Account Trade Systems: Achieving Data Truthfulness and Privacy Preservation in Data Markets Using Secure Blockchain-Based Medical Data Exchange Models

Dr. Lukas Meier[1]*
[1]University of Zurich, Department of Medical Informatics and Digital Health Security, Zurich, Switzerland

ABSTRACT

In this article, we proposed a set of accountable protocols denoted as AccounTrade for big data trading among dishonest consumers. For attaining secure big data trading environment. Bookkeeping and accountability are achieved against consumers throughout trading. Examines the consumer's responsibilities in the data trading, and then designed AccountTrade to achieve accountability against dishonest consumers that are likely to deviate from their responsibility. Specially uniqueness index is defined and proposed it is a new measure of data uniqueness. To avoid result from being manipulated by a false-name binding attack, we propose a Multi-round False-name Proof Auction (MFPA) scheme, which enables data trading among data owners (sellers) and data collectors (buyers).

KEYWORDS: AccontTrade, Uniqueness index, MFP, Auctioneer.

1. INTRODUCTION

A stock market (also known as an equity market or share market), is a collection of buyers and sellers of stocks. These stocks represent ownership interests in companies. These may include publicly or privately traded securities. The New York Stock Exchange (NYSE) is an example of a share market.

Usually, large companies will list their stock on a stock exchange because it makes their shares more liquid (i.e., easy to buy and sell), which investors love. This liquidity also attracts international investors. Many leading companies and corporations are traded in the stock market. Banks, airplane companies, online shopping sites, technology companies, fashion brands, the list goes on and on.

The role of stock brokers have evolved in a big way over the last few years. Now brokers are not just here to buy or sell stocks on behalf of their clients. They play a bigger role in helping an investor wade through whole investment process; providing research based advice on stocks to helping client to invest in alternative assets; and subscribing to IPOs and mutual funds schemes.

Most stocks are traded on physical or virtual exchanges. The New York Stock Exchange (NYSE), for example, is a physical exchange where some trades are placed manually on a trading floor (other trading activity is conducted electronically). NASDAQ, on the other hand, is a fully electronic exchange where all trading activity occurs over an extensive computer network, matching investors from around the world to each other at the blink of an eye.

Investors and traders submit orders to buy and sell stock shares, either through a broker or by using an online order entry interface (i.e., a trading platform such as E*Trade).

A buyer bids to purchase shares at a specified price (or at the best available price) and a seller asks to sell the stock at a specified price (or at the best available price). When a bid and an ask match, a transaction occurs and both orders will be filled. In a very liquid market, the orders will be filled almost instantaneously. In a thinly traded market, however, the order may not be filled quickly or at all.

A stock or share (also known as "equity") is a financial instrument that represents ownership in a company or corporation and represents a proportionate claim on its assets (what it owns) and earnings (what it generates in profits). Stock ownership implies that the shareholder owns a slice of the company equal to the number of shares held as a proportion of the company's total outstanding shares. For instance, an individual or entity that owns 100,000 shares of a company with 1 million outstanding shares would have a 10% ownership stake in it. Most companies have outstanding shares that run into the millions or billions.

While transaction between the buyer and seller huge volumes of data are created know as bigdata. Hence This article propose a set of accountable protocols know as AccountTrade for big data trading hosted by brokers.

AcountTrade enables broker to maintain accountability against dishonest consumers throughout the trading by detecting the misbehaviour. Misbehaviour defined in this article are tax evasion, denial of purchase and resell of others dataset. Also proposed to detect blatant copy in the dataset uploaded by the owners, by detecting whether the uploaded one is derived from already existing one. Here proposed an algorithm called uniqueness index to find out uploaded one is not same as existing.

Auction mechanisms have been applied, and have significant potential to facilitate data transactions in a fair, truthful, and secure way. The property of incentive compatibility is ensured by a truthful auction mechanism, defines that the bidders can obtain highest utility if and only if they submit their bids and asks truthfully. Furthermore, a truthful and fair auction should also secure the optimal auction results from being manipulated by false-name bidding attacks, where users (participants) utilize multiple identities or accounts to get the auction results. To address these issues, we propose a Multi-round False-name Proof Auction (MFPA) scheme, which enables data trading among data owners (sellers) and data collectors (buyers).

The main contributions of this paper can be summarized as follows:
First, we propose a one-side data auction market that consists of data owners, data collectors and the auctioneer. The data owner and multiple reliable data collectors are treated as the seller and buyers, respectively. The operation of the data trading market is controlled by the auctioneer in the cloud, which is capable of making decisions on big data allocation among buyers and the seller.

Second, we present a Multi-round False-name Proof Auction (MFPA) scheme that enables efficient big data trading. To defend against false-name bidding attacks, the MFPA scheme is designed to run in multiple rounds. In each round, the data from the seller is sold in bundles, and each buyer is able to achieve one bundle of the data at most. We carry out a theoretical analysis to prove that MFPA achieves the desired economic properties of incentive compatibility as well as computational efficiency. We also prove that the bidders cannot improve his/her utility by launching false-name bidding attacks in the MFPA market.

Third, we validate the effectiveness of the MFPA scheme on a data trading market with multiple buyers and one seller. Our experimental results show that the proposed MFPA scheme achieves good performance in terms of social surplus, buy- ers/seller satisfaction ratio, and computation overhead. We also show that MFPA is capable of defending against false- name bidding attacks, and compare this with the Generalized Vickrey Auction (GVA) scheme .

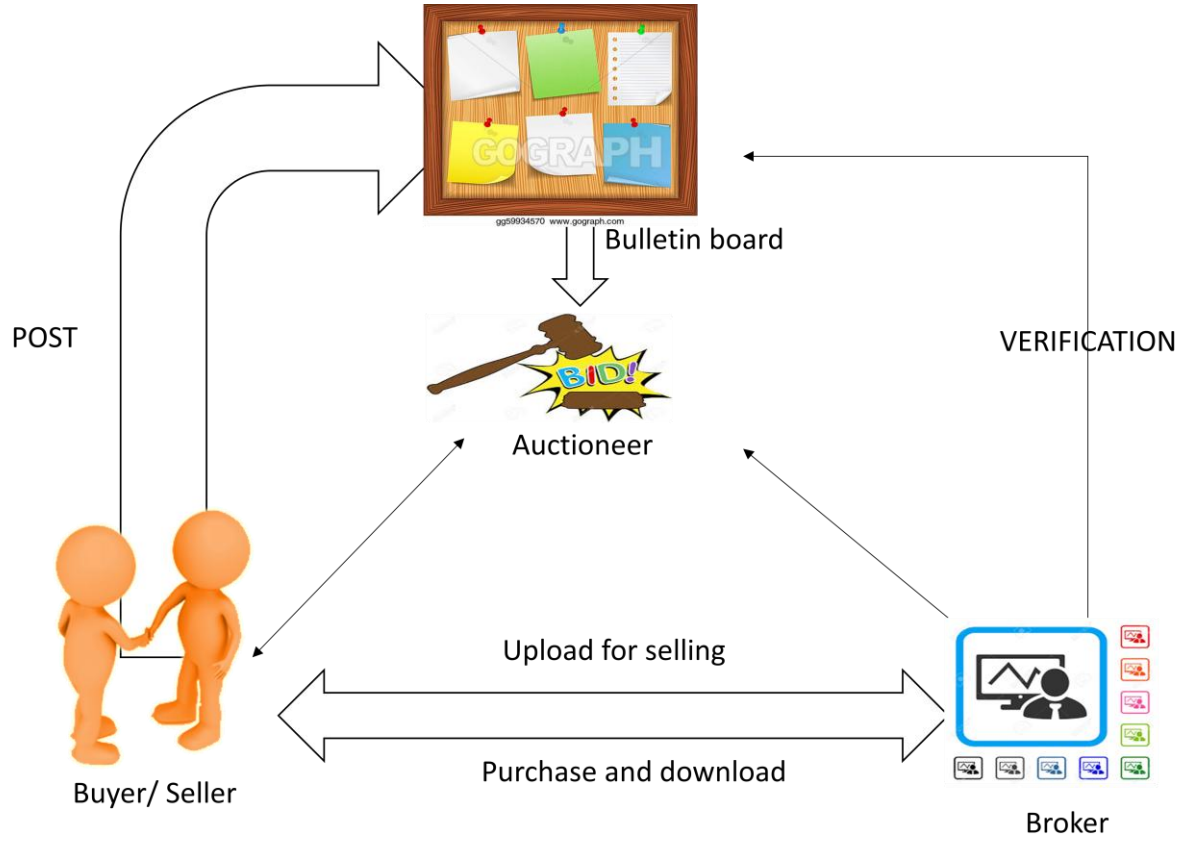
Fourth, We define formal models of accountability (symbolic and computational ones) for big data trading, and we design accountable protocols Upload, Examine, and Download that are provably accountable

2. SYSYEM MODEL

In this paper, we consider that the data trading market is composed of one data owner (seller), multiple data collectors (buyers), and an auctioneer in the cloud. Fig. 1 illustrates the system model.

we consider a practical scenario in which the trading process only occurs between the data owner (seller) and data collectors (buyers or users[1]), instead of between the data seller and users directly. The reason for this is the number of data buyers can be large, and once an individual buyer that is unreliable acquires the data, he/she can copy the data and sell it at a lower price. As a consequence, the utility of big data would be jeopardized. Thus, in this paper, the market is operated by the auctioneer, according to the bids and asks submitted by data owner (seller) and data collectors (buyers). In this paper, we consider to leverage cloud environment where the auctioneer is the trustworthy part to protect the purchased data will not be further resold to other parties.. **Adversary Model.** In the data trading market, buyers are induced to submit their demands for the data to the auctioneer. The auctioneer should always make an optimal decision to maximize the utilities of the seller and buyers. It is worth noting that the payment is the key determinant of the buyers' utility (i.e., the gain of using the purchased data), which is highly related to the other buyers' bids. In addition, we consider the auction process as a sealed-bid market, in which the buyers are not capable of knowing the bidding information of other buyers, and the auction process needs to be fair to all participants. Nonetheless, some "strategic" buyers would like to improve their utilities by colluding with eacho ther to manipulate the auction results, which is unfair to the non-colluding buyers. data will not be further resold to other parties.. In the data trading market, buyers are induced to submit their demands for the data to the auctioneer. The auctioneer should always make an optimal decision to maximize the utilities of the seller and buyers. It is worth noting that the payment is the key determinant of the buyers' utility (i.e., the gain of using the purchased data), which is highly related to the other buyers' bids. In addition, we consider the auction process as a sealed-bid market, in which the buyers are not capable of knowing the bidding information of

other buyers, and the auction process needs to be fair to all participants. Nonetheless, some "strategic" buyers would like to improve their utilities by colluding with eacho ther to manipulate the auction results,



System Model

which is unfair to the non-colluding buyers data will not be further resold to other parties.. **Adversary Model.** In the data trading market, buyers are induced to submit their demands for the data to the auctioneer. The auctioneer should always make an optimal decision to maximize the utilities of the seller and buyers. It is worth noting that the payment is the key determinant of the buyers' utility (i.e., the gain of using the purchased data), which is highly related to the other buyers' bids. In addition, we consider the auction

process as a sealed-bid market, in which the buyers are not capable of knowing the bidding information of other buyers, and the auction process needs to be fair to all participants. Nonetheless, some "strategic" buyers would like to improve their utilities by colluding with eacho ther to manipulate the auction results, which is unfair to the non-colluding buyers.

In our data trading market, the buyers are allowed to submit bids through the cloud-based platform, so it is difficult for them to collude with each other to obtain a larger utility. Nonetheless, it is easy for a "smart" buyer to create a new identity in the auction market. By doing this, the buyers could launch a collusion attack, but in this case are colluding with themselves via the new identities. As a result, efficiency and fairness of the auction can still be manipulated by such behavior, which is called multi-identity bidding or buyer false- name bidding . Unfortunately, commonly used auction schemes (e.g., VCG scheme, English auction scheme) are not able to resist this type of attack. To this end, we introduce a new auction scheme that is capable of defending against the false-name bidding attack, as well as satisfying several economic properties, including incentive compatibility and computational efficiency .

Malicious users: Users may try to deviate from the responsibilities described above. Namely, they may e.g., disrupt the brokers' data trading service, deny cleared transactions (i.e., paid and sold) and resell previously purchased datasets. A user is defined as a dishonest user if he avoided any of the trading related responsibilities, and such behaviour (either selling or buying) is denoted as misbehaviour. Note that, when illegally selling

previously purchased datasets, attackers may try to perturb the dataset to bypass copy detection mechanisms. Trusted brokers: We assume the brokers can be trusted, e.g., the role is played by the organizations that are strictly supervised with great transparency or commercial companies with high reputation. Similar assumptions can be found and the assumption that the brokers will be strictly supervised is also consistent with the FTC's recent action. Channel assumption: We assume both buyers and sellers interact with the broker via secure communication channels. The communication is encrypted and decrypted with pre distributed keys to guarantee that the dataset is not open to the public. This also implies authentication is in place since the broker needs to use the correct entity's key for communication.

3. SPECIFICATION AND ACCOUNTRADE

A. Upload for Sale

When a seller A wants to upload a dataset to sell it, she follows the Upload protocol and posts her declaration $post_t$ at the bulletin board at time t. Then, she sends the upload request along with H(d) to the broker. The broker finds the corresponding post from the bulletin board and blames A if none is found, because it is evident that she has tried to avoid being book-kept. If the broker sees the post, he accepts A's request and retrieves the dataset. Then, the broker checks whether the hash of received dataset is identical to the one posted at the bulletin board and blames A if not. Finally, the broker generates and publishes the description of the dataset d (e.g., its contents, price, H(d)).

B. Dataset Examination

If the upload is successful, the broker checks whether a similar dataset has been uploaded before. To do so, we propose uniqueness index, which is indicative of the amount of overlaps between a given set S and a set of sets $S = \{S_1, S_2, \ldots\ldots, S_n.\}$

Definition 3 (Uniqueness index). Given a set $S = \{S_1, S_2, \ldots\ldots, S_n\}$ of the uniqueness index of S_x over the set S is defined as $U_S(S_x) = 1 \max_{S2S} f(S; S_x)g$, where $(S; S_x)$ is a normalized similarity function describing how unique S_x is when compared to S, defined as:

$$\Delta(S, S_x) = J(S, S_x).\frac{\max(|S|, |S_x|)}{\min(|S|, |S_x|)}$$

$J(S_1; S_2)$ denotes Jaccard Index, which is statistical mea-surement of the similarity of two given sets, defined as $J(S1; S2) = \frac{S_1 \cap S_2}{S_1 \cup S_2}$ Then, we define selling of a dataset d asre-selling if $U_D(d) > \theta_{high}$ and as valid selling if $U_D(d) < \theta_{low}$, where D is the database of datasets the broker possesses, d is the dataset to be examined, and $_{high}$; $_{low}$ refer to two threshold values for decision making. If the uniqueness index is between the two threshold

values, the broker can manually inspect the dataset with human labor. The reason we define and use this uniqueness index in dataset re-selling detection is manifold. Firstly, it intuitively measures how many elements of S_x are similar to the ele-ments in the entire set S, and the multiplier after the Jaccard Index guarantees the index is equal to 1 when S_x is a subset/superset of any set in S. Secondly, in many existing similarity comparison approaches in information retrieval, the datasets are considered as sets of elements (k-grams for texts, feature descriptors for images, and key frames for videos), and therefore the proposed uniqueness index is consistent. Thirdly, there is no known similarity comparison mechanism for table-type datasets, and similarity comparison of JSON-like datasets are hardly scalable.

C. Download after Purchase

When a buyer B want to get access to certain dataset d (after reading the description provided by the broker), first he pay for it to the broker and then follows the Download protocol . he posts a declaration $post_t$ first at the bulletin board at time t, and then he initiates the download request by sending H(d) to the broker, where H(d) is available in the dataset provided by the broker. The broker finds the corresponding post from the bulletin board and blames B if none is found, because it is evident that he has tried to avoid being book-kept. If the broker sees the post, he accepts B's download request and sends the dataset to B.

D. Uniqueness index calculation

The flow is sketched below to calculate the uniqueness of the document. For a given dataset d, the user submitted data is first convert it to a membership vector. Then, we calculate the MinHash values of the membership vector, which will be used to estimate the uniqueness index .

E. MFPA: A Multi-Round False-Name Proof Auction Scheme

In this section, we introduce the MFPA scheme in detail, which is designed for the data trading market, but can be generalized to other types of markets as well. Particularly, we first present the workflow of the proposed auction scheme. We then conduct theoretical analysis and prove that our auction scheme satisfies several desired properties (e.g., incentive com- patibility, false-name bidding proofness, and computational efficiency). Finally, we show an illustrative example for better a understanding of our scheme.

Workflow

Recall that in our auction market, the data provider acts as the seller, and announces the total data supply capacity as well as the reserve price for selling each GB amount of the data. The auctioneer in the cloud determines the winning condition. The data collectors act as buyers, and submit their bidding information.

In proposed system there will be a multiple round of action in each round the data will be traded by a bundlesize which is set by the auctineer .he buyer is asked to submit the valuation for each bundle for each round. Once the valuation is processed the auction process take place round by round until no data left. At last the auctineer announce the result and the data will be transformed from the data provider to the data collectors.

F. Accountability Properties of AccountTrade

Upload

J1: where A is the one who sent the upload request. A is the owner

J2: If the posted hash H in $post_t$ is different from the calculated hash H^0, the broker states dis(A).

Examine

J3: If the calculated uniqueness index is very low, the dataset is derived from already-uploaded ones, the broker states dis(A) where A is the one who uploaded the dataset.

Download

J4: Same as J1 except that dis(B) is stated instead, where B is the one who sent the request. J1 detects a dishonest seller who tries to refuse a sale transaction, and J2 further prevents a dishonest seller from declaring a wrong dataset. J3 detects reselling, and J4 detects a dishonest buyer who tries to refuse a purchase transaction.

.J5: if the buyer want to get the data he has to send bids along with the request .

J6:The auctioneer will examine and grand access.

4. CONCLUSION

In this paper, The data trading problem in the big data market is addressed. Especially, to enable optimal data trading and defend against false-name bidding attacks, we proposed a novel Multi-round False-name Proof Auction (MFPA) scheme. To ensure false-name binding, The MFPA scheme runs in multiple rounds, while the data from owners are sold in bundles during each round. also Account Trade which assurance correct book-keeping and achieves accountability in the big data trading among dishonest consumers. In data transaction AccountTrade blames dishonest consumers if they deviate from their responsibilities. To achieve accountability against dishonest sellers who may resell others' datasets to find the uniqueness of document – uniqueness index – which is efficiently computable. We formally defined two accountability models. we also evaluated the performance and QoS using real-world datasets in our implemented test bed.

REFERENCES

[1] Data markets compared – a look at data market offerings from four providers. Goo.gl/k3qZsj.

[2] Ftc charges data broker with facilitating the theft of millions of dollars from consumers' accounts. Goo.gl/7ygm7Q.

[3] Ftc charges data brokers with helping scammer take more than $7 million from consumers' accounts.

[4] Goo.gl/kZMmXn.

[5] Ftc complaint offers lessons for data broker industry. goo.gl/csBYA3.

[6] Multimedia computing and computer vision lab. goo.gl/pbKeCj.

[7] R. Ara´ujo, S. Foulle, and J. Traor´e. A practical and secure coercionresistant scheme for remote elections. In Dagstuhl Seminar Proceedings.Schloss Dagstuhl-Leibniz-Zentrum f¨ur Informatik, 2008.

[8] D. Baltieri, R. Vezzani, and R. Cucchiara. Sarc3d: a new 3d body model for people tracking and re-identification. In ICIAP, pages 197– 206. Springer, 2011.

[9] B. Blanchet. Automatic verification of security protocols in the symbolic model: The verifier proverif. In FOSAD, pages 54–87. Springer, 2014.

[10] B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. Communications of the ACM, 13(7):422–426, 1970.

[11] S. Brin, J. Davis, and H. Garcia-Molina. Copy detection mechanisms for digital documents. In SIGMOD, volume 24, pages 398–409. ACM, 1995.

[12] Q. Yang, D. An, W. Yu, X. Yang, and X. Fu, "On stochastic optimal bidding strategy for microgrids," in 2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC), Dec 2015, pp. 1–8.
[13] [C. Yi, A. S. Alfa, and J. Cai, "An incentive-compatible mechanism for transmission scheduling of delay-sensitive medical packets in e-health networks," IEEE Transactions on Mobile Computing, vol. 15, no. 10, pp. 2424–2436, Oct 2016
[14] Q. Wang, B. Ye, B. Tang, T. Xu, S. Guo, S. Lu, and W. Zhuang, "Robust large-scale spectrum auctions against false-name bids," IEEE Transactions on Mobile Computing, vol. 16, no. 6, pp. 1730–1743, June 2017.
[15] M. Yokoo, Generalized Vickrey Auction. Springer US, 2008.
[16] M. Yokoo, Y. Sakurai, and S. Matsubara, "Robust double auction protocol against false-name bids," in Proceedings 21st International Conference on Distributed Computing Systems, Apr 2001, pp. 137–145.
[17] W. Wang and A. B. Whinston, Binary Vickrey auction - A robust and efficient multi-unit sealed-bid online auction protocol against buyer multi-identity bidding. Elsevier Science Publishers B. V., 2007.
[18] Z. Hidvgia, "Binary vickrey auction a robust and efficient multi-unit sealed-bid online auction protocol against buyer multi-identity bidding," Decision Support Systems, vol. 43, no. 2, pp. 301–312, 2007
[19] Z. Hidvgia, "Binary vickrey auction a robust and efficient multi-unit sealed-bid online auction protocol against buyer multi-identity bidding," Decision Support Systems, vol. 43, no. 2, pp. 301–312, 2007