面向大数据定价的委托拍卖方案

尹 鑫,田有亮,王海龙
(1. 贵州大学计算机科学与技术学院,贵州贵阳 550025;2. 贵州省公共大数据重点实验室,贵州贵阳 550025;
3. 贵州大学密码学与数据安全研究所,贵州贵阳 550025)

摘 要: 大数据合理定价是当前大数据交易中亟待解决的具有一定挑战性问题之一.本文针对大数据定价困难问题,基于Micali-Rabin的安全计算技术提出一种具有大数据定价功能的安全委托拍卖方案.在方案中首先基于Micali-Rabin的随机向量表示方法设计满足标价密封性的大数据拍卖及验证算法.其次,基于Merkle树和Bit承诺协议实现大数据交易中数据的完整性和底价的不可否认性,特别是在定价阶段,利用一种特殊的多方安全计算协议隐藏大数据的底价,以此保障了大数据交易的公平性.最后,方案安全性和性能分析表明,该方案特别适用于大数据交易场景下的数据委托拍卖.

关键词: 大数据定价; Micali-Rabin随机向量表示; 匿名性; 密封拍卖
中图分类号: TP309, TN918 **文献标识码:** A **文章编号:** 0372-2112 (2018)05-1113-08
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2018.05.014

Delegation Auction Scheme for Big Data Pricing

YIN Xin, TIAN You-liang, WANG Hai-long
(1. *College of Computer Science & Technology, Guizhou University, Guiyang, Guizhou* 550025, *China*;
2. *Guizhou Provincial Key Laboratory of Public Big Data, Guiyang, Guizhou* 550025, *China*;
3. *Institute of Cryptography & Data Security, Guizhou University, Guiyang, Guizhou* 550025, *China*)

Abstract: The reasonable pricing of big data is one of the most challenging problems in big data transaction, which needs to be solved urgently. This paper, based on Micali-Rabin's secure computing technology, presents a secure delegation auction scheme with the function of big data pricing. First, based on Micali-Rabin's random vector representation methods, we design an auction, as well as a verification algorithm with the property of the sealed bid. Second, the big data's integrity and non-repudiation of the reserve price are realized by Merkle tree and Bit commitment protocol respectively, especially in the pricing stage the use of a special multi-party security computing protocol to hide the reserve price of big data, whereby ensuring the fairness of the big data transaction. Finally, as the scheme's security and performance analysis indicate that it is particularly suitable for delegation auction in the context of the big data transaction.

Key words: big data pricing; Micali-Rabin's random vector representation; anonymity; sealed auction

1 引言

随着大数据、云计算、物联网、人工智能等计算机技术和通信技术的突破和融合,数据量呈现爆炸式增长.当前很多政府机构、组织和行业拥有大数据却不具备大数据分析的能力,因此把大数据交给需要的人,进一步释放大数据的价值变得尤为重要.大数据作为一种无形的资产,其价值具有不确定性、稀缺性和多样性,且买方和卖方的信息存在不对称问题,因此在大数据交易中,无法正确评估大数据的价值[1].同时大数据本身存在易拷贝的特点,一旦数据被任意拷贝给他人,就会损害数据拥有者的利益.这使得对大数据合理定价变得困难.目前大数据定价分为三种模式:协议定价、拍卖定价、集合定价[2].由于拍卖模式可以将数据使用权交到少数人的手中,保证了卖方的利益,同时也能够兼顾市场原则.因此,在拍卖模式下,研究大数据的合理定价

收稿日期:2017-03-27;修回日期:2017-05-15;责任编辑:马兰英
基金项目:国家自然科学基金(No. 61363068, No. 61662009, No. 61772008);教育部-中国移动科研基金(No. MCM20170401);贵州省科技重大专项计划(No. 20183001);贵州省教育厅科技拔尖人才支持项目(No. 黔教合KY字[2016]060);贵州大学引进人才科研项目(No. 贵大人基合字(2015)53号)

是当前大数据交易中亟待解决的问题之一.

随着上世纪 90 年代电子商务的兴起,设计安全、实用的拍卖方案一直成为学术界关注的焦点. 基于密码理论和方法的拍卖方案被大量提出. 2003 年王继林等[3]人基于单向函数,设计了一种具有投标者匿名、投标价保密、不可否认性和强可验证性等特性的密封式拍卖方案. 2008 年 Parkes 等[4]人提出一个基于同态加密技术的密封式拍卖的实用系统,该系统同时提供了拍卖结果正确性和可信度的验证. 2012 年王秀利等[5]人基于秘密共享和双重数字签名技术设计了一种安全的密封式电子拍卖方案. 2014 年 Micali 和 Rabin 扩展 Rabin[6,7]的方案,提出一种可拒绝泄漏秘密标价和不受控制的拒绝竞价,解决了长期存在于二价拍卖中合谋问题[8]. 由于交易所不希望将竞价者的名字和身份同时泄漏,因此需要做出多个承诺,进行多轮的交互证明,通信量比较大. 2015 年 Zhu 等[9]人提出了基于可比较的密码原语和多线性映射的第一价格密封投标方案,优化了拍卖者和投标者之间的通信轮数,减轻了通信工作量. 2016 年丛鑫等[10]人针对单个云服务提供商不能满足视频提供商所需的资源时,提出了基于拍卖的视频集分配方案,并证明所提策略是有效的.

近年来,大数据的研究成果集中在知识萃取和挖掘[11]、存储和管理[12]、安全和隐私[13]等方面. 关于大数据定价问题集中于云服务以及经济学领域. 2016 年 Niyato 等人[14]提出了基于数据量模式的最优定价方案,该方案允许服务供应商确定数据量以提供服务给用户. 同年 Li 等人[15]提出了一个具有多个云中介的方案,该方案以较低成本给用户提供流式大数据计算服务,同时能最大化多个云中介的收益. 刘朝阳[1]从大数据的基本特征与价值特征出发,分析如何确定大数据的合理价格. 王文平针对大数据交易,提出了五种交易定价策略,同时为大数据交易市场的完善提出了建议[2].

从以上的拍卖方案可以看出,目前的拍卖方案大多是针对物品的拍卖定价. 对于大数据定价问题,更多是从大数据在云服务定价和社会科学角度进行研究,未见操作性强的技术方案. 其实,不论是对物品的拍卖还是对大数据的拍卖,都要满足拍卖的匿名性、不可否认等基本特性. 对于物品拍卖,其价值相对固定,且一手交钱一手交货,所得即所有. 而大数据是一种价值不确定的新型资源,其价格定制没有任何先例参照,很难直接给出一个合理的价格. 拍卖是一种确定商品价格的基础且重要的方法. 但如何保证在拍卖中投标者自由竞价,同时兼顾卖家利益是本方案的难点;大数据最明显特征之一就是数据量大,保证大数据拍卖中数据的完整性变得尤为重要;此外,大数据具有易拷贝的特性,通过大数据拍卖将大数据直接交给需要的人,减少第三方拷贝数据的威胁,保证交易双方的利益是本方案亟待解决的问题.

本文针对上述问题,在大数据定价困难情况下,提出了一种新的面向大数据定价的二价委托拍卖方案,该方案不仅满足拍卖方案的底价隐藏性和不可否认性、匿名竞价性、密封投价性、抗合谋性、身份不可欺骗的性质,有效解决了大数据的完整性验证以及大数据价值的确定问题,且在一定程度上减少了大数据的拷贝. 在拍卖初始化阶段,卖方运用 Merkle 树[16]对大数据分块哈希获得根节点,在拍卖支付阶段,最高价者(获胜者)根据公布的 Merkle hash 函数对收到的数据进行完整性验证;鉴于大数据的价值不确定,在本方案中卖方运用 Bit 承诺协议[17],对底价进行承诺,使底价具有不可否认性和隐藏性. 这样在竞标阶段,投标者可以不用受底价影响自由投标,同时由卖方给出底价兼顾了卖家利益;在拍卖注册阶段,运用 RSA 盲签名[18],对参与拍卖的投标者的真实身份进行隐藏;在拍卖中,运用特殊的多方安全计算[19],对标价与底价进行保密比较,实现了拍卖中底价的隐藏性;在验证阶段,运用文献[8]中的 Micali-Rabin 随机向量表示方法,在交易以及验证时不会泄露标价的信息,同时由于交易所不知道投标者的真实身份,因此只需要进行两轮交互验证;大数据具有易拷贝的特性,本方案中大数据不经过拍卖中心,在一定程度上减少了数据的拷贝. 本文基于 Micali-Rabin 的安全计算技术,利用多种密码技术给出了一种委托拍卖方案,为解决大数据难以定价问题提供技术指导.

2 准备知识

本节简要介绍 Merkle 树和 Micali-Rabin 随机向量表示方法.

2.1 Merkle 树

Merkle 树是一棵 Hash 树,每一个叶子节点存放经过 Hash 后的散列值. 在构造 Merkle 树的过程中,首先每两个叶子节点的散列值相连接,然后通过 Hash 函数生成它们的父节点,如此递归计算直到生成最后的根节点.

在本文,假设某待定价大数据可以分为 8 块,对每块数据进行编号(0,1,…,7),哈希,成为叶子节点;接着,对于邻近的两个叶子节点再次哈希构成父节点,以此类推,得到根节点 root,如图 1 所示.

在验证该数据的完整性时,根据数据块的编号,用相同的哈希函数进行重构,获得根节点 root′,最后比较 root 和 root′是否相同. 若两者相同,则数据完整性未被篡改,否则反之.

2.2 Micali-Rabin 的随机向量表示

Micali-Rabin 利用随机向量表示值,采用 Pedersen 承

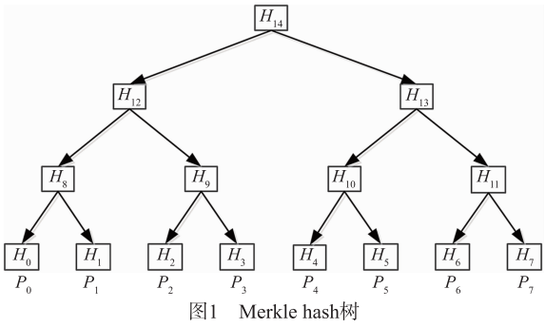

图1　Merkle hash树

诺[20],通过零知识证明等式的正确性. 这里,每个投标者与交易所共用一对不同的生成元 g,h. 其中 g,h 属于素数阶为 q 的群 G,F_p 为有限域,p 为 128bit 的素数,$q>p$.

定义 1　设 $\boldsymbol{X}$ 的随机向量表示是 $\boldsymbol{X}=(u,v)$,其中 $u,v\in F_p$,$\boldsymbol{X}$ 的值是 $Val(\boldsymbol{X})=(u+v)\ mod\ p$.

定义 2　令 $u\in F_p$,使用一个辅助值 $r,r\in F_p$,对 u 进行承诺 $Com(u)$,承诺值为 $Com(u)=Com(u,r)=g^u h^r$.

定义 3　对 $\boldsymbol{X}=(u,v)$ 的分量进行承诺,即 $Com(\boldsymbol{X})=(Com(u),Com(v))$.

定义 4　假设有两行承诺值,其中 $1\leqslant i\leqslant n$,

$$\begin{matrix} Com(X_1),\cdots,Com(X_n) \\ Com(Y_1),\cdots,Com(Y_n) \end{matrix} \tag{1}$$

如果 $Val(X_i)=Val(Y_i)$,$1\leqslant i\leqslant n$,那么这两行的承诺值是一致的.

引理 如果有超过 k 个承诺值是假的,那么验证者接受的概率为 $(1/2)^k$.

证明　对任意的 $Val(X_i)\neq Val(Y_i)$,这种情况不被发现的概率最多为 1/2. 因为独立随机的选择一个 $c\leftarrow\{1,2\}$,至少有 k 个等式不被发现的概率为 $(1/2)^k$.

3　大数据交易拍卖方案

3.1　拍卖模型

本方案包括 4 个拍卖主体:交易所,注册中心,卖方,若干投标者. 图 2 给出了本方案所使用的拍卖模型.

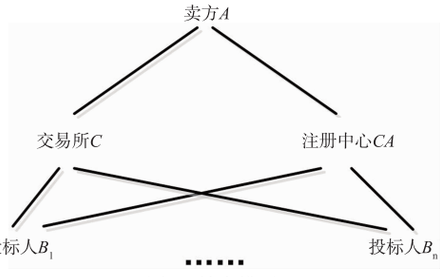

图2　拍卖模型

在该方案中,卖方首先对自己拥有的大数据通过 Merkle 树分块哈希. 拍卖开始前,投标者首先向注册中心提交申请,注册中心对参与拍卖的投标者进行身份核实,同时对投标者的身份进行 RSA 盲签名;交易所验证投标者身份的合法性;投标者秘密投标,交易所计算成交价;最后交易所公布获胜者的身份信息及获胜者要支付的第二高价,投标者可以验证自己标价的正确性.

3.2　拍卖方案描述

3.2.1　方案参数设置

(1)注册中心 CA:对参与拍卖的投标者进行身份核实,同时对投标者身份进行盲签名.

(2)交易所 C:负责发布拍卖信息,验证注册中心 CA 对投标者的盲签名.

(3)投标者 B:假设有 n 个投标者 $B_1,\cdots,B_n$.

(4)卖方 A:为合法用户,给交易所提供要拍卖数据的描述信息.

(5)电子公告栏 SBB:SBB 由交易所 C 控制,用来张贴数据;数据不能被删除;在拍卖中,SBB 对所有的投标者可见,且所有人看到的 SBB 上的内容是相同的.

(6)平板电脑 PC:每个投标者都有一个平板电脑来写自己的标价,并且这个平板电脑还可以为这个标价打印一个收据[21].

(7)同步时钟:交易所验证所有投标者的盲签名后,所有合法投标者与交易所同步时间.

3.2.2　拍卖方案

首先给出拍卖方案的总体流程图,如图 3 所示.

(1) 初始化阶段

(a)卖方 A 对自己的数据进行描述,将此数据的描述信息发送交易所,申请数据拍卖.

(b)交易所同意拍卖后,交易所在电子公告栏 SBB 上公布数据的描述信息. 卖方 A 自己选择数据的底价 r,对底价 r 进行 Bit 承诺 $H(R_1,R_2,r)$.

(c)卖方 A 将自己的数据进行分块编号,哈希,将最后的 Merkle 根节点 root、所采用的 Merkle Hash 函数、Bit 承诺 $(H(R_1,R_2,r),R_1)$ 以及 Bit 承诺 Hash 函数发送给交易所.

(d)交易所在电子公告栏 SBB 上公布拍卖时间、数据的根节点 root、Merkle Hash 函数、Bit 承诺 $(H(R_1,R_2,r),R_1)$ 以及 Bit 承诺 Hash 函数.

(2) 注册阶段

注册中心 CA 对参与拍卖的投标者 B_i 进行身份核实,同时对投标者 B_i 的身份进行 RSA 盲签名. 首先注册中心 CA 选取两个大素数 s,t,$n=st$,$\varphi(n)=(s-1)(t-1)$,随机选取 e 满足 $1<e<\varphi(n)$,且 $gcd(e,n)=1$,计算 d 满足 $de\equiv 1 \bmod \varphi(n)$ 和 $1<d<\varphi(n)$,CA 的公钥 (n,e),私钥为 d,$h(\cdot)$ 是 Hash 函数,同时注册中心 CA 秘密销毁 s,t.

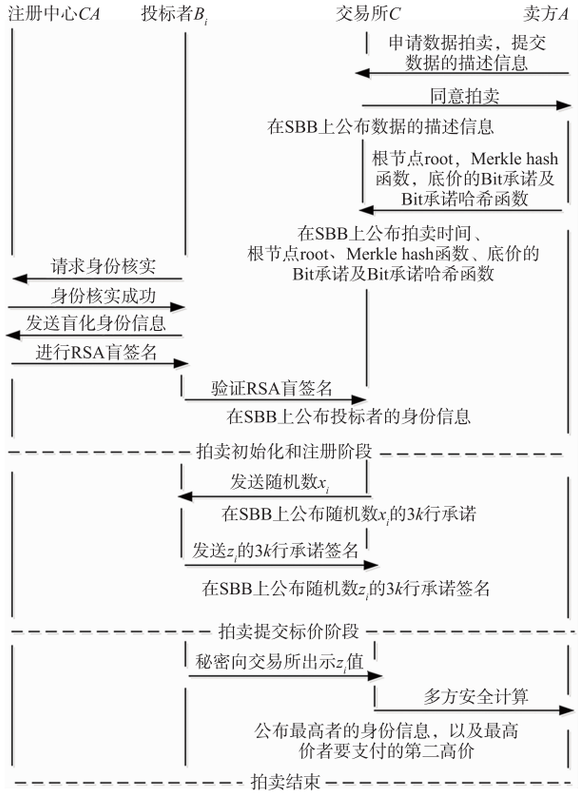

图3 数据拍卖流程图

(*a*)签名阶段

步骤 1 参与拍卖的投标者 B_i 选取自己的身份 $d_i \in F_p$ 以及大随机数 k_i，$gcd(k_i,n)=1$，计算 $d_i^{'}=h(d_i)k_i^e \bmod n$，并将 $d_i^{'}$ 发送给 *CA*.

步骤 2 注册中心 *CA* 接收到 $d_i^{'}$ 后，计算 $S_i^{'}=d_i^{'d} \bmod n$，并将 $S_i^{'}$ 发送给 B_i.

步骤 3 投标者 B_i 接收到 $S_i^{'}$ 后，计算 $S_i=k_i^{-1}S_i^{'} \bmod n$，$S_i$ 是注册中心 *CA* 对 d_i 的盲签名.

(*b*)验证签名

T_0 时间内，投标者 B_i 将 (d_i,S_i) 发送给交易所 *C*，交易所 *C* 计算 $v_i=S_i^e \bmod n$，验证 $v_i=h(d_i)$ 是否成立，若成立，则签名有效，可以进入拍卖；否则无效.

(3) 投标阶段

(*a*)T_1 时间内，交易所 *C* 发送一个随机数 $x_i \in F_p$ 给 B_i，$Val(X_i^j)=x_i$，随机数 x_i 只有交易所和投标者 B_i 知道. 交易所 *C* 接着在电子公告栏 SBB 上对 x_i 进行 $3k$ 行承诺 $(d_i,Com(X_i^j))$. 这 $3k$ 行承诺值采用 Micali-Rabin 的随机向量表示法，其中 $Com(X_i^j)=(Com(u_i^j),Com(v_i^j))$，$X_i^j=(u_i^j,v_i^j)$，$Val(X_i^j)=(u_i^j+v_i^j) \bmod p$，$1 \leqslant j \leqslant 3k$.

(*b*)T_2 时间内，投标者 B_i 确定标价 b_i. 投标者 B_i 选择 $z_i \in F_p$，满足 $x_i+z_i=b_i \bmod p$. 投标者对 z_i 进行 $3k$ 行承诺 $Com(Z_i^j)$，$1 \leqslant j \leqslant 3k$，将这 $3k$ 行承诺签名提交给交易所 *C*，同时平板电脑生成收据 p_i（任何人无法篡改）[21]. 收据由投标者 B_i 自己保存，上面有提交标价的时间 t_i. 交易所 *C* 在 SBB 上公布 z_i 的 $3k$ 行承诺签名.

(4)开标阶段

T_3 时间内，投标者秘密的向交易所 *C* 揭露值 z_i. 交易所 *C* 根据 $x_i+z_i=b_i \bmod p$ 得到每个投标者的标价，交易所 *C* 对收到的标价进行排序后，与卖方进行多方安全计算（本文采用文献[19]的多方安全计算方案 2）比较. 最后，在电子公告栏 SBB 上公布最高价位者的身份信息 d_i，以及要支付的第二高价. 在进行比较的过程中，有以下三种情况：

(*a*)如果投标者 B_i 的价格大于卖方 *A* 底价的个数 $num \geqslant 2$，在电子公告栏 SBB 公布最高价的身份信息 d_i，以及要支付的第二高价 b_j.

(*b*)如果投标者 B_i 的价格大于卖方 *A* 底价的个数 $num=1$，即只有一个投标者价格超过卖方底价. 此时身份信息为 b_i 的投标者将以底价 *r* 支付.

(*c*)如果投标者 B_i 的价格大于卖方 *A* 底价的个数 $num=0$，即没有人的价格超过卖方的价格，此次交易终止.

(5) 验标阶段

(1)T_3 时间结束后，投标者 B_i 秘密的向交易所 *C* 打开 z_i 的 $3k$ 行承诺. 此时，电子公告栏 SBB 上公布的信息有投标者的身份信息 d_i、随机数 x_i、z_i 的 $3k$ 行承诺值，$d_i,x_i,z_i \in F_p$，$1 \leqslant j \leqslant 3k$.

$$\begin{aligned}&<d_1,Com(X_1^j),Com(Z_1^j)>,\cdots,\\&<d_n,Com(X_n^j),Com(Z_n^j)>\end{aligned} \tag{2}$$

其中 $d_1{}^j$ 是第 1 个投标者身份 d_1 的第 *j* 个随机向量表示，$Com(X_1{}^j)$ 是交易所给第 1 个投标者的随机数 x_1 的第 *j* 个随机向量表示，$Com(Z_1{}^j)$ 是第 1 个投标者选的数 z_1 的第 *j* 个随机向量表示. 其余元组对第 2，…，*n* 投标者类似.

首先投标者 B_i 秘密的向交易所 *C* 打开 z_i 的 $3k$ 行承诺，交易所 *C* 可以根据 $Z_i^j=(s_i^j,t_i^j)$，$s_i^j,t_i^j \in F_p$，验证等式 $Val(Z_i^j)=(s_i^j+t_i^j) \bmod p$. 若等式成立，则说明在开标阶段，投标者 B_i 向交易所 *C* 秘密打开的 z_i 的值与投标时 B_i 选择的 z_i 是一致的.

接着投标者 B_i 可以与交易所 *C* 进行交互式证明，验证交易所 *C* 做出的随机数 x_i 的承诺值是否正确. 步骤如下：

(*a*)首先投标者 B_i 根据自己的身份信息 d_i，从电子公告栏 SBB 上任意选择一半的承诺值 $<d_i,Com(X_i^j),Com(Z_i^j)>$，投标者 B_i 秘密的与交易所 *C* 联系，由交易所 *C* 秘密向投标者 B_i 打开承诺值分量 $X_i^j=(u_i^j,$

v_i^j),其中 $u_i^j, v_i^j \in F_p$,投标者验证等式 $Val(X_i^j) = (u_i^j + v_i^j) \bmod p$ 是否成立.

(*b*)接着对于电子公告栏 SBB 上剩余的承诺值进行上下行一致性检验. 在电子公告栏 SBB 上,对于每个值的承诺值,每一列是一致的,$Val(X_i^j) = Val(X_i^{j+1}), 1 \leqslant j \leqslant 3k$. 投标者 B_i 根据自己的身份信息 d_i,对电子公告栏 SBB 剩余的一半承诺值 $< d_i, Com(X_i^j), Com(Z_i^j) >$ 中的 X 进行验证. 若投标者 B_i 要证明 X 的第 j 行承诺值与第 $j+1$ 行承诺值是一致的,即证明等式 $Val(X_i^j) = Val(X_i^{j+1})$ 成立. 设 X_i^j, X_i^{j+1} 的承诺分别是 $Com(X_i^j) = (Com(u_i^j), Com(v_i^j))$, $Com(X_i^{j+1}) = (Com(u_i^{j+1}), Com(v_i^{j+1}))$,其中 $X_i^j = (u_i^j, v_i^j), X_i^{j+1} = (u_i^{j+1}, v_i^{j+1})$. 交易所 C 根据 $X_i^j = (u_i^j, v_i^j), X_i^{j+1} = (u_i^{j+1}, v_i^{j+1})$ 的分量值,准备 t.

$$t = (u_i^{j+1} - u_i^j) \bmod p$$
$$t = (v_i^j - v_i^{j+1}) \bmod p \tag{3}$$

投标者 B_i 可以通过抛硬币选择值 $c \leftarrow \{1,2\}$. 如果 $c=1$,交易所 C 将打开承诺值 $Com(u_i^j)$、$Com(u_i^{j+1})$ 以及 $-t$ 的值验证等式 $u_i^j = (u_i^{j+1} + (-t)) \bmod p$. 反之 $c=2$,交易所 C 将打开承诺值 $Com(v_i^j)$、$Com(v_i^{j+1})$ 以及 t 的值,验证等式 $v_i^j = (v_i^{j+1} + t) \bmod p$. 从这个等式中可以看出,每次只打开承诺值的一半,因此交易所 C 失败的概率至少为 1/2. 若交易所 C 有超过 k 个承诺值是假的,那么投标者 B_i 接受的概率为 $(1/2)^k$.

通过证明过程(*a*)和(*b*),每个投标者都与交易所 C 进行交互式证明各自随机数 X 的承诺值是否正确.

(2)为了验证 $x + z = b \bmod p$ 等式的正确性,交易所 C 对投标者身份的按照大小进行排序,$d_{\pi(1)} < d_{\pi(2)}, \ldots, < d_{\pi(n)}$,其中 π 为随机置换,$1 \leqslant j \leqslant 3k$. 接着交易所 C 按照投标者身份的大小对 d, X, Z 进行另外的 $3k$ 行的承诺,并且公布在电子公告栏 SBB 上:

$$< d_{\pi(1)}, Com(X_{\pi(1)}^j), Com(Z_{\pi(1)}^j) >, \cdots,$$
$$< d_{\pi(n)}, Com(X_{\pi(n)}^j), Com(Z_{\pi(n)}^j) > \tag{4}$$

投标者 B_i 可以与交易所 C 进行交互式证明,验证自己的标价是否正确. 此时,在每个三元组 $< d_{\pi(i)}, Com(X_{\pi(i)}^j), Com(Z_{\pi(i)}^j) >$ 中,投标者的身份都已经秘密与价格 $Val(B) = Val(X) + Val(Z)$ 相联系. 交易所 C 对投标者 B_i 的标价做出 $3k$ 行承诺 $Com(b_{\pi(i)}^j)$. 对每一个验证者来说,标价的 $3k$ 行承诺的一致性关系已经建立,其中 $1 \leqslant j \leqslant 3k$.

$$< d_{\pi(1)}, Com(X_{\pi(1)}^j), Com(Z_{\pi(1)}^j), Com(b_{\pi(1)}^j) >, \cdots,$$
$$< d_{\pi(n)}, Com(X_{\pi(n)}^j), Com(Z_{\pi(n)}^j), Com(b_{\pi(n)}^j) > \tag{5}$$

首先投标者 B_i 根据自己的身份信息 d_i,从电子公告栏 SBB 上任意选择一半的承诺值 $< d_{\pi(i)}, Com(X_{\pi(i)}^j), Com(Z_{\pi(i)}^j), Com(b_{\pi(i)}^j) >$,投标者 B_i 秘密的与交易所 C 联系,由交易所 C 秘密向投标者 B_i 打开承诺值分量 $b_{\pi(i)}^j = (u_{\pi(i)}^j, v_{\pi(i)}^j)$,其中 $u_{\pi(i)}^j, v_{\pi(i)}^j \in F_p$,投标者验证等式 $Val(b_i^j) = (u_{\pi(i)}^j + v_{\pi(i)}^j) \bmod p$ 是否成立.

接着在每个四元组中,验证等式 $x + z = b \bmod p$ 的正确性,其中 $X = (u_1, v_1), Z = (u_2, v_2), B = (u_3, v_3)$, $Val(X) + Val(Y) = Val(B)$,同时存在 $w \in F_p$,使得 $X + Z = B + (w, -w)$. 其证明过程与步骤(*b*)类似.

验证者通过与交易所 C 进行交互式证明,可以验证自己的价格是否被交易所 C 修改.

(3)验证底价 r. 验证阶段,交易所 C 利用卖方发送 (R_1, R_2, r) 验证底价是否被更改. 首先交易所 C 通过公式 $(H(R_1, R_2, r), R_1)$ 验证底价承诺值以及随机数 R_1 与拍卖开始前卖方 A 发送的底价的承诺和 R_1 是否一致,若一致,卖方 A 的承诺是正确的. 其次,交易所 C 知道所有的标价,通过标价与 r 的对比,得到 num 值. 这可以防止在多方安全计算的比较中,卖方 A 改变参与比较的底价 r 欺骗交易所 C. 若发现欺骗,则终止此次交易,并惩罚卖方 A.

(4)拍卖支付. 首先最高价者(获胜者)将要支付的标价金额发送给交易所 C. 接着,卖方 A 根据 SBB 上公布的最高价信息,将大数据秘密发送给最高价者. 最高价者首先对收到的分块数据运用 SBB 上公布的 Merkle hash 函数进行计算,得到根节点 root'. 然后,最高价者将根节点 root' 与初始化阶段 SBB 上公布的根节点 root 进行对比,若结果一致,那么交易继续进行;若结果不一致,则说明数据完整性被破坏(与拍卖前描述的大数据不一样),那么最高价者可以向交易所提出终止交易的请求.

针对开标阶段的情况一,即最高价投标者不止一个. 这些最高价投标者可以向交易所 C 提交自己在投标阶段的收据,交易所 C 根据收据上标价的提交时间来确定获胜者,标价提交时间最早者为获胜者. 因为所有投标者是同时进入投标阶段,所以根据收据上标价的提交时间来确定获胜者是可行的. 举例:若有 B_1、B_2、B_3 三人都为最高价者,标价提交时间为 t_1、t_2、t_3,且 $t_2 < t_1 < t_3$,那么提交时间为 t_2 的投标者为获胜者.

4　方案分析

4.1　安全性分析

本拍卖方案实现了底价的隐藏性和不可否认性、匿名竞价性、密封竞价性、抗合谋性、身份不可欺骗性.

定理 1　该大数据交易方案具有底价隐藏性.

证明　在初始阶段,卖方 A 对底价进行 Bit 承诺 $H(R_1, R_2, r)$. 交易所 C 以及其他投标者要想获得底价,

就要逆向求解 $H(R_1,R_2,r)$,但密码哈希函数的单向性使得其他人无法通过承诺值获得底价.

在开标阶段,交易所 C 将收到的标价进行排序,与卖方通过构造 F 函数进行保密的多方安全计算,由于 F 函数具有易计算、难求逆的特性. 因此在投标中,交易所并不知道底价价格. 直到拍卖结束,卖方公布底价,这有效防止了交易所在拍卖前和拍卖中泄漏底价,造成大数据交易的不公平.

定理 2 该大数据交易方案具有底价不可否认性.

证明 交易所 C 利用卖方 A 发送 (R_1,R_2,r) 验证底价是否被更改. 首先交易所 C 通过公式 $(H(R_1,R_2,r),R_1)$ 验证底价承诺值是否与拍卖开始前卖方 A 发送的底价的承诺一致,若一致,则卖方 A 的承诺是正确的. 如果卖方 A 想欺骗所有的人,发送给交易所 C 错误的底价 $r*$,这几乎不会通过验证公式 $(H(R_1,R_2,r*),R_1)$,因为 *Bit* 承诺中的密码哈希函数 $H(R_1,R_2,r)$ 具有抗碰撞性,卖方 A 要想伪造底价是困难的. 如果卖方 A 拍卖开始前就准备了两个底价 r_1,r_2,先承诺 r_1 并发送给交易所 C. 但在开标阶段,卖方 A 与交易所 C 进行安全多方计算时却用了 r_2. 这样虽然在验证时继续发送 (R_1,R_2,r_1),通过公式 $(H(R_1,R_2,r_1),R_1)$ 的验证,但交易所 C 知道所有的标价,可以通过标价与 r_1 比较,得到 *num* 值. 若 *num* 值与开标时的值不同,则说明卖方 A 欺骗了交易所 C.

定理 3 该大数据交易方案具有匿名竞价性.

证明 在注册阶段的签名阶段,投标者 B_i 选择的随机数 $d_i \in F_p$ 代表自己的身份,选取随机数 k_i 作为盲化因子,计算 $d_i^{'} = h(d_i)\ k_i^e \bmod n$,注册中心 *CA* 对收到的 $d_i^{'}$ 进行签名,但注册中心 *CA* 并不知道 d_i. 在注册阶段的验证签名阶段,交易所 C 收到 (d_i,s_i),通过验证等式 $S_i^e = h(d_i) \bmod n$,判断投标者是否为合法用户,但却不知道投标者的真实身份. 即使注册中心 *CA* 与交易所 C 合谋,交易所 C 把收到的 d_i 全部告诉注册中心 *CA*,但由于投标者 B_i 选取随机数 k_i 作为盲化因子,因此注册中心 *CA* 无法将投标者 B_i 的真实身份与签名的 d_i 相对应.

定理 4 该大数据交易方案具有密封投标性.

证明 在投标阶段,交易所 C 分配一个随机数 $x_i \in F_p$ 给投标者 B_i,投标者 B_i 选取随机数 z_i,根据公式 $x_i + z_i = b_i \bmod p$. 得到自己的标价 b_i,投标者 B_i 向交易所提交对 z_i 的承诺签名. 在开标阶段,投标者 B_i 向交易所秘密地出示值 z_i. 此时,交易所才知道投标者 B_i 的标价 b_i. 从投标阶段到开标阶段,标价都是密封的,以分量形式被承诺.

定理 5 该大数据交易方案具有抗合谋性.

证明 (1)可以防止投标者之间合谋. 假设 B_1、B_2、B_3、B_4 之间存在合谋,投标者 B_2、B_4 分别为出价最高者和第二高价者,他需要付给其他人 $(b_2 - b_4)/3$. 由定理 3 可知,该大数据交易方案具有匿名竞价性,所有人都不知道对方的身份,也就不知道最高价者和第二高价者的真实身份. 假设 B_2 为最高价者,为了自己的利益,若他不想支付合谋者一定金额,会否定自己为最高价者. 因此,在这种情况下,投标者之间最好不要合谋.

(2)可以防止交易所 C 和投标者 B_i 之间合谋. 由定理 1 可知,该大数据交易方案具有底价隐藏性. 直到投标结束,交易所 C 才知道底价. 这有效防止了交易所 C 与投标者 B_i 相互合谋,泄漏底价,损害卖方 A 的利益.

定理 6 该大数据交易方案实现投标者身份不可欺骗性.

证明 投标者在投标前通过注册中心 *CA* 审核、RSA 盲签名. 进入投标阶段,交易所 C 通过 $S_i^e = h(d_i) \bmod n$,验证盲签名 (d_i,s_i) 的正确性. 若 D 想伪装成合法投标者企图欺骗交易所 C,给交易所 C 发送伪盲签名 $(d_i^{'},s_i^{'})$,使伪盲签名通过公式 $S_i^{'e} = h(d_i^{'}) \bmod n$,由于 RSA 盲签名安全性基于大整数分解的困难性,这是不可能的.

本文选取 3 篇采用二价拍卖的论文在底价的隐藏性和不可否认性、匿名竞价性、密封竞价性、抗合谋性、不可欺骗性这六方面进行了比较. 如表 1 所示.

表 1 方案比较

方案	底价隐藏性	底价不可否认	匿名竞价	密封竞价	抗合谋	不可欺骗
文献[8]	否	否	否	是	是	是
文献[22]	否	是	否	是	是	是
文献[4]	否	否	是	是	是	是
本方案	是	是	是	是	是	是

从表 1 中可以看出,文献[4]与本方案都可以实现匿名竞价性、密封竞价性、抗合谋和不可欺骗,但文献[4]无法实现底价隐藏和底价不可否认性;文献[8]和文献[22]与本方案相比,都能实现密封竞价、抗合谋和不可欺骗,但文献[8]和文献[22]都不能实现底价的隐藏性、匿名竞价性.

4.2 性能分析

本方案的通信复杂度是以在拍卖中(投标和开标

阶段)的通信轮数表示. 轮复杂度是指拍卖进行的轮数. 计算复杂度是指拍卖中(投标和开标阶段)交易所 C 的计算量. 本方案是一种二价拍卖方案,只需要进行 1 轮,在投标阶段,交易所 C 与卖方 A 进行保密的多方安全计算,共比较 n 次,计算复杂度为 $o(n)$.

下面给出本拍卖方案与其他方案的性能比较.

文献[5]提出的电子拍卖方案,轮复杂度是 1,计算复杂度为 $o(n)$,通信复杂度≥10. 该方案基于可信第三方,结合秘密共享技术和数字签名技术,设计了一种密封式的电子拍卖方案. 该方案实现了电子拍卖中投标者匿名性、标价保密性等安全性要求,但却需要投标者、拍卖服务器和注册中心三者之间进行多次通信.

文献[23]基于不可信第三方的电子拍卖方案,轮复杂度≥1,计算复杂度为 $o(n)$,通信复杂度≥5. 该方案基于不可信的第三方,运用密码技术,实现了密封式拍卖方案的安全需求,同时可以抵抗合谋攻击. 但该拍卖方案可能进行多轮,这需要投标者多次竞价,且拍卖结束后,中标者要公布自己的所有信息,在一定程度上是不可信的.

文献[24]提出的第一价位密封电子拍卖,轮复杂度是 1,计算复杂度为 $o(nk)$(n 为投标者的数目,k 为投标价的空间长度),通信复杂度为 4,与其他方案相比,效率较低. 该拍卖方案安全性基于离散对数困难性问题,最高价者以及要支付的金额是由投标者联合决定的,不依赖第三方机构,且要支付的金额只有卖方与获胜者知道,达到完全隐私性.

本方案是一种面向大数据定价的委托拍卖方案,轮复杂度是 1,计算复杂度为 $o(n)$,通信复杂度为 3. 本方案基于 Micali-Rabin 的安全计算技术,将标价采用了分量形式表示并承诺,在拍卖及验证中没有泄露任何标价信息. 竞价结束后,交易所只公布最高价的身份信息以及最高价要支付的第二高价的价格. 针对大数据这种特殊的商品,基于 Merkle 树,保障了其完整性;基于特殊的多方安全计算,隐藏了大数据的底价,保障了大数据交易的公平性.

5　结论

针对大数据交易中的定价困难问题,本文给出了一种新的面向大数据定价的二价拍卖方案. 方案采用 Micali-Rabin 的随机向量表示方法,在拍卖以及验证时不会泄露任何标价信息;基于 Merkle 树对大数据的分块哈希,保证了数据的完整性;基于 Bit 承诺对底价进行承诺,实现了底价不可否认性;基于 RSA 盲签名,实现了投标者的匿名性,同时由于竞价者的匿名性,拍卖验证只需进行两轮的交互验证,减少了通信量;基于特殊多方安全计算,对价格与底价大小进行了保密比较,隐藏了大数据的底价,保障了大数据交易的公平性. 最后,在安全性、性能方面与现有方案进行比较分析,本方案特别适用于大数据交易场景下的数据委托拍卖.

参考文献

[1] 刘朝阳. 大数据定价问题分析[J]. 图书情报知识,2016,0(1):57 - 64.
LIU Zhaoyang. Analysis on pricing of big data[J]. Document;Information& Knowledge,2016,0(1):57 - 64. (in Chinese)

[2] 王文平. 大数据交易定价策略研究[J]. 软件,2016,37(10):94 - 97.
WANG Wenping. Research on big data transaction pricing strategy[J]. Computer Engineering & Software,2016,37(10):94 - 97. (in Chinese)

[3] 王继林,陈晓峰,王育民. 一个安全的密封式电子拍卖方案[J]. 电子学报,2003,31(10):1578 - 1579.
WANG Jilin, CHEN Xiaofeng, Wang Yumin. A secure sealed-bid electronic auction scheme[J]. Acta Electronica Sinica,2003,31(10):1578 - 1579. (in Chinese)

[4] Parkes D C,Rabin M O,Shieber S M,et al. Practical secrecy-preserving, verifiably correct and trustworthy auctions[J]. Electronic Commerce Research and Applications,2008,7(3):294 - 312.

[5] 王秀利,王萌. 一种应用于双重数字签名的电子拍卖方案[J]. 计算机工程,2012,38(4):4 - 6.
WANG Xiuli,WANG Meng. Electronic auction scheme applied for double digital signature[J]. Computer Engineering,2012,38(4):4 - 6. (in Chinese)

[6] Rabin M O, Mansour Y, Muthukrishnan S, et al. Strictly-black-box zero-knowledge and efficient validation of financial transactions[A]. International Colloquium on Automata, Languages, and Programming[C]. Springer Berlin Heidelberg,2012:738 - 749.

[7] Rabin, M, Servedio, R, Thorpe, C. Practical secrecy preserving, verifiably correct and trustworthy auctions[A]. Proceedings of IEEE Symposium on Logic in Computer Science[C]. Wroclaw:IEEE,2007.

[8] Micali S, Rabin M O. Cryptography miracles, secure auctions, matching problem verification[J]. Communications of the ACM,2014,57(2):85 - 93.

[9] Zhu Y, Liu L, Chen X. Efficient first-price sealed-bid auction protocols from modified comparable encryption[A]. International Conference on Broadband and Wireless Computing, Communication and Applications(BWCCA)[C]. Krakow:IEEE,2015. 417 - 421.

[10] 丛鑫,訾玲玲,孙劲光. 基于可信拍卖机制的视频移植定价策略[J]. 通信学报,2016,37(4):53 - 63.
CONG Xin;ZI Lingling;SUN Jinguang. Pricing strategy

for video migration based on truthful auction mechanism [J]. Journal on Communications, 2016, 37(4): 53 – 63. (in Chinese)

[11] 郭迟,刘经南,方媛,等. 位置大数据的价值提取与协同挖掘方法[J]. 软件学报,2014,(4):713 – 730.
GUO Chi, LIU Jingnan, FANG Yuan, et al. Value extraction and collaborative mining methods for location big data[J]. Journal of Software, 2014, (4): 713 – 730. (in Chinese)

[12] Li J, Xu Z, Jiang Y, et al. The overview of big data storage and management [A]. International Conference on Cognitive Informatics & Cognitive Computing (ICCI&CC) [C]. London: IEEE, 2014. 510 – 513.

[13] 曹珍富,董晓蕾,周俊,等. 大数据安全与隐私保护研究进展[J]. 计算机研究与发展,2016,(10):2137 – 2151.
CAO Zhenfu, DONG Xiaolei, ZHOU Jun, et al. Research advances on big data security and privacy preserving[J]. Journal of Computer Research and Development, 2016, (10): 2137 – 2151. (in Chinese)

[14] Niyato D, Alsheikh M A, Wang P, et al. Market model and optimal pricing scheme of big data and Internet of Things (IoT) [A]. International Conference on Communications (ICC) [C]. Kuala Lumpur: IEEE, 2016. 1 – 6.

[15] Li H, Dong M, Ota K, et al. Pricing and repurchasing for big data processing in multi-clouds[J]. IEEE Transactions on Emerging Topics in Computing, 2016, 4 (2): 266 – 277.

[16] Merkle R C. Protocols for public key cryptosystems[A]. IEEE 1th Symposium on Security & Privacy [C]. Oakland: IEEE, 1980. 122 – 134.

[17] Schneider B. Applied Cryptography: Protocols, Algorithms, and Source Code in C[M]. John Wiley & Sons, 1996: 84 – 84.

[18] Chaum D. Blind signature for untraceable Payments[A]. Proc. Crypto'82[C]. New York: Plenum Press, 1983. 199 – 203.

[19] 李顺东,戴一奇,游启友. 姚氏百万富翁问题的高效解决方案[J]. 电子学报,2005,33(5):769 – 773.
LI Shundong, DAI Yiqi, YOU Qiyou. An efficient solution to Yao's millionaire s' problem[J]. Acta Electronica Sinica, 2005, 33(5): 769 – 773. (in Chinese)

[20] Pedersen T P. Non-interactive and information-theoretic secure verifiable secret sharing[A]. International Cryptology Conference on Advances in Cryptology[C]. Springer Berlin Heidelberg, 1991. 129 – 140.

[21] Pedroso M A L. The implementation of a split-value verifiable voting system[D]. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2015. 11 – 21.

[22] Montenegro J A, Fischer M J, Lopez J, et al. Secure sealed-bid online auctions using discreet cryptographic proofs[J]. Mathematical and Computer Modelling, 2013, 57(11): 2583 – 2595.

[23] 曹刚. 基于不可信第三方的电子拍卖方案[J]. 计算机工程,2010,36(20):140 – 141,144.
CAO Gang. Electronic auction scheme based on trustless third-party[J]. Computer Engineering, 2010, 36(20): 140 – 141, 144. (in Chinese)

[24] 张京良,马丽珍,王育民. 可达到完全隐私的密封电子拍卖方案[J]. 通信学报,2007,28 (11A):186 – 189.
ZHANG Jingliang, MA Lizhen, WANG Yumin. Fully private sealed—bid auction scheme[J]. Journal on Communications, 2007, 28(11A): 186 – 189. (in Chinese)

作者简介

尹 鑫 女,1992 年生,山东莱芜人,硕士研究生,主要研究方向为密码学与安全协议.
E-mail: csyxcryp@163.com

田有亮(通信作者) 男,1982 年生,贵州盘县人,博士、教授,主要研究领域为算法博弈论、密码学与安全协议.
E-mail: youliangtian@163.com