https://crad.ict.ac.cn
62(1): 269−280, 2025

面向电商联盟的区块链营销标签交易系统

代炜琦[1,2,3,4,5] 李 铭[1,2,3,4,5] 赵珂轩[3,4,5] 姜文超[8] 周蔚林[9] 邹德清[1,2,3,4,5] 金 海[1,2,6,7]

[1](大数据技术与系统国家地方联合工程研究中心(华中科技大学) 武汉 430074)

[2](服务计算技术与系统教育部重点实验室(华中科技大学) 武汉 430074)

[3](分布式系统安全湖北省重点实验室(华中科技大学) 武汉 430074)

[4](湖北省大数据安全工程技术研究中心(华中科技大学) 武汉 430074)

[5](华中科技大学网络空间安全学院 武汉 430074)

[6](集群与网格计算湖北省重点实验室(华中科技大学) 武汉 430074)

[7](华中科技大学计算机科学与技术学院 武汉 430074)

[8](广东工业大学计算机学院 广州 510006)

[9](数安时代科技股份有限公司 广东佛山 510100)

(wqdai@hust.edu.cn)

Blockchain Marketing Label Trading System for E-Commerce Alliance

Dai Weiqi[1,2,3,4,5], Li Ming[1,2,3,4,5], Zhao Kexuan[3,4,5], Jiang Wenchao[8], Zhou Weilin[9], Zou Deqing[1,2,3,4,5], and Jin Hai[1,2,6,7]

[1](*National Engineering Research Center for Big Data Technology and System* (*Huazhong University of Science and Technology*), *Wuhan* 430074)

[2](*Services Computing Technology and System Lab* (*Huazhong University of Science and Technology*), *Wuhan* 430074)

[3](*Hubei Key Laboratory of Distributed System Security* (*Huazhong University of Science and Technology*), *Wuhan* 430074)

[4](*Hubei Engineering Research Center on Big Data Security* (*Huazhong University of Science and Technology*), *Wuhan* 430074)

[5](*School of Cyber Science and Engineering*, *Huazhong University of Science and Technology*, *Wuhan* 430074)

[6](*Cluster and Grid Computing Lab* (*Huazhong University of Science and Technology*), *Wuhan* 430074)

[7](*School of Computer Science and Technology*, *Huazhong University of Science and Technology*, *Wuhan* 430074)

[8](*School of Computer Science and Technology*, *Guangdong University of Technology*, *Guangzhou* 510006)

[9](*Global Digital Cyber Security Authority Co.*, *Ltd.*, *Foshan*, *Guangdong* 510100)

Abstract In the era of big data e-commerce, data trading can enable collaborative sharing and value utilization of isolated data resources. As the main form of data trading in e-commerce business, marketing tags have enormous potential value. However, the traditional data trading market faces three main problems: 1) The opaque information of the centralized platform leads to trust crisis and malicious bidding ranking; 2) Lack of reasonable incentive mechanism to break the data island leads to data non-circulation and sharing difficulties; 3) Data security threats lead to privacy disclosure and data reselling and theft. In order to solve these problems, a blockchain marketing label trading mechanism for e-commerce alliance is designed, and the upper consensus incentive mechanism is designed based on decentralization, and all data transactions and computing businesses of the system are completed in combination with the trusted execution environment, thus realizing a safe and complete data transaction ecosystem. The authenticity verification mechanism is designed to ensure the effectiveness of marketing labels, the consensus

收稿日期: 2023-03-31; **修回日期**: 2023-12-11

基金项目: 国家重点研发计划项目(2019YFB2101700); 国家自然科学基金项目(62072202)

This work was supported by the National Key Research and Development Program of China (2019YFB2101700) and the National Natural Science Foundation of China (62072202).

通信作者: 邹德清(deqingzou@hust.edu.cn)

incentive mechanism is designed to enable users to actively share data, and the smart contract is used to effectively constrain the role behavior according to the system design specification; Then key transmission and data security storage are realized through SGX remote authentication, and smart contract security call is realized to ensure user privacy and data security; Finally, the reliable delivery of data transaction results is realized through the trusted computer system and system design idea. In order to verify the security and practicability of the system, 350 000 pieces of real data provided by an e-commerce company are used for performance testing. The test results show that the system can guarantee the security and performance requirements at the same time, and its additional costs mainly come from the remote authentication module and are within the acceptable range.

Key words blockchain; data transaction; e-commerce alliance; trusted execution environment; consensus mechanism

摘 要 大数据电商时代,数据交易可使彼此孤立的数据资源得到协同共享与价值利用,营销标签作为电商业务中数据交易的主要形式,拥有巨大的潜在价值.然而传统数据交易市场面临3个主要问题:1)中心化平台信息不透明导致信任危机和恶意竞价排名;2)缺乏合理的激励机制来打破数据孤岛导致数据不流通和共享困难;3)数据安全威胁导致隐私泄露和数据倒卖盗卖等问题.为解决这些问题,设计了一种面向电商联盟的区块链营销标签交易机制DSTS(decentralized data security transaction system),以去中心化为基础设计上层共识激励机制,结合可信执行环境完成系统各项数据交易和计算业务,从而实现了一个安全完备的数据交易生态.通过真实性验证机制确保营销标签有效性,设计共识激励机制使用户积极共享数据,利用智能合约对角色行为按照系统设计规范进行有效约束;随后通过SGX(software guard extensions)远程认证实现密钥传输和数据安全存储,实现了智能合约安全调用来保障用户隐私和数据安全;最后,通过可信计算机制和系统设计思想,实现了数据交易结果的可靠交付.为验证系统的安全性和实用性,采用某电商公司提供的35万条真实数据进行性能测试,测试结果表明系统能够同时保证安全和性能需求,其额外开销主要来自远程认证模块且在可接受范围内.

关键词 区块链;数据交易;电商联盟;可信执行环境;共识机制

中图法分类号 TP309.2

DOI: 10.7544/issn1000-1239.202330217 **CSTR:** 32373.14.issn1000-1239.202330217

大数据时代,各行各业应用数据大规模生产、分享和使用的方式已经深入到生产生活的每个角落,而大数据使用过程中的数据确权、安全交易、隐私保护等问题始终是制约当前大数据交易和数据价值合理利用和开发的重要因素.如何保证数据产权明晰、确保数据能够安全可靠地交易流通、充分发挥数据潜在价值是数据交易市场亟待解决的关键问题[1].

营销标签是客户特征或喜好的摘要[2],在电子商务中占有非常重要的地位.通过营销标签可以筛选客户特性,从而实现精准营销、市场热点分析等.但在实际营销过程中,用户信息的海量累积和对营销标签加工方式的不足,造成了用户数据难以流通和隐私泄露等问题.同时,随着数据安全和隐私保护意识越来越被人们关注和重视,当前传统电商平台的个性推荐、消息推送等功能总是被怀疑存在隐私泄露和侵权问题[3].此外,由于中心化平台掌握着关键数据,导致数据的使用不仅无法监管,还通过中心化排名和优惠活动导致商家和用户与电商平台绑得越来越紧.这种对数据的不合理使用引发中心化平台数据不流通、存储不安全、隐私泄露[4]等诸多弊端.中心化系统不公开透明且可被篡改,使得用户信息不对称、数据流动困难、买卖不诚信.所以需要一种更安全的去中心化数据交易系统,弥补传统数据交易平台的缺陷,从而保障电商行业数据交易的真实性和安全性,区块链技术为这一问题的解决提供了可行思路.

以区块链为底层架构搭建的去中心化平台大多利用智能合约实现用户功能,通过共识机制维护平台的公平运作,围绕大数据交易安全生命周期[5]展开工作.在此基础上结合数据交易流程及隐私保护需求,分析目前去中心化数据交易平台存在3个问题:

1)目前营销标签仅限于独立的电子商务平台或

商家，建设标准不统一且不具备作为数据商品用来交易的条件. 加上数据安全、信用风险等问题导致数据交换和信息共享困难，缺乏一种激励手段鼓励商家之间共享或交易数据[6]. 产品和服务难以实现附加价值[7]，甚至监管都会成为问题.

2）平台本身缺少对数据溯源和安全验证的能力，在数据共享过程中存在源数据造假问题，导致诚信危机. 假数据的流通会浪费平台资源，造成多方面损失.

3）在数据商品交易过程中，源数据的主体总是存在争议[8]，由于数据商品不具备实体特征的特殊性，恶意数据买家会转卖数据商品，或者对营销标签营销之后的收益进行抵赖.

针对以上 3 个问题，本文提出基于 Intel SGX 的去中心化数据安全交易系统，以保护数据交易过程以及智能合约的执行过程，防止原数据泄露并保护买卖双方交易过程的隐私信息. DSTS（decentralized data security transaction system）利用区块链搭建数据交易平台，通过智能合约规范平台中各角色的行为，设计共识机制打破平台数据孤岛，利用 SGX 保障数据传输和源数据计算过程，最终实现营销标签安全可靠交易.

1 相关工作

保证资源可靠共享是解决传统电商平台数据服务痛点的关键，当前电商平台各成一体，导致数据不流通、共享困难、操作不透明、信任危机、数据共享不安全等一系列问题. 实现电商营销标签安全交易是需要在数据真实的前提下，通过有效的共享激励机制，完成一系列数据安全计算和结果可靠交付的过程.

1）标签的价值评估和共享激励机制

区块链在营销标签交易过程中保证了数据的真实性并在交易双方间建立信任[9]，从而产生了交易的价值. 区块链共识机制在平台中可以使互不信任的多方用户对其数据、行为或流程达成一致，而由此制定的共识机制是保障区块链网络安全稳定运行的关键. 目前区块链领域的主流共识机制有多种，但不存在一套完美的机制满足所有人的价值观. 尤其是在去中心化系统中，目前主要的激励机制有工作量证明[10]和权益证明[11]，前者在达成共识过程中消耗大量能量，后者容易产生中心化问题，阻碍节点之间数据的流通.

2）交易真实性验证

在营销标签交易过程中，数据共享者可能存在营销标签造假问题，例如虚构一个客户及其标签来充当数据商品获取不正当收益. 但是面对虚构的客户数据购买者难以验证标签真实性，而交给第三方验证也可能存在数据泄露以及买家与数据交易平台合谋的诚信问题. 目前对于交易数据真实性保障工作比较少，传统的营销标签共享需要依赖中心化数据交易所，例如贵阳大数据交易所、Factual[12]为确保数据源真实有效，要提前筛选具备优质数据源的单位，费时费力且需要可信第三方背书. 此外，私人数据访问控制系统 PrivacyGuard[13]虽然保障了个人数据的隐私安全，但仍不能确认数据提供者的可信性. 针对社交数据外包场景，有研究者提出基于平衡哈希树的验证框架[14]对不诚实或恶意的数据提供商添加、删除和修改原始数据行为的验证，但该验证方法在数据交易场景中存在滞后性. Liu 等人[15]提出了一种新的用于双重认证的密码原语验证机制，并为数据赋予了价值，但是在电商大数据场景中，管理公钥证书不能保证效率，且频繁非对称加密计算会造成算力浪费. 区块链作为由多方协商一致进行维护的分布式账本，在数据真实性保障工作中展示了不可篡改、高可用性和多方协作支持的安全特性.

3）数据安全交易机制

对于数据交易安全，目前结合区块链技术的数据交易过程保障工作大致可以分为 3 类：第 1 类使用智能合约方法，Li 等人[16]提出将数据交易方案以智能合约形式部署在系统中，而待售的数据被加密存储在系统中，交易内容则是解密密钥. 这样实现了链下数据解密工作，虽然减轻了去中心化平台的压力，但仍无法防止买家对原始待售数据进行转卖. 第 2 类是通过委任仲裁的方法，该类方法根据仲裁手段可以进一步分为 2 种：一种为基于区块链的个人数据交易隐私保护方案[17]，该方案设计了一种去中心化个人数据交易系统，但是系统交易需要买家和卖家无条件信任一个仲裁方，而这种仲裁方式依然存在中心化隐患. 另一种分析了区块链系统中的共识机制并实际实现了一个基于区块链技术的物联网应用点对点交易平台，虽然从去中心化的物联网层面保障了数据真实性，但无法防止买家倒卖真实数据等问题[18]. 所以在可靠数据交易过程中，原始数据对包括平台本身在内的各个角色都是不可见的. 这就需要一种可信执行环境对该过程进行安全隔离，Intel SGX[19]技术的远程认证[20]和密封功能[21]可以实现对加密存储后的原始数据进行加解密以及对计算过程执行节点的安全保障. 第 3 类是基于 SGX 制定的一

系列应用于数据交易的可信计算方案[22]. SDABS[23]是在云平台存储数据过程中,通过结合区块链 SGX 安全审计方案保障远程数据完整性,使区块链公开透明特性和 SGX 密封功能在特定场景中得到完美融合. 此外, SDTE[24] 提出一个数据交易生态系统,通过 SGX 封装以太坊虚拟机,对数据交易系统中运行的智能合约进行有效隔离. SDTE 需要数据买家部署智能合约,也会增加一系列不确定的安全隐患或功能限制,交易行为不够灵活,所以数据交易需要实现权责明确、流通便捷和安全可靠的目标.

2 DSTS 系统设计

本文的目标是设计一个面向电商联盟的营销标签交易系统,致力于有效激励用户共享营销标签的同时保护隐私安全,保障标签有效性的同时确保交易过程安全可靠,并在交易完成后能够防止购买者可能存在的交易抵赖行为.

2.1 系统设计

从数据交易流程和平台运行全生命安全周期角度出发, DSTS 系统如图 1 所示在传统电商平台基础上搭建,并包含了数据交易平台和可信执行环境. 同时在传统电商平台的基础上引入 3 个重要角色,分别为数据买家、数据卖家和电商联盟. 此外,传统电商平台的用户作为标签的主体,既是数据提供者又是数据交易的推送目标,是 DSTS 运行过程中的输入来源和输出目标. 同时,数据买家和数据卖家又是传统电商平台中的商家,他们可以通过传统电商平台中的商品订单获取源数据,也可以通过数据交易平台出售或购买数据来增加积分收入. 其中,源数据在 DSTS 中是传统电商平台中所产生的订单信息,是数据交易平台的数据来源. 电商联盟则是一个可信的去中心化第三方,起到了数据共享激励、合约部署、可信计算以及数据真实性验证的作用. 电商联盟由平台中所有数据卖家共同组成,他们共享数据至电商联盟,同时也是买家和数据交易平台的媒介. 电商联盟作为系统核心角色,由于其去中心化特性,在平台中的各项功能都通过智能合约实现,下面将围绕电商联盟展开介绍 DSTS 的详细设计.

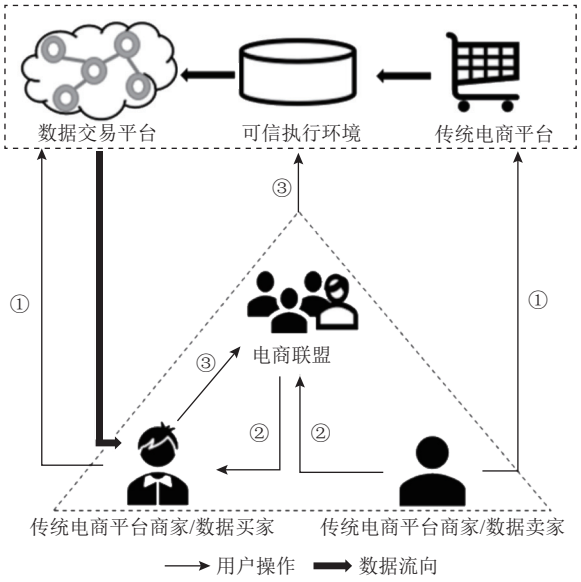


Fig. 1 DSTS process design

图 1 DSTS 流程设计

首先,数据在步骤①中从传统电商平台的交易中产生,系统源数据随着传统电商平台用户与传统电商平台商家之间的传统商品交易订单的确认产生,脱敏后的数据信息为该笔订单的商家所有. 此时,为激励传统电商平台商家积极共享自己手中的订单数据信息,需要引入共享积分作为系统中数据价值的评估指标. 共享积分由电商联盟经营,数据卖方通过共享自己手中的数据获取,同时买方通过支付共享积分获取电商联盟的数据,是系统中流通的价值主体并且在系统之外不具备任何意义. 通过共享积分对平台进行维护,解决了现有共识机制消耗大量资源的问题,同时激励商家共享数据.

其次,在步骤②系统有效的激励机制下,数据拥有者将作为数据卖家共享数据至电商联盟并获取相应的共享积分. 此时为防止数据共享者恶意共享虚假数据以及出现大量重复无用数据,电商联盟要对数据进行真实性验证来确认该数据是来自传统电商平台且交易来自该数据的共享者,若验证不通过则共享失败,并扣除一定的共享积分. 真实性验证通过后共享者获取一定共享积分,此时订单数据才可以作为营销标签在系统中流通. 电商联盟利用可信执行环境保管着整个平台的营销标签并密封存储保证数据安全,同时电商联盟还保证了数据的可靠计算,以数据集中保管的方式打破数据拥有者之间的信息孤岛,降低系统冷启动风险并提升标签计算的准确性,极大地提升了数据共享价值和标签计算的准确性. 然后,电商联盟通知数据买家.

最后,步骤③根据系统设计目标,数据买家需要作为传统电商平台商家以推送产品优惠券或广告形式向电商联盟提交数据交易请求. 买家根据电商联盟提供的数据交易规则进行标签属性选择,如用户属性或喜好选择以及推送数量等,该数量也就是本

次数据交易的数据量，而标签属性的复杂程度以及购买的数据量会作为本次交易计费的准则，最终以共享积分形式向买家收取. 此时电商联盟的交易合约做出响应，并验证此时交易环境的安全性，验证通过则按照电商联盟中的智能合约规则以及隐私保护方案进行标签计算. 同时利用可信执行环境封装以太坊虚拟机执行，且智能合约都由电商联盟提前部署至系统中，买家根据自身需求调用合约处理交易数据，可以防止恶意合约代码的入侵，并确保不泄露源数据和中间代码执行信息. 电商联盟将计算后得到的一系列满足相关属性的用户标识作为计算结果，交付至数据买家. 买家得到满足既定标签和数量的用户标识列表后，通过推送相应优惠券或广告的形式完成数据交易. 同时由于数据商品特殊的存在形式，利用区块链公开且不可篡改特性作为链上背书，设计交易防抵赖机制确保每一笔数据交易的有效性.

总体来说，DSTS 在传统电商平台基础上，形成了一个完整可靠的面向电商联盟营销标签交易的生态系统，其主要优势包括 3 点：

1)针对数据本身，数据来源于传统电商平台，通过数据买家发起数据交易和可信计算，整个过程源数据不会暴露给外界，且最终以差分隐私的形式作用于传统电商平台并不会暴露给数据买家.

2)针对平台及各角色的收益，引入共享积分概念. 平台用户通过共享数据获取共享积分的同时要通过消费共享积分购买数据，共享积分本身并没有实际意义且不会独立于数据交易平台存在，但会通过消费共享积分进行精准推送实现在传统电商平台上增加附加收益. 同时，若平台中存在恶意共享数据行为，还会触发惩罚机制扣除该用户的共享积分.

3)针对标签的可信计算，引入 SGX 作为平台可信执行环境，封装智能合约执行以及标签计算过程. 克服了去中心化平台公开透明特性在隐私保护环节的薄弱问题，且在平台智能合约执行时引入可信执行节点. 可信执行节点是在常规区块链网络节点基础上使用 SGX 保护的以太虚拟机(Ethereum virtual machine, EVM)，可信执行节点通过传统 EVM 运行常规合约，如：操作上链、数据验证合约执行. 使用 SGX 保护的 EVM 执行标签生成模块合约. 通过部署数据交易管理合约背书买家对营销标签的使用情况可以防止其对数据交易过程的抵赖和欺诈行为.

2.2 营销标签生成与匹配

本文设计了一套完备的营销标签生成机制，保证将原始订单信息转化为可以交易标签的同时有效降低隐私泄露风险. 具体来说，生成营销标签过程分为数据共享激励和共识机制、数据真实性验证 、营销标签生成与计算 3 个核心步骤.

1)数据共享激励和共识机制. 数据共享激励和共识机制流程图如图 2 所示. 其作为系统核心功能设计之一，主要负责交易打包、区块头信息计算以及根据共享积分出块，起到激励平台用户提供真实的交易源数据和维持平台稳定运行的作用. 由于用户在链上以智能合约方式实现数据共享、交易等行为，因此可以通过合约执行来同步用户交易获得的共享积分. 而共享积分在平台之外不具备任何价值，且存在平台冷启动和维持可信节点计算问题，所以共享积分价值评估将分为初期、启动期和运营期 3 个阶段进行并根据真实性验证模块信息将共享虚假数据的区块加入恶意区块. 根据共识机制提出数据共享激励机制和共识出块算法，改良平台交易处理过程的

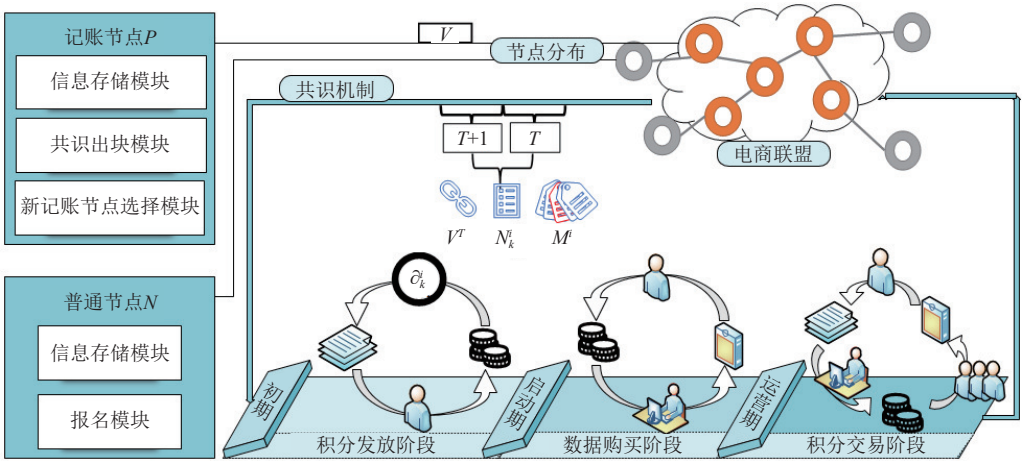


Fig. 2 Incentive and consensus mechanism for data sharing

图 2 数据共享激励和共识机制

同时保证交易公平. 在平台启动前需要在创世区块中写入第 1 批记账节点地址和初始阶段的系统信息, 随后在运行初期通过为共享数据的用户发放共享积分来获取一定的数据和用户积累, 此时电商联盟需要对共享积分进行价值评估并标定其价值, 使其能够体现数据价值和平台算力消耗的同时, 激励初期用户积极共享数据. 对此设计了基于共享积分的共识激励算法来维护平台的稳定运行, 具体如算法 1 所示.

算法 1. 共享积分共识激励算法.

输入: 当前时段 *Time*, 用户 *A*, 共享数据量 *Num*, 数据属性 *Attri*, 评估系数 *COE*, 全局变量 *gasUsed*;

输出: 共享数据所获得的共享积分 *R*.

① *A_transaction*←*txPool*(*Num*, *Attri*, *Time*); /*从交易池中取出用户 *A* 的交易信息*/

② *gas*←*EVM*(*A_transaction*); /*调用 *EVM* 执行交易获得所消耗的 *gas**/

③ *gasUsed* += *gas*;

④ if *gasUsed* ⩾ *gasLimit*

⑤ 返回到 ①;

⑥ else

⑦ *block*←*pack*(*A_transactions*); /*打包出块*/

⑧ end if

⑨ *COE*←*calculate*(*block*, *timestamp*, *t*); /*根据区块信息计算当前阶段评估系数*/

⑩ *R*←*calculate*(*COE*, *Num*, *Attri*). /*调用 *calculate* 计算积分*/

算法 1 通过当前交易的时间、交易的数据量及其属性作为输入, 调用当前系统评估系数 *COE* 来完成本次数据交易所消耗的共享积分. 首先计算本次交易所消耗的 *gas* 是否超过交易限制; 接着对符合条件的交易打包出块并计算评估系数; 最后用当前阶段计算得到的相关系数, 获取本次数据交易需要的共享积分. 根据系统运行规则和平台运行特点对步骤⑨中评估系数计算模块进行设计和实现. 在某一时段 *Time* 中, 商家通过共享 *Num* 条具有 *M* 个属性的数据共享积分, 此时价值评估系数为 ∂, 则该时段获得的积分奖励 *R* 可以由式(1)计算得出. 根据平台运行情况, 需要对评估系数 ∂_i 在不同阶段根据既定的运营时间周期 T 的发币数量 V 以及该时段额定数据交易量 N 进行调整. 根据式(2)所示, 下一阶段的评估系数 ∂_{k+1}^{i+1} 是利用该时段中具有属性 k 的 N_k^i 条数据与各属性数据的平均值 N^i/M^i 差的绝对值来调节电商联盟中各商品数据类型, 达到最终平台中各类数据的数量均衡, 从而完成各个运行周期指标.

$$R_i = N \times \sum_{k=1}^{M} \partial_k^i, \tag{1}$$

$$\partial_{k+1}^{i+1} = \frac{V^T}{V} \times \partial_k^i \Big/ \left(\left| N_k^i - \frac{N^i}{M^i} \right| \right). \tag{2}$$

如果该阶段电商平台共享糖果类数据较多, DSTS 则降低该类数据的评估系数, 并提升其他类型商品评估系数调节电商联盟数据类型的均衡, 从而实现运营期数据交易过程中数据的完整性和计算的准确性.

在平台启动期用户可以按照初期标定的营销标签价格来购买数据, 此时平台中出现数据买家消费共享积分向电商联盟购买营销标签的交易行为. 在平台积累一定用户量和数据量后, 平台进入运营期, 此时电商联盟通过售卖共享积分获取额外收益来维持平台算力消耗, 同时用户也可以通过购买共享积分购买更多更优质的营销标签.

全网区块链节点分为记账节点和普通节点, 其中, 记账节点通过共识计算来维护全局账本, 而普通节点通过查看完整共识过程监督平台运行, 同时也可以通过报名方式成为记账节点. DSTS 用户在平台中通过智能合约完成出售和购买行为, 所以合约对应的价值增量是当前商户所获利润. 同时设计基于价值评估的共享激励机制, 通过智能合约完成数据交易行为并对该过程产生的利润进行价值评估, 使主体自发地参与到计算和数据贡献过程中.

2)数据真实性验证. 为从电商联盟中获取可以

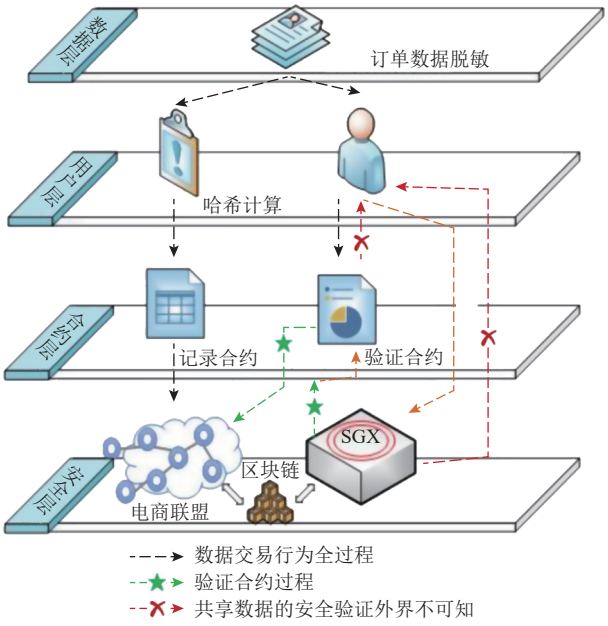


Fig. 3 Data authenticity verification

图 3 数据真实性验证

进行交易的共享标签，首先要确保标签是从真实的商品交易行为中产生的. 如图 3 所示，根据数据的流向和操作状态自上到下分为数据层、用户层、合约层以及安全层. 营销标签在上链之前需要一系列的真实性验证操作，避免共享者存在营销标签造假和依赖第三方验证的诚信问题. 由于数据交易中的数据来自传统商品订单，所以在订单确认时，将订单信息的哈希值以及脱敏信息上传到以太坊中，这样可初步实现分布式账本的信用背书，在链上实现真实性验证.

首先，平台需要分别维护验证合约和交易操作记录合约，验证合约对脱敏后的数据进行哈希并与链上相应的哈希值作比较，交易操作记录合约则是将脱敏后的原始数据和对应哈希值上链. 当卖家共享数据时，首先需要进行节点 SGX 安全性验证，确认环境安全后在节点中解密并进行计算，该过程对外界不可见. 得到数据哈希后，平台会将链上对应传统订单的哈希值进行对比，若哈希值相同，则说明该笔订单真实发生在该商家与买家之间，数据为真实订单产生的数据；若不同，则上传数据失败，说明存在虚假数据行为并扣除该共享者的共享积分. 通过标签真实性验证机制确保数据交易中每一条订单信息的真实性，且整个系统和数据买家都可以在不提供原始数据的情况下对数据真实性进行验证.

3）营销标签生成与计算. 订单数据包含客户属性和客户喜好 2 类标签，其中客户属性将客户分类，客户喜好代表客户与商品之间的关系. 根据目前的应用场景，对标签属性的分类及表示方法如表 1 所示. 目前平台选取如表 1 所示的 7 个属性作为当前每一笔订单的有效信息，分别从客户的性别、年龄段、地区以及该笔订单所包含的优惠情况、商品分类、价格区间和使用的套餐名称 7 个方面为用户贴上相应标签. 当前在标签计算过程中，7 种属性以向量作为表示形式，其中客户属性与客户喜好是与或运算的关系. 如客户 Z 的属性标签 $Cp_Z=\{C\_gen_Z, C\_age_Z, C\_addr_Z\}$，喜好标签 $Ci_Z=\{G\_d_Z, G\_class_Z, G\_pr_Z, G\_pn_Z\}$，则 Z 的营销标签 $Ml_Z=Ci_Z\cap Cp_Z$.

Table 1　Order Attribute ID

表 1　订单属性标识

标签类别	属性	标识方法	
		函数	函数的含义
客户属性	性别	*C_gen*	customer gender
	年龄段	*C_age*	customer age
	地区	*C_addr*	customer address
客户喜好	优惠情况	*G_d*	discount
	商品分类	*G_class*	goods classification
	价格区间	*G_pr*	goods price range
	套餐名称	*G_pn*	package name

客户属性是固定标签且一般不会改变，这里的标签生成只涉及到词条搜索与匹配，可以实现对具有某种属性的客户进行精准推送. 如需要对 20 岁左右男性用户推送某新品糖果广告，则在广告推送前从年龄段和性别 2 个表单中选择标签. 对于客户喜好标签的生成需要一定的用户订单数据积累，如用户 *A* 在 7 天内购买了 2 次商家 *B* 的巧克力组合套餐，那么 *B* 可以在推送优惠券时选择商品分类和套餐名称 2 个标签. 这种标签构造方法降低了计算复杂成本，同时提升了系统的计算效率和准确率.

2.3　隐私保护方案设计

隐私保护方案设计是在基于区块链的数据交易平台中使用 SGX 技术进行电商联盟数据及智能合约运行安全隔离的可信计算机制，并利用 SGX 远程认证功能实现平台数据真实性验证和数据交易过程中可信节点的认证和密钥安全传输，利用 enclave 的安全隔离特性实现对智能合约执行环境的保护. 隐私保护方案如图 4 所示，分为 3 个模块：1）密钥安全传输；2）合约安全执行；3）标签推送列表的交付. 其中，密钥的安全传输和存储是由 SGX 远程认证功能实现，智能合约安全执行通过 SGX 密封功能在智能合约执行环境 EVM 中实现 AES 算法加解密并通过 EDL 接口实现安全交互，最后标签列表交付模块使用差分隐私的思想输出需要推送的用户列表.

系统的安全传输由 SGX 远程认证功能实现. SGX 远程认证不仅能确认节点上正在运行的 enclave 代码的完整性和平台硬件环境的安全性，还能建立一条用户到 enclave 的加密信道，既能防止平台利用

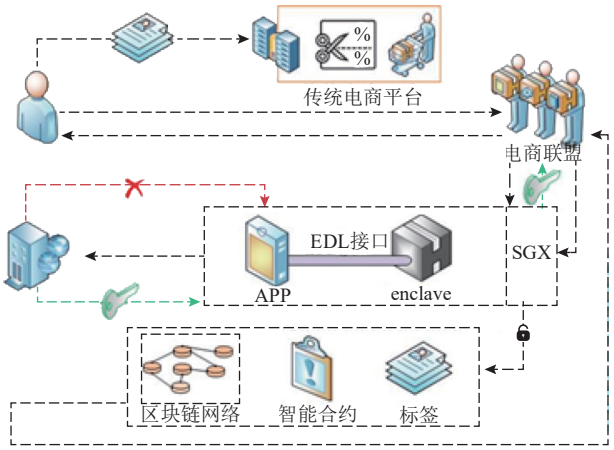


Fig. 4　Privacy protection scheme design

图 4　隐私保护方案设计

已有漏洞攻击 enclave, 也能保证数据和标签生成计算的私密性. 根据 SGX 使用规则, 在安全传输之前要进行远程认证来确认当前区块链节点的可信性. 从传统电商平台发起, 数据买家提交订单后, 电商联盟作为数据所有者需要通过 SGX 向 Intel 官方发送请求来确认当前执行交易的区块链节点是否安全. 此时会收到 Intel 官方发送的远程认证报告, 若报告表示环境可信则继续交易, 若不可信则放弃该节点并终止交易. 此时电商联盟具备了交易环境, 开始响应买家提交的订单申请, 并执行本次交易. 用于解密当前电商联盟源数据的密钥封装在远程认证报告中并送至 enclave 中, 用来解密源数据并进行下一步的计算, 具体如算法 2 所示.

算法 2. 智能合约安全执行算法.

输入: 合约地址 *addr*, 全局存储变量 *state*, SGX 挑战值 *C_value*, 合约输入值 *data*;

输出: 合约执行结果 *Results*.

① *T_contract*←*analysis*(*addr*, *state*); /*解析当前合约得到合约状态及合约调用情况*/

② if *T_contract* and *calling_contract* do not exist in EVM

③ 禁止调用外部智能合约并退出合约执行算法;

④ end if

⑤ *RA*←send *C_value* to Intel SGX; /*解析当前合约得到合约状态及合约调用情况*/

⑥ if *RA* is illegal/*如果得到不正确的远程认证报告*/

⑦ 退出并当作恶意节点;

⑧ end if

⑨ *Results*←*Trusted_node*(*addr*, *RA_publickey*, *data*). /*在可信节点内执行数据解密并输出计算结果*/

在步骤 3 中, 算法 2 禁用了外部合约的调用是为了防止恶意合约窃取中间数据, 系统在这一步骤中对合约调用函数的执行增加了安全筛查来禁止调用系统外部的 *CALL*, *CALLCODE* 等智能合约调用指令功能, 关闭了 SGX 内部执行的传参后门. 在步骤⑨中, 智能合约安全执行是通过 SGX 封装 EVM 实现, 即在 EVM 内部通过 AES 算法实现加解密. 智能合约提前部署在平台中, 在交易合约执行过程中, 使用上一步获取的密钥在 enclave 中获取源数据进行标签计算. 由于平台区块链网络采用基于价值的共识出块机制, 可信节点从记账节点中选拔, 并在执行可信计算过程中获取共享积分和 *gas* 奖励, 同时普通节点会监督记账节点的计算过程, 确保交易流程安全可靠.

安全交付是数据交易的最后一个阶段, 需要在源数据不泄露的前提下既满足数据买家的交易需求又能够实现平台数据交易的合理收益, 防止计算结果被盗取以及数据买家抵赖行为. 由于数据商品具有特殊的时效性和不稳定性, 难以评估其是否产生相应价值效益和可信性, 所以需要制定一套可靠的数据交付体系. 根据 DSTS 的安全交易机制, 买家在平台中以提交优惠券或广告的形式购买数据, 通过一系列平台流程, 交付给数据买家相应数量的推送列表, 使用交付合约上传计算结果和对应买卖双方的哈希值, 同时将买家收到推送列表后的推送操作记录上链, 说明其使用了交易数据, 保证其后续无法抵赖.

3 实验设计与结果分析

3.1 实验设置

3.1.1 实验数据集

本文采用了某电商公司提供的真实电商平台作为传统电商平台, 平台中所有商户可以通过注册区块链账户加入电商联盟, 成为去中心化数据交易平台的用户的一员. 数据集来自该电商平台 35 万条脱敏的真实订单数据信息, 包括订单、会员以及商品数据信息. 数据集包含的商品信息丰富, 会员数量庞大且相关的订单信息来源真实可靠, 因此具有较高使用价值和实验说服力, 具备在营销标签交易平台中流转的特征和要求. 实验数据集具体描述如表 2 所示.

Table 2 Experimental Data Set

表 2 实验数据集

标签类别	属性	数值
订单属性	成交商品数	348 875
	订单数	196 624
客户属性	客户数	75 359
	地址(精确到市)	64
商品属性	商品数	13
	价格区间	10
	套餐数	12
	促销优惠金额	18

实验数据满足营销标签生成计算所需要的订单属性信息, 在数据集筛选过程中, 根据数据属性分为 3 个类别, 分别是订单属性、客户属性以及商品属性. 订单属性包含了成交商品数和订单数, 其中成交商品数是每件商品的购买记录作为数据集中的一条数

据，订单数则是由于在实际交易环境中存在 1 笔订单中包含多件商品交易的情况，所以数据集中所展示的订单数少于成交的商品数. 利用订单数有助于推断商品之间的相似度，对于推断用户喜好有较大帮助. 客户属性来源于订单数据中会员预留在订单中的信息，其中收货地址精确到区县. 商品属性则根据订单中对于商品信息的描述、分类以及 SKU 码确定，并根据现有的商品信息进行价格区间的合理划分. 其中套餐也是商品的一种，属于不同商品组合销售的一种形式，在实验数据集中也有单独记录. 优惠是当前订单使用平台优惠券的情况，在原始数据处理中也可以当作商品进行标签计算和推送.

3.1.2 实验环境配置

DSTS 中有 3 个重要角色分别是数据买家、数据卖家和数据交易平台，共用 1 台机器进行测试. 此外，系统运行需要多个节点共同维护分布式的区块链网络，实验采用虚拟机节点模拟进行共识模块的功能和效率测试. 在同一台机器中开启 2 个系统终端来代表平台中数据买家和卖家，机器具备了运行 Intel SGX 应用的软硬件环境，以及以太坊客户端配置，如表 3 所示.

Table 3 Experimental Environment Configuration

表 3 实验环境配置

配置名称	版本
CPU	Intel Xeon E5-2630 v3 @ 2.40 GHz
内存	DDR3 64 GB
硬盘	1 TB 7 200 rpm
网络速度/Gbps	1
OS	Linux 18.04-64 位
SGX SDK	Intel SGX SDK for Linux v2.7.100.2
以太坊客户端	Geth 1.9.12-stable

在系统评估过程中，分别对密钥的安全传输、智能合约执行保护以及标签生成计算的执行效率进行数据统计和结果分析.

3.2 实验结果分析

由于在平台完成数据真实性验证和交易保障过程都需要 SGX 远程认证来确认当前执行计算的节点环境是否可靠，所以需要测试的第 1 个关键数据是远程认证时间. 并且平台需要执行智能合约实现用户功能，而合约的确认时间受共识机制影响，合约的执行时间受可信执行环境影响，所以还需要分别测试并对比智能合约在传统环境及当前平台中的时间开销，用于分析安全性提升对整个平台效率的影响.

3.2.1 远程认证时间

在实验环境机器中部署测试代码并进行单点远程认证时间测试来模拟营销标签交易平台在进行标签计算和区块链交易处理之前平台的安全环境准备，对实验机器环境进行 100 次远程认证，观察其认证代码执行稳定性和时间开销，测试结果如图 5 所示，最终平均远程认证时间开销为 1.28 s.

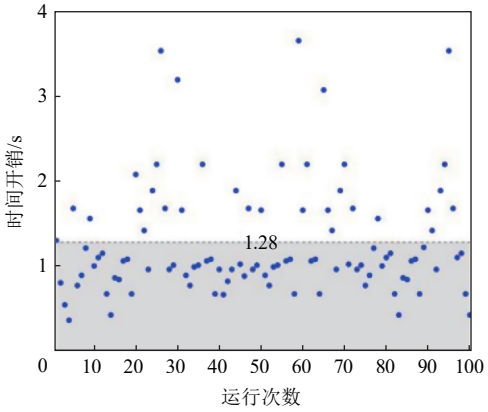


Fig. 5 Remote authentication time cost

图 5 远程认证时间开销

图 5 表明，远程认证时间总体维持在平均时间开销线上，分布较为稳定. 但是在测试结果中可以看出，极少数测试出现了时间开销超出 3 s 的情况，这是由于远程认证过程中需要向 Intel 官方确认测试机器中 enclave 生成的远程认证报告，受网络速率影响会出现时间波动和个别极值的情况. 接着分析该模块在整体数据交易过程中的时间开销占比，考虑在实际营销标签交易过程中，买家以推送优惠券或广告的形式提交标签交易请求，在获取推送列表后，还需要买家主动进行逐条推送，所以在这一交易周期中，本身就存在用户反馈的时间延迟，相对于安全保障和数据真实性而言，远程认证所需时间开销可以接受.

3.2.2 激励和共识机制性能测试

系统共识激励机制性能体现在合约执行过程中交易确认所需的时间开销，通过测试系统在处理不同数量交易过程中，每秒执行的交易量也用 TPS 来体现. 为控制测试过程中无关变量影响，测试采用常规的转账交易合约来避免额外的合约调用等函数参与. 在系统启动前将不同分组的交易打包存入交易池后签名，分别测得共识激励机制在以太坊和 DSTS 网络中的时间开销，如图 6 所示.

图 6 表明，共识激励机制运行的时间基本与交易数据量呈线性关系，且随着数据量的增加 DSTS 效率优势越明显. 由于系统是基于以太坊开发，相对于以

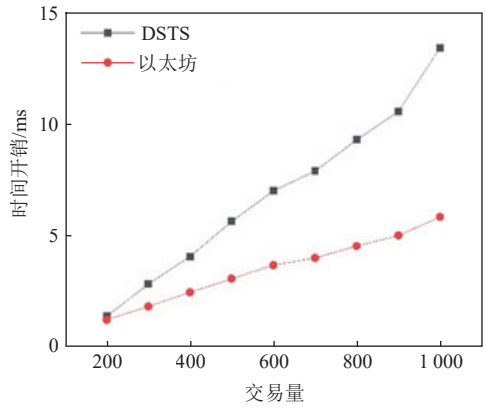


Fig. 6 Performance test of consensus incentive mechanism

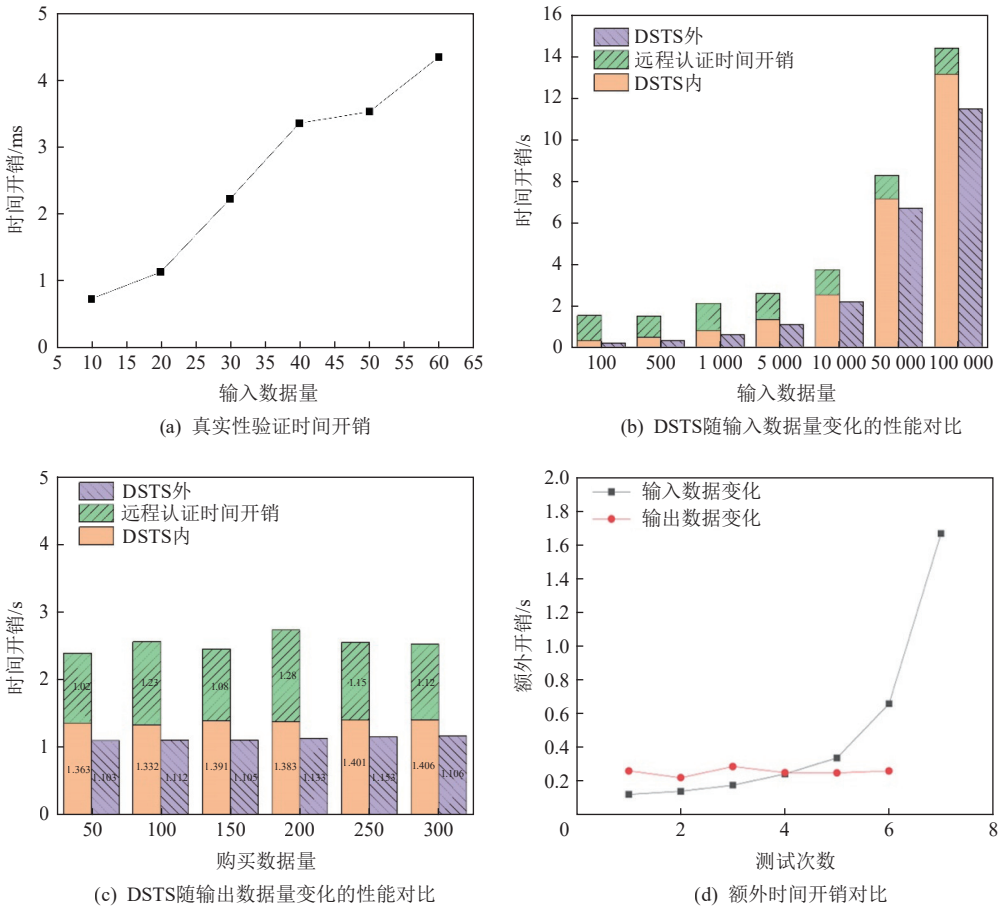
图 6 共识激励机制性能测试

太坊最多每秒处理 13 笔交易的效率来说, DSTS 在共识效率方面展示了良好性能的同时还避免了传统共识机制不必要的算力浪费.

3.2.3 标签可信计算时间

在标签可信计算前, 首先要通过数据真实性验证, 为了与数据在正常环境中的计算进行效率对比, 实验在不同体量的数据集中开展, 先测试真实性验证的时间开销随数据量增长的变化, 接着以选取收货地址为华中地区的用户标签来推送商家包邮券的数据交易为例, 对电商联盟数据量和数据交易量分别进行变量控制来得到正常环境和 DSTS 可信计算机制内部的时间开销对比.

如图 7(a)所示, 在数据卖家处随机产生不同数量的待共享数据, 包含真实数据和伪造数据. 真实性验证时间开销与数据量增加呈线性增长, 这与区块链网络的稳定运行有关. 控制输出数据量为 50 条不变, 即在数据买家需要推送 50 个包邮券的交易场景中, 测试 DSTS 计算性能随输入数据量的变化规律. 通过控制变量, 以同样的 C++代码对相同数据在 DSTS 外部测试, 分析其额外开销. 如图 7(b)所示, 除去远程认证的时间, 每次实验都会产生额外的时间消耗, 这部分额外开销主要来自可信计算中数据加解密操作. 由于 AES 加密算法对于大量数据的可信计算方面可以展示出良好的性能, 额外开销也随着输入数据的增加对整体性能的影响越来越小, 有利于平台的持续稳定运行. 随后通过控制输入数据量为 1 000 条, 即当前电商联盟具备的数据总数, 测试 DSTS 开销随输出数据量增长的变化, 测试结果如



(a) 真实性验证时间开销

(b) DSTS随输入数据量变化的性能对比

(c) DSTS随输出数据量变化的性能对比

(d) 额外时间开销对比

Fig. 7 Test results of DSTS performance

图 7 DSTS 性能测试结果

图 7(c)所示. 时间开销随数据量的增加缓慢增长, 同时从图 7(d)中可以看出随着输出数据量的变化对时间开销影响不大, 因为 DSTS 输出交付过程没有其他加解密计算的操作步骤, 只存在额外的操作上链步骤, 对整体效率影响不大.

由于数据交易与传统商品交易在交易流程和步骤中都不相同, 数据交易具备一定交易周期. 数据交易首先需要买家提交各项数据需求, 接着卖家要根据数据买卖提出的需求进行交易响应, 其次进行交易环境安全验证, 最终是在数据交付时, 买家对所购买数据的一系列操作完成后, 传统电商平台用户接收到广告或者优惠券时, 标志着整个数据交易过程完毕. 所以 DSTS 的额外时间开销对整个交易周期的影响较小, 均在可接受时间范围内.

4 总结与展望

本文提出了基于区块链的营销标签交易平台 DSTS, 平台基于传统电商平台组建电商联盟, 充分利用传统电商平台中的数据信息进行安全的标签计算以实现精确推荐, 解决了当前电商平台中由于诚信危机带来的买卖关系紧张问题, 可靠地发挥了数据价值. 此外, 所构建的数据安全交易机制, 可以为当前各类电商平台提供数据价值可靠利用的新方法和新体系提供思路. 根据去中心化数据交易平台特点设计了区块链共识激励和出块机制, 保证了平台运行的稳定性和持续性; 引入共享积分思想, 鼓励用户积极共享数据的同时, 保障数据交易市场可靠运行; 根据可信计算思想, 通过 SGX 保障数据在传输和计算过程中的安全性; 根据差分隐私思想, 在不泄露用户隐私的情况下保证数据交易结果的可靠交付, 保障用户隐私. 最后, 通过具有代表性的真实数据集对系统进行功能和性能测试, 结果表明系统开销在可接受的安全成本之内.

DSTS 的安全机制适用于当前市场中大多数类型的电商平台, 在真实电商平台运营过程中, 原始数据不限于订单数据, 未来的工作还将扩大销售的数据类型包括用户以浏览记录、加购记录等行为信息, 用作数据商品以提升标签准确性, 充分安全地发挥数据的最大价值.

作者贡献声明: 代炜琦提出系统整体设计方案, 修改论文及定稿; 李铭负责系统部分的搭建及完成实验并撰写论文; 赵珂轩负责系统开发和整理实验数据; 姜文超提出了系统设计思路并修改论文; 周蔚林提供系统应用环境和实验测试数据; 邹德清提出系统共识激励算法设计方案和测试方法; 金海提出区块链分布式系统设计方法和 SGX 应用开发方案.

参 考 文 献

[1] Zhu Liehuang, Gao Feng, Shen Meng, et al. Survey on privacy preserving techniques for blockchain technology[J]. Journal of Computer Research and Development, 2017, 54(10): 2170−2186 (in Chinese)
(祝烈煌, 高峰, 沈蒙, 等. 区块链隐私保护研究综述[J]. 计算机研究与发展, 2017, 54(10): 2170−2186)

[2] Sani A S, Bertino E, Yuan Dong, et al. SPrivAD: A secure and privacy-preserving mutually dependent authentication and data access scheme for smart communities[J]. Computers & Security, 2022, 115: 102610

[3] Yang Qi, Gong Nanning. Main problems and suggestions on China's big data trading[J]. Big Data, 2015, 1(2): 38−48 (in Chinese)
(杨琪, 龚南宁. 我国大数据交易的主要问题及建议[J]. 大数据, 2015, 1(2): 38−48)

[4] Zheng Xiao. Data trading with differential privacy in data market[C]//Proc of the 6th Int Conf on Computing and Data Engineering. New York: ACM, 2020: 112−115

[5] Lu Siqi, Zheng Jianhua, Cao Zhenfu, et al. A survey on cryptographic techniques for protecting big data security: Present and forthcoming [J]. Science China Information Sciences. 2022, 65(10): 201301

[6] Zhou Wei, Wang Chao, Xu Jian, et al. Privacy-preserving and decentralized federated learning model based on the blockchain[J]. Journal of Computer Research and Development, 2022, 59(11): 2423−2436 (in Chinese)
(周炜, 王超, 徐剑, 等. 基于区块链的隐私保护去中心化联邦学习模型[J]. 计算机研究与发展, 2022, 59(11): 2423−2436)

[7] Li Hui, Pei Lishuang, Liao Dan, et al. BDDT: Use blockchain to facilitate IoT data transactions[J]. Cluster Computing, 2021, 24: 459−473

[8] Shi Dan. Research on data ownership and protection path in the age of big data[J]. Journal of Xi'an Jiaotong University: Social Science Edition, 2018, 38(3): 78−85 (in Chinese)
(石丹. 大数据时代数据权属及其保护路径研究[J]. 西安交通大学学报: 社会科学版, 2018, 38(3): 78−85)

[9] Maesa D D F, Mori P. Blockchain 3.0 applications survey[J]. Journal of Parallel and Distributed Computing, 2020, 138(12): 99−114

[10] Wendl M, Doan M H, Sassen R. The environmental impact of cryptocurrencies using proof of work and proof of stake consensus algorithms: A systematic review[J]. Journal of Environmental Management, 2023, 326: 116530

[11] Ying Chenhao, Xia Fuyuan, Li Jie, et al. Incentive mechanism based on truth estimation of private data for blockchain-based mobile crowdsensing[J]. Journal of Computer Research and Development, 2022, 59(10): 2212−2232 (in Chinese)
(应臣浩, 夏福源, 李颉, 等. 区块链群智感知中基于隐私数据真值估计的激励机制[J]. 计算机研究与发展, 2022, 59(10): 2212−2232)

[12] Chernyavskiy A, Ilvovsky D. Extract and aggregate: A novel domain-independent approach to factual data verification[C]//Proc of the 2nd

Workshop on Fact Extraction and VERification (FEVER). Stroudsburg, PA: ACL, 2019: 69−78

[13] Xiao Yang, Zhang Ning, Li Jin, et al. PrivacyGuard: Enforcing private data usage control with blockchain and attested off-chain contract execution[C]//Proc of the 25th European Symp on Research in Computer Security. Berlin: Springer, 2020: 610−629

[14] Yao Xin, Yan Xiaoping, Zhao Ming. Authenticity verification for dynamic social data outsourcing[J]. IEEE Systems Journal, 2021, 16(2): 2325−2335

[15] Liu Jinhui, Yu Yong, Bi Hongliang, et al. Post quantum secure fair data trading with deterability based on machine learning[J]. Science China Information Sciences, 2022, 65(7): 170308

[16] Li Yanan, Feng Xiaotao, Xie Jan, et al. A decentralized and secure blockchain platform for open fair data trading[J]. Concurrency and Computation: Practice and Experience, 2020, 32(7): e5578

[17] Zhao Zhiwei. Research on the privacy protection of personal data trading based on blockchain [D]. Chengdu: University of Electronic Science and Technology of China, 2020 (in Chinese)
(赵志伟. 基于区块链的个人数据交易隐私保护研究[D]. 成都: 电子科技大学, 2020)

[18] Xia Changlin. Internet of things data trading system based on blockchain technology [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2020 (in Chinese)
(夏昌琳. 基于区块链技术的物联网数据交易系统[D]. 南京: 南京邮电大学, 2020)

[19] Knauth T, Steiner M, Chakrabarti S, et al. Integrating remote attestation with transport layer security [J]. arXiv preprint, arXiv: 1801.05863, 2018

[20] Sardar M U, Fetzer C. Towards formalization of enhanced privacy ID (EPID)-based remote attestation in Intel SGX[C]//Proc of the 23rd Euromicro Conf on Digital System Design (DSD). Piscataway, NJ: IEEE, 2020: 604−607

[21] Karande V, Bauman E, Lin Zhiqiang, et al. SGX-Log: Securing system logs with SGX[C]//Proc of the 2017 ACM on Asia Conf on Computer and Communications Security. New York: ACM, 2017: 19−30

[22] Kuang Boyu, Fu Anmin, Susilo W, et al. A survey of remote attestation in Internet of things: Attacks, countermeasures, and prospects[J]. Computers & Security, 2022, 112: 102498

[23] Lei Hong, Bao Zijian, Wang Qinghao, et al. SDABS: A secure cloud data auditing scheme based on blockchain and SGX[C]//Proc of the 2nd Blockchain and Trustworthy Systems. Berlin: Springer, 2020: 269−281

[24] Dai Weiqi, Dai Chunkai, Choo K K R, et al. SDTE: A secure blockchain-based data trading ecosystem[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 725−737

Dai Weiqi, born in 1984. PhD, associate professor. Member of CCF. His main research interests include blockchain security, privacy computing, and cloud security.

代炜琦, 1984 年生. 博士, 副教授. CCF 会员. 主要研究方向为区块链安全、隐私计算、云安全.

Li Ming, born in 1997. Master. His main research interests include blockchain security and trusted computing.

李　铭, 1997 年生. 硕士. 主要研究方向为区块链安全、可信计算.

Zhao Kexuan, born in 2001. PhD candidate. His main research interests include blockchain and privacy computing.

赵珂轩, 2001 年生. 博士研究生. 主要研究方向为区块链、隐私计算.

Jiang Wenchao, born in 1977. PhD, associate professor. Member of CCF. His main research interests include cloud computing, big data, knowledge mapping, and industrial artificial intelligence.

姜文超, 1977 年生. 博士, 副教授. CCF 会员. 主要研究方向为云计算、大数据、知识图谱、工业人工智能.

Zhou Weilin, born in 1978. Master, senior engineer. His main research interests include electronic authentication, information security, and distributed databases.

周蔚林, 1978 年生. 硕士, 高级工程师. 主要研究方向为电子认证、信息安全、分布式数据库.

Zou Deqing, born in 1975. PhD, professor. Member of CCF. His main research interests include big data security and AI security, cloud computing security, software definition security and active defense, and software vulnerability detection and network attack and defense.

邹德清, 1975 年生. 博士, 教授. CCF 会员. 主要研究方向为大数据安全与人工智能安全、云计算安全、软件定义安全与主动防御、软件漏洞检测与网络攻防.

Jin Hai, born in 1966. PhD, professor. Fellow of CCF. His main research interests include computer architecture, parallel and distributed processing, and cloud computing and big data.

金　海, 1966 年生. 博士, 教授. CCF 会士. 主要研究方向为计算机体系结构、并行与分布式处理、云计算与大数据.