要从事云计算数据安全、安全分布式计算等方面研究.

需要服务器知道用户的主密钥，显然是不合理的. 此外，该方案 Setup 阶段的参数设计方法是不安全的. Lakshmi 等[12]针对再生码存储系统提出了一种基于纠错码的同态加密方案，可以实现节点数据的加密和纠错，但该方案需要对存储数据进行预加密，同时审计和纠错过程中涉及大量的矩阵乘法计算，计算开销很大. 最近，Liang 等[13]将区块链技术与再生码技术进行融合，提出了一种区块链网络中的安全数据存储和恢复方案，有效地拓展了再生码的应用领域，但该方案并未考虑存储数据审计问题.

综上所述，现有面向再生码存储的数据审计研究工作虽具有一定的可行性，但在计算开销或安全性能上还很不理想，仍然没有克服分布式存储系统实现效率的性能瓶颈. 因此，如何利用代数编码方法能同时实现数据审计和在线隐私保护仍是当前基于再生码技术的分布式安全存储领域一个重要的挑战.

不同于现有离线加密实现隐私保护的云存储审计机制，本文的主要贡献是在审计过程中借助一种隐私计算部分外包的策略，采用基于随机线性掩码的隐私安全技术，提出了一种高效适用于分布式云存储系统具有隐私保护功能的云审计机制. 该方案有效地实现了质询响应数据的隐私保护，同时也给出了云存储节点隐私安全计算协同外包的审计策略，与现有方案相比，该方案可以在服务器端在线实施动态隐私加密，不仅具有完备的安全性，而且具有计算量小和通信开销少的特征，可以有效部署在用户资源有限的应用场景.

1 系统与安全模型

1.1 系统模型

通常基于再生码的分布式存储系统架构包含云存储服务提供商(包含多个分布式存储节点)(CSP, Cloud Storage Provider)、用户和审计者(TPA, Third-Party Auditor)3 个实体，如图 1 所示.

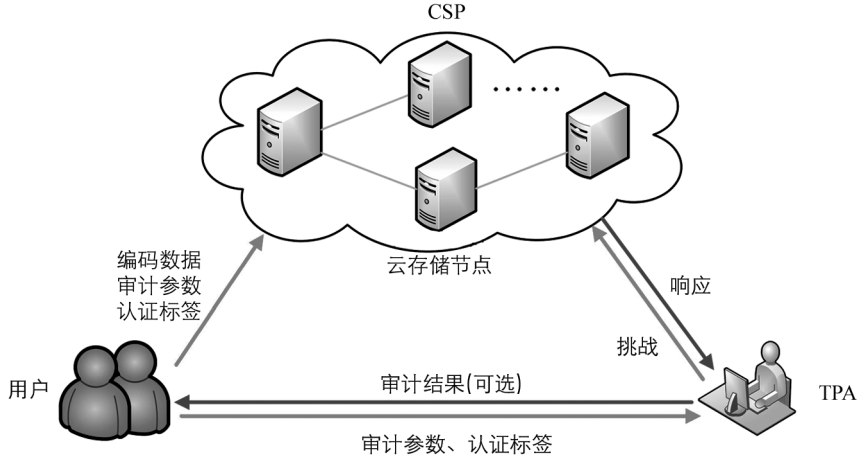

图 1 基于再生码的云存储系统模型

用户订购云存储平台的存储服务，可以将自己的数据分布式存放在云存储各节点 $\boldsymbol{N}_1$，$\boldsymbol{N}_2$，…，$\boldsymbol{N}_N$ 上，具体过程如下：

首先，用户将待存储文件数据分割成一个在有限域 $\mathbb{F}_q$ 上的 n 维初始消息向量序列 $\bar{\boldsymbol{v}}_1$，$\bar{\boldsymbol{v}}_2$，…，$\bar{\boldsymbol{v}}_m \in \mathbb{F}_q^n$；

其次，将初始消息向量 $\bar{\boldsymbol{v}}_i$ 扩展为 $n+m$ 维的初始编码向量 $\boldsymbol{v}_i=(\bar{\boldsymbol{v}}_i, \boldsymbol{e}_i) \in \mathbb{F}_q^{n+m}(n \gg m)$，其中 $\boldsymbol{e}_i$ 为第 i 个元素为 1 的 m 维单位向量，$i=1,2,\cdots,m$；

最后，用户对所有的扩展向量执行参数为(N, K)的 MDS 编码，即选取相应的编码系数 $\alpha_{isj} \in \mathbb{F}_q$，生成外包向量 $\boldsymbol{y}_{sj}=\sum_{i=1}^{m}\alpha_{isj}\boldsymbol{v}_i=(\bar{\boldsymbol{y}}_{sj}, \boldsymbol{g}_{y_{sj}})(j=1, 2, \cdots, M)$ 并发送给存储节点 $\boldsymbol{N}_s(s=1, 2, \cdots, N)$. 其中，$\boldsymbol{g}_{y_{sj}}=(\alpha_{1sj}, \alpha_{2sj}, \cdots, \alpha_{msj})$ 称为编码系数向量，由 y_{sj} 的最后 m 个元素组成.

根据上述编码方法，如果用户需要恢复某个文件，则在云存储系统中任选 K 个正常存储节点，下载对应该文件的存储数据，通过 MDS 译码规则进行解码. 如果云系统中某个存储节点数据损坏，则需要从任选的 d（这里 $K \leqslant d \leqslant N-1$）个正常存储节点上进行数据下载（每个节点的下载数据量为 β）来完成数据的修复. 再生码修复机制包含两类：精确修复或功能修复. 通过灵活选取参数 K、d、β、M 和 N，可以在分布式系统中构造最小存储再生码和最小带宽存储再生码，前者具有最优的存储效率，而后者在修复时具有最高的带宽利用率. 具体设计可见文献[14-15].

TPA 在系统中的作用是实时对云存储节点存储的数据进行完整性检测，TPA 一旦发现云平台中某个存储节点检测失败且该节点自身无法修复的情况下，将执行云存储节点分布式数据修复过程.

下文叙述中仍将采用本节使用的参数及其符号表示.

1.2　安全模型

在基于再生码的分布式云存储系统中，多个存储节点共同合作实现了 CSP 的功能. 假定系统中每个存储节点具有安全的系统防护和密钥管理机制，不会导致用户隐私信息的泄露. 同时，各存储节点能严格执行分布式云存储审计协议. 但是，CSP 可能会贪图节省自己存储开销而故意删除用户极少访问的数据，也有可能为了自己的商业信誉或利益而设法向用户隐瞒存储数据的毁坏. 基于再生码分布式修复功能，系统中各存储节点之间可以互相通信，但各实体之间的通信信道可能并不是安全的，因而传输中的消息需要进行认证加密处理以保证隐私数据的安全保护. 此外，TPA 是一个用户可信的实体，忠实执行审计操作，虽不能与 CSP 共谋，但有窥探用户隐私的动机和欲望.

2　分布式云存储隐私保护审计协议

与已有研究不同的是，为了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. 同时，引入 3 个伪随机函数（PRF，Pseudo-Random Function），即 F_1：$K \times ID \times \mathbb{Z}^+ \to \mathbb{F}_q$，$F_2$：$K \times \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{F}_q$，$F_3$：$K \times \mathrm{WID} \times \mathbb{Z}^+ \to \mathbb{F}_q$. 其中，记号 $\mathbb{Z}^+$ 表示由正整数集合，K 为 PRF 密钥集，ID 为文件标识符集合，WID 为审计任务的标识符集合.

2.1　协议过程描述

由于用户 U 对各存储节点的审计方式相同，本部分仅关注 U 与单存储节点 $\boldsymbol{N}_s$ 的协议执行过程.

2.1.1　Setup 阶段

1）系统初始化

系统选定安全参数 λ，确定伪随机函数 F_1，F_2 和 F_3，生成 CSP 与用户共享密钥 k_e 和 TPA 与用户共享密钥 k_v.

2）外包上传

① U 和 TPA 利用文件标识符 id 和 F_1，计算向量 $\bar{\boldsymbol{r}} \in \mathbb{F}_q^n$ 和 $\boldsymbol{r} \in \mathbb{F}_q^{n+m}$：

$$\begin{aligned}\boldsymbol{r} &= (F_1(k_v, id, 1), F_1(k_v, id, 2), \cdots, F_1(k_v, id, n+m)) \triangleq \\ &(\bar{\boldsymbol{r}}, F_1(k_v, id, n+1), \cdots, F_1(k_v, id, n+m)) \triangleq \\ &(r_1, r_2, \cdots, r_{n+m}) \in \mathbb{F}_q^{n+m}\end{aligned}$$

其中，向量 $\bar{\boldsymbol{r}}$ 由向量 $\boldsymbol{r}$ 的前 n 个元素组成.

② U 分别计算外包向量 $\boldsymbol{y}_{sj} = (\bar{\boldsymbol{y}}_{sj}, \boldsymbol{g}_{y_{sj}})$ 的认证标签，即

$$t_{sj} = \boldsymbol{y}_{sj} \cdot \boldsymbol{r} \in \mathbb{F}_q \tag{1}$$

其中 $j = 1, 2, \cdots, M$；

③ $\boldsymbol{N}_s$ 利用 F_2 计算

$$\bar{\boldsymbol{p}}_i = (F_2(k_e, i, 1), F_2(k_e, i, 2), \cdots, F_2(k_e, i, n-1), \zeta_i) \triangleq$$

$$(p_{i1}, p_{i2}, \cdots, p_{in}) \in \mathbb{F}_q^n (i=1, 2, \cdots, n+1)$$

其中,

$$\zeta_i = p_{in} = \begin{cases} r_{n+m}^{-1} \cdot \sum_{j=1}^{n-1} (r_j \cdot p_{ij}) & i=1,2,\cdots,n-1 \\ F_2(k_e, i, n) & i=n, n+1 \end{cases}$$

上述参数构造实际上要满足如下线性方程组 Σ_0:

$$\Sigma_0: \begin{cases} \bar{\boldsymbol{r}} \cdot \bar{\boldsymbol{p}}_1 = \varphi_1 \\ \vdots \\ \bar{\boldsymbol{r}} \cdot \bar{\boldsymbol{p}}_{n-1} = \varphi_{n-1} \\ \bar{\boldsymbol{r}} \cdot \bar{\boldsymbol{p}}_n = \varphi_n \\ \bar{\boldsymbol{r}} \cdot \bar{\boldsymbol{p}}_{n+1} = \varphi_{n+1} \end{cases}$$

其中, $\varphi_i = 0 (i=1, 2, \cdots, n-1)$, φ_n 和 φ_{n+1} 为用户 U 的专有私密信息.

④ U 随机选取 $\tau_1, \tau_2 \in \mathbb{F}_q \backslash \{0\}$, 计算 $\bar{\omega}_1 = \tau_1^{-1} \cdot \varphi_n$, $\bar{\omega}_2 = \tau_2^{-1} \cdot \varphi_{n+1}$.

最后, 用户 U 将 M 个向量 $\boldsymbol{z}_{sj} = (\boldsymbol{y}_{sj}, t_{sj})$ 和参数 $\bar{\omega}_1, \bar{\omega}_2$ 上传到 $\boldsymbol{N}_s$, 将 $\boldsymbol{g}_{y_{sj}}$, τ_1, τ_2 发送给 TPA. 最后, U 可以删除文件标识符为 id 的本地数据, 仅需保留 $\boldsymbol{g}_{y_{sj}}$ 和协议密钥即可.

2.1.2 Audit 阶段

1) 审计质询与响应

① TPA 任选一个待检测的外包向量索引集合 $\Delta \subseteq [1, M]$, 随机选取 $\varepsilon_i \in \mathbb{F}_q (i \in \Delta)$, 生成质询消息 chal $= \{\langle i, \varepsilon_i \rangle \mid i \in \Delta\}$, 并将 chal 发送给 $\boldsymbol{N}_s$;

② $\boldsymbol{N}_s$ 根据质询消息 chal 生成聚合响应消息

$$\boldsymbol{e} = \sum_{i \in \Delta} \varepsilon_i \boldsymbol{z}_{si} = (\bar{\boldsymbol{e}}, \boldsymbol{g}_e, t) \in \mathbb{F}_q^{n+m+1} \tag{2}$$

③ $\boldsymbol{N}_s$ 对 $\bar{\boldsymbol{e}}$ 进行随机加密, 生成密文 $\bar{\boldsymbol{c}}$, 产生响应消息 $\boldsymbol{V} = \langle \bar{\boldsymbol{c}}, t \rangle \cup \{o_1, o_2\}$, 具体步骤如下:

Step 1: 计算随机掩码系数 $\beta_z = F_3(k_e, wid, z) \in \mathbb{F}_q (z=1, 2, \cdots, n+1)$ 和掩码向量

$$\bar{\boldsymbol{m}} = \sum_{i=1}^{n+1} \beta_i \bar{\boldsymbol{p}}_i \in \mathbb{F}_q^n \tag{3}$$

Step 2: 计算

$$\bar{\boldsymbol{c}} = \bar{\boldsymbol{e}} + \bar{\boldsymbol{m}} \tag{4}$$

$$o_1 = \beta_n \bar{\omega}_1, \ o_2 = \beta_{n+1} \bar{\omega}_2 \tag{5}$$

最后, $\boldsymbol{N}_s$ 将消息 V 发送至 TPA.

2) 响应消息检测

TPA 利用响应消息计算编码系数向量 $\boldsymbol{g}_e$, 得到待检测消息 $\boldsymbol{c} = (\bar{\boldsymbol{c}}, \boldsymbol{g}_e)$, 计算

$$\kappa = \tau_1 o_1 + \tau_2 o_2 \tag{6}$$

进而判断等式

$$\boldsymbol{r} \cdot \boldsymbol{c} = t + \kappa \tag{7}$$

是否成立. 如果成立, 针对存储节点 $\boldsymbol{N}_s$ 的该次审计通过, 否则, 判断 $\boldsymbol{N}_s$ 数据存储出错.

上述过程中, TPA 利用公式(6)协作完成 $\boldsymbol{N}_s$ 的质询响应计算工作, 完成审计检测.

2.2 审计响应消息的外包协同计算

审计响应的外包协同计算是保证上述协议安全的重要操作. 用户首先使用 τ_1, τ_2 将 φ_n, φ_{n+1} 随机化为 $\bar{\omega}_1$, $\bar{\omega}_2$, 将 $\bar{\omega}_1$, $\bar{\omega}_2$ 发送到 $\boldsymbol{N}_s$. 由于 $\boldsymbol{N}_s$ 无法知道向量 $\bar{\boldsymbol{r}}$, 显然 $\boldsymbol{N}_s$ 无法知晓 φ_n, φ_{n+1}. 同时, 用户 U 将 τ_1,

τ_2 发送给了 TPA，确保 TPA 能正确参与审计协作安全计算. 审计响应时，服务器利用随机数 β_n，β_{n+1} 再次对 $\bar{\omega}_1$，$\bar{\omega}_2$ 进行了随机化，从而使 TPA 也无法获取 φ_n，φ_{n+1} 的值.

该方法不仅实现了对参数 φ_n，φ_{n+1} 的有效保护，而且能将 CSP 生成质询响应消息的部分计算外包给了 TPA，由 TPA 协作完成，从而给出了一种适用于数据审计的隐私计算安全外包的运行策略，这种策略在本方案中的作用有两个：一是实现了隐私保护审计参数的隐式安全传递，保证了方案的安全性；二是保证了 TPA 能正确地执行审计协议，且不会很大程度地影响 TPA 的工作效率.

3　方案正确性验证

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本方案的验证. 如果待审计文件数据已经受到损坏或被删除，存储节点返回的响应消息将无法通过审计检测.

定理 1　如果所有的存储节点严格遵守本方案的协议规则，则 TPA 可以有效地证明节点存储的安全可靠性.

证　令 $\boldsymbol{r}=(\bar{\boldsymbol{r}}, r_{n+1}, r_{n+2}, \cdots, r_{n+m})$，可得

$$\begin{aligned}
\boldsymbol{r}\cdot\boldsymbol{c} &= (\bar{\boldsymbol{r}}, r_{n+1}, r_{n+2}, \cdots, r_{n+m})\cdot(\bar{\boldsymbol{c}}, \boldsymbol{g}_e) = \\
&(\bar{\boldsymbol{r}}, r_{n+1}, r_{n+2}, \cdots, r_{n+m})\cdot(\bar{\boldsymbol{e}}+\bar{\boldsymbol{m}}, \boldsymbol{g}_e) = \\
&(\bar{\boldsymbol{r}}, r_{n+1}, r_{n+2}, \cdots, r_{n+m})\cdot((\bar{\boldsymbol{e}}, \boldsymbol{g}_e)+(\bar{\boldsymbol{m}}, \boldsymbol{0}_m)) = \\
&(\bar{\boldsymbol{r}}, r_{n+1}, \cdots, r_{n+m})\cdot(\bar{\boldsymbol{e}}, \boldsymbol{g}_e)+(\bar{\boldsymbol{r}}, r_{n+1}, \cdots, r_{n+m})\cdot(\bar{\boldsymbol{m}}, \boldsymbol{0}_m) = \\
&t+\bar{\boldsymbol{r}}\cdot\bar{\boldsymbol{m}} = \\
&t+\sum_{i=1}^{n+1}\beta_i\bar{\boldsymbol{r}}\cdot\bar{\boldsymbol{p}}_i = \\
&t+\sum_{i=1}^{n+1}\beta_i\varphi_i = \\
&t+\sum_{i=1}^{n-1}\beta_i\varphi_i+\sum_{i=n}^{n+1}\tau_{i+1-n}\cdot(\beta_i\bar{\omega}_{i+1-n}) = \\
&t+\sum_{i=n}^{n+1}\tau_{i+1-n}o_{i+1-n} = \\
&t+\kappa
\end{aligned}$$

其中，$\boldsymbol{0}_m$ 表示长为 m 的零向量.

4　安全性证明

本方案可以有效地保证云数据的安全审计和审计响应消息的隐私保护.

定理 2　如果 F_1 和 F_2 是理想安全的 PRF，CSP 能解出向量 $\bar{\boldsymbol{r}}$ 的最大概率为 q^{-1}.

证　根据参数设置方法，敌手可以构造一个关于向量 $\bar{\boldsymbol{r}}$（含 n 个未知量）的线性方程组 $\sum_0$，即

$$\sum\nolimits_1 : \bar{\boldsymbol{r}}\cdot(\bar{\boldsymbol{p}}_1^T, \bar{\boldsymbol{p}}_2^T, \cdots, \bar{\boldsymbol{p}}_{n-1}^T) \triangleq \bar{\boldsymbol{r}}\cdot\boldsymbol{A} = \boldsymbol{0}_{n-1} \tag{8}$$

由于 $\bar{\boldsymbol{p}}_i \in \mathbb{F}_q^n$，则 $\boldsymbol{A}$ 为一个维数为 $n\times(n-1)$ 的矩阵. 又由于 $\boldsymbol{A}$ 是由安全的 PRF 生成，所以矩阵 $\boldsymbol{A}$ 的秩最大值为 $n-1$，则 $\sum_0$ 的解空间的维数最小为 1. 所以，敌手获得向量 $\bar{\boldsymbol{r}}$ 的概率最高仅为 q^{-1}. 即证.

定理 2 保证了认证向量 $\bar{\boldsymbol{r}}$ 的安全性. 同理可知，TPA 也无法根据已有信息推导出认证向量 $\bar{\boldsymbol{r}}$ 的信息. 利用类似的方法，可以很容易发现方案 NC-Audit 是无法保证审计检测的安全性的. 根据方案 NC-Audit 的参数构造方法，CSP 可以很容易使用定理 2 和后文定理 4 的分析，构造出类似公式(8) 的攻击方程组，从而能

以很高的概率反解出用户的隐私密钥向量，导致该方案审计功能的丧失.

定理3 如果 F_1、F_2 和 F_3 是理想安全的 PRF，方案中的加密方法实现了完善的安全性.

证 记 $\overline{\boldsymbol{p}}_i=(p_{i1}, p_{i2}, \cdots, p_{i,n})(i=1, 2, \cdots, n+1)$，$\overline{\boldsymbol{m}}=(m_1, m_2, \cdots, m_n)$. 根据公式(3)和(4)，敌手可以构造如下以 $\beta_i(i=1, 2, \cdots, n+1)$ 为未知量，由 n 个线性方程联立组成的方程组 Σ_2：

$$\Sigma_2:\begin{cases}\beta_1 p_{11}+\beta_2 p_{21}+\cdots+\beta_{n+1} p_{n+1,1}=m_1\\ \cdots\\ \beta_1 p_{1,n-1}+\beta_2 p_{2,n-1}+\cdots+\beta_{n+1} p_{n+1,n-1}=m_{n-1}\\ \beta_1 p_{1n}+\beta_2 p_{2n}+\cdots+\beta_{n+1} p_{n+1,n}=m_n\end{cases}$$

在假设条件下，方程组 Σ_2 中所有的 β_i 和 p_{ij} 在敌手看来都是在 $\mathbb{F}_q$ 中随机选取的值. 令该方程组的系数矩阵的秩为 ρ，显然有 $\rho\leqslant n$.

任取 $t\in[1, n]$，我们来分析 m_t 的随机性. 除了 m_t 外，我们可以先固定方程组等号右边 $m_i(i\neq t)$ 的取值. 容易看出，无论 m_t 取 $\mathbb{F}_q$ 中的任意值，该方程组解空间的维数为 $n+1-\rho$，即可能的解个数都为 $q^{n+1-\rho}$，换句话说，m_t 的取值与其它 $m_i(i\neq t)$ 的取值完全独立，皆是在 $\mathbb{F}_q$ 中按均匀分布选取的随机值. 即证.

定理3保证了 CSP 质询响应消息 $\overline{\boldsymbol{e}}$ 的安全性.

定理4 如果 F_1、F_2 和 F_3 是理想安全的 PRF，敌手利用该方案成功选取一个伪造数据向量的合法认证标签的概率最多为 q^{-1}.

证 该定理可采用“质询—响应”游戏进行证明. 质询者可以随机选择 $r_i\in\mathbb{F}_q$，生成 $\boldsymbol{r}_{id}=(r_1, r_2, \cdots, r_{n+m})$. 敌手可以获得一个文件 id 值，适应性地选择文件消息向量 $\boldsymbol{y}_i\in\mathbb{F}_q^{n+m}$，向质询者询问该向量的认证标签. 随后，质询者将 $\boldsymbol{t}_i=\boldsymbol{y}_i\cdot\boldsymbol{r}_{id}(i=1, 2, \cdots, m)$ 返回给敌手.

假设敌手最终成功输出一个伪造的多元组$(id, \boldsymbol{y}', \boldsymbol{t}')$，则有

$$\boldsymbol{t}'=\boldsymbol{y}'\cdot\boldsymbol{r}_{id}$$

$$\boldsymbol{y}'\notin\operatorname{span}(y_1, y_2, \cdots, y_m)$$

此时，敌手完全可以构造一个关于向量 $\boldsymbol{r}_{id}$ 中个分量 $r_j(j=1, 2, \cdots, n+m)$ 的线性方程组 Σ_3：

$$\Sigma_3:\begin{cases}\overline{\boldsymbol{p}}_1\cdot\overline{\boldsymbol{r}}_{id}=0\\ \cdots\\ \overline{\boldsymbol{p}}_{n-1}\cdot\overline{\boldsymbol{r}}_{id}=0\\ \boldsymbol{y}_1\cdot\boldsymbol{r}_{id}=t_1\\ \cdots\\ \boldsymbol{y}_m\cdot\boldsymbol{r}_{id}=t_m\\ \boldsymbol{y}'\cdot\boldsymbol{r}_{id}=t'\end{cases}$$

令该方程组系数矩阵的秩为 ρ，则其解空间的维数为 $n+m-\rho$，可能解的个数为 $q^{n+m-\rho}$. 利用定理2的分析方法，可得值 t' 是 $\mathbb{F}_q$ 中按均匀分布选取的随机值. 由此可知，$\boldsymbol{t}'$ 为向量 $\boldsymbol{y}'$ 的正确认证标签的概率仅为 q^{-1}. 即证.

定理2和定理4共同保证了审计策略的安全性.

5 方案性能分析

目前适用于再生码云存储系统的审计方案中，方案 NC-Audit 在计算性能上具有十分显著的优势. 最近，文献[12]提出了一种用户在线审计的同类安全方案，但该方案利用一种基于对称密码体制的加法同态加密方案的可验证计算属性来检查外包数据的完整性，是该类方案的最新进展. 于是，在相同的安全强度

下，本节将本文方案与 NC-Audit 和文献[12]进行性能比较.

在隐私保护和审计安全性方面，根据定理 2 的分析，NC-Audit 并不能有效防止存储服务器反解出用户的私有审计向量，所以无法保证审计策略的安全性. 但是，本文提出的审计私有计算的协同外包机制避免了服务器获得多余的攻击辅助信息，从而有效地保证了用户隐私密钥的机密性，实现了完善的审计安全功能. 与本文不同，文献[12]采用数据预加密的方式实现了该方案的隐私安全性.

表 1　方案系统性能比较

方案名称	审计安全性	隐私保护	通信开销	TPA 计算量	CSP 计算量
NC-Audit	否	是	$m+n+2\xi+3$	$m+n+1$	$(n-1)^2+\xi(m+n+1)$
文献[12]	是	是	$2n+2\xi$	$O(n^2 s)$	$n\xi M s^{-1}+\xi$
本文方案	是	是	$m+n+2\xi+1$	$m+n+3$	$n(n+1)+\xi(m+n+1)$

注：$*\ \xi=|\Delta|$.

在计算和通信性能方面，本文方案具有良好的实现性能，如表 1 所示. 在整个审计过程中，本文方案在计算开销上仅比 NC-Audit 多了 $3n+1$ 次乘法计算，这在运行时间上完全可以忽略. 此外，本文方案无需在初始消息中填充随机字符，故在通信效率上比 NC-Audit 高. 本文方案也没有因安全功能的实现额外增加系统中各实体的存储开销. 由于文献[12]涉及到了较大规模的线性编码和信道译码操作，故计算和通信开销也较大.

为了比较上述 3 个方案的实际运行性能，本文利用 C 语言编程在 Intel(R) Core(TM) i5-7300HQ 2.50GHz@16G RAM 处理器 Win10-64bit 环境下对本文方案和 NC-Audit 以及文献[6]中的方案进行在线计算性能的实验测试. 实验设定方案安全强度为 80bits，参数 $q=2^8$，$n=2^{12}$，$m=200$，$M=300$(即每个节点存储 300 个编码数据包). 实验采用 AES 的 CTR 模式(分组长度为 128bits)来实现方案中的伪随机函数[16]. 为统一参数，实验将文献[12]中的 s 值(每个编码向量分割的子块数目)置为 300，且用户本身代替完成了 TPA 的计算任务. 实验忽略了加法运算耗时，重点关注有限域上的乘法计算时间. 表 2 给出了在审计运行过程中，当ξ取 100 和 200 时，TPA 和 CSP(每个服务器)在线计算总时间(实验次数均为 2000 次). 结果显示，本文方案在审计运行中达到了与方案 NC-Audit 同样高效的执行效率. 由于文献[12]需要进行复杂的译码操作，故用户端计算负载较大，但服务器端开销较小.

表 2　审计方案在线计算时间比较

方案名称	T/ms(TPA, 200)	T/ms(TPA, 300)	T/ms(CSP, 200)	T/ms(CSP, 300)
NC-Audit	7.453	7.456	2.927	3.002
文献[12]	12.50	15.74	1.310	1.694
本文方案	7.454	7.460	2.930	3.011

$*\ T(i, j)$表示当ξ取 j 时实体 i 的审计操作计算的平均时间.

需要说明的是，本文虽未考虑失效节点的数据修复问题，但方案的安全实现并没有抵消再生码系统的性能优势. 虽然方案的构造产生了一定的安全认证开销，但由此产生的通信开销在数据修复过程中消耗的带宽资源很少，因而本文方案的实施并不会显著影响再生码技术的带宽利用优势. 与之相比，文献[6-7，9，12]中的方案虽然具有更高的安全强度，但由于实现中会付出很大的计算和通信开销，其代价可能会造成再生码技术优势在分布式存储系统的严重丢失.

6　结　论

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议题之一. 虽然使用公钥密码技术能很容易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

再生码存储系统数据编码特性，并将其与审计计算协作外包的思想进行结合，构造了一种线性随机编码技术，成功实现了一种高效的具有动态隐私保护的云数据审计方案，实现了一种基于对称密码技术的隐私保护和数据认证的高效整合. 该方案的实现具有较低的计算复杂度，可在资源受限的环境下具有一定的实用价值.

在数据很少被读取或修改的情况下，例如长期归档、数据托管和监管存储等应用场景，基于再生码的云存储可能是首选的存储方案. 但是，如果基于再生码的云存储系统允许用户可以对外包数据进行增删和修改等操作，现有的静态数据审计机制可能将会面临着严重的可用性问题. 由于需要编码同步，所以，适用于传统云计算系统的动态数据审计方案都无法直接而又高效地应用于基于再生码的云存储系统. 因此，本文下一步的研究工作将致力于解决基于再生码存储的具有隐私保护特性的动态数据审计问题.

参考文献：

[1] DIMAKIS A G, GODFREY P B, Wu Y N, et al. Network Coding for Distributed Storage Systems [J]. IEEE Transactions on Information Theory, 2010, 56(9): 4539-4551.

[2] 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.

[3] THAKUR N, SINGH A, SANGAL A L. Data Integrity Authentication Techniques in Cloud Computing: A Survey [M]//Soft Computing: Theories and Applications. Springer, 2020: 1255-1267.

[4] GUDEME J R, PASUPULETI S K, KANDUKURI R. Review of Remote Data Integrity Auditing Schemes in Cloud Computing: Taxonomy, Analysis, and Open Issues [J]. International Journal of Cloud Computing, 2019, 8(1): 20-49.

[5] 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-CCSW '10. October 8, 2010. Chicago, Illinois, USA. New York: ACM Press, 2010: 31-42.

[6] 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.

[7] 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.

[8] SENGUPTA B, RUJ S. Publicly Verifiable Secure cloud Storage for Dynamic Data Using Secure Network Coding [C] //Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security-1ASIA CCS '16. May 30-1June 3, 2016. Xi'an, China. New York: ACM Press, 2016: 107-118.

[9] 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.

[10] LIU M P, JIANG R, KONG H F. Cryptanalysis and Countermeasures on Privacy-Preserving Public Auditing for Regenerating-Code-Based Cloud Storage [C] //Proceedings of the International Conference on Communication and Electronic Information Engineering (CEIE 2016). October 15-16, 2016. Guangzhou, China. Paris, France: Atlantis Press, 2016.

[11] LE A, MARKOPOULOU A, DIMAKIS A G. Auditing for Distributed Storage Systems [J]. IEEE/ACM Transactions on Networking, 2016, 24(4): 2182-2195.

[12] LAKSHMI V S, PP D. A Secure Regenerating Code-Based Cloud Storage with Efficient Integrity Verification [J]. International Journal of Communication Systems, 2019, 32(9): e3948.

[13] 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.

[14] RASHMI K V, SHAH N B, KUMAR P V. Optimal Exact-Regenerating Codes for Distributed Storage at the MSR and

MBR Points via a Product-Matrix Construction [J]. IEEE Transactions on Information Theory, 2011, 57(8): 5227-5239.

[15] HU Y, CHEN H C H, LEE P P C, et al. NCCloud: Applying Network Coding for the Storage Repair in a Cloud-of-Clouds [C] // Proceedings of the 10th USENIX Conf. File Storage Technol. (FAST), 2012: 265-272.

[16] MENEZES A J, VAN OORSCHOT P C, VANSTONE S A. Handbook of Applied Cryptography [M]. Boca Raton: CRC Press, 2018.

An Efficient Privacy-Preserving Data Auditing Scheme for Regenerating-Code-Based Cloud Storage

LIU Guang-jun[1], XIONG Jin-bo[2], ZHANG Li-ning[1], YANG Wei-qing[1], LI Shao-wu[3], XU Xun-lei[1]

1. *School of Information Engineering, Xi'an University, Xi'an 710065, China;*
2. *School of Mathematics and Informatics, Fujian Normal University, Fuzhou 350117, China;*
3. *North Automatic Control Technology Institute, Taiyuan 030006, China*

Abstract: The technology of regenerating codes is of important application value in big data distributed cloud storage. How to construct a mechanism of effective privacy-preservation and remote auditing by resorting to coding for a regenerating-code-based cloud storage system remains an unsolved issue. In this paper, by integrating a secure outsourcing strategy for challenge-response computation and a dynamic random linear mask technique, an efficient data auditing scheme is proposed with online privacy protection. The solution can not only achieve dynamic random real-time encryption to the response to the challenges from adversaries, but also construct an audit paradigm by which the privacy-computing can be cooperatively outsourced to the auditor. Theoretical analysis and experimental results show that this scheme achieves complete privacy protection and auditing security, and is more efficient in realization than the existing schemes.

Key words: data auditing; regenerating code; privacy protection; distributed storage; big data

责任编辑　汤振金