Enabling Privacy-preserving Auctions in Big Data

Taeho Jung[1], Xiang-Yang Li[12]
[1]Department of Computer Science, Illinois Institute of Technology, Chicago, IL
[2]Department of Computer Science and Technology, TNLIST, Tsinghua University, Beijing

***Abstract*—We study how to enable auctions in the big data context to solve many upcoming data-based decision problems in the near future. We consider the characteristics of the big data including, but not limited to, velocity, volume, variety, and veracity, and we believe any auction mechanism design in the future should take the following factors into consideration: 1) generality (variety); 2) efficiency and scalability (velocity and volume); 3) truthfulness and verifiability (veracity). In this paper, we propose a privacy-preserving construction for auction mechanism design in the big data, which prevents adversaries from learning unnecessary information except those implied in the valid output of the auction. More specifically, we considered one of the most general form of the auction (to deal with the variety), and greatly improved the the efficiency and scalability by approximating the NP-hard problems and avoiding the design based on garbled circuits (to deal with velocity and volume), and finally prevented stakeholders from lying to each other for their own benefit (to deal with the veracity). The comparison with peer work shows that we greatly improved the asymptotic performance of peer works' overhead from the exponential growth to a linear growth and from linear growth to a logarithmic growth, which greatly contributes to the scalability of our mechanism.**

I. Introduction

Increasingly many decisions are made based on the data because of the rich information hidden behind it, and more and more data is being collected almost everywhere nowadays, which will soon lead us to the big data era. Among many 'V's characterizing the big data, we focus on the 4'V's in this paper: variety, volume, velocity, and veracity. The starting point of this research is the observation that various auction mechanisms are adopted in different fields. Spectrum auction [1], [2], cellular networks [3], ad hoc networks [4], cloud computing [5], cognitive radio networks [6], web advertisement [7], and smart grids [8] are good examples. However, the large and diverse pool of the information available for attackers in the big data has increased the privacy concerns, and we present how to enable auctions in the big data context with 4Vs without privacy implications.

A. Variety

Different types of information is available from different sources for different parties in the big data, and the auctions may involve different types of goods. Existing solutions [9]–[12] only deal with single-good auctions and thus lack general applicability in the big data context. To deal with such a variety in the big data, we target at a more general form of the auction than the simple ones which sell only one good at each auction – Combinatorial Auction (CA hereafter). In a single-auctioneer CA, the auctioneer sells multiple heterogeneous goods simultaneously, and bidders bid on any combination of the goods instead of just one. Such auctions have been researched extensively recently [13]–[15], in part due to the generality of it, and in part due to growing applications in which combinatorial bidding is necessary [16], [17].

As further discussed in the following sections, the consideration of the combinatorial auction will bring great challenges to the auction design because of its inherent complexity.

B. Velocity and Volume

The velocity at which data is generated is at the different order of magnitude in the big data from the one in the traditional data, which pushed the volume of the processed data beyond PB, EB, and even ZB (10^9 TB). The velocity and the volume in the big data brings great challenges into the realization of the privacy-preserving auction design in the following two aspects.

Firstly, an early work [18] relies on the secure multi-party computation using garbled circuits [19] and oblivious transfers [20] to solve the CA in a privacy-preserving manner. Such works protect the private information due to the powerful secure multi-party computation, but the circuit size grows very fast *w.r.t.* the CA parameters (number of bidders, range of the bid value, maximum bid, number of goods) which leads to non-polynomial time computation time. Also, the oblivious transfer required for every gate in the circuit introduces a huge communication time as well. Therefore, the works based on garbled circuit are hardly applicable in the big data environment due to the inherent scalability and performance issue. Secondly, the combinatorial auction itself is a computationally hard problem. Even with assumptions which limit bidders' bidding behaviors (e.g., assuming *single-minded* bidders [13]), CA typically requires to solve one or more NP-hard optimization problems, which leads to infeasible generic theoretical designs [21], [22]. Consequently, several works [23], [24] avoiding garbled circuit or oblivious transfer remains impractical because those solutions rely on the dynamic programming to calculate the optimum solution, which leads to a super-polynomial run time.

To address the scalability and performance issue to deal with the volume and velocity of the big data, we exclude the garbled circuits and oblivious transfers in our design, and further replace the exact optimization with the approximated one. This raises another challenge: traditional mechanism designs in CA guarantees truthful bidding to potentially maximize the social welfare based on the assumption that the goods are allocated optimally. Then, those mechanisms do not provide the same guarantee in our setting because we seek for the approximated result. Therefore, we cannot simply implement an existing approximation algorithm in a privacy-preserving manner, and we also need to improve existing mechanisms to preserve the truthfulness (defined later) of the auction.

C. Veracity

Data source in the big data is *almost everywhere* in the world due to the proliferation of the data collection, and most of the data sources are not under strict quality control. Consequently, not all data in the big data era will be credible because of many reasons (*e.g.,* machine factors: errors/inaccuracy/noises; human factors: moral hazard, mistake, misbehavior). In the CA, the veracity issue has an especially great impact because 1) the lying bidder may negatively affect the social welfare or auctioneer's total revenue; 2) and the winners' payments are calculated by the auctioneer, who is well motivated to report a higher fake price. Obliviously (*i.e.,* without knowing the bid values or winners list) achieving a two-way verifiability against both untrusted bidders and auctioneers is another challenge in the privacy-preserving CA construction.

D. Contributions

The contribution of this work is prominent. This is the first paper to envision the privacy-preserving auction mechanism in the upcoming big data age, which is designed based on the four main characteristics (variety, volume, velocity, and veracity) of the big data, and the contributions can also be summarized based on the 4V's: considering the variety of the big data, we explore privacy-preserving constructions for one of the most general auctions, CA; we have designed a scalable and efficient privacy-preserving algorithm to deal with the volume and velocity; and our design also provides two-way verifiability against malicious bidders and auctioneer to be robust to the veracity issue in the big data.

Note that our research does not explicitly work for the anonymity of the bidder or the auctioneer, but in fact our work is the last step of the anonymization. Our work complements the simple anonymization which replaces users' personally identifiable information (PII) with pseudo-random PIIs in the following sense. Such anonymization is vulnerable to various de-anonymization attacks [25], [26] because published attributes can be fingerprinted or co-related with other datasets. By applying our on top of the sanitization, such de-anonymization becomes much more challenging because the attributes of any tuple is protected as well.

II. Preliminaries & Related Work

A. Backgrounds of Combinatorial Auction

Among various types of combinatorial auctions [27] [28], we shall consider the most common type in this work, one-stage, sealed-bid and single-sided CA. In such auctions, each bidder places several bids, the auction terminates and the results are announced (one-stage); no information about other's bids is released prior to the auction termination (sealed-bid); and one auctioneer is selling several goods to multiple bidders (single-sided).

Given such a CA, its mechanism design is composed of two parts. Firstly, winners of the auction are chosen based on their submitted bundles and bids **winner determination**, then each winner's payment is determined by some mechanism **payment determination**. Note that a winner's payment may not be equal to hid bid.

Winner Determination and Objective Function

The standard goal of the design is to maximize the social welfare [13], [29], [30], which is the sum of winners' reported bids on their allocated goods. An alternative goal is to maximize the auctioneer's revenue, which is the sum of winners' payments. Maximizing the revenue is closely related to the social welfare maximization, therefore we focus on how to maximize the social welfare. As aforementioned finding an allocation maximizing the social welfare is NP-hard [31], and it has been shown that the optimal allocation can be approximated within a factor of $O(m^{\frac{1}{2}})$ but not to a factor of $O(m^{\frac{1}{2}-\epsilon})$ for any $\epsilon > 0$ [13], [15], where m is the number of total goods.

Payment Determination and the Truthfulness

Each bidder's bid may not truly reflect his valuation of the bundle. The payment is determined by all bidders' bids, and therefore bidders may try to report a fake valuation to decrease their payment or win a chance to win the auction.

Definition 1. *An auction is truthful if reporting a true valuation is a weakly dominant strategy for every bidder, and utility of any honest bidder is non-negative.*

That is, no bidder can increase his benefit by lying no matter other bidders lie or not. Naturally, the payment mechanism determines whether the auction is truthful, and the one in the famous Generalized Vickrey Auction (GVA, [21]) guarantees the truthful auction, but determining one bidder's payment requires finding an optimum allocation without him, which is already shown to be NP-hard. Therefore, it is infeasible to implement the GVA in reality, and we study the truthfulness in conjunction with the aforementioned approximation.

A truthful mechanism for the approximated allocation is introduced in [13]. Let L be the sorted list of bundles in the greedy allocation (sorted by bidders' norm $\frac{b}{\sqrt{|S|}}$). For any bundle i, denote the first bundle j in L which would have been allocated if i were denied at first as **candidate** of i. Then, i's payment is $\frac{b'}{\sqrt{|S'|}}\sqrt{|S|}$ where b' is the bid of the candidate bundle, S' is the candidate bundle, and S is the allocated bundle i. This payment guarantees the truthfulness of the auction as proved in [13].

B. Privacy-preserving Combinatorial Auctions

Various approaches are proposed to achieve a private sealed-bid auction [9]–[12], [32], but much less attention is paid to the combinatorial auction. In general, recently proposed approaches for the secure multi-agent combinatorial auction can be divided into two classes: first class based on Secure Multi-party Computation (SMC) and the other class based on Homomorphic Encryption (HE).

To the best of our knowledge, [18], [23], [24], [33] are the only works solving CA in a privacy-preserving manner. [18] solves it by leveraging SMC, but as shown in Palmer's implementation, the solution based on SMC does not scale well because the circuit needs to implement all if-else branches in it in order to accept arbitrary input. Besides, [23], [24] designed a secure multi-agent dynamic programming based on HE, which is in turn used to design the privacy-preserving winner determination in CA. Pan *et al.* [33] also designed a combinatorial auction based on HE. However, these all target at solving the optimum solution for the winner determination problem, and their protocols cannot be run in polynomial time due to the inherent hardness of the problem.

Besides, our work has one more advantage: our auction scheme is the only one which presents a privacy-preserving payment determination mechanism to guarantee the truthfulness of the auction, while all aforementioned works only solve the winner determination problem in the combinatorial auction.

III. Secure Combinatorial Auction Model

A. Auction Model

A set of m goods $G = \{g_1, \cdots, g_m\}$ are auctioned to n bidders $\{\mathcal{B}_1, \cdots, \mathcal{B}_n\}$ in the CA, and $\mathcal{B}_i$ proposes his bid b_i (i.e., maximum willingness to pay) on the bundle S_i, and the bid might be different from his true valuation v_i if he wishes to lie. A set W of winners are chosen by selecting a group of conflict-free bidders whose social welfare is maximized. After the winners are chosen, each winner $\mathcal{B}_i$'s payment p_i is determined by the auction mechanism based on all bidders bids. Then, we assume a quasilinear utility for $\mathcal{B}_i$: $u_i = v_i - p_i$ if $\mathcal{B}_i$ is a winner, and 0 otherwise.

B. Adversarial Model

Two adversaries should be considered: adversarial auctioneer and adversarial bidders. The auctioneer is assumed to be **curious**, **malicious** and **ignorant**. He is interested in bidders' bids and bundles to improve his business (called "curious"). For example, he may try to infer bidders' preferences and rivalry relationship based on the bids and the bundles. The auctioneer may also report a fake payment to the winners to illegally increase his revenue (called "malicious"), but he is not aware of bidders' side information such as distribution of bid values or bidders' preferences on goods (called "ignorant"). Bidders are assumed to be **selfish**, **curious** and **non-cooperative**. Their objective is to maximize their own utilities, and bidders will report fake valuations if the utility is increased by doing so (called "selfish"). On the other hand, bidders are interested in others' bids and bundles to improve the decision making (called "curious"). However, they will not collude with other bidders or the auctioneer (called "non-cooperative").

C. Privacy Definitions

Definition 2. *Given all the communication strings $\mathcal{C}$ during the auction and the output of the auction* Output*, an adversary's advantage over the loser $\mathcal{B}_i$'s bid b_i is defined as*

$$adv_{b_i} = \Pr[b_i|\mathcal{C}, \mathsf{Output} \leftarrow \mathcal{A}_{our}(1^{\kappa})] - \Pr[b_i|\mathsf{Output} \leftarrow \mathcal{A}_{black}]$$

where $\Pr[b_i]$ *is the probability that b_i is correctly inferred, $\mathcal{A}_{our}$ is our algorithms, κ is the security parameter of $\mathcal{A}_{our}$, and $\mathcal{A}_{black}$ is a perfectly secure black-box algorithm.*

We focus on the confidentiality of auction losers' bids in this paper because winners' bids can be learned from the valid outputs of the auctions anyway (*e.g.,* claimed bundle and the payments).

Definition 3. *Given all the communication strings $\mathcal{C}$ during the auction and the output of the auction* Output*, an adversary's advantage over any bidder $\mathcal{B}_i$'s bundle S_i is defined as*

$$adv_{S_i} = \Pr[S_i|\mathcal{C}, \mathsf{Output} \leftarrow \mathcal{A}_{our}(1^{\kappa})] - \Pr[S_i|\mathsf{Output} \leftarrow \mathcal{A}_{black}]$$

where $\Pr[S_i]$ *is the probability that any information about S_i is inferred, and other notations are same as in Definition 2.*

Informally, these advantages measure how much side information an adversary gains during our privacy-preserving auction by measuring the increased probabilities. In other words, they reflect how much side information is disclosed other than what is derivable from the valid auction output.

IV. Building Blocks

A. Homomorphic Encryption

Homomorphic encryption allows specific computations to be directly carried on ciphertexts while preserving their decryptability. We employ the Paillier cryptosystem to implement a one-way privacy-preserving scalar product for efficient our winner determination. In short, Paillier cryptosystem has the following homomorphic property:

$$E(m_1) \cdot E(m_2) = E(m_1 + m_2) - \text{Addition}$$
$$E(m_1)^{m_2} = E(m_1 \cdot m_2) - \text{Multiplication}$$

B. Blind Signature

In our work, we employ a signer who is involved only to generate a signature of each bidder's value. In order to preserve user privacy, we employ the blinded Nyberg-Rueppel scheme in [34], where a signer can generate a signature of a value m without 'seeing' it. At a later time, the signature can be provided to recover the m, whose authenticity is guaranteed. In our work, we use the blind signature scheme to verify whether the payment is calculated correctly. Since the authenticity of the bids are guaranteed by the truthful mechanism, we do not need the signer to verify the authenticity of them, and the Blinded Nyberg-Rueppel scheme suffices. For the simplicity, we denote the signature of m as $\text{Sig}(m)$ hereafter.

V. Designing Auction Mechanism For Big Data

Before the auction proceeds, all bidders are asked to blindly sign their ψ_i and S_i via a third-party signer $\mathcal{T}$. Since the signature is blindly signed, $\mathcal{T}$ learns nothing. These signatures will be used to verify the authenticity of the bids later. Also, note that the entire protocol is performed in an anonymized network to enhance user privacy.

A. Privacy-preserving Winner Determination

Algorithm 1 Greedy Winner Determination

```
1: $A := \emptyset, W := \emptyset$. For each $\mathcal{B}_i$, computes $\psi_i = \frac{b_i}{\sqrt{|S_i|}}$.
2: Sort the instances in the non-increasing order of norm $\psi_i$. Denote the sorted list as $L$.
3: For each $\mathcal{B}_i \in L$ (in the sorted order), check whether $A \cap S_i = \emptyset$. If true, $A := A \cup S_i, W := W \cup \mathcal{B}_i$.
4: $A^* := A$. Announce $W$ as the winners. Finally allocated goods are $A^*$.
```

The above approximation algorithm for the winner determination guarantees an approximation ratio of at least $O(\sqrt{m})$ [13], where m is the number of total goods, and this has been proved to be the best approximation ratio that can be achieved [13], [15]. To guarantee each bidder's privacy, we cannot explicitly perform the sorting (Step 2.) because the order of all bidders' norms reveals excessive side information about the losers' b_i or S_i even if the norm ψ_i does not directly reveal either one. Therefore, we unravel Step 2. and Step 3. as follows. Firstly, among the *encrypted* bundles that can be allocated (*i.e.,* no overlap with already-allocated goods), find out the one whose corresponding norm ψ_i is the maximum (without revealing ψ_i's value). Then, find out the winner who owns the bundle (up to previous step, every one was anonymous). Finally, update A correspondingly and keeps looking for the next feasible bundle with maximum ψ_i. Note that the IDs are already anonymized either via sanitization or anonymized network such as Tor [35], therefore only the winners' identities are revealed to the auctioneer.

To further describe the privacy-preserving unraveled greedy algorithm more easily, we first elaborate a sub-procedure in it: feasibility evaluation.

Feasibility Evaluation

Given a bundle S_i, whether it is feasible (*i.e.,* does not overlap with already-allocated goods) must be evaluated in a privacy-preserving manner in order to keep the confidentiality of bids or bundles. Firstly, we use an m-dimension binary vector $\mathbf{A}$ represent the allocation status of all goods (*i.e.,* the goods in A), where the k-th bit $a_k = 1$ if the k-th good g_k is allocated already and 0 otherwise. Similarly, we use another vector $\mathbf{S}_i$ to represent $\mathcal{B}_i$'s bundle S_i, where $\mathbf{S}_i$'s k-th bit $s_{i,k} = 1$ if $g_k \in S_i$ and 0 otherwise. Then, $A \cap S_i = \emptyset$ if and only if $\mathbf{A} \cdot \mathbf{S}_i = \sum_{k=1}^{m} a_k s_{i,k} = 0$. If the scalar product is θ, that means $\mathcal{B}_i$'s bundle S_i includes θ already-allocated goods. In order to keep θ and $\{s_{i,k}\}$ secret to the auctioneer, and to keep θ and $\{a_k\}$ secret to $\mathcal{B}_i$, we propose the following protocol (Algorithm V-A) to let the auctioneer learn whether the above sum is equal to 0.

In the protocol, if $\delta_i \cdot \mathbf{A} \cdot \mathbf{S}_i = 0$, Auc learns S_i is feasible, and if S_i is not feasible, the outcome is $\delta_i \theta$ which is indistinguishable to a random number in $\mathbb{Z}_n$ from the auctioneer's perspective.

With this feasibility evaluation, we are ready to present our privacy-preserving winner determination algorithm which works as a black-box algorithm outputting the winning bundles and the winners only (Algorithm V-A). Essentially, the outcome of 3-b is 0 if and only if $\mathcal{B}_i$'s ψ_i is equal to the Auc's guess, and c at 3-c is equal to 0 if and only if $\mathcal{B}_i$'s S_i is feasible at the current allocation A. Then, the final outcome at 3-e is equal to 0 if and only if $\mathcal{B}_i$'s norm ψ_i is the maximum among all the remaining bidders and his bundle is also feasible.

Algorithm 2 Privacy-preserving Scalar Product

```
1: $Auc$ picks a pair of Paillier cryptosystem key: $PK'_{Auc} = (n, g)$, $SK'_{Auc} = \lambda$ (Section IV).
2: $Auc$ encrypts every bit $a_k$ homomorphically and sends its ciphertext $E_{Auc}(a_k)$ to the bidder $\mathcal{B}_i$ whose bundle is being checked.
3: Upon receiving $m$ ciphertexts, $\mathcal{B}_i$ first picks a random number $\delta_i \in \mathbb{Z}_n$ and performs following operations:
       $\forall k : c_k = E_{Auc}(a_k)^{\delta_i s_{i,k}} = E_{Auc}(\delta_i a_k s_{i,k})$
   Then, he computes the following and sends to the auctioneer:
       $c = \prod_{k=1}^{m} c_k = E_{Auc}(\delta_i \sum_{k=1}^{m} a_k s_{i,k})$
4: The auctioneer decrypts the received ciphertext using his secret key, which is the scalar product $\delta_i \cdot \mathbf{A} \cdot \mathbf{S}_i$.
```

The algorithm deserves further clarifications at the places terms or phrases are marked bold. Firstly, the way Auc guesses the maximum norm at step 3 is critical for the performance. Given the range of the possible values for the norms, Auc performs a binary search until finding a value ψ^* such that the final outcome at 3-e yields 0 at ψ^* but not at $\psi^* + 1$. If such values are discovered, the next binary search can be started from ψ^*. Secondly, given the outcome yielding 0 at 3-e, Auc must find out the winner first because every bidder is anonymous yet up to this point. This can be done by declaring the anonymous ID of the winner and asking him to reveal his $\psi_i, S_i, \text{Sig}(\psi_i)$, and $\text{Sig}(S_i)$ to auctioneer for the confirmation. Since everything was encrypted under Auc's keys, no bidders gain any information about the ψ^*. On the other hand, because the entire protocol is conducted in an anonymized network, declaring the anonymous ID does not breach winner privacy (anonymous ID is often a one time identity). If $psi_i = \psi^*$, Auc learns the bidder is the winner, and marks his goods as allocated in A (the authenticity of ψ_i and S_i can be verified with the signatures). Finally, Auc learns no more update is possible when he finds out the binary search is terminated but no one yielded 0 at 3-e.

B. Privacy-preserving Verifiable Payment Determination

In aforementioned truthful auction mechanism (Section II), the auctioneer determines a winner $\mathcal{B}_i$'s payment as follows. Among the bidders whose bundle would have been allocated if $\mathcal{B}_i$ were not the winner (i.e., the candidate of $\mathcal{B}_i$), the auctioneer finds out the one with the maximum ψ (say ψ_j of bidder $\mathcal{B}_j$). Then, $\mathcal{B}_i$'s payment is $p_i = \frac{b_j}{\sqrt{|S_j|}}\sqrt{|S_i|}$. Three

Algorithm 3 Privacy-preserving Winner Determination

```
1: $A := \emptyset, W := \emptyset, B = \{\mathcal{B}_i\}_i$. Every $\mathcal{B}_i$ computes $\psi_i = \frac{b_i}{\sqrt{|S_i|}}$ individually.
2: $Auc$ picks a pair of Paillier cryptosystem key: $PK = (n, g)$, $SK = \lambda$ (Section IV), and publishes $PK$.
3: $Auc$ **guesses** the maximum $\psi$ value $\psi^*$, and checks whether there exists a bundle with $\psi_i = \psi^*$ that can be allocated by performing the following procedure with every bidder $\mathcal{B}_i \in B$.
   3-a: $Auc$ sends $E_{Auc}(\psi^*)$ (ciphertext of $\psi^*$) to $\mathcal{B}_i$.
   3-b: $\mathcal{B}_i$ picks a random number $\delta'_i \in \mathbb{Z}_n$, then calculates:
        $\Big(E_{Auc}(\psi^*) \cdot E_{Auc}(-\psi_i)\Big)^{\delta'_i} = E_{Auc}\Big(\delta'_i(\psi^* - \psi_i)\Big)$
   3-c: $Auc$ sends out encrypted $\{a_k\}$'s to $\mathcal{B}_i$, and $\mathcal{B}_i$ calculates $c = E_{Auc}(\delta_i \sum a_k s_{i,k})$ as in the aforementioned scalar product calculation (Algorithm V-A).
   3-d: $\mathcal{B}_i$ sends the following to $Auc$:
        $E_{Auc}\Big(\delta'_i(\psi^* - \psi_i)\Big) \cdot E_{Auc}\Big(\delta_i \sum a_k s_{i,k}\Big)$
        $= E_{Auc}\Big(\delta'_i(\psi^* - \psi_i) + \delta_i \sum_{k=1}^{m} a_k s_{i,k}\Big)$
   3-e: $Auc$ decrypts it to see whether it is equal to 0.
4: Step 3 is repeated with different $\psi^*$ to find out the maximum $\psi^*$ yielding 0 in 3-e. If an anonymous bidder's outcome is discovered to yield 0 in 3-e, $Auc$ **finds out the winner**, mark the corresponding goods as allocated in $A$, and add $\mathcal{B}_i$ to $W$. Then, repeat 3. again with updated sets. This is repeated until **no more update is possible**.
5: Set $A^* = A$. Then, $Auc$ learns $W$ is the set of winners, and $A^*$ is the finalized allocation. Then, he proceeds to payment determination.
```

parties are engaged here: auctioneer Auc, winner $\mathcal{B}_i$ and $\mathcal{B}_i$'s candidate $\mathcal{B}_j$. The auctioneer Auc needs to know p_i without knowing $\mathcal{B}_j$'s bundle or bid; the winner $\mathcal{B}_i$ needs to know p_i without knowing $\mathcal{B}_j$'s bundle or bid, and he should not even know who is the $\mathcal{B}_j$; and finally, the bidder $\mathcal{B}_j$ does not need to know anything from this whole process. Furthermore, both the auctioneer and the winner should be able to verify the payment. We present the privacy-preserving verifiable payment determination (Algorithm V-B) which fulfills above requirements as in Algorithm V-B.

Since Auc uses aforementioned privacy-preserving feasibility evaluation, he does not learn about $\mathcal{B}_j$'s bundle, and therefore he does not learn b_j from $\psi_j = \frac{b_j}{\sqrt{|S_j|}}$. The winner $\mathcal{B}_i$ does not learn about b_j due to the same reason, and he also does not know who is $\mathcal{B}_j$ since he does not even communicate with $\mathcal{B}_j$. On the other hand, owing to the signature Sig(ψ_j) generated by $\mathcal{T}$, Auc is convinced that $\mathcal{B}_j$ did not report a fake lower $\hat{n}_j$ to harm Auc's business, and the winner $\mathcal{B}_i$ believes Auc did not tell a higher p_i to increase Auc's revenue.

Algorithm 4 $\mathcal{B}_i$'s Verifiable Payment Determination

```
1: The auctioneer $Auc$ excludes $B_i$ from $B$, and finds out the winner with $(A^* - S_i)$ by following the same procedure as the winner determination, where $A^*$ is the finally sold goods. If $\mathcal{B}_j$ is the winner, then he is the candidate of $\mathcal{B}_i$. Different from the original winner determination, $\mathcal{B}_j$ only reveals $\psi_j = \frac{b_j}{\sqrt{|S_j|}}$ and Sig$(\psi_j)$ to the auctioneer for the confirmation.
2: If a candidate is found, $Auc$ calculates $p_i = \psi_j \sqrt{|S_i|}$ and sends $p_i$ as well as the Sig$(\psi_j)$ to $\mathcal{B}_i$. Otherwise, $Auc$ sets $p_i$ as the reserve price (*e.g.,* pre-defined minimum price) and informs $\mathcal{B}_i$ that his payment is the reserve price.
3: If the payment is not the reserve price, $\mathcal{B}_i$ recovers $\psi_j$ from Sig$(\psi_j)$, and verifies whether $p_i = \psi_j \sqrt{|S_i|}$. If they are not equal to each other, he learns that the payment is incorrect.
```

VI. Performance Evaluation

A. Communication Overhead

The communication overhead in terms of the data transmission is depicted in the following table, where n is the total number of bidders, κ is the security parameter, $|W|$ is the number of winners, and $|\psi|$ is the size (bit-length) of the fixed-point representation of the norm values.

TABLE I
Communication Complexity

Winner Determination		
	Receive	Send
Auctioneer	$O(n \cdot \lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$	$O(n \cdot \lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$
Signer	$O(n \cdot \kappa)$	$O(n \cdot \kappa)$
Per bidder	$O(\lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$	$O(\lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$
Payment Determination		
	Receive	Send
Auctioneer	$O(n \cdot \lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$	$O(n \cdot \lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$
Per Winner	$O(\kappa)$	0
Per Loser	$O(\lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$	$O(\lvert W\rvert \cdot m \cdot \kappa \cdot \lvert\psi\rvert)$

B. Comparison with peer works

Due to the space limit, we do not plot the computation overhead in this paper. Instead, we compare the complexities of the algorithms used in peer works and ours, and present it in Table VI-B.

TABLE II
Growth of computation overhead

Variable	Ours	[18]	[23], [24], [33]
Maximum Bid	**Logarithmic**	Logarithmic	Linear
Bidder #	Linear	Linear	Linear
Goods #	**Linear**	Exponential	Exponential

Further, by comparing the actual run time of our implementation and the one in [18] (results are not presented in this paper due to the space limit), one can notice that our overhead grows with much smaller constant factors as well because our protocols do not involve oblivious transfer or garbled circuits.

Considering that our overhead can be dramatically reduced by replacing the old Paillier's cryptosystem with more advanced and faster additive homomorphic encryption (*e.g.,* [36]), the performance advantages over peer works is very prominent.

Note that our improvement in the asymptotic performance is a necessary step before privacy-preserving auction mechanisms' application in the big data context due to the large volume and velocity. A recent research [37] indicates that the polynomial time complexity, which used to be accepted as tractable one, is not tractable any more in the big data context, and therefore achieving the linear complexity is one of the necessary (yet not sufficient) conditions of the applications in the big data era.

VII. Conclusion & Future Direction

In this paper, we presented a privacy-preserving auction design for the big data context where volume, velocity, variety, and veracity may be challenging for the auction designers. We focused on the combinatorial auction for the variety; we achieved a much better asymptotic performance than peer works by approximating the NP-hard problem in the combinatorial auction for the volume and velocity; and for the veracity issue resulted from untrusted auctioneer and bidders, we designed an auction scheme that can guarantee the truthfulness bidding and price verifiability. We presented a construction where any adversary's view is same as the one in a black-box algorithm, and our analysis also shows that it greatly improved the asymptotic performance when compared to the peer works. Considering the exascale computing in the big data, our work is not yet perfectly suitable for the big data context where more than exabytes of data is involved. However, we firmly believe this work is a big step towards the auction design in the big data era, and we are more than convinced that successive research on our research will finally lead to the practical auction mechanisms for the big data applications.

References

[1] X. Feng, Y. Chen, J. Zhang, Q. Zhang, and B. Li, "Tahes: Truthful double auction for heterogeneous spectrums," in *INFOCOM*, IEEE, 2013.

[2] Q. Huang, Y. Tao, and F. Wu, "Spring: A strategy-proof and privacy preserving spectrum auction mechanism," in *INFOCOM*, IEEE, 2013.

[3] W. Dong, S. Rallapalli, R. Jana, L. Qiu, R. K.K., L. Razoumov, Y. Zhang, and T. W. Cho, "ideal: Incentivized dynamic cellular offloading via auctions," in *INFOCOM*, IEEE, 2013.

[4] M. Li, P. Li, M. Pan, and J. Sun, "Economic-robust transmission opportunity auction in multi-hop wireless networks," in *INFOCOM*, IEEE, 2013.

[5] W.-Y. Lin, G.-Y. Lin, and H.-Y. Wei, "Dynamic auction mechanism for cloud resource allocation," in *CCGrid*, IEEE, 2010.

[6] L. Chen, S. Iellamo, M. Coupechoux, and P. Godlewski, "An auction framework for spectrum allocation with interference constraint in cognitive radio networks," in *INFOCOM*, IEEE, 2010.

[7] C. Borgs, J. Chayes, N. Immorlica, K. Jain, O. Etesami, and M. Mahdian, "Dynamics of bid optimization in online advertisement auctions," in *WWW*, pp. 531–540, ACM, 2007.

[8] T. K. Wijaya, K. Larson, and K. Aberer, "Matching demand with supply in the smart grid using agent-based multiunit auction," in *COMSNETS*, pp. 1–6, Ieee, 2013.

[9] F. Brandt and T. Sandholm, "Efficient privacy-preserving protocols for multi-unit auctions," in *FC*, pp. 298–312, Springer, 2005.

[10] K. Suzuki, K. Kobayashi, and H. Morita, "Efficient sealed-bid auction using hash chain," in *ICISC*, pp. 183–191, Springer, 2001.

[11] S. G. Stubblebine and P. F. Syverson, "Fair on-line auctions without special trusted parties," in *FC*, pp. 230–240, Springer, 1999.

[12] M. Naor, B. Pinkas, and R. Sumner, "Privacy preserving auctions and mechanism design," in *EC*, pp. 129–139, ACM, 1999.

[13] D. Lehmann, L. I. Oćallaghan, and Y. Shoham, "Truth revelation in approximately efficient combinatorial auctions," *JACM*, vol. 49, no. 5, pp. 577–602, 2002.

[14] T. Sandholm, S. Suri, A. Gilpin, and D. Levine, "Cabob: A fast optimal algorithm for combinatorial auctions," in *IJCAI*, vol. 17, pp. 1102–1108, LAWRENCE ERLBAUM ASSOCIATES LTD, 2001.

[15] T. Sandholm, "Algorithm for optimal winner determination in combinatorial auctions," *AI*, vol. 135, no. 1, pp. 1–54, 2002.

[16] N. Fukuta and T. Ito, "Toward combinatorial auction-based better electric power allocation on sustainable electric power systems," in *CEC*, pp. 392–399, IEEE, 2011.

[17] S. Zaman and D. Grosu, "Combinatorial auction-based allocation of virtual machine instances in clouds," *Journal of Parallel and Distributed Computing*, vol. 73, no. 4, pp. 495–508, 2013.

[18] B. Palmer, K. Bubendorfer, and I. Welch, "Development and evaluation of a secure, privacy preserving combinatorial auction," 2011.

[19] A. C.-C. Yao, "How to generate and exchange secrets," in *FOCS*, pp. 162–167, IEEE, 1986.

[20] M. O. Rabin, "How to exchange secrets with oblivious transfer.," *IACR Cryptology ePrint Archive*, vol. 2005, p. 187, 2005.

[21] J. K. MacKie-Mason and H. R. Varian, "Generalized vickrey auctions," 1994.

[22] H. R. Varian, "Economic mechanism design for computerized agents," in *EC*, pp. 13–21, 1995.

[23] K. Suzuki and M. Yokoo, "Secure combinatorial auctions by dynamic programming with polynomial secret sharing," in *FC*, pp. 44–56, Springer, 2003.

[24] M. Yokoo and K. Suzuki, "Secure multi-agent dynamic programming based on homomorphic encryption and its application to combinatorial auctions," in *IFAAMAS*, pp. 112–119, ACM, 2002.

[25] G. Danezis and C. Troncoso, "You cannot hide for long: de-anonymization of real-world dynamic behaviour," in *WPES*, pp. 49–60, ACM, 2013.

[26] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *CVPR*, pp. 2360–2367, IEEE, 2010.

[27] P. Cramton, Y. Shoham, and R. Steinberg, "Combinatorial auctions," 2006.

[28] M. H. Rothkopf, "Bidding theory: the phenomena to be modeled," 1983.

[29] S. Dobzinski, N. Nisan, and M. Schapira, "Approximation algorithms for combinatorial auctions with complement-free bidders," in *STOC*, pp. 610–618, ACM, 2005.

[30] M. Yokoo and K. Suzuki, "Secure generalized vickrey auction without third-party servers," in *FC*, pp. 132–146, Springer, 2004.

[31] M. H. Rothkopf, A. Pekeč, and R. M. Harstad, "Computationally manageable combinational auctions," *Management science*, 1998.

[32] C. Cachin, "Efficient private bidding and auctions with an oblivious third party," in *CCS*, pp. 120–127, ACM, 1999.

[33] M. Pan, X. Zhu, and Y. Fang, "Using homomorphic encryption to secure the combinatorial spectrum auction without the trustworthy auctioneer," *Wireless Networks*, vol. 18, no. 2, pp. 113–128, 2012.

[34] J. L. Camenisch, J.-M. Piveteau, and M. A. Stadler, "Blind signatures based on the discrete logarithm problem," in *EUROCRYPT*, pp. 428–432, Springer, 1995.

[35] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," tech. rep., DTIC Document, 2004.

[36] J. H. Cheon, H. T. Lee, and J. H. Seo, "A new additive homomorphic encryption based on the co-acd problem," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 287–298, ACM, 2014.

[37] W. Fan, F. Geerts, and F. Neven, "Making queries tractable on big data with preprocessing: through the eyes of complexity theory," *Proceedings of the VLDB Endowment*, vol. 6, no. 9, pp. 685–696, 2013.