

区块链隐私众包中的数据验证与可控匿名方案

薛开平^{①②} 范茂^① 王峰^{*②} 罗昕怡^①

^①(中国科学技术大学网络空间安全学院 合肥 230027)

^②(中国科学技术大学网络信息中心 合肥 230026)

摘要: 针对隐私众包场景下出现的数据验证、匿名作恶检测和跨平台资源交互等需求, 该文基于区块链技术, 并结合零知识证明与环签名技术, 提出一种联盟链架构下的隐私众包方案。该方案依靠零知识证明实现对加密数据的验证, 依靠链接可撤销环签名改进方案实现工人身份的可控匿名, 引入联盟链架构实现众包实体之间的资源交互。在完成众包完整流程的同时, 实现隐私众包所需的数据保护与身份保护。安全性分析表明, 该方案具有隐私性、可验证性、可控匿名性与公平性。实验结果验证了方案在效率与性能方面的有效性。

关键词: 区块链; 众包; 隐私保护; 环签名; 零知识证明

中图分类号: TN918; TP309

文献标识码: A

文章编号: 1009-5896(2024)02-0748-09

DOI: 10.11999/JEIT230106

Privacy Crowdsourcing on Blockchain with Data Verification and Controllable Anonymity

XUE Kaiping^{①②} FAN Mao^① WANG Feng^② LUO Xingyi^①

^①(School of Cyber Science and Technology, University of Science and Technology of China, Hefei 230027, China)

^②(Network and Information Center, University of Science and Technology of China, Hefei 230026, China)

Abstract: Considering the requirements of data verification, anonymous malicious behavior detection and cross-platform resource interaction in privacy crowdsourcing, a scheme under the consortium chain architecture is proposed, basing on blockchain technology with zero-knowledge proof and ring signature technology. The proposed scheme relies on zero-knowledge proof to achieve encrypted data verification, relies on improved revocable-iff-linked ring signature to achieve controllable anonymity of workers, introduces consortium chain to realize resource interaction between crowdsourcing entities. In addition to completing the crowdsourcing process, the scheme also implements data protection and identity protection required for privacy crowdsourcing. Security analysis shows that the proposed scheme satisfies privacy, verifiability, controllable anonymity and fairness. Experimental results verify the advantages of the proposed scheme in efficiency and performance.

Key words: Blockchain; Crowdsourcing; Privacy protection; Ring signature; Zero-knowledge proof

1 引言

在海量数据、碎片信息的背景下, 许多复杂任务难以交由计算机独立完成, “众包”(crowdsourcing)这一新型计算范式逐渐成为工业界与学术界完成复杂任务的优先选择。“众包”这一概念

最初由Howe^[1]提出, 它通过整合互联网中分布式群体的解决方案, 来完成计算机难以单独完成的复杂任务。

现有的众包平台大多基于中心化的架构, 一个经典的中心化的众包模型如图1所示, 由请求者、工人和众包平台3个实体组成。一个完整的众包流程大致为: (1)请求者通过中心化的众包平台发布任务; (2)工人通过相同的平台接受任务; (3)工人根据任务信息和数据需求将解决方案提交到平台以供评估; (4)请求者从平台得到任务的解决方案, 工人从平台获取报酬。

然而, 类似的经典中心化众包架构存在着一些无法忽视的问题:

收稿日期: 2023-02-27; 改回日期: 2023-06-30; 网络出版: 2023-07-06

*通信作者: 王峰 wf0229@ustc.edu.cn

基金项目: 安徽省重点研发计划(2022a05020050), 中国科学院青年创新促进会优秀会员支持项目(Y202093)

Foundation Items: Anhui Province Key Technologies Research & Development Program (2022a05020050), Youth Innovation Promotion Association of the Chinese Academy of Sciences (CAS) (Y202093)

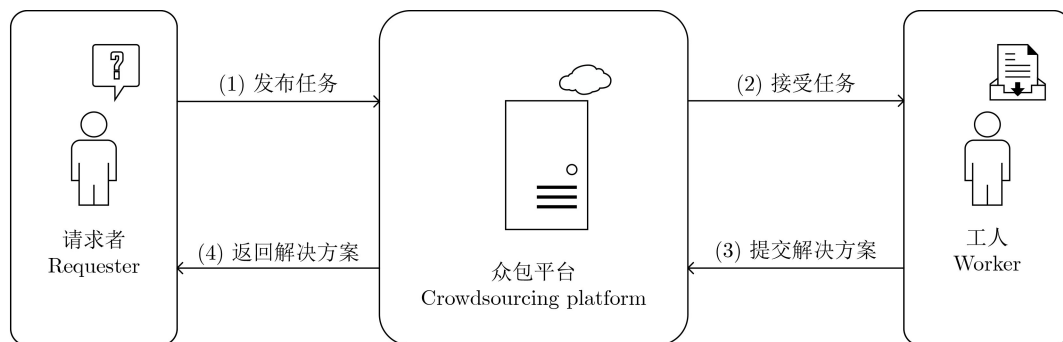


图1 经典的中心化的众包模型

(1)集中式问题：现有的众包平台如亚马逊土耳其机器人(Amazon Mechanical Turk, AMTurk)、滴滴出行(文中简称“滴滴”)等提供的众包服务都是基于集中式的众包系统实现的，请求者可以通过众包平台雇佣工人以获得任务的解决方案。集中式众包在面对分布式拒绝服务(Distributed Denial of Service, DDoS)攻击、女巫攻击等恶意行为时，容易导致服务器停机，从而使得众包服务陷入瘫痪状态。

(2)隐私性问题：众包平台会将流程中的用户信息与任务数据存储在中心化的服务器中，容易造成用户的隐私泄露，如滴滴违规收集用户数据、优步(Uber)遭到黑客攻击导致数据库被攻破泄露用户信息等。

(3)公平性问题：公平性问题主要分为两种，虚假报告(false-reporting)与搭便车(free-riding)^[2]。恶意的请求者可能通过虚假报告逃避对所发布任务应付报酬的支付，恶意的工人也可能通过搭便车骗取报酬而不提交解决方案。当用户出现争议时，众包平台也可能会做出恶意仲裁以保障某一方的权利，如文献[3]指出中心化的众包平台在进行仲裁时总是倾向于请求者。对此，文献[4]提出了基于声誉的激励机制来解决众包中存在的搭便车问题与虚假报告问题。然而，这些方案依旧采用的是传统的中心化的安全假设，以及无法抵御众包实体之间的共谋问题。

区块链技术具有去中心化、可追溯、不可篡改等特性，为上述挑战提供了新的解决路径。针对集中式问题，已有一些将区块链技术与众包结合的工作，如文献[5,6]提出了基于区块链的众包方案，避免了引入中心化实体进行监管，是区块链应用于众包场景的良好尝试。然而上述方案支持的众包场景较为单一，且缺乏对执行过程中众包数据的隐私保护。文献[7]在数据众包系统中设计了一种基于区块链的细粒度授权机制，解决用户之间的信任问题，使不受信任的用户以可验证的方式进行交易，去除

了不具有可信的中央授权机构。针对隐私性问题，Li等人^[8]提出了一个基于区块链的分布式众包框架(CrowdBC)。该框架基于轻量级方案设计，将请求者的任务分配给多个工人完成，并对用户的身份提供一定的隐私保护。然而，该方案所提供的隐私保护能力较低，且众包框架受限于单个众包平台。文献[9]基于区块链设计了智慧城市背景的众包系统，并设计了相应的智能合约算法实现众包流程，记录完整的众包服务过程。然而其隐私保护机制较为简单，数据完全存储在链上，用户的数据和用户行为都是公开透明的，具有一定的隐私风险。文献[10]提出了混合区块链众包平台(CHChain)，其核心是混合区块链结构，或称为跨链结构。它将每个任务的私有信息存储在私有任务链中，并将任务的其他公共信息记录到一个公共链中，在保证任务隐私的同时实现分布式透明存储，该方案缺少对存储数据的验证机制。针对公平性问题，文献[11]提出了一个匿名的众包系统，向解决方案增加前缀，通过不同解决方案的前缀是否一致来判断重复提交的发生，在有效实现用户隐私保护的同时，能够较为高效地检测出存在的用户搭便车行为，但是无法进一步在此基础上将作恶工人的实际身份揭示出来，而只是简单将重复提交的解决方案进行丢弃。Zhang等人^[12]提出了基于穿刺加密的区块链隐私众包方案以保护工人提交的信息，可以提供较强的前向安全性，然而方案直到众包流程结束才能对数据质量进行评估，无法较好地解决仲裁问题。

上述文献对于基于区块链的众包系统构建给出了一些可行的解决方案和性能优化思路，然而在区块链隐私众包场景下仍然存在一些尚待解决的问题：

(1)数据验证问题：由于区块链具有“公开性”和“透明性”，提交到链上的数据对所有用户都是可见的。而在隐私众包场景下，需要保证工人提交的解决方案对非授权用户以及众包平台的隐私性。相比于经典的众包架构，区块链隐私众包方案针对隐私保护的数据，往往更难进行验证与质量评估。

(2)匿名作恶问题: 由于区块链的“去中心化”特性, 在隐私众包场景下, 工人的匿名身份不受区块链中的任何实体的管理, 不同行为往往就很难关联到同一个实体, 因此更加容易带来公平性问题。如工人可以尝试以匿名身份多次提交同一个结果以骗取报酬, 或实施女巫攻击获得多个匿名身份等, 容易导致众包中的搭便车问题。

(3)资源交互问题: 目前大多数区块链众包方案的基本模式是直接使用区块链平台来代替传统的众包平台。在这种模式下, 请求者往往无法接触到单一区块链平台外的其他平台的众多潜在工人, 工人也无法以单一身份查询、接受其他众包平台上所发布的任务。现有的平台相互隔离的众包模式限制了平台之间的资源交互, 也限制了众包对分布式群体资源的整合。

针对上述问题, 本文提出一种基于联盟链架构的众包方案, 依靠零知识证明实现对加密数据的验证、依靠改进的链接可撤销环签名方案实现可控匿名, 即合法的工人的匿名身份信息得到保护, 恶意的工人的身份信息在被检测到恶意行为后可被揭示。该方案在完成众包基本流程的同时, 可以实现对全流程中解决方案的数据隐私保护和对工人匿名身份的可控保护。本文主要的贡献如下:

(1)在联盟链的架构下, 提出基于零知识证明和链接可撤销环签名的隐私众包方案, 能够在实现数据隐私保护与身份可控保护的前提下, 有效完成隐私众包完整流程。其中, 利用零知识证明实现对数据的有效性验证, 利用链接可撤销环签名改进方案实现工人身份的可控匿名。

(2)链接可撤销环签名的原始方案基于可信假设, 而这种可信假设与区块链的去中心化架构是矛盾的, 因此本文对原始方案进行了修改, 并提出一种用于联盟链架构的去中心化的链接可撤销环签名协议。

(3)对方案进行了安全性分析与实验验证, 结果表明与现有的代表性方案相比, 本文方案在满足隐私性、公平性的同时, 进一步提供可验证性与可控匿名性。同时, 与现有代表性方案相比, 所提方案在效率与性能方面具有一定的优越性。

本文第2节主要介绍区块链、环签名与零知识证明的基础知识; 第3节给出区块链隐私众包的整体架构与安全假设等; 第4节介绍方案的具体流程与细节; 第5节与第6节分别给出方案的安全性分析与性能分析; 第7节对全文进行总结。

2 预备知识

2.1 区块链与智能合约

区块链是一种分布式的数据结构, 是基于P2P

(Peer-to-Peer)网络依据共识协议维护的去中心化分布式账本。智能合约最初由文献[13]提出, 本质上是一段可执行程序, 是一个能在安全环境中自动执行的数字合约。目前, 已经存在多个支持智能合约的区块链平台, 如以太坊、hyperledger fabric^[14]等。通过将智能合约填入到区块链的交易中, 节点可以通过交易提供特定的输入使智能合约自动执行并产生对应结果, 并经由共识机制将结果写入区块链。

2.2 环签名与环签名变体

Rivest等人^[15]首次提出环签名的概念。在这之后, 一些工作关注于环签名的一些变体, 如可链接环签名与可撤销环签名。

可链接环签名^[16]中, 如果两个环签名由同一签名者签名, 则它们可以基于特定的方案被链接起来。在可链接环签名中, 可链接性被强制嵌入到签名中, 而不是签名者自愿添加到可链接环签名中。如果签名者拒绝添加正确的链接信息, 则整个签名都是无效的。可撤销环签名^[17]是环签名的另一种变体, 环中的任何成员都可以尝试揭示出某个签名者的实际身份, 撤销签名者的匿名性。与群签名相比, 可撤销环签名中的成员依旧可以自发形成一个自组织结构, 在撤销成员匿名性时也不需要特定的某个“管理者”提供信息。

除此之外, 还有一些工作, 如文献[18,19], 尝试将环签名的可链接性与可撤销性结合起来。在基于特定假设的情况下, 能够在实现环签名功能的同时, 对重复签名进行检测, 在此基础上进一步撤销对应签名者的匿名身份。

2.3 零知识证明

零知识证明^[20]是一个由证明者与验证者两方参与的协议。证明者向验证者证明某一个论断的正确性, 而不需要泄露除该论断是正确的以外的任何信息。此外, 零知识证明满足以下3个安全特性:

(1)完备性: 若论断为真, 则验证者总是接受证明;

(2)可靠性: 若论断为假, 则验证者总是拒绝证明;

(3)零知识性, 即验证者无法从该证明过程中获取额外的信息。

零知识证明可以分为交互式与非交互式两种, 本文方案采用简明非交互零知识证明(zero-knowledge Succinct Non-interactive ARGument of Knowledge, zk-SNARK)算法^[21]验证数据的有效性。算法主要由3个部分组成: (a)初始设置: 产生公共参数, 运行初始化算法 $\text{Gen}(1^k) \rightarrow \text{crs}$, 以得到公共参考串(Common Reference String, CRS);

(b)承诺：证明者通过承诺算法 $Com(crs, m) \rightarrow (com, \pi)$ 生成对其希望证明信息 m 的承诺 com 以及用于证明承诺的 π ；(c)验证：验证者通过验证算法 $Ver(crs, com, \pi, m)$ 判断证明的完备性与可靠性。

3 系统模型与设计目标

3.1 系统模型

本文所提联盟链架构的隐私众包系统模型如图2所示，以数据收集的众包场景为例，其包含3种实体，具体描述如下：

代理商联盟(brokers)：由不同众包平台中的代理节点(broker)组成的联盟，本质上是联盟链中的记账节点，需要负责用户的注册，需要对工人提交的数据进行有效性检测以及重复提交检测，在检测到重复提交发生后，需要能够披露作恶工人的真实身份。

请求者(requester)：为众包任务的发起者。请求者发布任务并通过工人计算来收集解决方案，同时需要支付一定的报酬。

工人(workers)：为解决方案的提交者。工人执行请求者发布的任务并提交解决方案，同时从请求者处获得报酬。

联盟链中主要部署了4种合约，分别为用户注册合约、公钥存储合约、任务发布合约与奖惩结算合约。其中，用户注册合约主要用于存储安全参数、生成用户的身份标识等；公钥存储合约则负责

记录工人在匿名条件下所对应的身份标识；任务发布合约用于请求者提供任务需求、发布任务并收集解决方案；奖惩结算合约负责管理报酬的发放与可控匿名。

3.2 设计目标

不同于传统的中心化众包方案，本文方案借助联盟链实现隐私众包，结合上述系统模型，方案需要满足如下设计需求：

(1)隐私性：除发布众包任务的请求者以及向其提交解决方案的工人外，其他用户以及系统中的代理节点均无法获取解决方案的真实信息。

(2)可控匿名性：匿名性要求指的是工人正常提交解决方案且不作出重复提交的恶意行为时，其身份(对应的 ID_w)对于整个联盟链系统是匿名的。在此基础上，可控匿名性还保证了当工人实施重复提交等恶意行为时，可以由代理商联盟恢复出作恶工人的真实身份。

(3)可验证性：在结果提交阶段，对于工人提交的加密的解决方案，代理商联盟能够在不知晓实际信息的情况下对数据的有效性进行验证；在奖惩结算阶段，对于请求者在任务发布阶段给定的激励政策，代理节点能够验证分配结果与激励政策是否匹配。

(4)公平性：请求者提供的任务奖励和工人给出的解决方案之间需要做到公平交换。在请求者发布任务后，任务中包含的激励政策应不可被随意篡

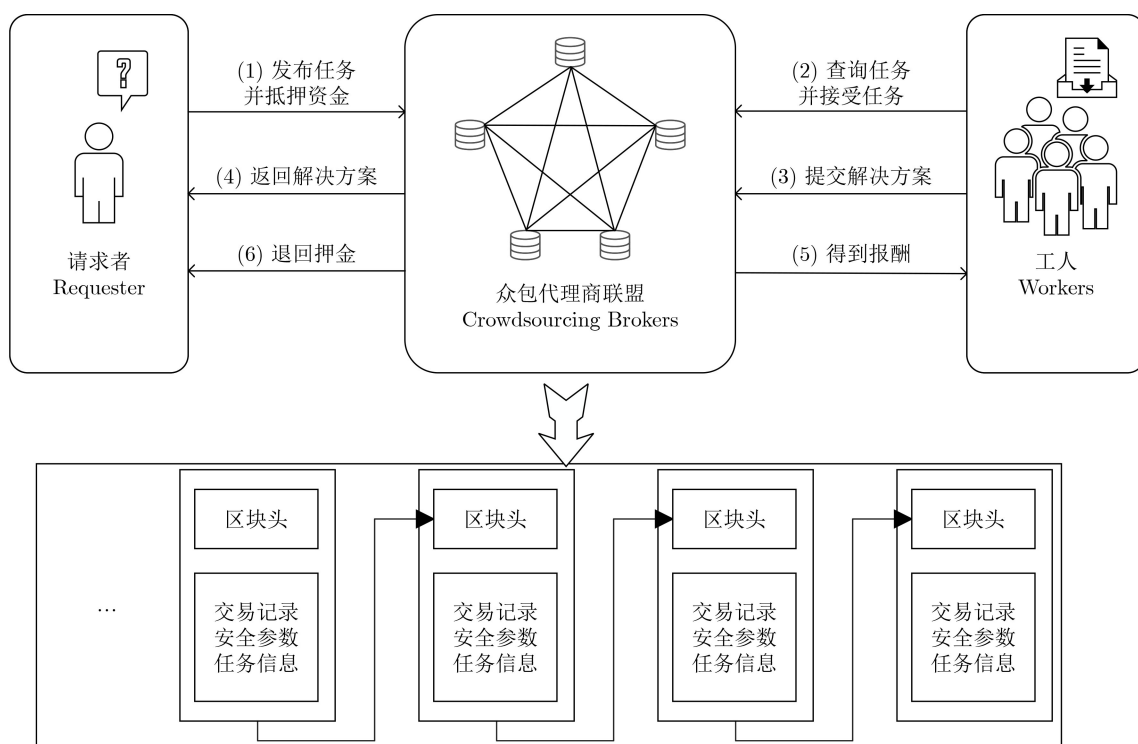


图2 联盟链众包模型

改；在工人提交解决方案后，需保证其解决方案的有效性与唯一性。

4 方案设计

4.1 初始化

初始化阶段前，随机选择一个代理节点 B_i 生成环签名安全参数 λ ，执行 $\text{Setup}(1^\lambda) \rightarrow \text{Param}$ ，并通过去中心化的链接可撤销环签名协议进行计算，将结果上传到区块链中。代理商联盟 B 中其他节点完成共识后，联盟链将用于众包过程的初始安全参数记录在链上。

本文提出的去中心化的链接可撤销环签名协议，基于 Au 等人^[19]的环签名方案改造得到。首先， B_i 选择双线性映射对 $(\mathbb{G}_1, \mathbb{G}_2)$ 满足 $|\mathbb{G}_1| = |\mathbb{G}_2| = p$ ，其中 p 是一个 λ 位的大素数， \mathbb{G}_1 的生成元为 (g_0, g_1, g_2) ， \mathbb{G}_2 的生成元为 (h_0, h_1, h_2) ，满足计算同构性即 $\psi(h_i) = g_i$ 。在此基础上， B_i 继续选择一个阶为 p 的循环群 \mathbb{G}_p ，并定义 $G_0: \{0, 1\}^* \rightarrow \mathbb{G}_p$ ， $G_1: \{0, 1\}^* \rightarrow \mathbb{G}_p$ ， $H: \{0, 1\}^* \rightarrow \mathbb{Z}_p$ 得到3个哈希函数。接着选择 \mathbb{G}_2 的某个生成元 h ，通过计数累加器选择 $q \in_R \mathbb{Z}_p^*$ ，计算 $q_i = h^{(q^i)}$ 。由双线性，可得对 $\forall x \in \mathbb{G}_1, y \in \mathbb{G}_2, a \in \mathbb{Z}_p, b \in \mathbb{Z}_p$ ，都有 $\hat{e}(x^a, y^b) = \hat{e}(x, y)^{ab}$ 。完成上述步骤之后， B_i 将 $\text{Param} = (H, G_0, G_1, \psi, \hat{e}, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_p, p, g_0, g_1, g_2, h_0, h_1, h_2, h, q)$ 上传到用户注册合约中，代理商联盟 B 的其他节点通过共识的方式完成环签名初始安全参数 Param 的生成。

在初始化阶段，众包系统中的某个请求者，记为 R ，首先通过 ElGamal 算法生成公私密钥对 (pk_R, sk_R) 用以在联盟链中发布交易。然后，该请求者需要将 pk_R 发送给注册机构以得到与公钥绑定的证书 $cert_R$ 。进而，通过与代理节点 B_i 进行交互，将 $(pk_R, cert_R)$ 发送给 B_i 进行注册， B_i 验证通过后为其生成联盟链中的身份标识 ID_R ，其中 ID_R 由 R 发送的公钥确定，通过对 pk_R 计算 SHA256 得到，具有唯一性，用于对之后众包流程中的身份进行标识。进一步地， B_i 查询用户注册合约中是否已经存在该身份标识，若不存在则由 B_i 将 ID_R 上传到用户注册合约中。按照相似的过程，众包系统中的某个工人，记为 W_i ，也要先生成自己的公私密钥对 (pk_{W_i}, rk_{W_i}) ，并与某一个注册机构进行交互得到证书 $cert_{W_i}$ ，之后将 $(pk_{W_i}, cert_{W_i})$ 发送给相应的代理节点完成注册，得到唯一的身份标识 ID_{W_i} 。

为了生成环签名， W_i 需要选择随机数 $\gamma, s', r_s \in_R \mathbb{Z}_p^*$ ，并且计算 $w = h_0^\gamma$ ， $C' = g_1^{s'} g_2^{r_s}$ 。接下来继续对自己的身份标识计算得到 $e = H(ID_{W_i})$ ，最后将 (C', e, γ) 这一3元组发送给 B_i 。

B_i 通过随机选择一个随机数 $s'' \in_R \mathbb{Z}_p^*$ ，并计算

$C = C' g_1^{s''}$ ， $A = (g_0 C)^{\frac{1}{e+s'}}$ ，将2元组 (A, s'') 发回给 W_i 。在验证满足关系 $\hat{e}(A, wh_0^e) = \hat{e}(g_0 g_1^s g_2^{r_s}, h_0)$ 后， W_i 最终得到 $rsk_{W_i} = s' + s''$ 。

在初始化阶段，代理商联盟通过共识得到用于后续签名公共参数 Param 并发布在用户注册合约中，将工人的身份标识 ID_{W_i} 发布在公钥存储合约中；请求者 R 得到 (ID_R, pk_R, sk_R) 用于后续的任务发布；工人 W_i 得到 $(ID_{W_i}, pk_{W_i}, sk_{W_i}, rsk_{W_i})$ 用于后续的结果提交。

4.2 任务发布

任务发布阶段开始时，请求者 R 首先需要生成一对用于加密解决方案的公私钥对 (pk_s, sk_s) 。然后， R 锁定一笔酬金 budget 作为任务预算，生成众包任务 T 的信息 taskinfo ， taskinfo 包括：加密解决方案的公钥 pk_s 、身份标识 ID_R 、任务描述、数据要求、任务提交的截止时间 t 、任务提交的最大数量 n 、任务预算 budget 等，通过任务发布合约以交易的形式上传到联盟链中。其中， pk_s 用于工人对解决方案进行加密； ID_R 用来标识用户的身份；任务描述用于工人查找适合自己的任务；数据要求用于代理商联盟对加密数据进行验证； t 规定工人交任务的最终期限，同时防止在任务收集数量不达标时一直等待而无法中止；最大数量 n 指 R 对任务收集量的需求，满足后立即停止任务收集过程，同时也会影响到激励分配。接着， R 计算 taskinfo 的哈希值。然后， R 需要对激励政策 P 通过零知识证明生成一个承诺 com_R ，一起上传到任务发布合约中。任务发布阶段详细步骤如下：

(1) R 生成一对用于加密解决方案的公私钥对 (pk_s, sk_s) ；

(2) R 生成众包任务的信息 taskinfo ， $\text{taskinfo} = (pk_s || ID_R || t || n || \text{budget} || \dots)$ ；

(3) R 计算 $ID_T = H(\text{taskinfo})$ ；

(4) R 通过零知识证明算法 $\text{Gen}(1^k)$ 生成一个公共参考串 crs_R ；

(5) R 对激励政策 P 生成一个承诺，计算 $\text{Com}(\text{crs}_R, P) \rightarrow (\text{com}_R, \pi_R)$ ，得到承诺 com_R 和证明 π_R ，然后将 $\text{taskinfo} || ID_T || \text{crs}_R || \text{com}_R$ 上传到任务发布合约，将其作为交易内容进行广播；

(6) 代理商联盟 B 通过查询用户注册合约验证 R 的合法性，并检查交易的合法性，在共识后将包含任务发布合约的交易添加到区块中完成上链。

任务发布阶段结束后， R 的众包任务 T 以智能合约的形式发布到了联盟链中。工人可以通过对任务发布合约进行查询，确定希望接取的众包任务。

4.3 结果提交

假设有 n 个工人 $W = \{W_1, \dots, W_i, \dots, W_n\}$ 通过查询任务发布合约的内容选择接受众包任务 T 。在截止时间 t 前, W 需要通过本地计算、数据聚合或其他方式得到 R 所需要的解决方案并提交以获得报酬。以某个工人 W_i 为例, W_i 需要用 pk_s 加密解决方案 sol_i 得到解决方案的密文 C_i 。

提交加密得到的 C_i 时, W_i 需要对加密的结果 C_i 执行算法 $Sign(ID_T, v_i, L, s, C_i)$ 得到签名 σ_i 。为此, W_i 需要依次计算 $e = H(ID_W), u_0 = G_0(ID_T), S = u_0^e$ 并通过公钥存储合约获得其他工人的身份标识, 以计算 v 和 v_w

$$\begin{aligned} v &= \psi \left(h^{\prod_{k=1}^{k=n} (q+H(ID_k))} \right), \\ v_w &= \psi \left(h^{\prod_{k=1, k \neq w}^{k=n} (q+H(ID_k))} \right) \end{aligned} \quad (1)$$

完成上述步骤, 继续验证是否满足关系 $\hat{e}(v_w, q_1 h^e) = \hat{e}(v, h)$ 后, W_i 计算 $Q = H(S || C_i || L || ID_T)$, $u_1 = G_1(ID_T)$, $T_i = u_0^e (u_1^Q)^s$, 以得到签名 $\sigma_i = (S, T_i)$ 。最后, W_i 还需要对 C_i 生成一个零知识证明, 与 σ_i 一起提交给任务发布合约, 由合约对工人的解决方案进行验证。结果提交阶段详细步骤如下:

(1) W_i 根据任务描述得到任务的解决方案 sol_i ;

(2) W_i 用 R 在任务发布合约中的加密公钥 pk_s 对 sol_i 进行加密, 得到密文形式的 C_i ;

(3) W_i 需要从用户注册合约获得安全参数 $Param$, 从公钥存储合约中选择大小为 v_i 的工人身份标识符集合 L_i , 对加密的结果 C_i 签名得到 σ_i ;

(4) W_i 通过零知识证明算法生成公共参考串 crs_{W_i} , 对 sol_i 给出承诺 com_{W_i} 与证明 π_{W_i} ;

(5) W_i 将 $C_i, \sigma_i, crs_{W_i}, com_{W_i}$ 和 π_{W_i} 一起发送给任务发布合约, 由代理商联盟 B 进行验证;

(6) 任务发布合约每收集到一个结果, 首先执行算法 $Verify(ID_T, v_i, L_i, C_i, \sigma_i)$, 计算 $u_1 = G_1(ID_T)$, $Q = H(S || C_i || L || ID_T)$ 以判断签名的合法性, 若通过则继续进行可链接性检测算法 $Link(ID_T, v_i, v_j, L_i, L_j, C_i, C_j, \sigma_i, \sigma_j)$ 判断此次收到的结果与合约之前收到的结果的 S 是否相同, 以验证重复提交是否发生。在此基础上, 代理节点继续验证零知识证明是否符合 R 在任务发布合约中的数据要求。上述过程中若有某个工人的结果认证不通过, 则直接跳转到奖惩结算阶段。

当满足任务发布合约中的截止时间 t 的要求或任务收集数量 n 要求时, 结果提交阶段结束。

4.4 奖惩结算

奖惩结算阶段主要分为两种情况, 在未检测到恶意行为时, 当解决方案数量收集到任务发布时的

预先设定值 n 时, R 从任务发布合约获取所有加密的 C_i , 在本地自行解密后得到解决方案 sol , 接着需要按照任务发布合约中已定的激励政策 P 对发送过解决方案的 W_i 发放奖励。

此时, R 需要对承诺进行证明, 向奖惩结算合约发送 π_R 。为了防止 R 篡改激励政策的分配方式, B 需要对承诺进行验证, 判断激励政策与分配结果是否一致, 并将承诺与验证结果发布到奖惩结算合约中, 使得所有用户都可以在奖惩结算合约中查询激励政策与分配结果的一致性。最后, 将 R 在合约中锁定的酬金的余额退还给 R 。

当检测到重复提交的恶意行为时, 代理商联盟 B 通过环签名算法 $Revoke(v_i, v_j, L_i, L_j, \sigma_i, \sigma_j)$ 完成对作恶工人 W_i 的身份揭示。为此, B 首先需要计算参数 $Q = H(S || C_i || L_i), Q' = H(S || C_i' || L_j)$ 。接着通过计算以式(2)尝试恢复作恶工人的匿名身份

$$u_0 = G_0(ID_T), U = \left(\frac{T_i^{Q'}}{T_j^Q} \right)^{\left(\frac{1}{Q' - Q} \right)} \quad (2)$$

对于链接可撤销环签名协议, 当且仅当工人使用相同的身份标识对两个相同任务进行签名时, 可以得到 $U = u_0^{H(ID)}$, 此时 B 可以披露工人 W_i 的 ID_{W_i} , 并将其记录在合约中杜绝 W_i 再次作恶。

5 安全分析

5.1 隐私性分析

方案中的隐私性指的是, 除发布众包任务的请求者 R 以及向其提交解决方案的工人 W 外, 其他用户以及系统中的代理节点均无法获取其解决方案 sol 的真实信息。在整个众包流程中, 与 sol 有关的链上公开信息只有加密得到的密文 $\{C_1, \dots, C_i, \dots, C_n\}$ 以及对应的 W 为其生成的零知识证明。由于上传到联盟链的解决方案密文基于 Paillier 公钥加密体制完成, 基于公钥体制的语义安全性, 密文不会向任何计算能力为多项式有界的对手泄露任何相关有用信息。

以选择明文攻击为例, 假设对手从区块链中获得请求者的加密算法公钥 $pk_s = (n, g)$ 与收到的解决方案的密文 sol 。由于加密算法为一个单向陷门置换函数, 对于解决方案密文 $sol \in Z_{n^2}^*$, 存在唯一的 $r, m \in Z_{n^2}^* \times Z_n$ (其中 r 为随机数, m 为 sol 所对应的明文) 满足关系 $sol = g^m r^n \pmod{n^2}$, m 为 sol 相对于 $pk_s = (n, g)$ 的剩余类。对于任意概率多项式算法, 都存在可忽略函数 $v(k)$ 使得

$$\begin{aligned} \text{prob}[pk_s = (n, g), m \in Z_n, r \in Z_{n^2}^*; \\ sol = g^m r^n \pmod{n^2}; \text{Dec}(n, g, sol) = m] \leq v(k) \end{aligned} \quad (3)$$

在 n 的分解情况未知的条件下, 计算合数的剩余类是很困难的。此外, 用于加密sol的公私钥对由请求者R自行生成, 没有第三方的参与, 生成过程也不需要链上公布。因此敌手在没有 sk_s 的情况下, 通过解密获得解决方案的概率是可以忽略的。综上, 本方案保障了隐私性。

5.2 可控匿名性分析

方案中提供的匿名性基于环签名方案的匿名性。本方案中, 工人 W_i 得到的 $(ID_{W_i}, pk_{W_i}, sk_{W_i}, rsk_{W_i})$ 由工人通过区块链中的参数自行生成。代理节点 B_i 通过验证满足关系 $\hat{e}(A, wh_0^e) = \hat{e}(g_0 g_1^s g_2^{r_s}, h_0)$ 之后才能将对应的工人的 ID_{W_i} 添加到区块链, 任何敌手都不能篡改。因此, 任何概率多项式时间敌手, 都无法通过伪造签名来获得代理节点的验证, 无法根据相应的公钥或密文计算出私钥。综上所述, 本文提出的方案提供了匿名性。

此外, 方案具有可链接性, 如果签名者拒绝添加正确的链接信息, 则整个签名都是无效的。因此借助可链接性检测算法Link($ID_T, v_i, v_j, L_i, L_j, C_i, C_j, \sigma_i, \sigma_j$), 若存在某个工人对同一个众包任务提交了两次解决方案, 代理节点一定能够检测出来。若工人不对同一众包任务恶意提交两次由自己签名的解决方案, 则依旧可以通过环签名保证其匿名性。当且仅当某工人 W_i 对同一任务使用两个相同签名时, 才可以通过Revoke($v_i, v_j, L_i, L_j, \sigma_i, \sigma_j$)恢复出该工人的 ID_{W_i} 。综上, 本方案保证了可控匿名性。

5.3 可验证性分析

本方案中的可验证性主要针对请求者与工人两类实体, 具体指的是: (1)在结果提交阶段, 对于工人提交的加密的解决方案, 代理节点能够在不知晓实际信息的情况下对数据的有效性进行验证; (2)在奖惩结算阶段, 对于请求者在任务发布阶段给定的激励政策, 代理节点能够根据承诺验证分配结果与激励政策是否匹配。

上述两个不同阶段的验证都是基于零知识证明实现的。由于零知识证明的完备性, 对于一个断言为真的承诺, 验证者总是能够接受证明; 由于零知识证明的可靠性, 对于一个断言为假的承诺, 验证者总是拒绝证明。以对零知识证明 $\pi_0 = \text{NIZK}\{(x, y) | X = g^x, Y = g^y\}$ 进行可验证性证明为例:

首先选择随机整数 $(\alpha, \beta) \in \mathbb{Z}_p$ 并计算 $\tilde{X}' = g^\alpha$, $Y' = g^\beta$, $\tilde{Y}' = \tilde{g}^\beta$, $\tilde{Y} = \tilde{g}^y$, 计算挑战值 $h = H(g, \tilde{g}, Y, \tilde{Y}, \tilde{X}', Y', \tilde{Y}')$ 与响应值 $s = \alpha + h \cdot x$ 和响应值 $s' = \beta + h \cdot y$, 并生成零知识证明 $\pi_0 = (h, s, s')$ 。

因为 $s = \alpha + h \cdot x$, $X = g^x$, $X' = g^\alpha$, 所以有 $X'' = \frac{g^s}{X^h} = \frac{g^{\alpha+h \cdot x}}{g^{xh}} = g^\alpha = X'$ 。

因为 $s' = \beta + h \cdot y$, $Y = g^y$, $\tilde{Y} = \tilde{g}^y$, $Y' = g^\beta$, $\tilde{Y}' = \tilde{g}^\beta$, 所以有 $Y'' = \frac{g^s}{Y^h} = \frac{g^{\beta+h \cdot y}}{g^{y \cdot h}} = g^\beta = Y'$, $\tilde{Y}'' = \frac{\tilde{g}^s}{\tilde{Y}^h} = \frac{\tilde{g}^{\beta+h \cdot y}}{\tilde{g}^{y \cdot h}} = \tilde{g}^\beta = \tilde{Y}'$ 。

综上可得 $h' = H(g, \tilde{g}, Y, \tilde{Y}, X'', Y'', \tilde{Y}'') = h$, 即可验证 π_0 是正确的。因此众包方案具有可验证性。

5.4 公平性分析

本方案中的公平性指的是请求者提供的任务奖励和工人给出的解决方案之间需要做到公平交换。在请求者发布任务后, 任务中包含的激励政策应不可被随意篡改; 在工人提交解决方案后, 需保证其方案的有效性与唯一性。

考虑敌手采取如下两种方式进行攻击:

(1)请求者尝试以新的激励政策 P' 分配奖励, 以减少对工人奖励的发放;

(2)工人尝试通过复制其他工人已提交的解决方案以骗取报酬。

对于攻击1, 请求者将taskinfo||ID_T||crs_R||com_R上传到任务发布合约后, 联盟链中已经存在承诺com_R和公共参考串crs_R的情况下, 请求者找到一个新的激励政策 P' 和 π' 来证明承诺com_R的概率是可以忽略的。

对于攻击2, 智能合约接收的结果都是解决方案加密后的密文, 与5.1节中隐私性分析相似, 基于公钥体制的语义安全性, 在没有请求者私钥的情况下, 工人通过他人的提交的密文直接解密出解决方案的概率也是可以忽略的。对于工人而言, 也无法尝试直接复制他人的密文进行提交, 因为其无法对解决方案给出有效的零知识证明。综上, 本方案保证了公平性。相关方案安全性能对比情况如表1所示。

6 性能分析

本文设计了一个基于联盟链的可验证与可控匿名的隐私众包方案。为验证方案的有效性, 对众包流程中涉及到的不同阶段编写了实现对应功能的智能合约, 并将其部署在以太坊Ropsten测试网络上。进一步, 对合约性能消耗进行了测试, 如图3所示, 包括用户注册合约和任务发布合约的部署, 以及相关合约的调用。可以看出, 主要开销集中于

表1 相关方案安全性能对比

安全性能	文献[8]	文献[11]	文献[12]	本文方案
隐私性	否	是	是	是
可控匿名性	否	否	否	是
可验证性	是	否	否	是
公平性	是	是	是	是

合约的部署中，这主要是由于用户在注册以及任务发布时需要在链上存储较多的信息。一旦完成合约部署，用户此后调用合约时所需要的开销明显更小。相比于请求者，工人由于需要在注册时同时返回用于环签名的参数，所需的开销相对更大。

本方案中涉及的环签名算法可通过在链下执行从而减少链上操作的处理开销。在链下对环签名算法性能进行了测试，实验部署在Intel i5-10 400 CPU @2.9 GHz的主机上，测试了100个工人以内的情况下，生成环签名所需的签名时间与验证时间，结果如图4所示。可以看出链接可撤销环签名的时间趋于线性，参与环签名工人数量为100时，签名时间和验证时间仅为40 ms，具有较好的性能开销。

此外，针对结果提交步骤中可能存在的重复提交检测，本文测试了对于同一个众包任务，在待收集的解决方案数量 n 不同的情况下，统计检测出存在重复提交所需的时间，结果如图5所示。尽管检测时间呈现 $O(n^2)$ 的开销，但在提交数量为10 000时，最大检测时间开销也仅为110 ms左右，而当解决方案中存在较多重复提交方案时，所需要的检测开销将会更小，可以满足实际应用的需求。

方案中使用到的零知识证明算法部署在Intel i5-10 400 CPU @2.9 GHz的主机上，运行在版本为

Ubuntu-20.04的虚拟机中，也在链下进行了性能测试。基于libsnark完成零知识证明的生成与验证，图6给出了零知识证明算法的执行开销。以10次为增量分别执行该算法，一直执行到100次，并对结果取平均值。对单个工人而言，生成一个证明的时间约为4.8 s，用于验证对应证明的时间约为33 ms。

7 结论

本文针对隐私众包场景下出现的数据验证问题、匿名作恶问题和跨平台资源交互等需求，基于区块链技术并结合零知识证明与链接可撤销环签名改进方案，提出了一种联盟链架构下的隐私众包方案。依靠零知识证明实现加密数据验证、依靠链接可撤销环签名改进方案实现匿名与可控匿名，在完成众包流程的同时，实现隐私众包所需的数据保护与身份可控保护。本文借助智能合约实现隐私众包各个实体之间的交互，采用链下和链上结合的方式，在联盟链中完成任务发布、结果提交、身份验证等过程，在链下完成公私密钥生成、环签名生成、零知识证明生成与验证、结果解密等流程。安全性分析表明，基于联盟链架构的方案满足隐私性、可验证性、可控匿名性与公平性。最后，通过实验分析了方案的性能开销，验证了方案的有效性。

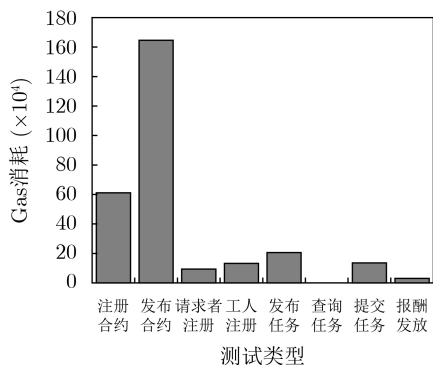


图3 合约部署与合约调用过程的gas消耗

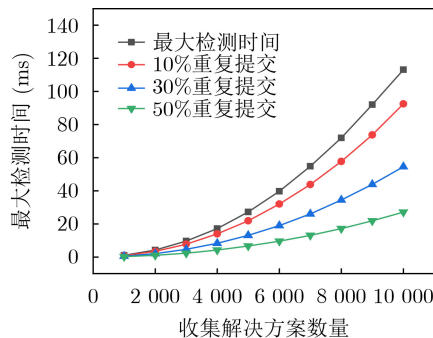


图5 检测重复提交方案的时间开销

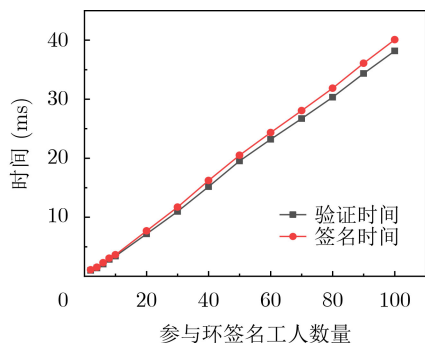


图4 环签名的验证时间与签名时间开销

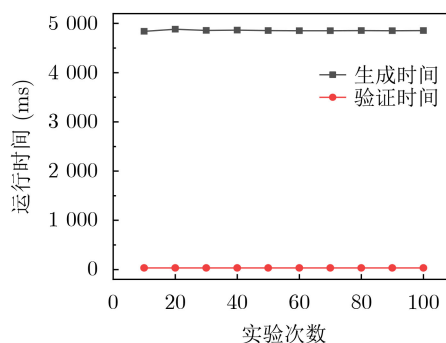


图6 零知识证明的生成时间与验证时间开销

参考文献

- [1] HOWE J. The rise of crowdsourcing[J]. *Wired*, 2006, 14(6): 1–4.
- [2] ZHANG Xiang, XUE Guoliang, YU Ruozhou, *et al.* Keep your promise: Mechanism design against free-riding and false-reporting in crowdsourcing[J]. *IEEE Internet of Things Journal*, 2015, 2(6): 562–572. doi: [10.1109/JIOT.2015.2441031](https://doi.org/10.1109/JIOT.2015.2441031).
- [3] ZHANG Chen, GUO Yu, DU Hongwei, *et al.* PFcrowd: Privacy-preserving and federated crowdsourcing framework by using blockchain[C]. 2020 IEEE/ACM 28th International Symposium on Quality of Service, Hang Zhou, China, 2020: 1–10. doi: [10.1109/IWQoS49365.2020.9212891](https://doi.org/10.1109/IWQoS49365.2020.9212891).
- [4] ZHANG Yu and VAN DER SCHAAR M. Reputation-based incentive protocols in crowdsourcing applications[C]. 2012 Proceedings IEEE INFOCOM, Orlando, USA, 2012: 2140–2148. doi: [10.1109/INFOCOM.2012.6195597](https://doi.org/10.1109/INFOCOM.2012.6195597).
- [5] JACYNYCZ V, CALVO A, HASSAN S, *et al.* Betfunding: A distributed bounty-based crowdfunding platform over ethereum[M]. OMATU S, SEMALAT A, BOCEWICZ G, *et al.* Distributed Computing and Artificial Intelligence, 13th International Conference. Cham: Springer, 2016: 403–411. doi: [10.1007/978-3-319-40162-1_44](https://doi.org/10.1007/978-3-319-40162-1_44).
- [6] ZHU Huasheng and ZHOU Z Z. Analysis and outlook of applications of blockchain technology to equity crowdfunding in China[J]. *Financial Innovation*, 2016, 2(1): 29. doi: [10.1186/s40854-016-0044-7](https://doi.org/10.1186/s40854-016-0044-7).
- [7] MA Haiying, HUANG E X, and LAM K Y. Blockchain-based mechanism for fine-grained authorization in data crowdsourcing[J]. *Future Generation Computer Systems*, 2020, 106: 121–134. doi: [10.1016/j.future.2019.12.037](https://doi.org/10.1016/j.future.2019.12.037).
- [8] LI Ming, WENG Jian, YANG Anjia, *et al.* CrowdBC: A blockchain-based decentralized framework for crowdsourcing[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2019, 30(6): 1251–1266. doi: [10.1109/tpds.2018.2881735](https://doi.org/10.1109/tpds.2018.2881735).
- [9] TAN Liang, XIAO Huan, YU Keping, *et al.* A blockchain-empowered crowdsourcing system for 5G-enabled smart cities[J]. *Computer Standards & Interfaces*, 2021, 76: 103517. doi: [10.1016/j.csi.2021.103517](https://doi.org/10.1016/j.csi.2021.103517).
- [10] TONG Wei, DONG Xuewen, SHEN Yulong, *et al.* CHChain: Secure and parallel crowdsourcing driven by hybrid blockchain[J]. *Future Generation Computer Systems*, 2022, 131: 279–291. doi: [10.1016/j.future.2022.01.023](https://doi.org/10.1016/j.future.2022.01.023).
- [11] LU Yuan, TANG Qiang, and WANG Guiling. ZebraLancer: Private and anonymous crowdsourcing system atop open blockchain[C]. 2018 IEEE 38th International Conference on Distributed Computing Systems, Vienna, Austria, 2018: 853–865. doi: [10.1109/ICDCS.2018.00087](https://doi.org/10.1109/ICDCS.2018.00087).
- [12] ZHANG Chen, GUO Yu, JIA Xiaohua, *et al.* Enabling proxy-free privacy-preserving and federated crowdsourcing by using blockchain[J]. *IEEE Internet of Things Journal*, 2021, 8(8): 6624–6636. doi: [10.1109/JIOT.2021.3051295](https://doi.org/10.1109/JIOT.2021.3051295).
- [13] SZABO N. Formalizing and securing relationships on public networks[J]. *First Monday*, 1997, 2(9). doi: [10.5210/fm.v2i9.548](https://doi.org/10.5210/fm.v2i9.548).
- [14] ANDROULAKI E, BARGER A, BORTNIKOV V, *et al.* Hyperledger fabric: A distributed operating system for permissioned blockchains[C]. The Thirteenth EuroSys Conference, Porto, Portugal, 2018: 30. doi: [10.1145/3190508.3190538](https://doi.org/10.1145/3190508.3190538).
- [15] RIVEST R L, SHAMIR A, and TAUMAN Y. How to leak a secret[C]. The 7th International Conference on the Theory and Application of Cryptology and Information Security, Security Gold Coast, Australia, 2001: 552–565. doi: [10.1007/3-540-45682-1_32](https://doi.org/10.1007/3-540-45682-1_32).
- [16] LIU J K, WEI V K, and WONG D S. Linkable spontaneous anonymous group signature for ad hoc groups[C]. The 9th Australasian Conference on Information Security and Privacy, Sydney, Australia, 2004: 325–335. doi: [10.1007/978-3-540-27800-9_28](https://doi.org/10.1007/978-3-540-27800-9_28).
- [17] LIU D Y W, LIU J K, MU Yi, *et al.* Revocable ring signature[J]. *Journal of Computer Science and Technology*, 2007, 22(6): 785–794. doi: [10.1007/s11390-007-9096-5](https://doi.org/10.1007/s11390-007-9096-5).
- [18] ZHANG Xinyu, LIU J K, STEINFELD R, *et al.* Revocable and linkable ring signature[C]. The 15th International Conference on Information Security and Cryptology, Nanjing, China, 2019: 3–27. doi: [10.1007/978-3-030-42921-8_1](https://doi.org/10.1007/978-3-030-42921-8_1).
- [19] AU M H, LIU J K, SUSILO W, *et al.* Secure ID-based linkable and revocable-iff-linked ring signature with constant-size construction[J]. *Theoretical Computer Science*, 2013, 469: 1–14. doi: [10.1016/j.tcs.2012.10.031](https://doi.org/10.1016/j.tcs.2012.10.031).
- [20] GOLDWASSER S, MICALI S, and RACKOFF C. The knowledge complexity of interactive proof-systems[M]. GOLDREICH O. Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali. New York: ACM, 2019: 203–225. doi: [10.1145/3335741.3335750](https://doi.org/10.1145/3335741.3335750).
- [21] GROTH J. On the size of pairing-based non-interactive arguments[C]. The 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, 2016: 305–326. doi: [10.1007/978-3-662-49896-5_11](https://doi.org/10.1007/978-3-662-49896-5_11).
- 薛开平: 男, 教授, 研究方向为下一代网络体系结构与网络安全。
范茂: 男, 硕士生, 研究方向为区块链与信息安全。
王峰: 男, 中级实验师, 研究方向为身份认证和授权管理、访问控制、区块链系统。
罗昕怡: 女, 博士生, 研究方向为隐私计算和区块链技术。