A Game Theory based Repeated Rational Secret Sharing Scheme for Privacy Preserving Distributed Data Mining

Nirali R. Nanavati and Devesh C. Jinwala
Sardar Vallabhbhai National Institute of Technology, Surat, India

Keywords: Privacy, Game Theory, Secure Multiparty Computation, Rational Secret Sharing, Distributed Data Mining.

Abstract: Collaborative data mining has become very useful today with the immense increase in the amount of data collected and the increase in competition. This in turn increases the need to preserve the participants' privacy. There have been a number of approaches proposed that use Secret Sharing for privacy preservation for Secure Multiparty Computation (SMC) in different setups and applications. The different multiparty scenarios may have parties that are semi-honest, rational or malicious. A number of approaches have been proposed for semi honest parties in this setup. The problem however is that in reality we have to deal with parties that act in their self-interest and are rational. These rational parties may try and attain maximum gain without disrupting the protocol. Also these parties if cautioned would correct themselves to have maximum individual gain in the future. Thus we propose a new practical game theoretic approach with three novel punishment policies with the primary advantage that it avoids the use of expensive techniques like homomorphic encryption. Our proposed approach is applicable to the secret sharing scheme among rational parties in distributed data mining. We have analysed theoretically the proposed novel punishment policies for this approach. We have also empirically evaluated and implemented our scheme using Java. We compare the punishment policies proposed in terms of the number of rounds required to attain the Nash equilibrium with eventually no bad rational nodes with different percentage of initial bad nodes.

1 INTRODUCTION

The utility and sharing of huge amounts of data has become possible today with the development of network, data collection and storage technologies. The huge amount of extracted knowledge among different parties does have the issue of loss of privacy of the participants. Hence a number of algorithms have been proposed to resolve the issue of privacy in distributed data mining. These algorithms are actually based on the concept of coopetation(Pedersen et al., 2007) among the parties.

[m,m] Secret Sharing for the PPDARM semi honest model in a mesh topology which is our focus of study has been proposed recently by (Nanavati and Jinwala, 2012); (Ge et al., 2010). However in a realistic formulation of PPDARM these parties would be rational. According to (Abraham et al., 2006); (Maleka et al., 2008); (Miyaji and Rahman, 2011) these rational agents will have an inclination to not send their shares as each of them would first prefer getting the secret and secondly prefer that fewer of the other agents that get it, the better. Hence we propose an extension to the [m,m] Secret Sharing scheme for Privacy preservation in Distributed Data Mining (PPDDM) for such rational agents.

(Kargupta et al., 2007) does explain multiparty PPDDM as a game but it does not cater to Secret sharing and only explains the Secure sum protocol in a ring topology unlike the mesh topology used by us. (Maleka et al., 2008) introduces the concept of Repeated Rational [m,n] Secret Sharing but cannot be used for SMC. It is a model with mediators unlike our model which is without mediators. (Miyaji and Rahman, 2011) is a recent work that explains the application of game theory but only to the set intersection protocol in PPDDM. (Abraham et al., 2006) has a novel version of Rational Secret Sharing that explains collusion and cheap talk but does not consider secret sharing as a repeated game and does not give details of the punishment strategy or the application to PPDDM.

Hence we introduce a novel algorithm to the best of our knowledge which models Secret Sharing (Shamir and Adi, 1979) in PPDARM as a Repeated

R. Nanavati N. and C. Jinwala D..
A Game Theory based Repeated Rational Secret Sharing Scheme for Privacy Preserving Distributed Data Mining.
DOI: 10.5220/0004525205120517
In *Proceedings of the 10th International Conference on Security and Cryptography* (SECRYPT-2013), pages 512-517
ISBN: 978-989-8565-73-0

Rational Secret Sharing scheme using cheap talk. It is modeled for rational parties without mediators. This model is applicable to all the recent work done in PPDARM and distributed clustering using [m,m] secret sharing(Nanavati and Jinwala, 2012); (Ge et al., 2010)(Patel, Garasia and Jinwala, 2012) without mediators to compute the secure sum in a mesh topology unlike most other [m,n] – threshold secret sharing algorithms that aim to share a secret and not perform secure sum computation.

The proposed model is advantageous as it avoids the use of techniques like homomorphic encryption and zero knowledge proofs that incur a high cost. Our punishment policies do not aim at removing the players but aim at getting the maximum possible participation in the game.

2 PROBLEM FORMULATION

Our problem setup involves a co-opetative setup of vertically or horizontally partitioned databases where PPDARM is required (depending on the application (Nanavati and Jinwala, 2012); (Ge et al., 2010); (Nanavati and Jinwala, 2013) with 'p' rational parties collaborating to find the secure global sum in their data privately. This problem extends the [m,m] threshold scheme proposed for rational agents.

We model secret sharing as a game among rational agents represented by Γ(m,m). We also consider it as a repeated game (Maleka et al., 2008) where the same set of players come to play the same game repeatedly. Our model is a model without mediators unlike(Maleka et al., 2008). The players are connected together in a mesh topology.We assume that the players are normally concerned about their future utilities. Among the three main attacks possible in [m,m] secret sharing(Nanavati and Jinwala, 2013), we try to resolve the two attacks by a rational adversary who is not disrupting the protocol by sending wrong shares but is trying to selfishly get his own gain by witholding his shares or sending them only to the collusion. Our punishment policies resolve these attacks and motivate the player to abide by the protocol. The attack where wrong shares are send by malicious parties can be resolved using the Verifiable Secret Sharing Scheme.

3 GAME THEORY AND NASH EQUILIBRIUM

Game Theory has today developed into an umbrella term encompassing various scenarios in the real world where the individuals want maximum benefit after taking into account the actions of the other parties involved in the setup. 'Cheap talk' in a game theoretic framework refers to the direct and costless communication among players

In game theory, the Nash equilibrium is a solution concept of a game that comprises of two or more players, and none of them has anything to gain by changing only his own strategy alone. If each player has chosen a strategy and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium.

4 PROPOSED ALGORITHM

The main aim of PPDDM is to be able to compute the global function on all the parties and hence they want to have maximum participation. For the [m,m] Repeated Secret Sharing game proposed by us; we assume that all rational parties broadcast their shares and then their sum of shares simultaneously to know the value of the global sum $\sum_{i=1}^{p} S_{ij}$ of the secret values at each party P_i. It is an extension of the algorithm based on Shamir's secret sharing technique(Shamir and Adi, 1979); (Ge et al., 2010).

We improve on the punishment strategies of (Kargupta et al., 2007); (Maleka et al., 2008) as mentioned in (Nanavati and Jinwala, 2013) using [m,m] Secret Sharing so that all parties will have the maximum utility and will attain the Nash equilibrium state.

Considering the p rational partitions (where $0 < i \leq p$) and each transaction has a subset of 'N' items. Consider D is the no. of defaulters; y_i is the number of rounds for which the defaulter 'i' is ousted in Policy 1 and $incr_i$ is the increment added to y_i. Consider c_i is the maximum number of chances given to the defaulter to enter the setup, n_i is the no. of rounds that the party i has defaulted, $rating_i$ is the current rating of the i^{th} player in policy 3 and $life_i$ is the number of lives left with the i^{th} player. The algorithm proposed by us is given below in Figure 1.

The proper administration of the punishment policies listed in section 4.1 will be monitored by a moderator who is not a part of the protocol. The

moderator will just ensure that the parties are ousted properly and are given their due punishment.

Input: p, N, n_i, policy used, y_i, $incr_i$, c, $rating_i$
Output: Secure Sum using Repeated Rational Secret Sharing for [m,m] scheme

```
1:for each transaction j = 1 to N do
2: for each party Pi, (i = 1 to p)do
3: each party computes the share of other party Pi, where share(Sij,Pi)= qi(x)(random polynomial)
4:   for t = 1 to p do
5:   send share(Sij, Pi) to party Pt
6:   receive the shares share(Sij, Pt)
7:   end for
8:compute                Sum(xi)=q1(xi)+ q2(xi)..+qn(xi)
9:  for t = 1 to p do
10:  send Sum(xi) to party Pt
11:  If a party does not send a sum of shares; initiate cheap talk.
12:   If policy 1 is chosen by the setup
13:  If (ni = 0)
14: defaulter(d) is removed from the setup for 'yi' rounds; ni++;
15:  Goto Step 3 for 'p – D' parties
16:  Else
17: defaulter(d) is removed from the setup for 'gi' rounds where gi = yi + incri; yi=gi; ni++;
18:  if ni =c; the party Pi is permanently ousted
19:  Goto Step 3 for 'p – D' parties
20:   If policy 2 is chosen by the setup
21:  If (ni = 0 )
22: defaulter is removed for 1 round until a limit of 'c' rounds; ni++;
23:  if ni =c; the party i is permanently ousted
24: Goto Step 3 for 'p – D' parties
25: If policy 3 is chosen by the setup
26:   Lifei = c;
27:  Party with lowest rating sends share first
28:  If (party sends share)
29:  Increment the ratingi by 1 Else
30:  {Decrement ratingi and Lifei by 1 and no shares are sent to that party in that round.
31:  If Lifei = 0
32:  Party 'i' is permanently ousted
33:  Goto Step 3 for 'p – D' parties}
```

Figure 1: Proposed Repeated Secret Sharing Algorithm applicable to our scenario of rational parties.

```
35: solve the set of equations to find the sum of . ∑_{i=1}^{p} S_ij secret values.
36: end for
37:  Each party gets the value of ∑_{i=1}^{p} S_ij
38: If policy 1 or 2 is followed
39: After yi or 1 round respectively if party initiates cheap talk to enter
40:  Allow it and goto 3
41: If policy 3 is followed
42:  After each round there is a ratings agreement which if satisfied; the party with lowest rating is made to send the shares first
43:If pi does not send its shares first
44: Decrement ratingi and Lifei by 1 and no shares are sent to that party in that round.
45:end for end for
```

Figure 1: Proposed Repeated Secret Sharing Algorithm applicable to our scenario of rational parties (cont.).

4.1 Punishment Policies

Policy 1: Incremental Punishment Strategy for repeated Rounds upto c Times: If a party defaults; remove him from the game for 'y' rounds. He is again given a chance to enter the setup after 'y' rounds and if he defaults again; he is removed for an incremental 'g' rounds. This defaulting behavior is allowed for only upto 'c' attempts.

Policy 2: Punishment Strategy allowing 'c' Attempts: Remove the party for 1 round ; if he again defaults remove him again and give him 'c' chances where 'c' is the no. of chances you want to give a defaulter and is decided by the coalition.

Policy 3: Ratings based Punishment Strategy simulating a Real Game: The rating of the party that follows the protocol is incremented by one and the party that does not follow the protocol gets its rating decremented by 1. These ratings are maintained at all parties since we do not have a mediator in our protocol. There is a round of 'rating agreement' which is more of a cheap talk that needs to take place before each round of secret sharing which ensures that the bad rational parties do not change the ratings. Now based on the ratings; the party with the lowest rating is supposed to send the shares first so as to ensure corrective behaviour. If the party with lowest ratings fails to send its shares first; its lives and ratings are decremented by 1 and it does not receive the shares for that round. Also like a real game; there are 'c' lives or chances which

means that only 'c' decrements in rating are allowed after which the party is ousted from the game.

If a new party wants to enter the setup; that party is granted the minimum rating of the setup and is made to send its shares first as the setup does not trust this new party entirely.

To summarize if there is no penalty for cheating, rational participants tend to behave dishonestly. Hence this new scheme does help control the bad behaviour of dishonest parties without a heavy cost.

5 THEORETICAL ANALYSIS

In a distributed setup where Secret sharing is undertaken for PPDARM; the parties have different roles like carrying out computations at their end, sending the messages and receiving the messages. For the [m,m] secret sharing scheme we do not discuss the attack where the parties send wrong information as the aim of rational parties is not to sabotage the protocol. Instead they may default by withholding their sum of shares and/or sending them just to the coalition. $u_S(S_{i,t})$, $u_R(R_{i,t})$ are the utilities or payoff where the i^{th} $(0 < i \leq p)$ party sends and receives corresponding t^{th} sum of shares. $w_{i,s}$ and $w_{i,R}$ are the weights of these utilities respectively. $u_i(\{\sigma_i, \sigma_{-i}\})$ denotes the utility of the party's strategy which is interdependent on the other parties. Hence the formulation of multiparty PPDDM as a strategic game (Kargupta et al., 2007) for our approach is proposed as:

$$u_i(\{\sigma_i, \sigma_{-i}\}) = w_{i,s} * u_s(S_{i,t}) + w_{i,R} * u_R(R_{i,t}) \quad (1)$$

Our game is based on the goal of getting maximum shares and thus maximum participation and not on the communication cost. Also if there is no penalty; the gain would be maximum when no sum of shares is sent by that bad rational party.

Hence the punishment policies we propose will increase the utility of all the parties and the Nash equilibrium is attained when the parties cooperate honestly making it the optimum strategy.

5.1 Collusion Analysis

This algorithm can also effectively prevent collusive behavior if the number of collusive parties C < p.

If the no. of collusive parties = p then the protocol has no meaning. Now we consider D as the no. of defaulters and C as the no. of colluding nodes who send the sum of shares only in the colluding group. For the collusion analysis; if C < D; then the defaulting parties will not get the result back and so that scenario has no meaning and is not a rational behaviour. However if C = D = 1, then the party can get the sum of values of the other parties. On the other hand if C = p-1 or the general case where C=D then the colluding parties can definitely get the value of the sum polynomial $Sum(x_i)$ and hence can predict the value of the non-collusive party. They can have a high utility as they are not sending any shares outside the collusion.

This can be avoided if we follow our punishment strategy so that fewer parties will collude so as to attain the maximum gain in future rounds. The punishment policies are applied to the entire collusion by assuming that if more than 1 party is defaulting; there is possibility of a collusion.

6 RESULTS AND IMPLICATIONS

This section gives a brief summary of our implementation. The platform for the prototype is Java 7 SE on a 2.5 GHz i5 processor with 64 bit Operating system and 6 GB RAM.

As a case study we show a comparison between the no. of rounds of secret sharing and the percentage of bad nodes for all our policies. We have evaluated our results with a total of 10 nodes. We assume that about at least 50% of the bad nodes after being penalized prefer to join the game in the coming round of which about at least 50% would behave honestly.

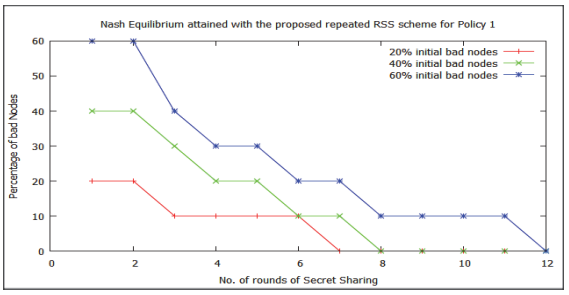

Figure 2: Nash Equilibrium attained with the proposed repeated RSS scheme for Punishment Policy 1.

The comparison shown above is based on Equation (1). Also we consider the $u_s(S_{i,t})$ and $u_R(R_{i,t})$ as the number of shares sent and received where the $u_s(S_{i,t})$ is negative as none of the rational parties want to send their shares willingly and the $w_{i,R}$ is twice that of sending $w_{i,S}$. according to the preference of the rational agents. It validates the fact that the parties after being ousted try and attain better utilities. Also it compares the percentage of

bad nodes with the rounds of secret sharing.

Percentage of bad nodes = No. of nodes not sending shares/ Total participating nodes (does not include the permanently ousted nodes) (2)

The comparison shown in Figure 3 and Figure 4 below is based on punishment policy 2 and policy 3.

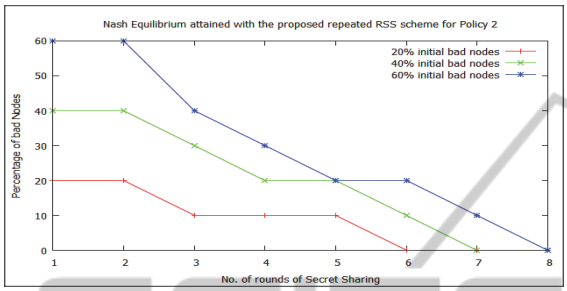

Figure 3: Nash Equilibrium attained with the proposed repeated RSS scheme for Policy 2.

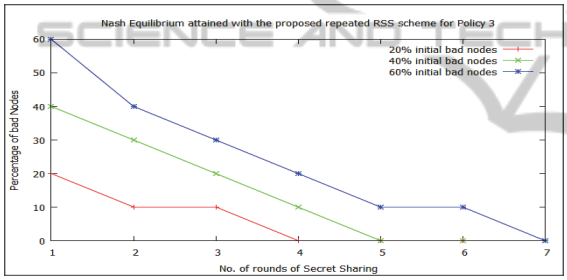

Figure 4: Nash Equilibrium attained with the proposed repeated RSS scheme for Policy 3.

Our experiments indicate that irrespective of the initial percentage of bad nodes; with the proposed policies all the non cooperating nodes converge and we get a Nash equilibrium where all the nodes are honest. If the bad nodes do not change their behavior; they are permanently ousted. This graph would vary if the nodes behaved in a different manner than the assumptions would still converge to zero bad nodes after certain rounds.

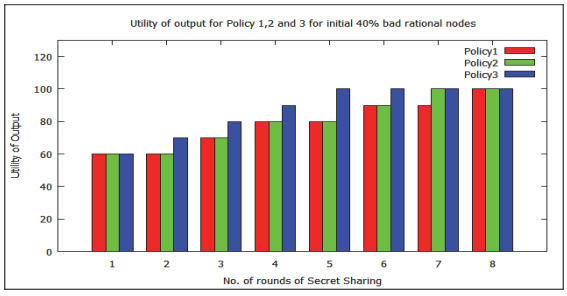

Figure 5: Utility of output for Policy 1, 2 and 3 for initial 40% bad rational nodes.

Finally we have another metric which evaluates the "utility of the output" with respect to the number of rounds in Figure 5. This metric varies inversely with respect to the percentage of bad nodes. The utility of output is the maximum when we have the maximum participation from the parties. We observe that the utility of output attains the maximum value earlier for Policy 3 than Policy 1 and 2.

Utility of output = No. of good nodes/Total No. of participants(does not include the permanently ousted nodes) (3)

Further we are considering an example where we are considering four participants in the game theoretic setup for our policies so that we can validate the attainment of the Nash equilibrium with our policies. We use the same assumptions as in the 10 node scenario. Hence based on this, our payoff matrix for repeated secret sharing is in Table 1 below where the Nash equilibrium state is denoted by (NE) and the good and the bad rational nodes are denoted by (G) and (B) respectively.

It is clear from the Table 1 that if a party defaults in the first round then; it only gets a good payoff if it undertakes corrective behaviour. We also observe from the Figure. 2-5 and Table 1 that Policy 1 needs more rounds to attain the Nash equilibrium but the advantage is that it is a harsher incremental policy. Hence it might inspire more nodes to correct themselves initially itself rather than not getting shares for a large no. of rounds. Policy 2 is less harsh as it ousts the party temporarily only for 1 round and helps attain the Nash equilibrium faster. Finally Policy 3 replicates an actual game scenario and instead of ousting temporarily; it decreases the ratings of that player. Hence the punishment is not harsh and the Nash equilibrium is attained faster.

7 CONCLUSIONS

In this paper we propose a novel scheme for repeated rational secret sharing using the game theoretic approach and cheap talk for various approaches in PPDARM(Nanavati and Jinwala, 2012);(Ge et al., 2010). Our scheme can also be applied to other privacy preserving approaches that use Secret Sharing to undertake SMC among rational agents without mediators.

We have proposed three novel and effective punishment strategies for our game theoretic approach. We also give the theoretical analysis and empirical evaluation of our proposed scheme. We compare the different policies proposed in terms of

Table 1: Payoff Matrix for 4 party scenario showing Nash Equilibrium state as well.

Policy Used	A,B,C,D	R_1	R_2	R_3	R_4	R_5	R_6	R_7
1,2,3	G,G,G,G	3,3,3,3 (NE)	3,3,3,3	3,3,3,3	3,3,3,3	3,3,3,3	3,3,3,3	3,3,3,3
1,2	G,G,G,B	1,1,1,6	2,2,2,0	3,3,3,3 (NE)	3,3,3,3	3,3,3,3	3,3,3,3	3,3,3,3
3	G,G,G,B	1,1,1,6	3,3,3,3 (NE)	3,3,3,3	3,3,3,3	3,3,3,3	3,3,3,3	3,3,3,3
1	G,G,B,B	-1,-1,5,5	1,1,0,0	2,2,2,0	1,1,1,6	2,2,2,0	2,2,2,0	3,3,3,3 (NE)
2	G,G,B,B	-1,-1,5,5	1,1,0,0	2,2,2,0	1,1,1,6	2,2,2,0	3,3,3,3 (NE)	3,3,3,3
3	G,G,B,B	-1,-1,5,5	2,2,2,0	3,3,3,3 (NE)	3,3,3,3	3,3,3,3	3,3,3,3	3,3,3,3

the no. of rounds taken to attain the Nash equilibrium. For all these policies we have considered different percentage of initial bad rational nodes that refrain from sending their shares. Among these policies our results show that the Nash equilibrium is attained in the least no. of rounds for Policy 3 which simulates an actual game setting.

Finally we conclude that our protocol works in the favour of all the rational parties to attain a state where all parties are honest ultimately and has the maximum utility considering that they take their future utilities in account.

However there are a few open research challenges which include extending our scheme using game theory to verifiable secret sharing as well as to prevent the malicious adversaries. Another extension would be to analyse the application of our punishment policies to other privacy preserving schemes in distributed data mining.

ACKNOWLEDGEMENTS

I would like to thank my student Neeraj Sen for the discussions on this problem and also for assisting me in the implementation. I would also like to thank my husband Rohan for the endless talks and motivation.

REFERENCES

Abraham, I., Dolev, D., Gonen, R., and Halpern, J. (2006). 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*, PODC '06, pages 53–62, NY, USA. ACM.

Ge, X., Yan, L., Zhu, J., and Shi, W. (2010). Privacy-preserving distributed association rule mining based on the secret sharing technique. In *Software Engineering and Data Mining (SEDM), 2010 2nd International Conference on*, pages 345 –350.

Kargupta, H., Das, K., and Liu, K. (2007). A game theoretic approach toward multi-party privacy-preserving distributed data mining. In *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Warsaw, Poland, 2007, pages 523–531, Springer.

Maleka, S., Shareef, A., and Rangan, C. P. (2008). Rational secret sharing with repeated games. *In Proceedings of the 4th international conference on Information security practice and experience*, ISPEC'08, pages 334–346, Berlin, Springer-Verlag.

Miyaji, A. and Rahman, M. S. (2011). Privacy-preserving data mining: a game-theoretic approach. In *Proceedings of the 25th IFIP WG 11.3 conference on Data and applications security and privacy*, DBSec'11, pages 186–200, Berlin. Springer-Verlag.

Nanavati, N. R. and Jinwala, D. C. (2012). Privacy preserving approaches for global cycle detections for cyclic association rules in distributed databases. In *SECRYPT*, pages 368–371.

Nanavati, N. R. and Jinwala, D. C. (2013). A Novel Privacy preserving game theoretic repeated rational secret sharing scheme for distributed data mining. In *Security and Privacy Symposium*, Kanpur.

Patel, S., Garasia, S., and Jinwala, D. C. (2012). An efficient approach for privacy preserving distributed k-means clustering based on shamir's secret sharing scheme. In *IFIPTM*, pages 129–141.

Pedersen, T. B., Saygin, Y., and Savas, E. (2007). Secret Sharing vs. Encryption-based Techniques For Privacy Preserving Data Mining. *Sciences*-New York, 17–19.

Shamir, A. (1979). How to share a secret. Commun. ACM, 22:612–613.