基于区块链的感知数据交易隐私保护方案

李云辉，陈家辉
广东工业大学计算机学院，广东 广州 510006

摘要
感知数据交易能将感知数据转化为经济价值，促进数据的有效利用和共享。为了确保感知数据交易的可靠性和隐私安全，提出了一个基于混洗差分隐私的区块链感知数据交易方案。该方案设置了审计节点进行用户筛选和任务执行，混洗节点进行争议处理和奖励分发，并使用混洗模型下的差分隐私技术对用户的数据进行加噪。此外，还使用加法秘密共享技术划分数据到r个混洗器，以隐藏用户和数据的映射关系。该方案不需要可信的第三方，数据消费者可通过区块链交易平台发布任务并进行广播，进行安全隐私的数据交易。同时，根据隐私放大定理，该方案可获得接近中心化差分隐私的隐私保护效果。最后通过实验验证了方案的可行性，对比相关算法，该方案得到的数据准确性更高。

关键词
感知数据；数据共享；区块链；隐私保护；差分隐私

中图分类号：TP181 文献标志码：A doi: 10.11959/j.issn.2096-0271.2023071

A blockchain-based privacy protection scheme for sensing data trading

LI Yunhui, CHEN Jiahui
School of Computers, Guangdong University of Technology, Guangzhou 510006, China

Abstract
Sensing data trading is to transform sensory data into economic value and promote the utility and sharing of data. To ensure the reliability and privacy of data transaction, a blockchain sensing data transaction scheme based on shuffle differential privacy was proposed. In our scheme, we set an audit node to supervise users and perform tasks, a shuffle node to deal with disputes and reward distribution. We used the differential privacy technology under the shuffle model to add noise to the user's data. In addition, we supplied additive secret sharing divide the data into r shufflers to prevent the mapping relationship between users and data. Our scheme does not require a trusted third party, while data consumers could publish tasks and broadcast data through the blockchain trading platform for secure and private transactions. According to the privacy amplification theorem, the proposed scheme could obtain similar privacy

protection with the centralized differential privacy. Finally, we gave experiments to verify the feasibility of the scheme. Compared with related algorithms, the data accuracy obtained by our scheme was better.

Key words
sensing data, data sharing, blockchain, privacy protection, differential privacy

0 引言

随着物联网技术、5G技术以及大数据分析技术等的发展，个人感知数据的收集和交易日益普遍。在感知数据交易中，个人用户可以通过出售自己的感知数据来获取奖励。购买者可以通过感知数据进行研究、产品开发和智能系统的训练等。然而个人感知数据涉及个体隐私信息，因此需要确保数据的安全。传统的隐私保护方法（如加密和脱敏等）虽然有效，但在某些情况下仍然会受到攻击和破解。因此，设计一种更加高效和安全的隐私保护方案变得越来越重要。

差分隐私（differential privacy，DP）是Dwork C[1]于2006年针对统计数据库的隐私泄露问题提出的一种新的隐私定义，用于在发布统计信息时保护数据库中个人的隐私信息。差分隐私一般指中心化差分隐私（central differential privacy，CDP），即中心服务器收集用户的数据，将噪声添加到聚合结果中，然后发布结果。然而中心服务器可能会泄露隐私数据，为此，本地差分隐私（local differential privacy，LDP）被提出。LDP与CDP的不同之处在于，在将数据发送到中央服务器之前，每个用户都会添加随机噪声。因此，用户不需要信任服务器。然而本地化差分隐私加入大量噪声，会使数据的可用性下降。为此，2017年Bittau A等人[2]提出ESA(encode-shuffle-analyze)框架，该框架主要由编码器(encoder)、混洗器(shuffler)和分析器 (analyzer)三部分组成。编码器运行在客户端，对用户数据进行本地化的编码、分割、扰动等处理；混洗器运行在一个半诚信的第三方，它可借助现有的安全混洗协议对数据完成安全的混洗操作；分析器运行在数据收集者端，对收集的数据进行校正与分析。该框架中，混洗器完成了对用户数据完全匿名的操作，使用户可以在尽可能对数据本身进行较小扰动的情况下，获得较多的隐私保护。随后，Balle B等人[3-6]根据该框架介绍了混洗模型下的差分隐私，对隐私放大理论进行了严格的数学证明。隐私放大理论是指用户在客户端通过本地化差分隐私的方法对数据进行扰动，使扰动后的数据经混洗后，可以接近于中心化方法获得的数据统计结果。

此外，区块链作为一种新技术，基于区块链的感知数据交易可以实现去中心化的数据存储和管理，避免了单点故障的风险，提高了数据的安全性。此外，区块链的智能合约功能可以实现数据交易的自动化和可编程性，增加了数据交易的透明度和可信度。然而，基于区块链的感知数据交易仍然面临一些挑战和隐私保护需求。例如，如何在保证交易不可篡改性的同时保护感知数据的隐私、如何实现匿名化的交易过程、如何确保数据的可控性和合规性等。因此，需要开展深入的研究，提出创新的、基于区块链的感知数据交易隐私保护方案，为感知数据交易的安全可靠提供有效的技术支持。

本文提出的方案使用混洗差分隐私，用户上传数据前在本地加入少量噪声，然后经

过混洗器的处理，最终数据收集者得到数据并进行分析校正。本文的贡献如下。

- 针对数据的安全性，本文方案利用区块链技术，在几乎不影响系统性能的情况下，所设计的系统具有鲁棒性、不可否认性以及可追踪性等特性。
- 针对数据的隐私性，本文方案使用混洗差分隐私技术，根据不同的数据特性选择相应的处理算法实现。同时为了防止单一混洗器的不安全性，本文设计了r个混洗器来对用户数据进行混洗操作。在具有更高数据可用性的情况下，达到中心化差分隐私技术的隐私保护效果。
- 本文对所提方案的安全性进行了分析，证明了方案在抵抗合谋攻击、拒绝服务攻击和篡改攻击方面的保护能力。
- 本文对设计的方案和对比方案进行了仿真实验，实验结果表明，在数据隐私保护上，本文方案所得均方误差更好。同时，在以太坊私链上实际部署了本文方案，实验结果验证了方案的可用性与有效性。

1 相关工作

文献[7]提出一个用于人群感知数据市场的利润驱动型数据采集框架，实现了群体感知数据交易模式的确定、多项式计算复杂性的利润最大化以及战略环境中的支付最小化。文献[8]设计了一个移动众感数据市场架构，提出了一种基于在线查询的众感数据定价机制来确定众感数据的交易价格，优于最先进的定价机制，实现了约90%的最佳收入，并且以公平的方式在数据提供者之间分配奖励，激励数据提供者贡献数据。然而中心化的数据存储和管理模式可能存在单点故障和数据滥用的风险，并且一些安全问题也无法得到保证。

另外，DP与区块链结合的隐私保护方案也是一个比较前沿的研究方向。例如，文献[9]使用DP为工业物联网构建了一个基于区块链的隐私保护架构，提出的架构依赖于一个被称为“优化服务器”的集中实体，该实体负责分配任务、收集数据并使用DP向数据添加噪声。该方案能实现隐私保护，但一旦受信任的中心化实体被攻击，数据则会全部泄露。Liu Z W等人[10]提出了一种基于差分隐私的安全电力数据交易区块链方案，该方案利用零知识证明和区块链在不泄露数据的情况下实现了数据的可用性和数据交易的可靠性，同时提出一种差分隐私保护方案来保护电力数据中的隐私信息。然而该方案使用的是CDP，数据加噪过程在中心化服务器中实现，一旦被攻击，数据也将会全部泄露。Fotiou N等人[11]利用LDP来保护数据提供者的隐私免受数据消费者和系统运营商的侵害，构建了一个基于区块链的解决方案来确保公平交换和不可变的数据日志。但该方案加入了大量噪声，影响数据的可用性，同时使用的RAPPOR[12]方法需要很大的通信开销。

2 预备知识

2.1 区块链与智能合约

区块链技术于2008年首次作为加密货币比特币的技术出现，比特币也是第一个使用区块链的P2P数字货币系统[13]。区块链实际上是一种分布式数字账本技术，它由不断增长的被称为块的记录列表组成。区块链技术具有去中心化、可追溯、防篡改和数据信息公开等特性，区块链技术在相关领域应用广泛，包括金融、人工智能、物联网和医疗保健等[14]。

智能合约是在区块链中确定执行的去中心化应用程序。智能合约被广泛用于实现以公平方式交换数字商品的托管服务。这种托管智能合约允许“买方”存入数字货币，当向合约提供数字商品交换的证明时，该数字货币的一部分被转移给“卖方”。

2.2 加法秘密共享

加法秘密共享指用户可以把一个秘密值 $v\in\{0,1,2,\cdots,d-1\}$ 拆分为r份，随机选择其中的$r-1$份（$a_1+a_2+\cdots+a_{r-1}$），计算最后一份a_r使（$a_1+a_2+\cdots+a_{r-1}+a_r$）mod $d=v$。随后，这些被发送给r个参与方，每个参与方只拥有一个随机值。在该技术中，只有所有r个参与方合作，v才能被恢复。

2.3 混洗差分隐私

设 $M=R\circ S$ ，每个用户u_i在本地客户端利用满足ε_1的本地化差分隐私算法$R:V\to Y$扰动$v_i:y_i=R(v_i)$，得到$\{y_1,y_2,\cdots,y_n\}$为n个用户的扰动结果，$S:Y^n\to Y^n$为混洗器对n个用户的输出结果进行随机混洗操作。对于任意相邻数据集D和D'（n个用户中仅有一个用户数据不同），任意输出集合 $y'\subset Y^n$，满足式（1），则M满足（ε_c,δ）-混洗差分隐私：

$$\Pr[M(D)\in y']\leqslant \mathrm{e}^{\varepsilon_c}\Pr[M(D')\in y']+\delta \quad (1)$$

其中，ε表示隐私预算，$\delta(\delta\in(0,1])$为隐私泄露风险概率。

2.4 随机响应机制

基本的机制被称为随机应答[15]，它是为二进制状态（$D=\{0,1\}$）引入的，但很容易得到扩展。在随机响应机制（generalized randomized response，GRR）中，每个具有私有值$v\in D$的用户将GRR（v）发送到服务器，其中GRR（v）以概率P输出真实值v。以概率$1-P$随机选择 $v'\in D$ 代替真实值v且 $v'\neq v$ ，域的大小表示为 $d=|D|$ ，即有式（2）：

$$\Pr[\mathrm{GRR}(v)=y]=\begin{cases}p=\dfrac{\mathrm{e}^{\varepsilon}}{\mathrm{e}^{\varepsilon}+d-1}, & y=v\\ q=\dfrac{1}{\mathrm{e}^{\varepsilon}+d-1}, & y\neq v\end{cases} \quad (2)$$

2.5 k值随机响应机制的隐私放大定理

给定n个用户，每个用户对应1条记录 $v_i\in\{1,2,\cdots,k\}$ 且在本地运行协议R。对于任意的k和 $\gamma\in[0,1],\varepsilon_c\in(0,1]$，如果协议$R$以 γ 的概率均匀得到 $\{1,2,\cdots,d\}$ 中的随机值，以 $1-\gamma$ 的概率得到真实值，则当 γ 满足式（3）时，协议 $R\circ S$ 对应混洗之后的n个输出满足 (ε_c,δ) -DP：

$$\gamma=\max\left\{\frac{14k\ln\left(\dfrac{2}{\delta}\right)}{(n-1)\varepsilon_c^{\,2}},\frac{27k}{(n-1)\varepsilon_c}\right\}<1 \quad (3)$$

当 $\gamma=k/(\mathrm{e}^{\varepsilon_l}+k-1)$ 时，k 值随机响应机制满足 ε_l -LDP，代入可得 $\varepsilon_c\leqslant\sqrt{14\ln(2/\delta)(\mathrm{e}^{\varepsilon_l}+k-1)/(n-1)}$ ，即放大为$\left(\sqrt{14\ln(2/\delta)(\mathrm{e}^{\varepsilon_l}+k-1)/(n-1)},\delta\right)$-DP。

文献[3]中使用的技术被称为毯子分解，将用户对查询给出的随机答复称为隐私毯子，基于LDP模型生成的分布可以分为两部分，一部分依赖于真实值的分布，另一部分是独立随机的分布，此过程被称为隐私毯子分解。因此GRR的输出分布可以分解为式（4）：

$$\begin{aligned}&\Pr[\mathrm{GRR}(v)=y]=(1-\gamma)\\&\Pr[y\mid v]+\gamma\Pr[\mathrm{Uni}([k])=y]\end{aligned} \quad (4)$$

其中，$\Pr[y|v]$表示依赖于v的真实值形成的分布，$\mathrm{Uni}([k])$是均匀随机分布，并且$\Pr[\mathrm{Uni}([k])=y]=1/k$，$n$个用户中，除第$n$个用户外，其余$n-1$个用户的输出可以看作包含一些均匀噪声，这些噪声使输出具有不确定性，$v\in[k]$噪声服从$\mathrm{Bin}(n-1),\gamma/k)$，即服从$\mathrm{Bin}\left(n-1,\frac{1}{\mathrm{e}^{\varepsilon_l}+k-1}\right)$。

3 系统模型

本模型包含以下实体：审计节点（auditing node，AN）、数据消费者（data consumer，DC）、区块链交易平台（blockchain trading platform，BTP）、数据提供者（data submitter，DS）、混洗节点（shuffler node，SN）。具体模型如图1所示。

AN：主要负责注册参与系统事务的实体，为注册实体生成必要的公共和私有参数；负责筛选符合条件的DS和SN，为它们设置一个信誉值并且公开；可接受处理DC的投诉与争议；负责数据提供者和混洗节点的奖励发放。

DC：该实体希望获取大量用户感知数据，获得数据后进行解密以及分析校正，从而得到自己所需的统计数据并支付费用。

BTP：负责存储交易信息，参与实体查询区块链上数据进行验证交易的有效性；实现交易的不可篡改和可追溯性，智能合约负责交易规则的制定和交易公平。

DS：感知数据可直接在传感设备中对用电数据加入少量噪声；利用加法秘密共享技术将数据分为r份并发送给r个给定的SN，并将各部分数据的哈希值和该事务的信息存储在区块链上。

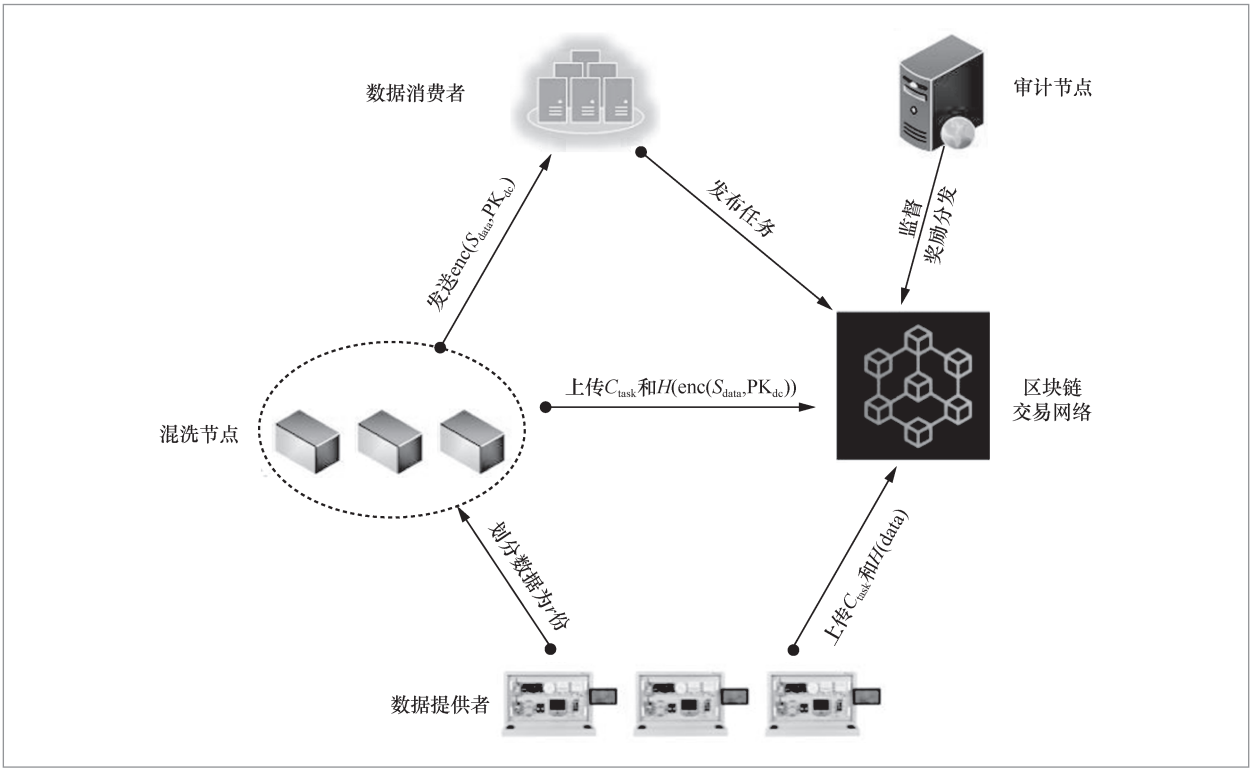

图1 系统模型

SN：数据消费者发布任务后，各SN根据自身情况和任务属性进行报名，经过AN筛选后的r个SN得到感知数据，之后进行混洗操作，最后将各自的数据进行加密并发送给数据消费者，在区块链中记录数据哈希值和该事务信息。

方案主要包括初始化、各实体注册、任务发布、任务执行准备、上传数据、获得数据和奖励分发7个阶段。本方案符号说明见**表1**。

3.1 初始化阶段

DC通过自己的ID_{dc}得到公私钥对（PK_{dc}，SK_{dc}）。

3.2 各实体注册阶段

DS、DC、SN在区块链交易平台以各自身份进行注册，得到区块链地址。SN也可以是用户，但用户需要再以此身份进行注册，不同的身份有不同的权限。各实体都有一个信誉值属性，由AN管理，主要解决交易的争议。

表1 符号说明

符号	含义
ID_{dc}	DC的身份ID信息
（PK_{dc}，SK_{dc}）	DC的公私钥对
DC_{addr}	DC的地址
Deposits	发布任务所需押金
$Info_{task}$	任务信息
$Status_{sc}$	智能合约状态
DC_IP	DC的IP地址
TSC	时间戳
AN_{addr}	AN的地址
SN_{addr}	SN的地址
n_SN	SN表
List_credit	SN表的信誉值
ε	中心端隐私预算
δ	隐私泄露风险概率
ε_l	本地端隐私预算
DS_{addr}	DS的地址

3.3 任务发布阶段

DC在区块链平台发布任务，将任务信息$Info_{task}$和自己的IP地址广播出去，DC在发布任务时会支付少许费用，防止拒绝服务攻击。

算法1 创建任务合约

输入 DC_{addr}，deposits，$Info_{task}$，$Status_{sc}$，DC_IP

输出 TSC

1. If $Status_{sc}$ == Running then
2. If msg.balance >= deposits&& msg.sender == DC_{addr}
3. SC←deposits;
4. Set $Info_{task}$;
5. TSC = block.timestamp;
6. emit $Info_{task}$, DC_IP;
7. Else
8. Revert $Status_{sc}$ && Display errors
9. End if
10. Else
11. Revert $Status_{sc}$ && Display errors
12. End if

如果合约状态$Status_{sc}$正在运行，将确定调用合约的地址是否为DC_{addr}。DC向合约账户发送存款，设置任务信息$Info_{task}$，记录时间戳TSC，并广播任务信息$Info_{task}$和DC_IP。

3.4 任务执行准备阶段

在执行任务之前，AN需要选择适当的

用户和SN，如算法2所述。DC和所选节点需要通过可信的安全通道传输消息。

空闲的SN和符合要求的用户向AN申请执行任务并上传自己的IP地址。

AN根据这些应用节点的信誉值选择DC所需的节点数量和用户数量，并广播所选r个节点的IP地址，以通知DC、DS和其他$r-1$个节点，以便它们在交易过程中相互通信。

r个选择的SN和DC相互认证，并且DC通过可信安全信道向每个SN发送（$Info_{task}$，PK_{dc}）。

算法2 分配任务合约

输入 AN_{addr}，$Info_{task}$，$Status_{sc}$，SN_{addr}，r，DC_{addr}，n_SN

输出 满足任务的SN

```
1. If msg.sender == AN_addr
2.     Sort (n_SN, val_credit);
3.     Select r top SN;
4.     SN←DC_addr, Info_task;
5.     If SN_i not honest
6.       val_credit of SN_i--;
7.       Choose SN_j;
8.     End if
9.     Updata List_credit;
10. End if
```

如果合约地址为AN_{addr}，则合约从信誉表中选择信誉值最高的前r个SN，并将任务信息和DC的地址发送给SN。如果某个SN有不诚实行为，则其信誉值将降低，选择其他SN来完成任务，并更新信誉表。

3.5 上传数据阶段

感知设备收集到数据v，根据v的取值范围k、用户数n和需要得到差分隐私保证的参数ε和δ计算出相应的本地端差分隐私参数ε_1。特别地，针对不同的数据特性，有以下两个算法。首先是使用随机响应机制对数据进行扰动处理，即混洗随机应答（shuffler randomized response，SRR）算法[3]。该算法在值域较小的情况下，效率和数据可用性较高，然而当数据值域较大时，GRR中真实值的输出概率P将会变小，得到的数据可用性较差。其次是文献[6]提出用哈希编码的方式来降低数据的值域，即SOLH（shuffler-optimal local hash）算法。因此针对不同的值域，可根据两个算法的方差进行对比，选出相应的算法中的机制进行差分隐私处理。一般地，算法中的RR（randomized response）机制适用于值域较小的情况，OLH（optimal local hash）机制适用于值域较大的情况。

算法3描述了本文的数据处理方法。输入参数为感知设备采集的n个用户数据（第i个数据为v_i）、数据的取值范围k、需采集的用户数量n以及需要获得差分隐私保证的参数。输出是n个混洗后的结果。数据处理的具体实现步骤如下。

- 使用参数k、n、ε和δ自动计算本地端隐私预算参数ε_1。之后计算两个算法机制的方差，选择较小的机制进行处理数据。
- 每个用户使用加法秘密共享技术为扰动数据y选取$r-1$个随机值，然后计算第i个值。这些值分别发送到r个SN。
- 用户将r个向量的哈希值$H(a_i)$和交易的相关信息上传到区块链。
- 节点收到用户发送的数据后，查询区块链上存储数据的哈希值，验证其真实性。验证通过后，他们将根据文献[16]的混洗方法对数据进行混洗。从r个SN中随机选取$t=\lfloor r/2 \rfloor+1$作为“隐藏者”的数量，将$r-t$作为“搜寻者”的数量。经过$\binom{r}{t}$轮之后，将获得n个混洗后的结果。
- SN使用PK_{dc}对混洗后的数据进行加密，并通过可信的安全通道将其发送到DC。然后，它们将加密数据的哈希值和交易相关信息上传到BTP。

加噪声的步骤可以嵌入感知设备中，并且无法更改，从而使随机扰动的概率相同。

算法3 数据处理算法

输入 data of n users, k, ε_c, δ

输出 the n shuffled results

1. $\varepsilon_1 = \text{Calculate}(\varepsilon_c, \delta)$;
2. If var(RR) < var(OLH)
3. y=RR(data)
4. Else
5. y=OLH(data)
6. For users i=1 to $n$do
7. U_i divide y_i into r
8. For$m$=1 to $r$do
9. U_i send $r_m$to r SNs and upload H(Infotask, r_m) to Blockchain
10. Endfor
11. Endfor
12. Randomly select $t=\lfloor r/2 \rfloor+1$ as the number of "hiders" and r−t as the number of "seekers"
13. Fori=1 to $\binom{r}{t}$do
14. The vector of each seeker is divided into t parts and then sent to the t hiders
15. An agreed arrangement is used by the hiders to shuffle their vectors
16. After shuffling, the vectors are divided into r shares and distributed among all of the r shufflers
17. Endfor
18. For j=1 to r
19. Shuffler$j$ encrypt data$_j$, send enc(data$_j$) to DC and upload H(Infotask, enc(data$_j$)) to Blockchain
20. Endfor

3.6 获得数据阶段

DC获得数据后，查询区块链中各SN存储的数据哈希值进行验证，验证通过后，组合收到的数据，根据 $(a_1+a_2+\cdots+a_{r-1}+a_r)\bmod d=y$ 得到各用户的扰动数据，得出频率估计 $l_{\{y_i=v\}}$，再根据式(5)：

$$\tilde{f}_v=\frac{1}{n}\sum_{i\in[n]}\frac{l_{\{y_i=v\}}-q}{p-q} \qquad (5)$$

得到最终频率估计，进行支付费用。

3.7 奖励分发阶段

AN根据算法4将奖励分配给DS和SN。

算法4 奖励分配合约

输入 AN_{addr}, $Status_{sc}$, SN_{addr}, DC_{addr}, DS_{addr}

输出 reward information

1. If msg.sender == AN_{addr}&& $Status_{sc}$ = completed task
2. If msg.balance >= deposits
3. SC←payment (from DCaddr);
4. SN_{addr}←reward (from SC);
5. DS_{addr}←reward (from SC);
6. Else
7. Reward $Status_{sc}$ && Display errors
8. End if
9. End if

如果合约地址为AN_{addr}，且合约状态为任务已完成，则合约将自动从DC地址的余额中扣除费用，并在任务完成时向SN和DS发放奖励。

4 安全性分析

4.1 隐私保护分析

与现有方案相比，本文方案不仅可以保护参与者的身份隐私，还可以确保数据隐私。

首先，在身份隐私上，注册阶段各实体参与者的身份将通过区块链由AN进行严格审查，避免恶意用户，确保所有参与者都是合法的，然后区块链将为每个参与者生成一个假名。参与者的隐私将得到保护，因为在后续过程中其使用的是假名而不是他们的真实身份。

其次，在数据隐私上，交易阶段用户的原始数据在各自本地通过(ε_l,δ)差分隐私加入噪声，并且通过秘密共享的方式将数据分别发给r个SN。确保原数据只有自己拥有，不会被任何参与者得到。混洗器对用户报告的数据进行随机排列，DC接收到数据之后无法将用户链接到数据，因为数据被打乱了。交易过程中只是将传输数据的哈希值上传到区块链上，公开的交易信息只是用来验证数据是否篡改，不包含传输数据。

4.2 对常见攻击的防护能力

(1) 合谋攻击

如果用户与DC勾结，DC可以得到除被攻击者之外的所有用户的报告。通过从最终结果中减去每个用户的数据，DC可得到受害者的LDP报告；因此隐私也会退回到(ε_l,δ)本地化差分隐私，得到本地差分隐私的保护。

当SN互相串通时，没有放大隐私。当服务器与辅助服务器勾结时，隐私保证退回到原来的LDP模型。在使用混洗模型时，需要减少这种共谋的可能性，例如，通过引入更多的辅助服务器来实现。

(2) 拒绝服务攻击

为了防御拒绝服务攻击，系统运营商首先在DC发布任务时收取部分额外费用作为广播费。最低收费标准将由AN设定，确保合约代码的正常运行。在发布任务时对DC进行收费，一个作用是保证事务的正常运行，另一个原因是防止恶意的DC消耗资源。攻击者得到的回报远远少于付出，因此可以防御拒绝服务攻击。

(3) 篡改攻击

攻击者可能会恶意篡改存储的数据。然而在区块链上发布的交易是不能被篡改的，如果要修改它，则需要重新发布。参与者可以通过验证区块链上数据的哈希值来检查数据是否被篡改。此外，他的数据在发送到DC之前由密钥加密。没有相应的私钥，内部攻击者和外部攻击者都无法破解密文。

另外，如果攻击者是系统中的恶意节点，则可能会故意伪造执行结果。对于这种情况，设置的AN会不定时对每名参与者进行抽查。如果发现恶意行为，恶意节点将受到严厉惩罚，扣除一定的信誉值属性。此外，如果DC对结果不满意并提出异议，那么系统将跟踪交易，并且恶意行为也会被检测到。因此，当弊大于利时，参与者通常不会有恶意行为。

4.3 相关系统特性分析

本文的方案可以保证系统的鲁棒性、不可抵赖性和可追溯性。

(1) 鲁棒性

首先，任何一方都可以尝试中断交易过程，但这很容易解决。如果用户拒绝上传数据，则会找到其他用户上传数据。如果

SN拒绝该服务，AN可以找到另一个SN从之前应用的节点中替换它，并降低拒绝服务的SN的信誉值。

其次，SN可能会偏离协议，这样将不会执行混洗操作，因此DC得到原始的LDP报告。在这种情况下，DC可以获取更多信息，但SN除了节省一些计算能力外，没有任何好处。恶意节点可能会被AN随机检查，这样就会受到严厉的惩罚，信誉值属性就会降低。因此，本文假设SN不会偏离协议操作。由于DC只能查看和评估最终报告，因此DC无法从用户那里获取更多信息。

(2)不可抵赖性

数据交易是通过区块链进行的，区块链的透明度可以确保交易的不可否认性。对相应实体分配奖励由智能合约执行，如果任何实体存在非法操作，例如用户和SN通过伪造数据或篡改数据来获取不正当利益。AN会不定时进行抽查，并接受参与者的投诉，如果用户和SN被发现，他们将受到严厉的惩罚。如果DC收到数据后拒绝支付相应费用，也将被追究责任。因此，任何实体都不能通过拒绝存储在区块链上的交易来获得非法利润。

(3)可追溯性

由于具有某些争议或恶意参与者，AN有权追踪参与者的身份。区块链上发布的交易可以追溯到具体细节。本方案利用区块链的特性，将传输数据的哈希值存储在区块链上，这可以让AN有效地追踪交易消息，从而进行责任确定。

5 实验评估

5.1 实验环境

首先，对数据处理过程进行了仿真实验，测试了隐私效果以及不同参数对均方误差（MSE）的影响。实验是在配备Intel(R) Core(TM) i7-1065G7 CPU @ 1.30GHz，16GB内存的系统上进行的。数据隐私保护算法由运行在Pycharm 2022.1.3上的Python语言编写，Python版本为 Python 3.9。分别执行了10次，并取平均值进行比较分析。

智能合约的实验是在AMD R5 4600H@ 3.00GHz and 16GB of RAM running 64bit Ubuntu 16.04上进行的。本文构建了以太坊中的私有链来模拟该方案，均采用哈希函数SHA256()进行测试，测试了智能合约主要函数的gas消耗和调用函数的系统时间开销；分别执行了20次，并取平均值进行比较分析。

5.2 实验结果

由于本文实际应用场景中，感知数据的值域以及数据量一般较大，因此设置参数ε_c=1.0,k=3 000,δ=10^{-6}，随机生成 $\mu=k/2$，$\sigma=k/6$ 的正态分布。测试了n=500 000的处理结果，显示了原始数据和隐私处理后数据的分布对比，结果如**图2**所示。

从**图2**可以看出，经过差分隐私算法处理后，数据的分布并没有太大的差异。n越大，统计样本越大，得到的统计数据更接近原始统计数据。

设置参数ε_c=1.0,k=3 000,δ=10^{-6}，随机生成 $\mu=k/2$、$\sigma=k/6$ 的正态分布。可以观察到n从100 000到1 000 000的均方误差（MSE）变化。本方案针对LDP算法HR（hadamard response）[17]，RAPPOR[12]和基于shuffler的差分隐私算法SOLH[6]和SRR[3]进行比较。MSE的计算式如下：

$$\mathrm{MSE}_{\mathrm{frea}}=\frac{1}{k}\sum_{i\in[k]}(f_i-\tilde{f}_i)^2 \qquad (6)$$

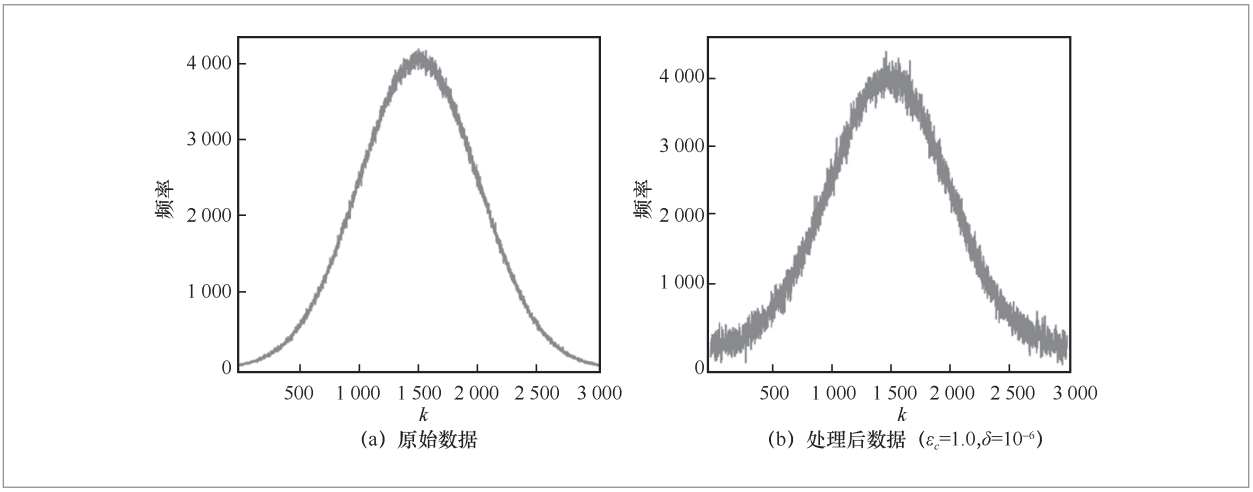

图 2 数据处理前后数据分布对比

其中，f_i 是原始数据 v_i 的频率，$\tilde{f}_i$ 是最终得到 v_i 的估计频率。

图3显示，本文的混洗差分隐私使用的SOLH和SRR算法的MSE明显比HR和RAPPOR要低。与LDP算法相比，本方案的MSE更小。与SRR相比，SOLH的MSE仍然较小。因为实验采取的k为3 000，值域较大的情况下，SOLH更能够提高数据的可用性。此外，当n的值增加时，MSE会降低，并且由此产生的数据可用性更高。

设置参数n=1 000 000，k=100，δ=10^{-6}，随机生成 $\mu = k/2$、$\sigma = k/6$ 的正态分布。可以观察　从0.1到2.0变化时，MSE的变化。

图4显示，LDP算法HR和RAAPPOR的MSE比本方案中的两个算法明显高几个数量级。可以观察到与SRR相比，SOLH的MSE更小，是因为取的k值为3 000。最后还发现，当ε_c的值增加时，MSE减小，并且获得的数据更接近原始值。

为了观察频率MSE对k从100到5 000的变化，设置参数n=1 000 000，ε_c=1.0，δ =10^{-6}，对于变量k每次随机生成 $\mu = k/2$、$\sigma = k/6$。

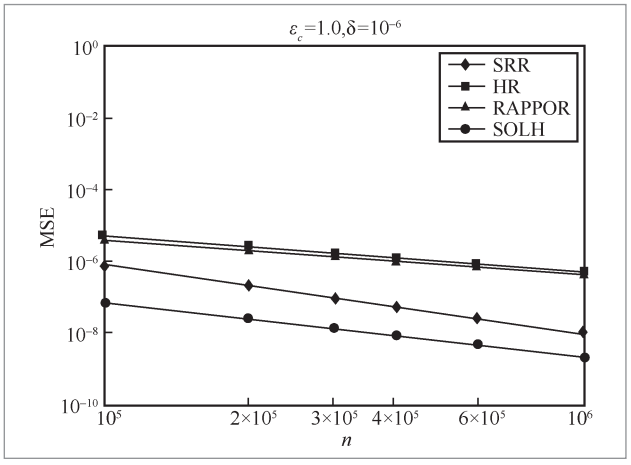

图 3 n 变化时 MSE 的变化

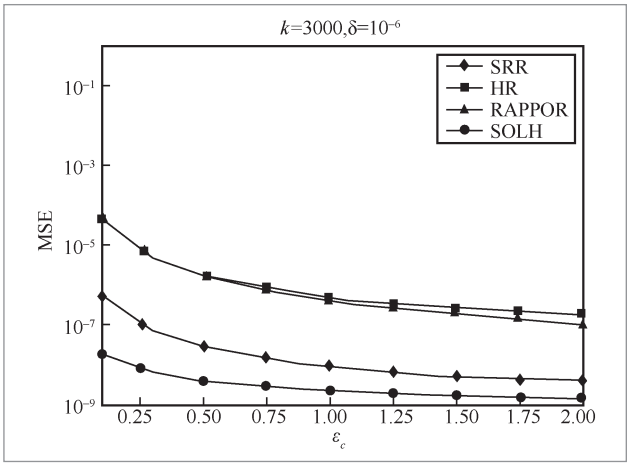

图 4 ε_c 变化时 MSE 的变化

图5显示，当k的值从100到5 000变化时，本方案中的两个算法的MSE仍然很小。还可以观察到，k值较小时，SRR算法的MSE较小，随着k值的增加，SOLH算法的MSE小于SRR算法，并且二者之间的差距变大。随着k值的变化，SOLH的MSE几乎不变，因为SOLH算法适用于较大的k值且较为稳定。因此根据感知数据的值域大小，选择相适应的算法是非常有必要的。

表2显示了部署智能合约以及调用其函数的成本，以gas为单位。每个操作的成本不受DS数量的影响，也不受值范围的影响。该方案调用合约函数的gas成本和时间成本都很小，均属于正常消耗范围内。虽然部署合约会消耗大量gas，但合约只需要部署一次，因此该时间是可以接受的。

表3显示了隐私保护前的原始数据上传与隐私保护后的数据上传效率对比。可以看到，在加隐私保护处理之前与之后的数据上传的比较上，由于进行了隐私保护，智能合约上传数据的消耗时间和gas消耗均有所提升。

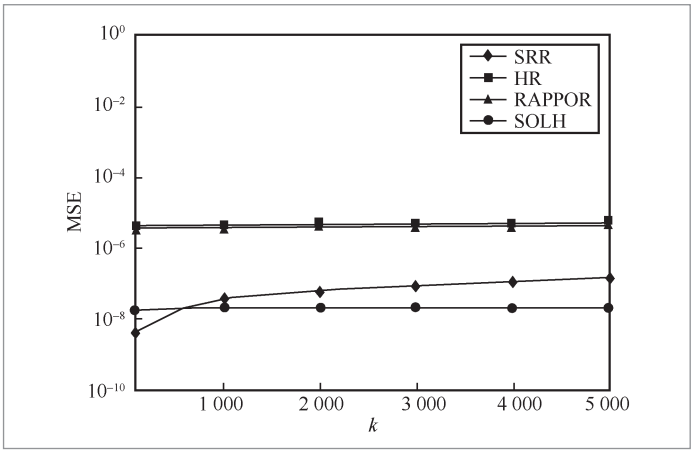

图 5 k 变化时 MSE 的变化

表 2 部署智能合约以及调用其函数的成本和消耗时间

操作	gas消耗	消耗时间/ms
合约部署	1 164 642	178
创建查询	90 158	70
用户数据哈希值上传	85 044	81
混洗数据哈希值上传	85 044	85
奖励分发	37 330	53

6 结束语

本文提出了一种基于混洗差分隐私的区块链感知数据交易方案。在该方案中，数据需求者可以下达任务并通过BTP广播购买数据，使用区块链来确保交易的公平性和可追溯性，并使用智能合约进行奖励分配。在收集数据时，使用随机应答机制模型下的差分隐私对用户的数据进行加噪，可根据不同的数据特性选择相应的处理算法，不需要可信的第三方就可以获得接近CDP的隐私保护效果。最后通过实验验证了该方案的可行性和隐私保护效果，并将几种相关算法进行了比较，获得了更好的结果。由于区块链系统的效率问题，在真实数据集下的实际应用部署仍然是一个挑战，在未来的工作中笔者将会继续关注如何提高区块链隐私保护系统的效率，并在真实数据集上进行性能实验测试；此外，研究设计混洗节点使其效率更高以及隐私性更强，探索将方案应用到其他应用场景并进行改进也是未来的工作方向之一。

参考文献:

[1] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[M]//Theory of cryptography. Heidelberg: Springer, 2006: 265-284.

[2] BITTAU A, ERLINGSSON Ú, MANIATIS P, et al. Prochlo: strong privacy for

表 3 隐私保护前的原始数据上传与隐私保护后的数据上传效率对比

大小	操作	gas消耗	消耗时间/ms
256	隐私保护前数据上传	85 044	87
	隐私保护后数据上传	170 088	166
512	隐私保护前数据上传	126 075	90
	隐私保护后数据上传	252 150	173
768	隐私保护前数据上传	167 251	92
	隐私保护后数据上传	334 502	178
1 024	隐私保护前数据上传	208 427	102
	隐私保护后数据上传	416 854	216

analytics in the crowd[C]//Proceedings of the 26th Symposium on Operating Systems Principles. New York: ACM, 2017: 441–459.

[3] BALLE B, BELL J, GASCÓN A, et al. The privacy blanket of the shuffle model[C]//Proceedings of Annual International Cryptology Conference. Cham: Springer, 2019: 638–667.

[4] CHEU A, SMITH A, ULLMAN J, et al. Distributed differential privacy via shuffling[C]//Proceedings of 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques. Darmstadt: Springer, 2019: 375–403.

[5] ERLINGSSON Ú, FELDMAN V, MIRONOV I, et al. Amplification by shuffling: from local to central differential privacy via anonymity[C]//Proceedings of the Thirtieth Annual ACM–SIAM Symposium on Discrete Algorithms. New York: ACM, 2019: 2468–2479.

[6] WANG T, DING B, XU M, et al. Improving utility and security of the shuffler–based differential privacy[EB]. arXiv preprint, 2019, arXiv: 1908.11515.

[7] ZHENG Z Z, PENG Y Q, WU F, et al. Trading data in the crowd: profit–driven data acquisition for mobile crowdsensing[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(2): 486–501.

[8] ZHENG Z Z, PENG Y Q, WU F, et al. ARETE: on designing joint online pricing and reward sharing mechanisms for mobile data markets[J]. IEEE Transactions on Mobile Computing, 2020, 19(4): 769–787.

[9] GAI K K, WU Y L, ZHU L H, et al. Differential privacy–based blockchain for industrial Internet–of–things[J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 4156–4165.

[10] LIU Z W, HU C Q, XIA H, et al. SPDTS: a differential privacy–based blockchain scheme for secure power data trading[J]. IEEE Transactions on Network and Service Management, 2022, 19(4): 5196–5207.

[11] FOTIOU N, PITTARAS I, SIRIS V A, et al. A privacy–preserving statistics marketplace using local differential privacy and blockchain: an application to smart–grid measurements sharing[J]. Blockchain: Research and Applications, 2021, 2(1): 100022.

[12] ERLINGSSON Ú, PIHUR V, KOROLOVA A. RAPPOR: randomized aggregatable privacy–preserving ordinal response[C]//Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2014: 1054–1067.

[13] NAKAMOTO S. Bitcoin: a peer–to–peer

electronic cash system[J]. Decentralized Business Review, 2008: 21260.

[14] 李懿, 王劲松, 张洪玮. 基于区块链与函数加密的隐私数据安全共享模型研究[J]. 大数据, 2022, 8(5): 33-44.
LI Y, WANG J S, ZHANG H W. Research on privacy data security sharing scheme based on blockchain and function encryption[J]. Big Data Research, 2022, 8(5): 33-44.

[15] WARNER S L. Randomized response: a survey technique for eliminating evasive answer bias[J]. Journal of the American Statistical Association, 1965, 60(309): 63-66.

[16] LAUR S, WILLEMSON J, ZHANG B S. Round-efficient oblivious database manipulation[C]//Proceedings of the 14th International Conference on Information Security. New York: ACM, 2011: 262-277.

[17] ACHARYA J, SUN Z, ZHANG H. Hadamard response: estimating distributions privately, efficiently, and with little communication[EB]. arXiv preprint, 2018, arXiv: 1802.04705.

作者简介

李云辉(1996-), 男, 广东工业大学计算机学院硕士生, 主要研究方向为区块链技术及隐私保护。

陈家辉(1986-), 男, 博士, 广东工业大学计算机学院副教授, 主要研究方向为后量子密码学、区块链技术及人工智能安全。

收稿日期: 2023-06-06
通信作者: 陈家辉, csjhchen@gmail.com
基金项目: 国家自然科学基金资助项目(No.61902079)
Foundation Item: The National Natural Science Foundation of China (No.61902079)