隐私数据验证场景下的隐私保护研究

石 侃[1], 陈 洁[2]

[1] 华东师范大学 计算机科学与技术学院 上海 中国 200241
[2] 华东师范大学 软件工程学院 上海 中国 200241

摘要 隐私数据验证场景是信息验证服务下的一类特殊场景, 其实用性要求数据在第三方数据库进行存储、发布且有能力处理任意形式声明的验证, 其安全性要求数据在存储、更新与证明期间提供有效的隐私保护手段。目前该场景下的隐私保护研究尚且处于空白阶段, 因此本文引入可证明数据加密策略的概念, 以满足隐私数据验证场景下的实用性与安全性需求。本文主要有三个贡献: (1)对可证明数据加密策略进行讨论并给出形式化定义; (2)基于非交互零知识证明构造出首个可证明数据加密方案, 并同时支持高效的数据更新操作; (3)基于承诺方案、非交互零知识证明与全同态加密, 提出可证明数据加密策略的两种通用构造框架并给予相关性质证明。

关键词 隐私数据验证场景; 隐私保护; 可证明数据加密策略; 非交互零知识证明
中图法分类号 TP309 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2020.11.03

Research on privacy protection in privacy data verification scenarios

SHI Kan[1], CHEN Jie[2]

[1] School of Computer Science and Technology, East China Normal University, Shanghai 200241, China
[2] School of Software Engineering, East China Normal University, Shanghai 200241, China

Abstract The privacy data verification scenario is a special scenario under the information verification service. Its practicability requires data to be stored and published in third-party database and have the ability to handle any form of statement verification. Its security requires effective privacy protection during data storage, update and certification. At present, the privacy protection research in this scenario is still blank, so this paper introduces the concept of provable data encryption scheme to meet the practicality and security requirements in the privacy data verification scenario. This paper has three main contributions: (1) discussing the provable data encryption scheme and giving its formal definition; (2) constructed the first provable data encryption scheme based on non-interactive zero-knowledge proof, which also supports efficient data update operations; (3) based on the commitment scheme, non-interactive zero-knowledge proof and fully homomorphic encryption, we propose two general construction frameworks for provable data encryption scheme and proves their relevant properties.

Key words privacy data validation scenario; privacy protection; provable data encryption scheme; non-interactive zero-knowledge proof

1 引言

随着信息技术的发展, 大量的私人信息(如个人档案、旅游消费记录、商业数据、生理数据、学历信息、职业信息等)被存储在第三方数据库中, 为公民提供便捷的信息验证服务(如个人档案审核, 学历审核, 医学检测, 商业信息审查等)。新的技术方法取代了旧时代纸质保存、人工审核的原始方法, 但在其加持下, 重大隐私泄露事件仍旧频发, 如第一美国金融公司 8.85 亿个人交易记录泄漏、华住酒店 5 亿条房客入住信息泄露等。

上述场景可以统称为数据验证场景, 该场景下数据所有方能提前生成、处理并发布目标数据, 且证明方可以随时提取所发布内容, 用以证明对应数据满足特定声明。数据验证场景的安全性要求数据在发布、存储以及证明阶段皆存在有效的隐私保护手

通讯作者: 陈洁, 博士, 教授, Email: jchen@cs.ecnu.edu.cn。
本课题得到国家自然科学基金项目(No.61972156, No. U1705264), 中国科协青年人才托举工程(No.2017QNRC001)资助。
收稿日期: 2020-03-13; 修改日期: 2020-05-25; 定稿日期: 2020-09-23

段, 以防止无关的隐私信息泄露。

数据验证场景本身可描述成一类特殊的零知识证明场景, 该场景下秘密 w 可作为数据独立进行加密、发布、存储与共享, 且其加密数据 ϕ_w 具备某类声明的可证明能力(即证明者可以通过使用 ϕ_w 来证明 w 满足某类声明)。

1.1 相关工作

目前针对数据验证场景中特定声明类型的隐私保护技术已有相关研究成果: Micali, Rabin 和 Kilian 于 03 年首次提出 Zero-knowledge Sets/Databases [1-2] (ZKS)的概念, ZKS 允许证明者对秘密有限数据集 S 进行承诺, 承诺密文本身具备成员证明(即证明任意元素 x 是否属于 S)的可证明能力, 后基于 ZKS 衍生出 Zero-knowledge Elementary Databases[3]、Zero-knowledge Accumulator[4]、Zero-knowledge Range Proof[5]等相关概念。朱岩等[6]于 11 年基于可恢复性证明模型(POR)与零知识证明提出了 Zero-knowledge POR 的概念, 以得到同时防止证明者欺骗与验证信息泄露的有效 POR 协议。SNARGs /SNARKs 的概念起始于 1992 年 Kilian[7]的工作, 其属于一种高效的非交互零知识论证系统, 在区块链中的数据验证上有着广泛的应用, 如认证、信息分享、公有/私有链的起源证明、匿名交易、计算验证等。

1.2 研究目标

本文关注隐私数据验证场景下的隐私保护问题。隐私数据验证场景是需要支持任意声明形式的数据验证场景, 其针对的是一类“泛用型”隐私数据——如公民个人信息, 会随个人活动被生成、发布与存储, 并在未来需要应对任意可能出现的验证要求(个人资质审核、入职审查、遗传病史调查等)。该类数据的生成不以特定声明类型为目标, 但需要有应对任意声明形式的可证明能力。

隐私数据验证场景具有如下特点。

1) 独立性: 隐私数据的生成与验证是相互独立的场景, 其数据的生成发生在验证场景之前, 不以某类特定声明类型为目标。

2) 不确定性: 用户在上传隐私数据时, 无法提前获知验证场景, 如验证场景的次数、时间、地点、目标数据、声明内容等。

3) 流动性: 为方便在任意验证场景下提取目标数据, 隐私数据可能需要存储在第三方数据库中, 并支持多个数据存储方之间进行数据发布、整合与共享。

以上三点使得隐私数据验证场景具有更强的实用性需求与安全性需求。

1) 实用性需求: 隐私数据本身应具有支持任意声明的可证明能力。

2) 安全性需求: 除了验证结果, 还需要保证数据所包含的隐私信息不被泄露。

上述特点造成了该场景下隐私保护的困境: 一方面, 明文存储虽然能满足隐私数据验证场景下的实用性需求, 但无论使用何种存储、验证方式, 其依然存在重大隐私泄露风险; 另一方面, 使用相关加密技术虽然能保障隐私数据验证场景下的安全性需求, 但如何同时保证其实用性需求, 将成为构造具体加密方案的技术难点。

综合上述问题并同时结合不诚实证明者等实际因素, 隐私数据验证场景下的加密方案应解决如下问题:

必要问题:

1) 密文生成问题: 隐私数据的密文是否可以通过不依赖声明而独立生成。

2) 密文可证明问题: 使用密文数据的证明者是否可以与使用明文数据的证明者具有相同的证明能力。换句话说, 密文是否可以代表明文完成任意声明的证明。

3) 密文可靠性问题: 对于明文无法满足的声明, 其密文是否不具备欺骗能力。

4) 密文二义性问题: 密文是否具有唯一对应的明文, 否则不诚实的证明者可利用二义性进行欺骗证明(如使用具有多个身份的加密个人档案, 在面对不同的声明时使用不同的身份进行选择性欺骗证明)。

5) 多次证明的隐私泄露问题: 在多次使用相同密文证明不同声明后, 是否会泄露与声明无关的隐私。

非必要问题:

6) 证明效率问题: 加密是否会导致证明过程效率低下。

7) 更新效率问题: 加密是否导致多个密文的整合与更新过程效率低下。

8) 同态计算问题: 密文是否支持(全)同态计算。

1.3 本文结果

针对隐私数据验证场景的要求与问题, 本文基于零知识证明提出可证明数据加密策略的全新概念。可证明数据加密策略允许数据拥有者对明文数据 w 进行可证明加密以得到密文数据 ϕ_w, ϕ_w 可以代替 w 进行发布与存储, 并具有与 w 等同的可证明能力。

本文通过改进 Gorth[8]的方案得到首个可证明数据加密方案, 其效率与 Gorth[8]一致。除此之外, 本文利用不同的构造思路, 基于承诺方案、全同态加密和零知识证明提出可证明数据加密策略的两种通用构造框架。上述工作指出了可证明数据加密方案的三种构造方向。

1.4 本文组织

本文组织如下: 第二章介绍预备知识和在方案构造中所使用到的关键技术; 第三章对可证明数据加密策略进行描述并给出形式化定义; 第四章提供首个可证明数据加密策略原型方案, 并给出相应证明与效率分析; 第五章介绍可证明数据加密策略的两种通用构造框架并给出性质证明; 第六章对全文进行总结。

2 预备知识与关键技术

2.1 完美绑定的承诺方案

承诺方案(Commitment schemes)是一类两方交互协议, 承诺方通过发布对消息 m 的承诺 $com(m)$ 以保证 m 的私密性(承诺阶段的隐藏性), 并在公开 $com(m)$ 的承诺内容时, 无法做到对 m 的篡改(公开阶段的绑定性)。

定义 1. 完美绑定的承诺方案. 称 $\langle G,C,V\rangle$ 是完美绑定的承诺方案, 如果满足:

(1)有效性: G,C,V 都是多项式时间算法。

(2)完备性: 对于所有的 m 有

$$\Pr\left[\begin{array}{l}\sigma\leftarrow G(1^{\kappa});(c,d)\leftarrow C(\sigma,m):\\ V(\sigma,c,d,m)=1\end{array}\right]=1 \tag{1}$$

(3)(完美)绑定性: 对于任意的多项式时间算法 S 有

$$\Pr\left[\begin{array}{l}\sigma\leftarrow G(1^{\kappa});\\ (c,m_0,m_1,d_0,d_1)\leftarrow S(\sigma):\\ m_0\neq m_1\wedge\\ V(\sigma,c,d_0,m_0)=V(\sigma,c,d_1,m_1)=1\end{array}\right]=0 \tag{2}$$

(4)隐藏性: 对于任意敌手 $\mathcal{A}$, 存在一个可忽略函数 ν, 对于任意满足 $|m_0|=|m_1|$ 的 m_0,m_1 有

$$\Pr\left[\begin{array}{l}\sigma\leftarrow G(1^{\kappa});b\leftarrow\{0,1\};\\ (c,d)\leftarrow C(\sigma,m_b):\\ b\leftarrow\mathcal{A}(c)\end{array}\right]<\frac{1}{2}+\nu(\kappa) \tag{3}$$

2.2 非交互零知识

零知识证明(Zero-knowledge poofs, ZKP)由 Goldwasser, Micali 和 Rackoff 最早提出[9], 其允许证明者可以向验证者证明某个声明的正确性, 并且验证者无法获取除声明正确性以外的其他任何信息。随后 Blum 等人[10]提出了非交互零知识(Non- interactive zero-knowledge, NIZK)的概念, 使证明者与验证者可以利用公共参考串(Common reference string, CRS)用以代替证明阶段的交互行为, 大幅度提高了零知识证明系统的适用性。本文中非交互零知识的定义取自于文献[8-11]。

定义 2. (自适应多理论)计算零知识. 设 $\langle K,P,V\rangle$ 是多项式时间可计算二元关系 R 的证明系统, 称 $\langle K,P,V\rangle$ 是非交互零知识的, 如果存在一个模拟器 $S=(S_1,S_2)$, 对于所有的非均匀多项式时间敌手 $\mathcal{A}$ 有

$$\begin{aligned}&\Pr\left[\sigma\leftarrow K(1^{\kappa}):\mathcal{A}^{P(\sigma,\cdot,\cdot)}=1\right]\\ &\approx\Pr\left[(\sigma,\tau)\leftarrow S_1(1^{\kappa}):\mathcal{A}^{S'(\sigma,\tau,\cdot,\cdot)}=1\right]\end{aligned} \tag{4}$$

其中 S' 满足

$$S'(\sigma,\tau,x,w)=\begin{cases}S_2(\sigma,\tau,x), & (x,w)\in R\\ \perp, & (x,w)\notin R\end{cases} \tag{5}$$

2.3 双线性映射

定义 3. 假设 p 为素数, $\mathbb{G}_1,\mathbb{G}_2$ 为两个 p 阶乘法循环群 e 是一个双线性映射 $e:\mathbb{G}_1\times\mathbb{G}_1\rightarrow\mathbb{G}_2$ 如果其满足:

(1)双线性: 对于所有 $P,Q\in\mathbb{G}_1$ 与 $a,b\in\mathbb{Z}_p$, 有 $e(aP,bQ)=e(P,Q)^{ab}$。

(2)非退化性: 存在 $P,Q\in\mathbb{G}_1$, 使得 $e(P,Q)$ 不是 $\mathbb{G}_2$ 的单位元。

(3)可计算性: 存在一个有效算法, 对于任意的 $P,Q\in\mathbb{G}_1$ 都可以计算 $e(P,Q)$ 的值。

2.4 子群判定性问题

子群判断性问题(The subgroup decision problem)是一类计算困难性假设。

首先定义一个生成算法 $\mathcal{G}$, 在输入安全参数 κ 后, 输出 $(p,q,\mathbb{G},\mathbb{G}_1,e)$ 保证: (1) p,q 是素数; (2) $\mathbb{G},\mathbb{G}_1$ 是两个 n 阶循环群; (3) e 是一个双线性映射 $e:\mathbb{G}\times\mathbb{G}\rightarrow\mathbb{G}_1$。然后令 $\mathbb{G}_q$ 为 $\mathbb{G}$ 的 q 阶子群, 而子群判定性问题就是对 $\mathbb{G}$ 的群元素与 $\mathbb{G}_q$ 的群元素进行区分。

定义 4. 如果生成算法$\mathcal{G}$满足子群判定性困难假设, 那么存在一个可忽略函数$\nu_{SD}:\square\to[0,1]$, 对于任意非均匀多项式时间敌手$\mathcal{A}$都有

$$\Pr\left[\begin{array}{l}(p,q,\mathbb{G},\mathbb{G}_1,e)\leftarrow\mathcal{G}(1^\kappa);n=pq;\\ g,h\leftarrow\mathbb{G}_{gen}:\mathcal{A}(n,\mathbb{G},\mathbb{G}_1,e,g,h)=1\end{array}\right]$$
$$-\Pr\left[\begin{array}{l}(p,q,\mathbb{G},\mathbb{G}_1,e)\leftarrow\mathcal{G}(1^\kappa);n=pq;\\ g\leftarrow\mathbb{G}_{gen},h\leftarrow\mathbb{G}_q\setminus\{1\}:\\ \mathcal{A}(n,\mathbb{G},\mathbb{G}_1,e,g,h)=1\end{array}\right]$$
$$<\nu_{SD}(k) \tag{6}$$

其中$\mathbb{G}_{gen}$为$\mathbb{G}$的有限生成群。

2.5 Boneh-Goh-Nissim 加密系统

Boneh 等人[12]在 2005 年提出了能支持加法同态运算与一次乘法同态运算的 Boneh-Goh-Nissim 加密系统(BGN), 它由如下三个算法组成:

(1)密钥生成算法: 输入安全参数κ, 运行$\mathcal{G}(1^\kappa)$得到$(p,q,\mathbb{G},\mathbb{G}_1,e)$, 计算$n=pq$并取出$\mathbb{G}$中的随机生成元$g$与$\mathbb{G}_q$中的随机生成元$h$, 输出公钥$PK=(n,\mathbb{G},\mathbb{G}_1,e,g,h)$与私钥$SK=q$。

(2)加密算法: 输入明文m与公钥PK, 选取随机数$r\leftarrow\mathbb{Z}_n^*$, 计算并输出密文$c=g^m h^r$。

(3)解密算法: 输入密文c与公钥SK, 计算$c^q=(g^m h^r)^q=(g^q)^m$, 使用 Pollard's lambda 算法计算以g^q为底的离散对数, 得到并输出明文m。

2.6 BGN 密文的 01 内容 NIZK 证明

Gorth 等[8]构造了一个零知识证明方案(具有完美完备性, 完美可靠性以及多理论计算零知识), 用于证明一个 BGN 密文的对应明文是否属于 0 或 1, 该方案步骤如下:

2.6.1 初始化阶段

对于安全参数κ, 数据拥有者分别执行如下操作:

(1)运行$\mathcal{G}(1^\kappa)$得到$(p,q,\mathbb{G},\mathbb{G}_1,e)$。

(2)计算$n=pq$。

(3)取出$\mathbb{G}$中的一个随机生成元g。

(4)取出$\mathbb{G}_q$中的一个随机生成元h。

(5)返回公共参考串$\sigma=(n,\mathbb{G},\mathbb{G}_1,e,g,h)$。

2.6.2 声明阶段

证明者与验证者相互确认验证和公共参考串σ与 BGN 密文$c\in\mathbb{G}$。证明者声明存在$(m,w)\in\mathbb{Z}^2$满足$m\in\{0,1\}$且$c=g^m h^r$。

2.6.3 证明阶段

对于公共参考串σ, BGN 密文c与满足要求的(m,w), 证明者分别执行如下操作:

(1)选取随机数$r_i\leftarrow\mathbb{Z}_n^*$。

(2)分别计算$\pi_1=h^r$, $\pi_2=\left(g^{2m-1}h^w\right)^{wr^{-1}}$和$\pi_3=g^r$。

(3)发送证明$\pi=(\pi_1,\pi_2,\pi_3)$。

2.6.4 验证阶段

对于公共参考串σ, BGN密文c与证明π, 验证者分别执行如下操作:

(1)检查是否$c\in\mathbb{G}$且$\pi\in\mathbb{G}^3$;

(2)检查是否$e(c,cg^{-1})=e(\pi_1,\pi_2)$;

(3)检查是否$e(\pi_1,g)=e(h,\pi_3)$;

(4)如果检查都通过则输出 1, 否则输出 0。

3 可证明数据加密策略

本章提出了可证明数据加密策略的全新概念, 以应对隐私数据验证场景下的相关问题。全章分为三部分: 首先对隐私数据验证场景和(知识的)零知识证明场景进行区分, 以表明新概念提出的必要性, 然后提供可证明数据加密策略的形式化定义, 最后通过对应用场景的描述来展示可证明数据加密策略的实用价值。

3.1 隐私数据验证场景与零知识证明场景

隐私数据验证场景可视为一类特殊的(知识的)零知识证明场景——其需要在特定时间节点内, 提前对等待证明的秘密进行完美绑定承诺, 证明者在后续证明中, 需要同时使用该承诺进行证明。对比(知识的)零知识证明场景, 隐私数据验证场景具有以下两个特点:

1) 独特验证目标: 隐私数据验证场景的验证目标为“提前承诺的秘密是否满足声明”, 而(知识的)零知识证明场景的验证目标为“证明者是否拥有满足声明的秘密”。

2) 复数证明场景: 在隐私数据验证场景下, 允许证明者使用同一组承诺(同一个秘密)进行多次不同声明的证明, 而(知识的)零知识证明场景并不考虑复数证明场景之间的关联性。

基于上述两点, 普通(NP 语言)零知识证明方案

在隐私数据验证场景中难以直接使用:

1) 方案本身不具有对秘密进行提前承诺的功能, 或其承诺不具备完美绑定性, 使得证明者在证明阶段可以使用其他秘密进行欺骗证明;

2) 部分方案的承诺生成需要提前获取声明信息, 从而导致秘密的承诺在要求时间节点内无法生成, 或生成后不具有任意声明的可证明能力;

3) 使用同组承诺进行多次不同声明的证明, 可能会导致除声明正确性以外的秘密信息泄露, 因此需要对方案进行额外性质证明以保证其满足“整体的零知识”;

4) 在隐私数据验证场景中, 单次欺骗证明所导致的损失是无法估量的, 因此方案本身应具备完美可靠性。

3.2 形式化定义

本文基于零知识证明提出可证明数据加密策略概念, 其具有如下四个算法:

初始化算法(K **)**:输入安全参数 κ, 产生公共参考串 σ。

数据加密算法(E **)**:输入明文 w, 公共参考串 σ 以及加密密钥 μ, 输出密文 ϕ_w。

证明生成算法(P **)**:输入公共参考串 σ, 声明 x, 密文 ϕ_w, 明文 w 以及加密密钥 μ, 输出证明 π。

证明验证算法(V **):**输入公共参考串 σ, 声明 x, 密文 ϕ_w 以及证明 π, 如果 w 满足声明 x 且 π 是 (x,ϕ_w) 的对应证明则输出 1, 否则输出 0。

定义 5. 可证明数据加密策略. 设 $\langle K,P,V\rangle$ 是多项式时间可计算二元关系 R 的非交互证明系统, E 是一个完美绑定的承诺算法, 称 $\langle K,E,P,V\rangle$ 是多项式时间可计算二元关系的可证明数据加密策略, 如果满足:

- 有效性: K,E,P,V 都是多项式时间算法。
- 重用完备性: 对于所有的 $(w,x)\in R$,

$$\Pr\left[\begin{array}{l}\sigma\leftarrow K\left(1^{\kappa}\right),\phi_w\leftarrow E(\sigma,w;\mu);\\ \pi\leftarrow P(\sigma,x,\phi_w,w,\mu):\\ V(\sigma,x,\phi_w,\pi)=1\end{array}\right]=1 \tag{7}$$

- 重用可靠性: 对于所有的 $(w,x)\notin R$ 以及非均匀多项式时间敌手 $\mathcal{A}$,

$$\Pr\left[\begin{array}{l}\sigma\leftarrow K\left(1^{\kappa}\right),\phi_w\leftarrow E(\sigma,w;\mu);\\ \pi\leftarrow \mathcal{A}(\sigma,x,\phi_w):\\ V(\sigma,x,\phi_w,\pi)=1\end{array}\right]=0 \tag{8}$$

- 重用零知识: 存在模拟器 $S=(S_1,S_2)$, 对于所有的非均匀多项式时间敌手 $\mathcal{A}$ 以及多项式函数 q 有

$$\Pr\left[\begin{array}{l}\sigma\leftarrow K\left(1^{\kappa}\right),\\ \phi_w\leftarrow E(\sigma,w;\mu):\\ \mathcal{A}^{P(\sigma,\cdot,\phi_w,w,\mu)}=1\end{array}\right]\approx\Pr\left[\begin{array}{l}(\sigma,\tau)\leftarrow S_1\left(1^{\kappa}\right),\\ \phi_w\leftarrow E(\sigma,w;\mu):\\ \mathcal{A}^{S'(\sigma,\tau,\cdot,\phi_w,w,\mu)}=1\end{array}\right] \tag{9}$$

其中 S' 满足

$$S'(\sigma,\tau,x,\phi_w,w,\mu)=\begin{cases}S_2(\sigma,\tau,x,\phi_w), & (x,w)\in R\\ \perp, & (x,w)\notin R\end{cases} \tag{10}$$

该定义下的方案解决了隐私数据验证场景下的五个必要问题: 明文数据使用数据加密(承诺)算法进行独立加密(密文生成问题); 并借助其完美绑定性以防止二义性(密文二义性问题); 重用完备性保证密文的可证明能力(密文可证明问题); 重用可靠性确保密文的可靠性(密文可靠性问题); 重用零知识针对多次证明下的隐私泄露问题(多次证明的隐私泄露问题)。

3.3 应用场景——可证明数据库

随着医疗数据信息化工作的展开, 公民医疗活动的便捷性正在逐步提升, 通过在第三方数据库整合并存储个人医疗数据, 可以实现患者远程医疗、快速获取医疗数据信息等便捷功能, 大幅度减少了患者的时间成本和经济成本。然而, 医疗信息化也带来了前所未有的隐私泄露风险, 因过失、技术薄弱或恶意行为所导致的医疗数据泄露事件不断频发。

本文通过对可证明数据加密策略与非关系型数据库进行结合, 提出可证明医疗数据库的概念。可证明医疗数据库是一类由可证明密文所构成的非关系型数据库, 负责存储公民医疗隐私数据。作为可证明数据加密策略的一类特殊应用, 其证明方由两部分构成: (1)多个可信的数据生成方(如官方认证的医疗职能部门)拥有部分医疗隐私数据的生成、更新、整合能力, 以确保对应数据(承诺秘密)的准确性; (2)数据拥有方(公民)作为实际证明者, 其持有全部加密相关信息(明文、加密密钥、承诺使用随机值等), 并具备自身医疗隐私数据的获取、解密、证明能力。

可证明医疗数据库使得用户可以在任意场合获取自己的可信医疗隐私数据, 并且能向验证方提供任意相关声明的证明, 同时在证明过程中将隐私信息泄露问题最小化(只提供声明正确性的相关信息), 如:

1) 药物开具/购买: 医生/药师需要得到患者是否满足特定药物使用条件的结论, 因此患者可能需

要提供过敏史、相关病史、近期服药史以及相关生理数据等隐私信息。通过使用可证明医疗数据库, 双方可获取目标药物的使用条件声明与患者对应医疗数据, 然后由患者提供相应证明。这使得患者即能证明自身达到药物使用条件, 又避免过多暴露医疗隐私信息。

2) 遗传学研究: 公民愿意向非可信研究机构提供其特定遗传学分析结果, 却不愿透露具体遗传信息。通过使用可证明医疗数据库, 双方确定分析声明与患者遗传信息数据, 然后由公民提供相应证明。这使得公民可以在主动提供个人可信遗传分析结果的同时防止其遗传信息泄露。

类似的应用方法可以推广到更多场景, 譬如学历信息验证、出行记录验证、商业信息验证、个人档案验证等。在具体场景下, 可证明数据库通过结合其他隐私保护技术以达到更高安全性和实用性需求, 如通过访问控制技术以限制数据获取、通过签名技术以增加数据可信度、利用全同态加密增加数据的灵活性、使用可恢复性证明以保证数据完整性等。

4 可证明数据加密策略的设计与原型实现

本章对 Gorth 等[8]提出的 NP 语言的非交互零知识论证系统 Ω_{CS} 进行相应的改进, 构造出首个可证明数据加密方案 Ω_{ori}。Ω_{ori} 仍属于零知识证明系统, 并继承了 Ω_{CS} 的完美正确性, 完美可靠性和计算零知识, 且支持任意 NP 语言。

本章分为四个部分, 第一部分对 Ω_{ori} 方案进行描述; 第二部分对 Ω_{ori} 方案的相关性质进行证明; 第三部分考虑隐私数据验证场景下的非必要问题——数据更新效率问题, 并对其进行相关介绍与讨论; 第四部分提供 Ω_{ori} 的效率分析结论。

4.1 可证明数据加密策略原型 Ω_{ori}

Ω_{ori} 具有直观的改进方法——将 Ω_{CS} 中部分证明(关于秘密 w 的部分)的生成步骤提前到声明生成阶段之前, 并将其视为可证明数据加密策略的数据加密阶段 E。Ω_{ori} 分为如下阶段:

4.1.1 初始化阶段

对于安全参数 κ, 执行如下操作(视为初始化算法 K_{ori}):

(1)运行 $\mathcal{G}\left(1^{\kappa}\right)$ 得到 $\left(p,q,\mathbb{G},\mathbb{G}_1,e\right)$;

(2)计算 $n=pq$;

(3)取出 $\mathbb{G}$ 中的一个随机生成元 g;

(4)取出 $\mathbb{G}_q$ 中的一个随机生成元 h;

(5)返回公共参考串 $\sigma=\left(n,\mathbb{G},\mathbb{G}_1,e,g,h\right)$。

4.1.2 数据加密阶段

对于原始数据 w 和公共参考串 σ, 数据拥有者执行如下操作(视为数据加密算法 E_{ori}):

(1)将 w 拆分为多个比特位 $w=\left(w_1,\cdots,w_{|w|}\right)$;

(2)对于每个比特 w_i, 选取随机数 $r_i\leftarrow\mathbb{Z}_n^*$ (令 $\mu_1=\left(r_1,\cdots,r_{|w|}\right)$)进行 BGN 加密, 即 $c_{w_i}=g^{w_i}h^{r_i}$;

(3)对于每个新密文 c_{w_i}, 做一个 01 内容的 NIZK 证明;

(4)发布加密数据 $\phi_w=\left(c_{w_1},\cdots,c_{w_{|w|}}\right)$ 与相关证明。

4.1.3 声明阶段

证明者与验证者相互确认验证和公共参考串 σ, 加密数据 ϕ_w 与声明电路 $x=Cir_{(R,x)}^{NAND}$, 证明者声明 ϕ_w 中的原始数据 w 满足 $Cir_{(R,x)}^{NAND}\left(w\right)=1$。

4.1.4 证明阶段

对于声明电路 $Cir_{(R,x)}^{NAND}$, 加密数据 ϕ_w, 原始数据 w, 加密密钥 μ_1 以及和公共参考串 σ, 证明者分别执行以下操作(视为证明算法 P_{ori}):

(1)将 w 作为输入, 计算出电路 $Cir_{(R,x)}^{NAND}$ 中每一个非输入输出线上的值 μ_i。

(2)对于每一个非输入线上的值 μ_i, 选取一个随机数 $r_i\leftarrow\mathbb{Z}_n^*$ (令 $\mu_2=\left(r_1,\cdots,r_{|u|}\right)$)进行 BGN 加密,即 $c_{u_i}=g^{u_i}h^{r_i}$。

(3)对于每一个新密文 c_{u_i}, 证明者做一个 01 内容的 NIZK 证明。

(4)用 g 来表示输出线上的密文, 保证其明文为 1。

(5)将 ϕ_w 放置进对应的输入线上, 至此 $Cir_{(R,x)}^{NAND}$ 上的每一条线都有一个对应的密文。

(6)对于每一个 NAND 电路门上的输入线密文 c_{i_1},c_{i_2} 和输出线密文 c_{i_3}, 证明者需要做一个有效性证明, 保证存在 w_{i_1},w_{i_2},w_{i_3} 与 r_{i_1},r_{i_2},r_{i_3} 满足 $c_{i_j}=g^{u_{i_j}}h^{r_{i_j}}$ 且 $w_{i_3}=\neg\left(w_{i_1}\wedge w_{i_2}\right)$。为此证明者可以对 $g^m h^r=c_{i_1}c_{i_2}c_{i_3}^2g^{-2}$ 做一个 01 内容的 NIZK 证明来代替。

(7)证明者发送证明 $\pi_{x,w}$，其中包含了所有非输入线密文 c_{u_i}，以及 $Cir_{(R,x)}^{NAND}$ 中所有电路线上密文的 01 内容的 NIZK 证明和所有 NAND 电路门有效性的 NIZK 证明。

4.1.5 验证阶段

对于电路 $Cir_{(R,x)}^{NAND}$，加密数据 ϕ_w 和证明 $\pi_{x,w}$，验证者分别执行以下操作(视为验证算法 V_{ori}):

(1) 将 ϕ_w 和 $\pi_{x,w}$ 的所有密文放入 $Cir_{(R,x)}^{NAND}$ 中对应的电路线上，并将所有的 NIZK 证明对应;

(2) 检查 $Cir_{(R,x)}^{NAND}$ 中所有线路都有对应的密文且输出线上的密文是 g;

(3) 检查 $Cir_{(R,x)}^{NAND}$ 中所有电路线上密文的 01 内容的 NIZK 证明;

(4) 检查 $Cir_{(R,x)}^{NAND}$ 中所有 NAND 电路门的有效性 NIZK 证明;

(5) 如果检查全部通过，则输出 1，否则输出 0。

4.2 Ω_{ori} 的性质证明

本节对 Ω_{ori} 的相关性质进行证明。

首先假设数据拥有者与证明者为同一方，即将数据加密算法 E_{ori} 与证明算法 P_{ori} 合并为新的证明算法 P_{ori}^*:

$$\left(\phi_w,\pi_{x,w}\right)\leftarrow P_{ori}^*\left(\sigma,x,w;\mu_1,\mu_2\right) \tag{11}$$

引理 1. Ω_{CS} 具有完美完备性，完美可靠性与(自适应多理论)计算零知识。

证明. 引理 1 在 Gorth 等[8]已给出相关证明，不再赘述。

引理 2. Ω_{ori} 具有完美完备性和完美可靠性。

证明. Ω_{ori} 与 Ω_{CS} 初始化过程与验证过程完全一致，只是将其中一部分证明的生成步骤提前到声明之前(因此其初始化过程与验证过程完全一致)，并不会影响操作结果，因此 Ω_{ori} 的证明者 P_{ori}^* 与 Ω_{CS} 的证明者 P_{CS} 会有着相同的行为，可以视为同一个算法，即对于任意的 $\pi=\left(\phi_w,\pi_{x,w}\right)$ 和 $\mu=\left(\mu_1,\mu_2\right)$ 有

$$K_{ori}\left(1^\kappa\right)=K_{CS}\left(1^\kappa\right) \tag{12}$$

$$P_{ori}^*\left(\sigma,x,w;\mu_1,\mu_2\right)=P_{CS}\left(\sigma,x,w;\mu\right) \tag{13}$$

$$V_{ori}\left(\sigma,x,\phi_w,\pi_{x,w}\right)=V_{CS}\left(\sigma,x,\pi\right) \tag{14}$$

这表明在相同参数下，Ω_{ori} 与 Ω_{CS} 的交互视图将会完全一致，所以 Ω_{ori} 会继承 Ω_{CS} 完美完备性和完美可靠性，否则可以将 Ω_{ori} 中的反例转换成 Ω_{CS} 中对应的情况，从而推翻引理 1。

引理 3. Ω_{ori} 是(自适应多理论)计算零知识的。

证明. 由引理 1 可知存在模拟器 $S=\left(S_1,S_2\right)$ 满足

$$\begin{aligned}&\Pr\left[\sigma\leftarrow K_{CS}\left(1^\kappa\right):\mathcal{A}^{P_{CS}(\sigma,\cdot,\cdot)}=1\right]\\&\approx\Pr\left[\sigma\leftarrow S_1\left(1^\kappa\right):\mathcal{A}^{S'(\sigma,\tau,\cdot,\cdot)}=1\right]\end{aligned} \tag{15}$$

其中 S' 满足

$$S'\left(\sigma,\tau,x,w\right)=\begin{cases}S_2\left(\sigma,\tau,x\right),\left(x,w\right)\in R\\\perp\quad,\left(x,w\right)\notin R\end{cases} \tag{16}$$

因为 P_{ori}^* 与 P_{CS} 有相同的行为，可以视为同一个算法，所以有

$$\begin{aligned}&\Pr\left[\sigma\leftarrow K_{ori}\left(1^\kappa\right):\mathcal{A}^{P_{ori}^*(\sigma,\cdot,\cdot)}=1\right]\\&=\Pr\left[\sigma\leftarrow K_{CS}\left(1^\kappa\right):\mathcal{A}^{P_{CS}(\sigma,\cdot,\cdot)}=1\right]\end{aligned} \tag{17}$$

结合公式(15)(17)得到

$$\begin{aligned}&\Pr\left[\sigma\leftarrow K_{ori}\left(1^\kappa\right):\mathcal{A}^{P_{ori}^*(\sigma,\cdot,\cdot)}=1\right]\\&\approx\Pr\left[\sigma\leftarrow S_1\left(1^\kappa\right):\mathcal{A}^{S'(\sigma,\tau,\cdot,\cdot)}=1\right]\end{aligned} \tag{18}$$

这意味着用于证明 Ω_{CS} 中的零知识模拟器 S 同样适用于 Ω_{ori} 中，所以 Ω_{ori} 是(自适应多理论)计算零知识的。

定理 1. Ω_{ori} 满足重用完备性与重用可靠性。

证明. 为通过证明，证明者提供了所有电路线上密文的 01 内容证明与所有电路门上的有效性证明，以保证密文中明文在电路中进行了正确地计算。如果 $\left(w,x\right)\in R$ ($Cir_{(R,x)}^{NAND}\left(w\right)=1$)，那么诚实的证明者(拥有 w 与 μ_1)可以使用 ϕ_w 在不违反上述证明的条件下，让输出电路的输出密文为 g。如果 $\left(w,x\right)\notin R$ ($Cir_{(R,x)}^{NAND}\left(w\right)=0$)，那么其输出一定不为 g，否则与上述证明矛盾。所以有，

$$\Pr\begin{bmatrix}\sigma\leftarrow K_{ori}\left(1^\kappa\right),\\\phi_w\leftarrow E_{ori}\left(\sigma,w;\mu_1\right);\\\pi\leftarrow P_{ori}\left(\sigma,x,\phi_w,w,\mu_1\right)\\:V_{ori}\left(\sigma,x,\phi_w,\pi\right)=1\end{bmatrix}=1 \tag{19}$$

$$\Pr\left[\begin{array}{l}\sigma \leftarrow K_{ori}\left(1^{\kappa}\right),\\ \phi_w \leftarrow E_{ori}\left(\sigma, w; \mu_1\right);\\ \pi \leftarrow \mathcal{A}\left(\sigma, x, \phi_w\right):\\ V_{ori}\left(\sigma, x, \phi_w, \pi\right)=1\end{array}\right]=0 \tag{20}$$

定理 2. Ω_{ori} 是重用零知识的。

证明. 由于 Ω_{CS} 是多理论零知识的, 根据 Gorth 等[8]的证明思路, 存在模拟器 $S=\left(S_1, S_2\right)$, 使得 S_2 模拟的证明与实际证明是不可区分的, 且其模拟的每一个 BGN 密文(电路线上密文 c_i)与 BGN 密文的 01 内容证明(c_i 的 01 内容证明 π_i^{wire} 以及各电路门的有效性证明 π_j^{gate})都与实际部分不可区分。

相较于 Ω_{CS}, 在同一个 ϕ_w 的多个证明场景下 Ω_{ori} 只是提前固定了输入线的密文与其对应 01 内容证明 π_{ϕ_w}, 在同时去掉这部分证明后, P_{ori} 的证明与 S_2 模拟的证明显然是不可区分(否则与 Ω_{CS} 的多理论零知识冲突)。

接下来构造模拟器 S_{ori_2}, 其模拟过程与 S_2 基本一致, 只是在模拟输入线密文与其对应的 01 内容时, 直接使用 ϕ_w 与 π_{ϕ_w} 替换。因为 S_{ori_2} 与 P_{ori} 生成的证明在输入线部分完全一致(全为 ϕ_w 与 π_{ϕ_w}), 非输入线部分不可区分(这部分 S_{ori_2} 与 S_2 是一致的), 所以它们是不可区分的, 于是 $S'_{ori}=\left(S_1, S_{ori_2}\right)$ 满足

$$\Pr\left[\begin{array}{l}\sigma \leftarrow K_{ori}\left(1^{\kappa}\right),\\ \phi_w \leftarrow E_{ori}\left(\sigma, w; \mu_1\right):\\ \mathcal{A}^{P_{ori}\left(\sigma, \cdot, \phi_w, w, \mu_1\right)}=1\end{array}\right]=\Pr\left[\begin{array}{l}(\sigma, \tau) \leftarrow S_1\left(1^{\kappa}\right),\\ \phi_w \leftarrow E_{ori}\left(\sigma, w; \mu_1\right):\\ \mathcal{A}^{S'_{ori}\left(\sigma, \tau, \cdot, \phi_w, w, \mu_1\right)}=1\end{array}\right] \tag{21}$$

其中 S'_{ori} 满足

$$S'_{ori}\left(\sigma, \tau, x, \phi_w, w, \mu\right)=\begin{cases}S_{ori_2}\left(\sigma, \tau, x, \phi_w\right), & (x, w) \in R\\ \perp, & (x, w) \notin R\end{cases} \tag{22}$$

4.3 数据更新

在隐私数据证明场景中, 高频率更新、低频率证明的情景普遍存在(如个人档案, 医疗数据, 学历数据等, 用户往往需要长期高频率地维护隐私信息以应对未来极少数的验证场景), 因此讨论并改进各方案下的数据更新效率是至关重要的。本节将讨论数据更新的种类, 并结合特定场景, 讨论数据更新的有效性验证方式。

4.3.1 数据更新操作与内容更新模式

隐私数据证明场景下的数据更新操作主要分为三种:

(1)内容更新操作: 例如生理数据更新, 学历数据更新, 消费记录更新等, 这些更新操作不改变整体数据的结构, 只对其中的单项数据进行修改。例如: 对个人档案中的学籍信息进行更新。

内容更新又分为三种更新模式: 全部更新, 即对数据进行全部替换; 单项更新, 即替换变化过的单项数据; 比特更新, 即只对数据中变化的比特位进行更新。虽然三种模式的更新开销有显著差距, 但在特定场景仍需选择适当的更新方式以应对潜在攻击, 如通过对比新旧数据获取数据更新项目信息, 或通过比特位变化推测数据变化范围。

(2)结构更新操作: 例如生理数据整合, 个人档案提取, 商业数据重组等, 这些数据更新操作不改变数据内容, 但需要对数据进行结构上重组。例如: 将病人的多份生理数据资料整合为一份, 或将商业数据资料按照新的格式进行重组。

(3)混合更新操作: 由上述两种更新操作组合而成。

表 1 Ω_{ori} 数据更新操作类型

Table 1 The Ω_{ori}'s operations of data modification

种类	操作	示例
结构更新	重组	$(w_1, w_2, w_3) \rightarrow (w_3, w_1, w_2)$
	拆分	$(w_1, w_2, w_3) \rightarrow (w_1, w_2)(w_3)$
	组合	$(\{w_1, w_2\})(\{w_3\}) \rightarrow (\{w_3\}, \{w_1, w_2\})$
内容更新	全部	$(w_1, w_2, w_3) \rightarrow (w_1^{\bullet}, w_2^{\bullet}, w_3^{\bullet})$
	单项	$(\{w_{1,1}, w_{1,2}\}, \{w_{2,1}\}) \rightarrow (\{w_{1,1}^{\bullet}, w_{1,2}^{\bullet}\}, \{w_{2,1}\})$
	比特	$(w_1, w_2, w_3) \rightarrow (w_1, w_2^{\bullet}, w_3)$

(注: • 表示内容更新后的数据, {} 表示单项数据, () 表示数据集合)

4.3.2 数据更新的有效性验证

在一些场景下, 数据更新方可能需要提供数据更新的有效性证明, 即保证加密数据按照实际要求进行更新(如正确地对某项资产数据, 按当前国际汇率进行货币单位换算, 或按约定增加/减少固定值), 而不是欺骗性更新。针对该情况, 证明者可使用可证明数据加密方案进行一次额外证明以达到等效证明结果——针对新旧隐私数据 ϕ_{new}、ϕ_{old}, 证明其承诺的内容 w_{new}、w_{old} 满足 $f\left(w_{old}\right)=w_{new}$。

4.4 效率分析

本小节只展示Ω_{ori}的效率分析结果, 其分为计算开销与通讯开销(由公共参考串规模, 证明规模与数据规模构成)两部分。其中表 2 为单次证明中Ω_{ori}各阶段的效率, 表 3 为ϕ_w在不同种类更新下的效率, 表 4 为多次证明与更新后Ω_{ori}的总效率。

表 2 Ω_{ori}各阶段效率

Table 2 The efficiency of Ω_{ori}'s each step

步骤	公共参考串规模	数据与证明规模	计算开销
初始化阶段	$O(\kappa)$	–	$O(\kappa)$
数据加密阶段	–	$O(\lvert w\rvert\cdot\kappa)$	$O(\lvert w\rvert\cdot\kappa)$
证明阶段	–	$O(\lvert Cir\rvert\cdot\kappa)$	$O(\lvert Cir\rvert\cdot\kappa)$
声明阶段	–	–	–
验证阶段	–	–	$O(\lvert Cir\rvert\cdot\kappa)$

(注: κ 表示安全参数, $\lvert w\rvert$ 表示数据 w 的比特位数, $\lvert Cir\rvert$ 表示声明电路 Cir 的电路门数)。

表 3 Ω_{ori}数据更新效率

Table 3 The efficiency of Ω_{ori}'s data update

更新种类		数据规模	计算开销
结构更新		–	–
内容更新	全部	$O(\lvert w\rvert\cdot\kappa)$	$O(\lvert w\rvert\cdot\kappa)$
	单项\比特	$O(m\cdot\kappa)$	$O(m\cdot\kappa)$

(注: κ 表示安全参数, $\lvert w\rvert$ 表示数据 w 的比特位数, m 表示更新比特的总数), 由于结构更新可以直接验证, 因此无需额外的证明与计算。

表 4 Ω_{ori}总效率

Table 4 The efficiency of Ω_{ori}

	Ω_{ori}
公共参考串规模	$O(\kappa)$
证明规模	$O\left(\sum\lvert Cir_i\rvert\cdot\kappa\right)$
证明者计算开销	$O\left(\left(\lvert w\rvert+m+\sum\lvert Cir_i\rvert\right)\cdot\kappa\right)$
验证者计算开销	$O\left(\sum\lvert Cir_i\rvert\cdot\kappa\right)$

(注: κ 表示安全参数, $\lvert w\rvert$ 表示数据 w 的比特位数, m 表示更新比特的总数, $\lvert Cir_i\rvert$ 表示每个声明电路(包括证明更新有效性的声明电路) Cir_i 的电路门数)

虽然Ω_{ori}具有高效的数据生成与更新效率, 并且支持单项/比特更新, 以及数据更新的有效性验证, 但其声明规模$\sum\lvert Cir_i\rvert$会导致其证明规模与计算规模十分庞大。为得到更高效的方案, 下一章将提出可证明数据加密策略的全新构造思路与框架。

5 可证明数据加密策略的两种构造框架

本节基于承诺方案、全同态加密与零知识证明提出两种可证明数据加密策略的通用构造框架并给予相应性质证明。与Ω_{ori}的构造思路不同, 两种构造框架皆采取将ϕ_w与旧声明x整合成新声明x_{ϕ_w}的方法, 将隐私数据验证场景退化为普通零知识证明场景。该方法导致了额外的计算与证明开销, 但其性质证明相对直观, 且能通过结合优异的加密组件(如文献[13-16])以获得更加高效与强大的可证明数据加密方案。

5.1 构造框架 1——承诺方案+零知识证明

第一种构造框架由完美绑定承诺方案$\langle Gen_1, Com, Ver_1\rangle$与具有完美可靠性的 NP 语言的非交互零知识证明系统$\langle Gen_2, Pro, Ver_2\rangle$构成(将构造出的可证明数据加密方案称为$\Omega_\alpha$)。该框架具有直观的构造与证明思路——通过将加密数据ϕ_w与原声明x的约束关系进行整合, 以此获得一个等价的新声明x_{ϕ_w}, 并通过证明x_{ϕ_w}来达到等价的证明效果。

由于该方法将隐私数验证场景退化为普通零知识证明场景, Ω_α的相关性质(重用完备性、重用可靠性、重用零知识)将由非交互零知识组件$\langle Gen_2, Pro, Ver_2\rangle$中的性质(完备性、可靠性、非交互零知识)继承而来。

5.1.1 过程描述

Ω_α的大致过程如下:

(1)对于安全参数κ, 数据拥有者使用初始化算法Gen_1与Gen_2分别生成公共参考串σ_1与σ_2。

(2)数据拥有者使用公共参考串σ_1, 随机数组$\mu=\left(r_{w_1},\cdots,r_{w_{|w|}}\right)$与承诺算法$Com$对原始数据$w$的各比特位进行加密(承诺), 得到加密数据(承诺集合)$\phi_w=\left(c_{w_1},\cdots,c_{w_{|w|}}\right)$。

(3)对于公共参考串σ_1, 原声明x和加密数据ϕ_w, 证明者与验证者构造应对的新声明x_{ϕ_w}, 使得存在$w_{\phi_w}=w\,|\,\mu$满足$\left(w_{\phi_w}, x_{\phi_w}\right)\in R$当且仅当$\phi_w=Com\left(w;\mu_1\right)$和$(w,x)\in R$。视该构造过程为算法$Tra$, 即$x_{\phi_w}\leftarrow Tra\left(\sigma,x,\phi_w\right)$。

(4)证明者与验证者使用证明算法 *Pro* 与验证算法 Ver_2 对新声明 x_{ϕ_w} 进行零知识证明, 即证明存在 w_{ϕ_w} 满足 $\left(w_{\phi_w}, x_{\phi_w}\right) \in R$。

5.1.2 策略描述

根据过程描述, Ω_α 可以总结为四个算法:

初始化算法(K_α **)**:输入安全参数 κ, 运行 Gen_1 与 Gen_2 产生并输出公共参考串 $\sigma=\left(\sigma_1,\sigma_2\right)$。

数据加密算法(E_α **)**:输入明文 w, 公共参考串 σ 以及加密密钥 μ, 运行 *Com* 产生并输出密文 ϕ_w。

证明生成算法(P_α **)**:输入公共参考串 σ, 声明 x, 密文 ϕ_w, 明文 w, 加密密钥 μ, 先运行 *Tra* 产生 x_{ϕ_w}, 然后运行 *Pro* 产生并输出证明 π。

证明验证算法(V_α **):**输入公共参考串 σ, 声明 x, 密文 ϕ_w 以及证明 π, 先运行 *Tra* 产生 x_{ϕ_w}, 然后运行 Ver_2 产生并验证结果。

5.1.3 性质证明

本小节对 Ω_α 的性质进行证明。

定理 3. 如果 $\langle Gen_2, Pro, Ver_2\rangle$ 具有完美完备性与完美可靠性, 那么 $\langle K_\alpha, E_\alpha, P_\alpha, V_\alpha\rangle$ 满足重用完备性与重用可靠性。

证明. 根据 Ω_α 的算法描述有

$$P_\alpha\left(\sigma,x,\phi_w,w,\mu\right)=Pro\left(\sigma,x_{\phi_w},w_{\phi_w}\right) \tag{23}$$

$$V_\alpha\left(\sigma,x,\phi_w,\pi\right)=Ver_2\left(\sigma,x_{\phi_w},\pi\right) \tag{24}$$

由于 ϕ_w 在整合进声明 x_{ϕ_w} 中只使用了公共参考串 σ_2, 结合公式(24)以及 $\langle Gen_2, Pro, Ver_2\rangle$ 的完美完备性有

$$\begin{aligned}
&\Pr\left[\begin{array}{l}\sigma\leftarrow K_\alpha\left(1^\kappa\right),\\ \phi_w\leftarrow E_\alpha\left(\sigma,w;\mu\right);\\ \pi\leftarrow P_\alpha\left(\sigma,x,\phi_w,w,\mu\right):\\ V_\alpha\left(\sigma,x,\phi_w,\pi\right)=1\end{array}\right]\\
&=\Pr\left[\begin{array}{l}\sigma_2\leftarrow Gen_2\left(1^\kappa\right);\\ \pi\leftarrow Pro\left(\sigma_2,x_{\phi_w},w_{\phi_w}\right):\\ Ver_2\left(\sigma_2,x_{\phi_w},\pi\right)=1\end{array}\right]\\
&=1
\end{aligned} \tag{25}$$

同理有

$$\begin{aligned}
&\Pr\left[\begin{array}{l}\sigma\leftarrow K_\alpha\left(1^\kappa\right),\\ \phi_w\leftarrow E_\alpha\left(\sigma,w;\mu\right);\\ \pi\leftarrow \mathcal{A}_\alpha\left(\sigma,x,\phi_w\right):\\ V_\alpha\left(\sigma,x,\phi_w,\pi\right)=1\end{array}\right]\\
&=\Pr\left[\begin{array}{l}\sigma_2\leftarrow Gen_2\left(1^\kappa\right),\\ \pi\leftarrow \mathcal{A}\left(\sigma,x_{\phi_w}\right):\\ Ver_2\left(\sigma,x_{\phi_w},\pi\right)=1\end{array}\right]\\
&=0
\end{aligned} \tag{26}$$

因此 $\langle K_\alpha, E_\alpha, P_\alpha, V_\alpha\rangle$ 满足重用完备性与重用可靠性。

定理 4. $\langle K_\alpha, E_\alpha, P_\alpha, V_\alpha\rangle$ 是重用零知识的。

证明. 由于 $\langle Gen_2, Pro, Ver_2\rangle$ 是非交互零知识的, 所以存在模拟器 $S=\left(S_1,S_2\right)$ 使得

$$\Pr\left[\begin{array}{l}\sigma_2\leftarrow Gen_2\left(1^\kappa\right):\\ \mathcal{A}^{Pro\left(\sigma_2,\cdot,\cdot\right)}=1\end{array}\right]\approx\Pr\left[\begin{array}{l}\sigma_2\leftarrow S_1\left(1^\kappa\right):\\ \mathcal{A}^{S'\left(\sigma_2,\tau,\cdot,\cdot\right)}=1\end{array}\right] \tag{27}$$

其中 S' 满足

$$S'\left(\sigma_2,\tau,x,w\right)=\begin{cases}S_2\left(\sigma_2,\tau,x\right), & \left(x,w\right)\in R\\ \perp, & \left(x,w\right)\notin R\end{cases} \tag{28}$$

假设模拟器 S_α' 与 S_{α_2} 满足

$$S'_\alpha\left(\sigma_2,\tau,x,\phi_w,w,\mu\right)=S'\left(\sigma_2,\tau,x_{\phi_w},w_{\phi_w}\right) \tag{29}$$

$$S_{\alpha_2}\left(\sigma_2,\tau,x,\phi_w\right)=S_2\left(\sigma_2,\tau,x_{\phi_w}\right) \tag{30}$$

结合公式(24)(28)(29)(30)(31)有

$$\begin{aligned}
&\Pr\left[\begin{array}{l}\sigma\leftarrow K_\alpha\left(1^\kappa\right),\\ \phi_w\leftarrow E_\alpha\left(\sigma,w;\mu\right):\\ \mathcal{A}^{P_\alpha\left(\sigma,\cdot,\phi_w,w,\mu\right)}=1\end{array}\right]\\
&=\Pr\left[\begin{array}{l}\sigma_2\leftarrow Gen_2\left(1^\kappa\right):\\ \mathcal{A}^{Pro\left(\sigma_2,Tra\left(\sigma_2,\cdot,\phi_w\right),w_{\phi_w}\right)}=1\end{array}\right]\\
&\approx\Pr\left[\begin{array}{l}\left(\sigma_2,\tau\right)\leftarrow S_1\left(1^\kappa\right):\\ \mathcal{A}^{S'\left(\sigma_2,\tau,Tra\left(\sigma_2,\cdot,\phi_w\right),w_{\phi_w}\right)}=1\end{array}\right]\\
&=\Pr\left[\begin{array}{l}\left(\sigma_2,\tau\right)\leftarrow S_1\left(1^\kappa\right),\\ \phi_w\leftarrow E_\alpha\left(\sigma,w;\mu\right):\\ \mathcal{A}^{S'_\alpha\left(\sigma_2,\tau,\cdot,\phi_w,w,\mu\right)}=1\end{array}\right]
\end{aligned} \tag{31}$$

其中 S'_α 满足

$$S'_\alpha(\sigma_2,\tau,x,\phi_w,w,\mu)=\begin{cases}S_{\alpha_2}(\sigma_2,\tau,x,\phi_w),(x,w)\in R\\ \perp\quad,(x,w)\notin R\end{cases}\tag{32}$$

因此$\langle K_\alpha,E_\alpha,P_\alpha,V_\alpha\rangle$是重用零知识的。

5.1.4 分析结论

Ω_α 有两个优势: 首先对构造组件没有过多限制, 因此可进行自由选择并组合, 即使中途替换零知识组件也不影响 ϕ_w 的可证明能力; 其次是 Ω_α 的性质已经被证明, 无需在组合后进行额外的性质证明。

Ω_α 对声明的改动造成了其证明规模的增长, 从而导致额外的通讯与计算开销——如果将 Ω_{ori} 中的 BGN 加密系统(承诺方案组件)与 Ω_{CS} (零知识证明组件)按该构造框架直接组合, 其声明规模相较 Ω_{ori} 会增加 $O(|w|\cdot|Cir_{BGN}|)$(Cir_{BGN} 为 BGN 密文内容为 0 或 1 的单比特验证电路, 输入为明文的 w_i 与对应随机数 r_i)。

5.2 构造框架 2——全同态加密+零知识证明

受 Gentry[17-18]与 Gorth 等人[8]的启发, 第二个构造框架由具有完美绑定性密文的全同态加密方案 $\langle Gen_1,Enc,Dec,Eval\rangle$ 与证明其密文内容为 0 或 1 的非交互零知识证明系统 $\langle Gen_2,Pro,Ver\rangle$ 组合构成(将构造出的可证明数据加密方案称为 Ω_β)。Ω_β 同样具有直观的构造与证明思路——将声明看作计算电路 Cir, 通过计算 w 的全同态密文 $\phi_w=(c_{w_1},\cdots,c_{w_{|w|}})$ 以获得 $Cir(w)$ 的全同态密文 $c_{Cir(w)}$, 并通过证明 $c_{Cir(w)}$ 的明文为 1 来达到等价的证明效果。

由于该方法将隐私数验证场景退化为普通零知识证明场景, 因此可以使用与 Ω_α 类似的证明方式对 Ω_β 的相关性质进行证明(将 $Eval$ 视为 Ω_α 中的构造算法 Tra)。

5.2.1 过程描述

Ω_β 的大致过程如下:

(1)对于安全参数 κ, 数据拥有者使用初始化算法 Gen_1 与 Gen_2 分别生成公钥 pk, 私钥 sk 与公共参考串 σ_2。

(2)数据拥有者使用加密算法 Enc 与加密密钥 $\mu=(sk,r_{w_1},\cdots,r_{w_{|w|}})$ (r_i 为噪音)对原始数据 w 的各比特位进行全同态加密, 得到加密数据(全同态密文集合) $\phi_w=(c_{w_1},\cdots,c_{w_{|w|}})$。

(3)对于公钥 pk, 原声明 $x=Cir$ 和加密数据 ϕ_w, 证明者与验证者同时运行密文计算算法 $Eval$ 得到全同态密文 $c_{Cir(w)}$。

(4)证明者与验证者使用证明算法 Pro 与验证算法 Ver 对全同态密文 $c_{Cir(w)}$ 进行零知识证明, 即证明者使用秘密 $w_{\phi_w}=w|\mu$ 证明声明 $x_{\phi_w}=c_{Cir(w)}$ 满足 $(x_{\phi_w},w_{\phi_w})\in R$。

5.2.2 策略描述

根据上述过程描述, Ω_β 可以总结为四个算法:

初始化算法(K_β):输入安全参数 κ, 运行 Gen_1 与 Gen_2 产生并输出公共参考串 $\sigma=(pk,\sigma_2)$ 与私钥 sk 。

数据加密算法(E_β):输入明文 w, 公共参考串 σ 以及加密密钥 μ, 运行 Enc 产生并输出密文 ϕ_w。

证明生成算法(P_α):输入公共参考串 σ, 声明 $x=Cir$, 密文 ϕ_w, 明文 w, 加密密钥 μ, 先运行 $Eval$ 产生 x_{ϕ_w}, 然后运行 Pro 产生并输出证明 π 。

证明验证算法(V_α):输入公共参考串 σ, 声明 $x=Cir$, 密文 ϕ_w 以及证明 π, 先运行 $Eval$ 产生 x_{ϕ_w}, 然后运行 Ver 产生并验证结果。

5.2.3 性质证明

Ω_β 具有与 Ω_α 相似的证明思路。

定理 5. 如果 $\langle Gen_2,Pro,Ver\rangle$ 具有完美完备性与完美可靠性, 那么 $\langle K_\beta,E_\beta,P_\beta,V_\beta\rangle$ 是数据可重用的。

证明. 根据 Ω_β 的算法描述有

$$P_\beta(\sigma,x,\phi_w,w,\mu)=Pro(\sigma,x_{\phi_w},w_{\phi_w})\tag{33}$$

$$V_\beta(\sigma,x,\phi_w,\pi)=Ver(\sigma,x_{\phi_w},\pi)\tag{34}$$

由于证明过程中只使用了公共参考串 σ, 结合公式(24)以及 $\langle Gen_2,Pro,Ver\rangle$ 的完美完备性有

$$\Pr\left[\begin{array}{l}\sigma\leftarrow K_\beta(1^\kappa),\\ \phi_w\leftarrow E_\beta(\sigma,w;\mu);\\ \pi\leftarrow P_\beta(\sigma,x,\phi_w,w,\mu):\\ V_\beta(\sigma,x,\phi_w,\pi)=1\end{array}\right]$$

$$=\Pr\left[\begin{array}{l}\sigma_2\leftarrow Gen_2(1^\kappa);\\ \pi\leftarrow Pro(\sigma_2,x_{\phi_w},w_{\phi_w}):\\ Ver(\sigma_2,x_{\phi_w},\pi)=1\end{array}\right]$$

$$=1\tag{35}$$

同理有

$$\Pr\begin{bmatrix}\sigma\leftarrow K_\beta\left(1^\kappa\right),\\ \phi_w\leftarrow E_\beta\left(\sigma,w;\mu\right);\\ \pi\leftarrow \mathcal{A}_\beta\left(\sigma,x,\phi_w\right):\\ V_\beta\left(\sigma,x,\phi_w,\pi\right)=1\end{bmatrix}$$

$$=\Pr\begin{bmatrix}\sigma_2\leftarrow Gen_2\left(1^\kappa\right),\\ \pi\leftarrow \mathcal{A}\left(\sigma,x_{\phi_w}\right):\\ Ver\left(\sigma,x_{\phi_w},\pi\right)=1\end{bmatrix}$$

$$=0 \tag{36}$$

因此$\left\langle K_\beta,E_\beta,P_\beta,V_\beta\right\rangle$满足重用完备性与重用可靠性。

定理 6. $\left\langle K_\beta,E_\beta,P_\beta,V_\beta\right\rangle$是重用零知识的。

证明. 由于$\left\langle Gen_2,Pro,Ver\right\rangle$是非交互零知识的, 所以存在模拟器$S=\left(S_1,S_2\right)$使得

$$\Pr\begin{bmatrix}\sigma_2\leftarrow Gen_2\left(1^\kappa\right):\\ \mathcal{A}^{Pro(\sigma_2,\cdot,\cdot)}=1\end{bmatrix}\approx\Pr\begin{bmatrix}\sigma_2\leftarrow S_1\left(1^\kappa\right):\\ \mathcal{A}^{S'(\sigma_2,\tau,\cdot,\cdot)}=1\end{bmatrix} \tag{37}$$

其中S'满足

$$S'\left(\sigma_2,\tau,x,w\right)=\begin{cases}S_2\left(\sigma_2,\tau,x\right),\left(x,w\right)\in R\\ \perp\quad,\left(x,w\right)\notin R\end{cases} \tag{38}$$

假设模拟器S'_β满足

$$S'_\beta\left(\sigma_2,\tau,x,\phi_w,w,\mu\right)=S'\left(\sigma_2,\tau,x_{\phi_w},w_{\phi_w}\right) \tag{39}$$

$$S_{\beta_2}\left(\sigma_2,\tau,x,\phi_w\right)=S_2\left(\sigma_2,\tau,x_{\phi_w}\right) \tag{40}$$

结合公式(34)(38)～(41)有

$$\Pr\begin{bmatrix}\sigma\leftarrow K_\beta\left(1^\kappa\right),\\ \phi_w\leftarrow E_\beta\left(\sigma,w;\mu\right):\\ \mathcal{A}^{P_\beta(\sigma,\cdot,\phi_w,w,\mu)}=1\end{bmatrix}$$

$$=\Pr\begin{bmatrix}\sigma_2\leftarrow Gen_2\left(1^\kappa\right):\\ \mathcal{A}^{Pro\left(\sigma_2,Eval\left(\phi_w,\cdot\right),w_{\phi_w}\right)}=1\end{bmatrix}$$

$$\approx\Pr\begin{bmatrix}\left(\sigma_2,\tau\right)\leftarrow S_1\left(1^\kappa\right):\\ \mathcal{A}^{S'\left(\sigma_2,\tau,Eval\left(\phi_w,\cdot\right),w_{\phi_w}\right)}=1\end{bmatrix}$$

$$=\Pr\begin{bmatrix}\left(\sigma_2,\tau\right)\leftarrow S_1\left(1^\kappa\right),\\ \phi_w\leftarrow E_\beta\left(\sigma,w;\mu\right):\\ \mathcal{A}^{S'_\beta\left(\sigma_2,\tau,\cdot,\phi_w,w,\mu\right)}=1\end{bmatrix} \tag{41}$$

其中S'_β满足

$$S'_\beta\left(\sigma_2,\tau,x,\phi_w,w,\mu\right)$$

$$=\begin{cases}S_{\beta_2}\left(\sigma_2,\tau,x,\phi_w\right),\left(x,w\right)\in R\\ \perp\quad,\left(x,w\right)\notin R\end{cases} \tag{42}$$

因此$\left\langle K_\beta,E_\beta,P_\beta,V_\beta\right\rangle$是重用零知识的。

5.2.4 分析结论

Ω_β有三个优势: 首先因为Ω_β的新声明为单个全同态密文, 所以其具有更小且稳定的声明与证明规模; 其次零知识证明组件并不需要支持 NP 语言; 最后由于使用全同态加密组件, 其密文可以支持全同态计算。

另外, 由于Ω_β需要在声明阶段进行至少$\sum\left|Cir_i\right|$个电路门的全同态计算, 其计算效率会受到全同态加密组件全同态计算效率影响。

6 总结

本文针对隐私数据验证场景下的隐私泄露问题, 首先提出了可证明数据加密策略的概念, 讨论其性质并给出了形式化定义, 然后基于 Gorth 等[8]的方案构造出首个可证明数据加密策略原型方案Ω_{ori}, 最后本文提出实现高效可证明数据加密策略的两种构造框架并给出相应性质证明, 利用两种框架构造出更加高效的方案是我们下一步的研究重点。

参考文献

[1] Silvio Micali, Michael Rabin, Joseph Kilian. Updatable zero-knowledge sets[C]. *The IEEE Symposium on Foundations of Computer Science*, 2003:80-91.

[2] Moses Liskov. Updatable Zero-Knowledge Databases[C]. *ASIACRYPT*, 2005:174-198.

[3] Benoît Libert, Khoa Nguyen, Benjamin Hong, et al. Zero-Knowledge Elementary Databases with More Expressive Queries[C]. *Public Key Cryptography*, 2019:255-285.

[4] Zhang Yupeng, J. Katz, C. Papamanthou. An Expressive (Zero-Knowledge) Set Accumulator[C]. *2017 IEEE European Symposium on Security and Privacy*, 2017:158-173.

[5] Chaabouni R, Lipmaa H, Zhang B. A non-interactive range proof

with constant communication[C]. *Financial Cryptograph*, 2012:179-199.

[6] Zhu Yan, Wang Huaixi, Hu Zexing, et al. Zero-knowledge proof of data retrievability[J]. *Science China Information Sciences*, 2012. (朱岩, 王怀习, 胡泽行, 等. 数据可恢复性的零知识证明[J]. *中国科学:信息科学*, 2012.)

[7] Joe Kilian. A note on e□cient zero-knowledge proofs and arguments[C]. *Symposium on Theory of Computing,* 1992: 723-732.

[8] Groth J, Ostrovsky R, Sahai A. Perfect Non-interactive Zero Knowledge for NP[M]. *Theory and Application of Cryptographic Techniques*, 2006: 339-358.

[9] S Goldwasser, S Micali, C Rackoff. The Knowledge Complexity of Interactive Proof Systems[J]. *SIAM Journal on Computing*, 1989: 186-208.

[10] Manuel Blum, Paul Feldman, Silvio Micali. Non-interactive zero-knowledge and its applications[C]. *Symposium on Theory of Computing,* 1988: 103-112.

[11] Chase, Melissa Erin. Efficient Non-Interactive Zero-Knowledge Proofs For Privacy Applications[D]. Brown University, 2008.

[12] Dan Boneh, Eu-Jin Goh, Kobbi Nissim. Evaluating 2-dnf formulas on ciphertexts[C]. *International Conference on Theory of Cryptography*, 2005: 325-341.

[13] Groth J. Short pairing-based non-interactive zero-knowledge arguments[C]. *ASIACRYPT*, 2010: 321-340.

[14] Groth J. On the Size of Pairing-Based Non-interactive Arguments[M]. *Theory and Application of Cryptographic Techniques*, 2016: 305-326.

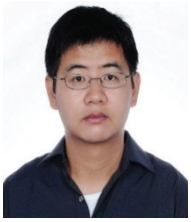
[15] Eli Ben-Sasson, Alessandro Chiesa, Eran Tromer, et al. Scalable zero knowledge via cycles of elliptic curves[C]. *Algorithmica*, 2017: 1102-1160.

[16] Zvika Brakerski, Craig Gentry, Vinod Vaikuntanathan. (Leveled) Fully homomorphic encryption without bootstrapping[J]. Acm Transactions on *Computation Theory*, 2014, 6(3): 1-36.

[17] Craig Gentry. Fully homomorphic encryption using ideal lattices[C]. *Symposium on Theory of Computing,* 2009: 169-178.

[18] Craig Gentry. A fully homomorphic encryption scheme[D]. Stanford University, 2009.

石侃 于 2016 年在湘潭大学软件工程专业获得学士学位。现在华东师范大学计算机科学与技术专业攻读硕士学位。研究领域为零知识证明、隐私保护。Email: shi_kan@qq.com



陈洁 于2013年在新加坡南洋理工大学数学系获得博士学位。现任华东师范大学软件工程学院研究员、博士生导师。研究领域为公钥密码学。Email: jchen@cs.ecnu.edu.cn