IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust

Incentive Compatible Distributed Data Mining

Murat Kantarcioglu and Robert Nix
Jonsson School of Engineering and Computer Science
The University of Texas at Dallas
800 West Campbell Road
Richardson, Texas, USA
Email: {muratk,rcn062000}@utdallas.edu

***Abstract*—In this paper, we propose a game-theoretic mechanism to encourage truthful data sharing for distributed data mining. Our proposed mechanism uses the classic Vickrey-Clarke-Groves (VCG) mechanism and does not rely on the ability to verify the data of the parties participating in the distributed data mining protocol. Instead, we incentivize truth telling based solely on the data mining result. Under reasonable assumptions, we prove that these mechanisms are incentive compatible for distributed data mining. In addition, through extensive experimentation, we show that they are applicable in practice.**

***Index Terms*—game theory, data mining, cryptography, mechanism design**

I. Introduction

Today, we live in an information age. Information has become a power currency in our society. As such, people treat information with care and secrecy. There are times, however, that information needs to be shared among owners for the betterment of society, or simply for their own profit. Data mining seeks to take information and aggregate it into models that are more useful than the original information. Since people are cautious and do not wish to give up their private information the need for *privacy-preserving* data mining has arisen. In addition to the simple desire for privacy, certain government regulations, such as the Health Insurance Portability and Accountability Act (HIPAA)[3] require that certain data be kept private.

Techniques for privacy preserving data mining are many in number. They include anonymization of data, noise addition techniques, and cryptographic techniques, in addition to countless others. The cryptographic techniques have the distinction of being able to compute models based on unperturbed data, since the cryptography ensures that the data will not be revealed. However, they make no guarantees that participants will not use false data anyway.

Consider the following scenario: Suppose that the different intelligence agencies around the world wish to share their information on terrorist networks, in order to increase global knowledge about terrorists and terrorist organizations. This, of course, is a noble goal, and would benefit mankind as a whole. Intelligence agencies, however, wish to receive credit for capturing terrorists, and to this end, may provide false information in hopes of having the best information to themselves. However, several agencies could have this plan. Even if the agencies compute the overall terrorist information model securely and privately, this would not change the fact that the end result would not be an accurate model based on real data. Because of this, the intelligence agencies get no closer to finding terrorists, and the terrorists could go free.

Granted, this is a rather extreme example, but it illustrates the failure of traditional cryptographic secure multi-party computation to ensure that players use truthful data. The discipline of cryptography can be used to create provably secure protocols which guarantee the privacy of the data of all parties in data mining. What then does this say about the correctness of the result of the calculation? It is true that in many situations, it can be proved that the calculation will be correct with respect to the data supplied by the players for the calculation. This is usually based on commitments that must be made by each player, ensuring that no player can change their input at any time during the calculation. However, this does not ensure that the player would provide true data for the calculation! In particular, if the data mining function is *reversible*, that is, given two inputs, x and x', and the result $f(x)$, it is simple to calculate $f(x')$, then a player might wish to provide false data in order to exclusively learn the correct data mining result![21]

In order to combat this problem, scholars have attempted to mesh game theory with cryptography to deal with the actions of players who act in their own self interest. Given that one can verify, after the fact, albeit with some cost, that a player used their true data, it is quite simple to ensure that players use true data. We simply audit the process with a high enough frequency, and stiff enough penalty, that players will think twice about lying about their data. The classic IRS game[20] is a typical example of this: a taxpayer can be motivated to be truthful on his return by both the magnitude of the penalty for cheating, and the frequency of audits. The higher the penalty, the less frequent audits need to be. However, in most cases, the ability to audit the data defeats the purpose of privacy-preserving data mining, in that it requires a trusted auditor to be able to access each player's data. The main question we address in this paper is: What guarantees can we make about the truthfulness of players' data when we have no way of verifying the data used by a given player?

We tackle this problem by using a monetary mechanism to encourage players to be truthful about their data without being able to verify the truthfulness of the data that players provide. It is important to be able to do this without verifying

978-0-7695-4211-9/10 $26.00 © 2010 IEEE
DOI 10.1109/SocialCom.2010.114

IEEE computer society

data, because the verification of the data could violate privacy! To illustrate the effectiveness of an after-the-fact mechanism, consider the following scenario: Several passengers are flying on a chartered cross-country flight, and the flight passes on fuel costs to the passengers. In order to board, the charter airline requires all passengers to report their weight, so that the airline can calculate the necessary fuel to get to their destination. In this case, passengers have the incentive to tell the turth about their weights, since if they under-report, the plane could crash from lack of fuel, and no amount of money (or embarrassment) saved is worth their lives. In addition, if they over-report, they are simply increasing their cost. Therefore, there is no reason to verify each passenger's weight by means of a scale, since each passenger will give their correct weight (unless, of course, they do not *know* their weight).

In a similar vein, our data mining mechanism does not require the verification of the data, it simply encourages truthfulness through extrinsic incentives. Namely, it provides monetary incentives which subsidize the calculation, and these, in turn, motivate truthful behavior. We invoke a Vickrey-Clarke-Groves (VCG) mechanism based on the accuracy of the result itself in order to encourage correct data reporting. We show that, to the risk-averse player, the mechanism will encourage true data sharing, and for the risk-neutral player, the mechanism gives a close approximation that encourages minimal deviation from the correct data.

Our contributions can be summarized as follows:

- We develop a mechanism to encourage truthful data sharing which does not require the ability to audit or verify the data.
- We prove that this mechanism is incentive compatible under reasonable assumptions.
- We provide extensive experimental data which shows the viability of the mechanism in practice.

In the next section, section 2, we survey the previous work that has been done related to this. In section 3, we provide some background in game theory and mechanism design. In section 4, we describe the game theoretic model we use to represent the data mining process, the *assured information sharing game*. In section 5, we outline our mechanism, and prove its incentive compatibility. In section 6, we show experimental data on different kinds of data mining problems, indicating the practical use of this mechanism. Finally, in section 7, we give our conclusions and outline future research directions.

II. Related Work

Cryptography and game theory have a great deal in common, in terms of the goals they try to achieve. The problems tackled by cryptography generally seek to assure that participants in certain activites are forbidden to deviate (profitably) from the prescribed protocol by rendering such actions detectable, impossible, or computationally infeasible. Similarly, mechanism design seeks to forbid deviations, but it does so by rendering the deviations unprofitable. It is understandable, therefore, that a fair amount of work has been done to use the techniques of one to solve the problems of the other. Most of this work is not directly related to ours, since a fair amount of the game theoretic security work deals with specific functions, and the individual steps of the computations of those functions.

Shoham and Tennenholtz [21], define the class of *NCC*, or *non-cooperatively computable* functions, and define specifically the boolean functions which are NCC. In addition, the paper defined two additional classes, p-NCC and s-NCC, which stand for probabilistic-NCC and subsidized-NCC, respectively. p-NCC are the functions which are computable with some probability non-cooperatively, and s-NCC are the functions which are computable when external monetary motivation is allowed. This was expanded to consider different motivations [16], and coalitions [4]. While our work does involve making functions computable in a competitive setting, it involves more complicated functions, and specifies a mechanism to ensure computability.

In addition to this, much work seeks to include a game-theoretic model in standard secure multi-party computation. Instead of considering players which are honest, semi-honest, or malicious, these works simply consider players to be rational, in the game theoretic sense. Much of this work concentrates on the problem of *secret sharing*, that is, dividing a secret number among players such that any quorum (sufficiently large subset) of them can reconstruct the secret number. This was first studied by Halpern and Teague [8], and later re-examined by Gordon and Katz, [6]. Other protocols for this problem were outlined in [1] and [15]. The paper by Ong, et al.[18], hybridizes the two areas, within the realm of secret sharing, by considering some players honest and a majority of players rational. Other work seeks a broader realm of computation, such as [9], and [13], which build their computation model on a secret sharing model. There is other work that attempts to combine game theoretic and cryptographic methodologies, many of which are surveyed in [12]. Many of these rational secure computation systems could be used to ensure privacy in our mechanism. However, like other secure computation systems, they make no guarantees about the truthfulness of the inputs.

More closely related to the work in this paper, several works have attempted to enforce honest behavior among the participants in a data sharing protocol. This paper builds on the work of Agrawal and Terzi[2], who present a model which improves enforces honesty in data sharing through use of auditing mechanisms. Layfield, et al., in [14], present strategies which enforce honesty in a distributed computation, without relying on a mediator. Jiang, et al., in [10] integrate the auditing mechanism with secure computation, to convert existing protocols into rationally secure protocols. Finally, the work of Kargupta, et al.[11], analyzes each step of a multi-party computation process in terms of game theory, with the focus of preventing cheating withing the process, and removing coalitions from gameplay. Each of these deals with the problem of ensuring truthfulness in data mining. However, each one requires the ability to verify the data after the calculation. Our mechanism has no such requirement.

III. Game Theoretic Background

Game theory is the study of competitive behavior among multiple parties. A game contains four basic elements: *players*, *actions*, *payoffs*, and *information* [20]. Players have actions which they can perform at designated times in the game, and as a result of the actions in the game, players receive payoffs. The players have different pieces of information, on which the payoffs may depend, and it is the responsibility of the player to use a profitable *strategy* to increase his or her payout. A player who acts in such a way as to maximize his or her payout is termed *rational*. Games take many forms, and vary in the four attributes mentioned above, but all games deal with them. The specific game we describe in this paper is a finite player, single round, simultaneous action, incomplete information game, with payouts based on the final result of players' actions.

Before proceeding with a discussion of mechanism design, it is convenient to define a common notation used within the literature and within this paper.

Given a vector, $X = (x_1, x_2, ..., x_n)$, we define:

$$X_{-i} = (x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n)$$

Or, intuitively, X_{-i} is the vector X without the ith element.

Mechanism design is a sub-field of game theory, and deals with the construction of games for the purpose of achieving some goal, when players act rationally. A *mechanism* is defined, for our purposes[1], as:

Definition 1: Given a set of n players, and a set of outcomes, A, let V_i be the set of possible valuation functions of the form $v_i(a)$ which player i could have for an outcome $a \in A$. We then define a *mechanism* as a function $f : V_1 \times V_2 \times ... \times V_n \rightarrow A$, which given the valuations claimed by the players, selects an outcome, and n payment functions, $p_1, p_2, ..., p_n$, where $p_i : V_1 \times V_2 \times ... \times V_n \rightarrow \Re$, that is, given the valuations claimed by the players, selects an amount for player i to pay [17].

Thus, the overall payout to a player in this mechanism is his valuation on the outcome, $v_i(a)$, minus the amount he is required to pay, $p_i(v_i, v_{-i})$. A mechanism is said to be *incentive compatible* if rational players would prefer to give the true valuation rather than any false valuation. Or, more formally:

Definition 2: If, for every player i, every $v_1 \in V_1$, $v_2 \in V_2$, ..., $v_n \in V_n$, and every $v_i' \in V_i$, where $a = f(v_i, v_{-i})$ and $a' = f(v_i', v_{-i})$, then $v_i(a) - p_i(v_i, v_{-i}) \geq v_i(a') - p_i(v_i', v_{-i})$, then the mechanism in question is *incentive compatible* [17].

Thus, a player would prefer to reveal his true valuation rather than any other valuation, assuming all other players are truthful.

Another important term is *individual rationality*, which is intuitively whether a player would desire to participate in a game in the first place. The utility a player receives in the event that they choose not to participate is called the *reservation utility*. In order for a strategy to be considered an equilibrium, for all players, it must be individually rational and incentive compatible.

The specfic mechanism used in our data mining is the Vickrey-Clarke-Groves (VCG) mechanism. The VCG mechanism, in general, seeks to maximize the social welfare of all participants in a game. The social welfare can be defined as the sum of the valuations of all players. Thus, VCG wishes to cause rational players to act in such a way that the sum of the valuations each player has of the outcome is maximized. In mathematical notation, this is where the outcome chosen is $\text{argmax}_{a \in A} \sum_i v_i(a)$, where A is the set of possible actions, and v_i is the valuation function for player i. The VCG mechanism is defined as follows:

Definition 3: A mechanism, consisting of payment functions $p_1, p_2, ..., p_n$ and a function f, for a game with outcome set A, is a Vickrey-Clarke-Groves mechanism if

$$f(v_1, v_2, ..., v_n) = \text{argmax}_{a \in A} \sum v_i(a)$$

(f maximizes the social welfare) and for some functions $h_1, h_2, ..., h_n$, where $h_i : V_{-i} \rightarrow \Re$ (h_i does not depend on v_i), for all $(v_1, v_2, ..., v_n) \in V, p_i(v_1, v_2, ..., v_n) = h(v_{-i}) - \sum_{j \neq i} v_j(f(v_1, v_2, ..., v_n))$ [17].

Since p_i is the amount paid by player i, this ensures that each player is paid an amount equal to the valuation of all the other players. This means that each player would have incentive to make actions to maximize the social welfare. The formal proof that the VCG mechanism is incentive compatible can be found in [17].

IV. Our Model: The Assured Information Sharing Game

In order to analyze data mining tasks in terms of game theory, we now describe a game scenario outlining the process for some data mining task. This model is a simple model in which a mediator does the data mining calculations. This may not be necessary, but for now, we use this to simplify our calculations. In terms of doing the calculation, the mediator can be removed using the cryptographically secure techniques outlined in [13] or [9], however, it may or may not be possible to remove the mediator for payments. We examine this further in section 7. We also consider only individual actions, rather than coalitions, for simplicity.

Definition 4: Mediated Information Sharing Game

Players: $P_1, P_2, ..., P_n$, and a mediator P_t.

Preconditions: Each player $P_i \in \{P_1, ..., P_n\}$ has x_i, a piece of data which is to be shared for the purposes of computing some function of the data. P_t is another party who is bound to compute a data mining model from the players' data in a secure fashion. P_t is also in possession of a small independent test data set. It is reasonable that P_t could have such a set through observation of a small amount of public data, though this amount of data may not be enough to build an accurate model.

Game Progression:

[1]Technically, this is only a *direct revelation* mechanism, but we will have no need to generalize this.

1. Each player $P_i \in \{P_1, ..., P_n\}$, selects x'_i, which may or may not equal x_i, or chooses not to participate. These inputs are committed. Define X to be the vector of original values x_i, and X' to be the vector of chosen values x'_i.

2. Players send X' to P_t for secure computation of the data mining function. The function which builds the model will be referred to as D.

3. All players receive the function result, $m = D(X')$.

Payoffs: For each $P_1 ... P_n$, define the utility of a participating player $u_i(x_i, D(X')) = \max\{v_i(m) - v_i(D(x_i)), 0\} - p_i(X', m) - c(D)$. $v_i(m)$ is the intrinsic utility of the function result, which we approximate as the accuracy of a data mining function. Thus, $v_i(m) = acc(m)$ where acc is some accuracy metric applied to the data mining model. This will, of course, vary based on the truthfulness of each player. We normalize each player's reservation utility, that is, the utility received if the player chooses not to participate, to zero. This can be done without loss of generality by subtracting the reservation utility (which is $v_i(D(x_i))$, based on the accuracy of the model based only on one's own data), from the valuations in the mechanism. Note that a player will always recieve at least this much utility, so we obtain the expression $\max\{v_i(m) - v_i(D(x_i)), 0\}$. $p_i(X', m)$ is the amount paid by P_i, based on the inputs and the results. Note that if p_i were to be negative, it would mean that P_i receives money instead. $c(D)$ is the computational cost of computing D. Since D is securely computed, there will be some cryptography involved in the computation of the model, hence computational cost should be considered.

V. Our Solution

To motivate players to truthfully reveal the information, we propose the following:

1. In addition to computing the data mining model, P_t also computes $D(X'_{-i})$ for each P_i, that is, the data mining function without using the data provided by player i.

2. For each P_i, we let $p_i(X', m) = \sum_{j \neq i} v_j(D(X'_{-i})) - \sum_{j \neq i} v_j(m) - c(D)$, where v_i is determined by measuring the accuracy of the data mining model on the independent test set which P_t has. This pays each player an amount equal to the difference in accuracy between the overall data mining model and the data mining model without his input, essentially rewarding each player based on their own contribution to the model. We include the $-c(D)$ term in order to balance out the cost of the calculation. Figure 1 shows the process used to calculate the payment for a given player i.

Theorem 5.1 The above mechanism motivates players to truthfully reveal their inputs, under the following assumption:

Assumption: For each player i, the probability of an increase in the classifier's accuracy decreases significantly with the distance between the player's actual data and the data the player provides to the classifier building process. More formally, we may state that the expected value of the classifier's accuracy does not increase with said distance. Mathematically, for $X = x_i \cup X_{-i}$ and $X' = x'_i \cup X_{-i}$, this can be written as

$$E[acc(D(X))] \geq E[acc(D(x'))] + f(dist(X, X'))$$

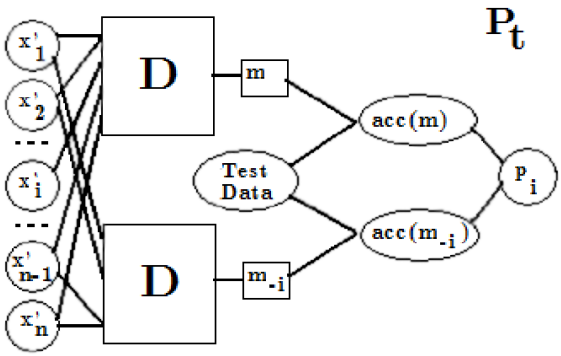


Fig. 1. Payment calculation for player i

where f is a non-negative, increasing function for all i, x_i, x'_i and X_{-i}.

This is essentially the implicit assumption used by any data miner: that deviating from the true data makes a bad classifier more likely. We feel that this assumption is, while not always true, always reasonable. Raw data mining processes, in practice, use true data unless they are trying to combat the problem known as "overfitting". Overfitting is when the data model is too well tuned to training data, and this causes accuracy on practical data to fall. In such instances, outliers are removed, or irrelevant dimensions are reduced away, but the data otherwise remains true. Usually, if the data is to be doctored in any way, it would be done before the data mining process would even take place. Another way to think of this assumption might be to say that we assume all players' data is relevant to the data mining task.

Proof (Incentive Compatibility): We proceed in a similar fashion to the proof of VCG incentive compatibility. For any given i, x_i, X_{-i}, and x'_i, we must show that

$$E[u_i(X = x_i \cap X_{-i})] \geq E[u_i(X' = x'_i \cap X_{-i})].$$

The utility of i for X is given by

$$\begin{aligned} u_i(x_i, D(X)) \quad &= \quad \max\{v_i(D(x)) - v_i(D(x_i)), 0\} \\ &\quad -p_i(X, D(X)) - c(D) \end{aligned}$$

where

$$p_i(X, D(X)) = \sum_{j \neq i} v_j(D(X_{-i})) - \sum_{j \neq i} v_j(D(X)) - c(D).$$

Likewise,

$$\begin{aligned} u_i(x'_i, D(X')) \quad &= \quad \max\{v_i(D(X')) - v_i(D(x_i)), 0\} \\ &\quad -p_i(X', D(X')) - c(D) \end{aligned}$$

where

$$p_i(X', D(X')) = \sum_{j \neq i} v_j(D(X_{-i})) - \sum_{j \neq i} v_j(D(X')) - c(D).$$

Over expectation, in order for incentive compatibility to exist, this requires that

$$E[\max\{v_i(D(X)) - v_i(D(x_i), 0\}] + E[\sum_{j \neq i} v_j(D(X))] \geq E[\max\{v_i(D(X')) - v_i(D(x_i)), 0\}] + E[\sum_{j \neq i} v_j(D(X'))].$$

By our assumption that the expected value of $v_k(D(X'))$ (for all k decreases as X' differs from X, we know that $E[\sum_{j \neq i} v_j(D(X))] \geq E[\sum_{j \neq i} v_j(D(X'))]$. We also know that $E[\max\{v_i(D(X)) - v_i(D(x_i), 0\}] \geq E[\max\{v_i(D(X')) - v_i(D(x_i)), 0\}]$, since either the last expression is zero, in which case the first expression is greater than or equal to zero, the last expression is greater than zero, in which case the first expression is greater than or equal to the last expression by our assumption. Therefore, the mechanism is incentive compatible.

Proof (Individual Rationality): To show that the mechanism is individually rational, we need only show that the mechanism has a utility of at least zero (since we have normalized the reservation utility to zero). Note, once again, that the utility of player i is given by

$$u_i(x_i, D(X)) = \max\{v_i(D(x)) - v_i(D(x_i)), 0\} - p_i(X, D(X)) - c(D)$$

Since $\max\{v_i(D(x)) - v_i(D(x_i)), 0\}$ is at least zero, and $-c(D)$ is offset by the term in $p_i(X, D(X))$, we need only show that $E[\sum_{j \neq i} v_j(D(X_{-i})) - \sum_{j \neq i} v_j(D(X))] \leq 0$. Note that, X_{-i} has a nonzero distance from X. Therefore, by our assumption, $E[v_j(D(X_{-i}))] \leq E[v_j(D(X))]$ for all j. Because of this, $E[\sum_{j \neq i} v_j(D(X_{-i})) - \sum_{j \neq i} v_j(D(X))] \leq 0$, and the mechanism is individually rational.

VI. Experiments

Having proven that the mechanism is incentive compatible under reasonable assumptions, we now set out to show how the mechanism performs in practice. As previously mentioned, the assumption that the best model is given by the true data is not always correct. This can happen when the data is stacked in particular ways, or due to the simple overfitting phenomenon. However, most of data mining relies on this assumption when aggregating results. We therefore ran a series of experiments on real data to show the mechanism's practical viability.

A. Methodology

We tested the mechanism on the following three different data mining models: naive Bayes classification, ID3 decision tree classification, and support vector machine (SVM) classification. For the decision tree and SVM, we used the Weka data mining library[7]. We used three different data sets from the UC Irvine Machine Learning Repository[5].

Adult(census-income) This data set contains census information from the United States, each row corresponding to a person. Each row of the data set is one of two classes, indicating the gross income of each person. A positive class (50000+) indicates that the person has a gross income greater than $50,000, while a negative value (-50000) indicates an income of $50,000 or lower. For our purposes, we included only 20,000 randomly selected rows of this data set, 18,000 for training, and 2,000 for the independent test set. In addition, certain fields were omitted due to their continuous nature, and others (such as age) were generalized to more discrete values to prevent overfitting.

German-credit This data set contains credit applications in Germany, and classifies people as either a good credit risk (+) or a bad credit risk (-). There were two continuous attributes (duration and amount) which we generalized to avoid overfitting.

Car-evaluation This data set takes the characteristics of cars and classifies them as unacceptable, acceptable, good, or very good. Since we wished to deal only with binary classification problems for the purposes of this experiment, we generalized the class into simply unacceptable (unacc) or acceptable (acc) with those vehicles which were originally evaluated as good or very good being listed as acceptable. No attributes were adjusted.

We chose to use real data, rather than fabricated data, because the mechanisms in question deal with the actions of real people. The incentive to lie for an individual row of the data set is not in play here. We are looking at the incentive for the owner of several pieces of data to lie about his input to a classification process. It is the potential for knowledge discovery, and the exclusive discovery thereof, which would drive someone to lie about the data they have.

In each case, 10% of the data was set aside as an independent test set (to be used by the mediator).

Each training data set was partitioned vertically into three pieces, each piece having as close to the same number of fields as possible. Each of these pieces was designated as belonging to a player. Thus, all the experiments involve three parties, for simplicity.

For each data set and data mining method, we first ran 50 trials to determine the overall accuracy using the truthful data, and the estimated payouts to each player in this case. In order to combat overfitting, each trial consisted of the classification of 20 separate bootstrap samples of the test data (that is, a sample with replacement). The size of these samples was 25% of the test set size.

After this, for each player, we varied the truthfulness of that player's data. Any choice of x'_i is either honest or dishonest. However, the dishonest choices may have varying degrees of dishonesty, with some applying merely a small perturbation to the input, and some blatantly dishonest about every data row. We classify moves by the amount of dishonesty in them. Let $x'_i[k]$ refer to an input for which k times the total number of rows in the input are falsified, that is, k is the fraction of falsified rows in the data. Thus, $x'_i[.01]$ would be an input for which a mere 1% of the data would be falsified. $x'_i[1]$, on the other hand, would essentially be a random set drawn from the domain.

In order to test the results of the falsification (or, equivalently, the perturbation) of the data, we tested the model with several different perturbation values. For each player i, we used $x'_i[.01]$, $x'_i[.02]$, $x'_i[.04]$, $x'_i[.08]$, $x'_i[.16]$, $x'_i[.32]$, $x'_i[.64]$, and $x'_i[1]$. Note that only one player's data was perturbed at any given time. This was because we wished to determine what

a player's unilateral deviation would do when other players were truthful.

To calculate the expected payout for player i, we would subtract the overall accuracy for the model without the data belonging to player i from the overall accuracy of the full model.

B. Results

Figures 2 through 4 show the overall accuracy and estimated payouts to each player for each model, data set, and perturbation. For the estimated payouts, each line shows the payout to the player that is lying, for each perturbation value.

In the vast majority of cases, deviation from the truth produces a lower payout, on the average. Some cases produce a small payout increase on the average, however. Smaller deviations have a higher probability of increasing payout than larger deviations. In practice, a small (1-4%) deviation from the truth has the effect of reducing the impact of overfitting, and can result in a slightly more accurate classifier. However, rarely is the amount gained significant.

It is worth mentioning that in several cases, the calculation would not qualify as individually rational without further subsidy. For example, the Adult data set, under naive Bayes classification and ID3 decision tree classification, produces payouts for each player which are negative. This means that the addition of a third player's data decreases the accuracy of the classifier. This is likely due to the presence of many fields in the data. While each player's fields perform well, combining the fields results in a slight reduction in accuracy due to redundant or irrelevant fields.

There are a few exceptions to the generalization about small deviations and small payout increases, such as the volatile looking graph for the estimated payouts for the SVM data mining on the Adult data set, as the graph moves up and down very quickly, and does appear to increase sharply in a few places. However, the scale of this graph shows that this fluctuation is actually very small. The difference between any two points on this graph is no more than 0.6% in terms of the overall accuracy of the classifier.

While a risk-neutral player might attempt to perturb the data slightly to gain a slight average profit, a risk-averse player would certainly never perturb the data. In all cases, at least one bootstrap sample produced a lower classifier accuracy for any perturbed data. Therefore, if the player is risk-averse, then the player would provide true data, since otherwise there would be a risk of losing profit.

VII. Conclusions

We have shown that, under a reasonable assumption, our mechanism which rewards players based on their contribution to the model is incentive compatible. We then determined the usefulness of the mechanism in practice by running our mechanism using real data. This shows that, while the assumption used in the incentive compatibility proof is not always strictly true, the mechanism yields proper motivation for the vast majority of cases.

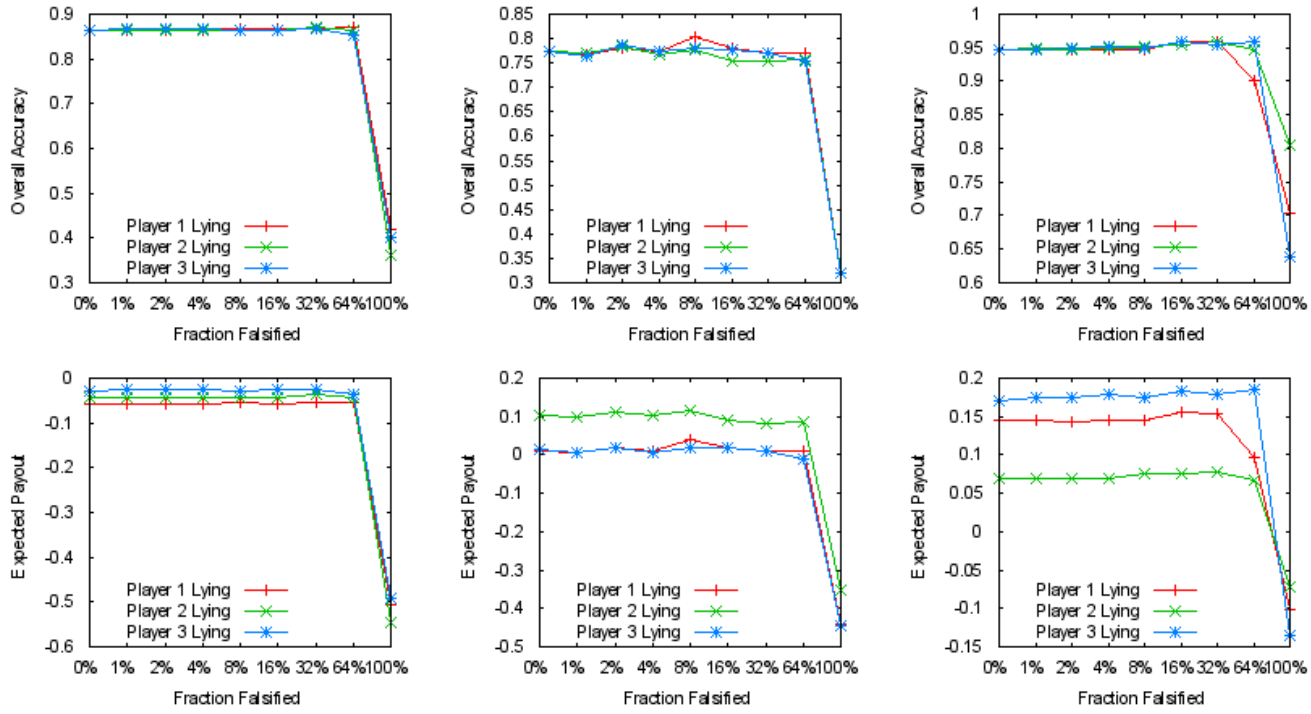
While our primary goal has been to ensure that players truthfully reveal their data, one could also take a different approach to the problem. If a deviation from the truth affords a player a payout advantage, then this means that the deviation has necessarily increased the overall accuracy of the final classifier. So, in the cases where it is advantageous to lie, we have created a better classifier than the truthful data would provide! Thus, while the mechanism does not *guarantee* truthfulness *every* time, in the cases where it does not, it results in a better classifier. If the goal of the process is then changed to the creation of the best model, rather than ensuring truth, the mechanism works even better.

A. Future Work

There are several other questions which can be asked about this process. First of all, is the mediator necessary? In traditional secure multi-party computation scenarios, it has been shown that a mediator is not necessary in order to ensure privacy. However, it is less intuitive to believe that the mediator, who subsidizes the computation, is not necessary to ensure honesty. The work by Parkes and Shneidman [19], however, may shed some light on this possibility. The work outlines methods for implementing VCG mechanisms in distributed environments, which looks promising for the possibility of removing the central subsidizing entity. Naturally, the details would need to be worked out, however, it is entirely possible that some method, be it this or another, can effectively remove the need for the mediator in both the computation and subsidization of the data mining results.

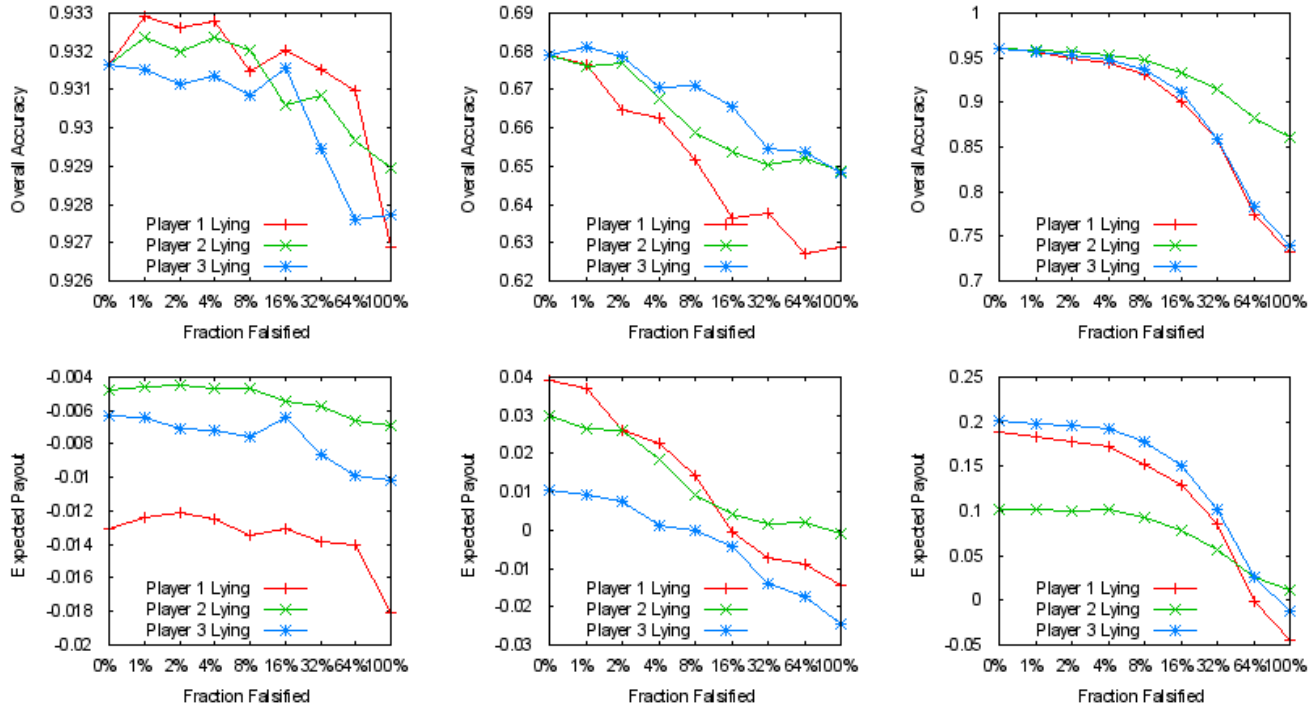
In addition, our experimentation raises some interesting questions about distributed data mining and the assumptions behind it. In some scenarios, the final data mining model performed better with slightly (and in some cases, greatly) falsified data from one of the players. While much has been done in the areas of noise reduction and dimension reduction, these methods assume that all the data is available for the process. These two facts in mind, we pose the question: What kinds of noise reduction techniques can be used on parts of the data effectively? In addition, which of these methods can be applied in an efficient manner, so that players will still have incentive to use them (since computation is costly)? Such a method would need to have its eventual payoff exceed the computation cost of the noise reduction. A simple random perturbation, as used in our experiment set, is very low-cost, but also affords a very small advantage. Could an efficient method exist for predicting which parts of the data need to change in order to increase the overall effectiveness of the final classifier?

Finally, it remains to be seen how these mechanisms work in a coalitional setting. Other game theoretic techniques such as the Shapely value might be used to incentivize coalitional gameplay. It is most likely that such mechanisms would only be secure against a non-majority coalition, and possibly smaller. We leave such work for another time.



Census-Income German-Credit Car-Evaluation

Fig. 2. Results for Naive Bayes Classification



Census-Income German-Credit Car-Evaluation

Fig. 3. Results for ID3 Decision Tree Classification

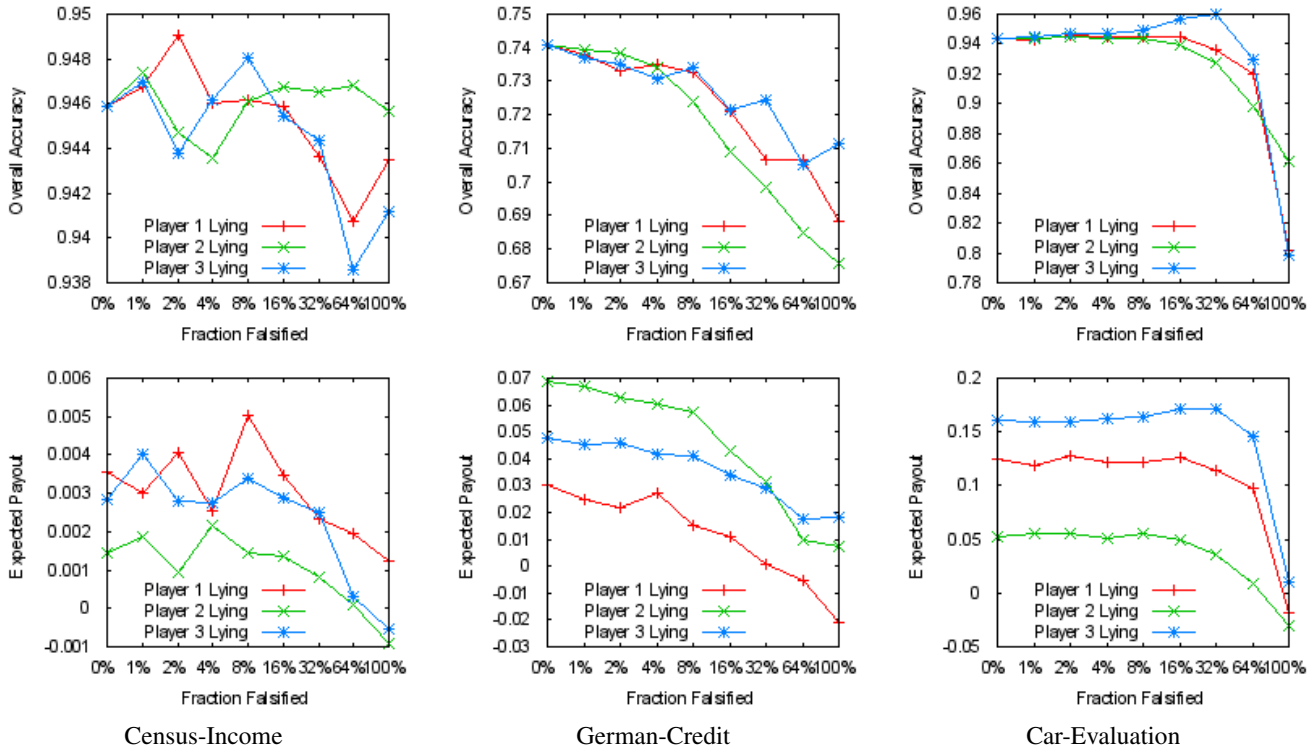


Fig. 4. Results for SVM Classification

VIII. Acknowledgements

This work was partially supported by Air Force Office of Scientific Research MURI Grant FA9550-08-1-0265, National Institutes of Health Grant 1R01LM009989, National Science Foundation (NSF) Grant Career-0845803, and NSF Grant 0964350.

References

[1] I. Abraham, D. Dolev, R. Gonen, and J. Halpern. Distributed computing meets game theory: Robust mechanisms for rational secret sharing and multiparty computation. In *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pages 53–62. ACM New York, NY, USA, 2006.

[2] R. Agrawal and E. Terzi. On honesty in sovereign information sharing. *Lecture Notes in Computer Science*, 3896:240, 2006.

[3] G. Annas. HIPAA Regulations–A New Era of Medical-Record Privacy? *The New England Journal of Medicine*, 348(15):1486, 2003.

[4] I. Ashlagi, A. Klinger, and M. Tenneholtz. K-NCC: Stability Against Group Deviations in Non-Cooperative Computation. *LECTURE NOTES IN COMPUTER SCIENCE*, 4858:564, 2007.

[5] A. Asuncion and D. Newman. UCI machine learning repository, 2007.

[6] S. Gordon and J. Katz. Rational secret sharing, revisited. *Lecture Notes in Computer Science*, 4116:229, 2006.

[7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The WEKA Data Mining Software: An Update.

[8] J. Halpern and V. Teague. Rational secret sharing and multiparty computation: extended abstract. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 623–632. ACM New York, NY, USA, 2004.

[9] S. Izmalkov, S. Micali, and M. Lepinski. Rational secure computation and ideal mechanism design. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 585–594, 2005.

[10] W. Jiang, C. Clifton, and M. Kantarcıoğlu. Transforming semi-honest protocols to ensure accountability. *Data & Knowledge Engineering*, 65(1):57–74, 2008.

[11] H. Kargupta, K. Das, and K. Liu. A Game Theoretic Approach toward Multi-Party Privacy-Preserving Distributed Data Mining.

[12] J. Katz. Bridging game theory and cryptography: Recent results and future directions. *Lecture Notes in Computer Science*, 4948:251, 2008.

[13] G. Kol and M. Naor. Cryptography and game theory: Designing protocols for exchanging information. *Lecture Notes in Computer Science*, 4948:320, 2008.

[14] R. Layfield, M. Kantarcioglu, and B. Thuraisingham. Incentive and Trust Issues in Assured Information Sharing. In *Collaborative Computing: Networking, Applications and Worksharing: 4th International Conference, CollaborateCom 2008, Orlando, FL, USA, November 13-16, 2008, Revised Selected Papers*, page 113. Springer, 2009.

[15] A. Lysyanskaya and N. Triandopoulos. Rationality and adversarial behavior in multi-party computation. *Lecture Notes in Computer Science*, 4117:180–197, 2006.

[16] R. McGrew, R. Porter, and Y. Shoham. Towards a general theory of non-cooperative computation. In *Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge*, pages 59–71. ACM New York, NY, USA, 2003.

[17] N. Nisan. Introduction to mechanism design (for computer scientists). *Algorithmic Game Theory*, pages 209–242, 2007.

[18] S. Ong, D. Parkes, A. Rosen, and S. Vadhan. Fairness with an honest minority and a rational majority. In *Sixth Theory of Cryptography Conference (TCC)*. Springer, 2009.

[19] D. Parkes and J. Shneidman. Distributed implementations of vickrey-clarke-groves mechanisms. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 261–268. IEEE Computer Society Washington, DC, USA, 2004.

[20] E. Rasmusen. *Games and information: An introduction to game theory*. Blackwell Pub, 2007.

[21] Y. Shoham and M. Tennenholtz. Non-cooperative computation: Boolean functions with correctness and exclusivity. *Theoretical Computer Science*, 343(1-2):97–113, 2005.