

# A Game-based Method for User Privacy Protection in Big Data

Xiaohui Li<sup>1</sup>, Chaoyang Chen<sup>1</sup>, Feng Liu<sup>2</sup>, Bo Li<sup>1</sup> and Yajun Wang<sup>1</sup>

1, College of Electrical and Information Engineering, Liaoning University of Technology, Jinzhou, Liaoning, 121001, China

2, Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang, Liaoning, 110168, China

Email: lhx1xh@163.com

**Abstract.** In order to protect the user privacy in the big data environment, we propose a security method based on game theory through trust management mechanism. The main contribution of this paper includes: (1) analysis of service game relationship in the big data environment, (2) combine trust evaluation with decision management, and apply game theory to design a user-centric scheme to protect privacy, (3) proposing a service selection method. By using game theory to restrict the service interaction between the users and services in big data, we introduce some factors corresponding to the game model. We then apply the trust evaluation and decision management to describe the relationship and trust quantify in which we propose a method for service selection to implement user privacy protection. Simulation results show that the proposed method is effective and can apply the dynamic game of incomplete information relationship between user and service to make decisions on the service selection, thus providing customized privacy protection.

## 1. Introduction

In the big data environment, the need for data protection (especially in a shared multi-tenant environment) is the main problem of users to consider, which is the main fetters to carry out big data, user data protection in big data environment has become a well studied hot spot. How shared storage, computing and network resources and the environment effectively isolate and protect workloads and data need to rely on some technical, many researchers have introduced trusted tools, to try to solve the user data in the big data environment protection issues, and made some research. Studies have shown that trust in the big data environment should be trusted service centers, the trust is a two-way behavioral data and storage services have side trust between the parties, by means of both credibility and dual constraints together contributed data reasonable and lawful use without being abused. For users, the service can select trusted party, agreed a mutually satisfactory security mechanisms in order to achieve maximum security. On the service side, once integrity will have no place to lose profits than the integrity of an important enterprise. Therefore, big data environments can use data protection and service end user behavior trust bidirectional way to resolve it.

## 2. Service Game Analysis in Big Data

### 2.1. Related Theory

The basic elements of the game are:

(1) participants: that is decision-making body of the game selecting actions to maximize the effectiveness, who can be individuals or organizations.



(2) strategies: that person's participation in decision-making variables and the selective rules of decision variables, which tells what kind of action involved under what conditions.

(3) orders: game party action priorities order, game can be divided into static game and dynamic game.

(4) payoffs: Results benefit or payment, corresponding to the decision-making game party, the proceeds or the failure of the game to participate in what people are most concerned about.

Divided from different angles, the game has a variety of classification game: game action according to the order, the game can be divided into static and dynamic. If all the players actions while selecting called static game, players' actions have successively divided, is called dynamic game; according to people involved in the game to other participants understand the extent of information that can be divided into complete information game and non-complete information game. If each participant to have an accurate knowledge of the characteristics of all the other participants, strategies is called complete information game, otherwise it's not complete information game; according to the demand side of the interests of the game can be divided into cooperative game and non-cooperative game, organizations seeking to maximize the interests of cooperative game, participants get the most out of the first collaborative community interest groups and then assign interests rather than the interests of individuals seeking to maximize cooperative game, not sign the agreement or the existence of collaboration between participants.

## 2.2. Mechanism Design

Mechanism design theory is one of the core problems of modern economics, it is for any discussion of a given economic or social goals, in a free choice of decentralized decision-making conditions, whether and how to design an economic mechanism that allows economic activity participants consistent with personal interests and goals. Designers can be as large as the entire economic and social system designers, small to have only two participants (in their own best interests as the goal) of the principal economic organization and management.

People face with an incomplete information society, all personal information cannot be completely one person to master, so the use of decentralized decision-making, with incentives to motivate people to do the designers want to do, or to realize the designers want to achieve goal, it is a matter of economic mechanism design theory to the study. Incentive mechanisms, that is, under the mechanism designed so that each participant while in pursuit of personal interests to achieve goals set by the designer. To achieve a certain goal, the first to make this goal within the scope of the technical feasibility; secondly, to make it meet the individual rationality, namely participation, if you do not participate in the game provided, because there is a better choice, then the mechanism design is invalid; third, to meet incentive compatible constraint, to make voluntary self-interested behavior to achieve personal goals.

For a variety of technical integration of big data environment, the user's choice to face multiple problems, mechanism design theory this has important practical significance. services and users as a resource requester is also the provider of resources, are the dual roles, but the identity is not fixed, access control permissions granted can not fully guarantee safety. From the user's privacy protection, we need a variety of services to compare and study the merits of the degree of basic standards and a unified framework, the need for a general theory can research and compare various service providers to consider the service selection problem, and mechanism design theory is undoubtedly a good choice in this regard.

## 2.3. Dynamic Service Analysis

The service interaction big data environment is described as the game process between users and service: the service both quantify and make service selection based on trust, because the trust itself is a psychological expectation, both services and interactive process cannot fully grasp the information with each other, so the process is non-complete information game.

The two sides were based on trust game, if the user trust services, the service is accepted, otherwise rejected; services provide users may take integrity or fraud. Four factors corresponding game model are as follows:

(1) Participants: big data environment service parties, collectively referred to users  $P-c$  and services  $P-s$ ;

(2) Strategy: In the service interaction process, the strategy space users  $S_c = \{\text{accept}, \text{reject}\}$ ; strategy space for services  $S_s = \{\text{honest}, \text{dishonest}\}$ ;

(3) the order of the game: the interactive process based on service, services are the first actors, users are after actors;

(4) payoffs: the service side of honest or dishonest policy, the service participants benefits or losses obtained.

In the big data environment to benefit form the game to make the following assumptions, when the user chooses to accept strategy, users benefit  $C - \text{Income}$ , services benefit  $S - \text{Income}$ ; users are dishonest to provide services, user loss  $C - \text{Loss}$ , services proceeds  $S_d - \text{Income}$ ; users to opt out strategy, service benefits both sides are 0. Payoffs matrix shown in Table 1.

**Table 1.** Payoffs matrix table

	accept	refuse
honest	$S - \text{Income}, C - \text{Income}$	0,0
dishonest	$S_d - \text{Income}, C - \text{Loss}$	0,0

### 3. Method Design

#### 3.1. Design Idea

The first is the user's information release and services to provide a conflict of interest exists honesty, integrity and provide services need to pay the cost, from the game theory analysis, for each participant, it will choose strategies to maximize their effectiveness. services in order to maximize revenue, will choose the acts of bad faith, that the participants did not have any normative constraints under service commitment is not credible; In a big data environment, trust between users and services is unequal, service users' trust in the is a service that users can choose the premise, the purpose is to protect the environment of services are not affected by the user, and the users trust services demand security services. In other words, although users and services while both the dual role of resource requests and resources provided, but the service selection cooperation is to establish a good reputation, there is no users to establish a good reputation problems. Game theory analysis between users and service information is asymmetric. No matter how the user incentive services, services would choose to maximize the effectiveness of their strategies. When users to maximize the effectiveness of the policy will be made to receive services, but services in order to maximize revenue, will choose non-faith conduct, even if such an impact on reputation.

Hence the need to make services through incentives to make the integrity of the service provider, an increase of user data release granular, increasing the income of services, users with the release of more data, the more we need to protect the integrity of the service. Using trust to balance the user information release and conflicts of interest between services, services and its service providers trust quantify historical behavior are closely related.

In the dynamic game of incomplete information, the involvement of people with the revaluation of action sequences and correction of information is an important factor in the decision game results. Trust in the -based service acquisition game, services are the first actors to provide the type of service, service users to take action by observing the catalog, which is determined by the order of the game. So this game have information on the results of the game is important. Before the game, from the observation, the accumulation of pre-game or subjective inference, this information is called a priori information (prior probability) is also known as a priori probabilities. In the dynamic game, the game is the process involved in the process who initially released and reveal private information, and is the process by participants observed results of prior information be corrected, and a priori information corrected for a posteriori information (posterior probability), also known as the posterior probability.

Game information correction process includes two dynamic decision-making processes and their integration interact and influence each other. Trust in the -based service acquisition for analysis, used to calculate the credibility and trust services directly belong to the history of trust and other information prior information on services user feedback evaluation belong posteriori information.

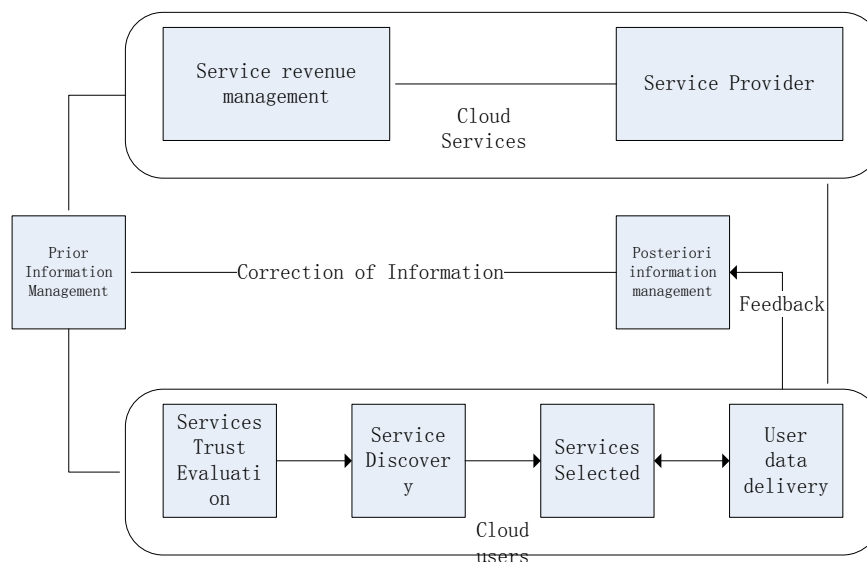
### 3.2. Users Data Protection Method Based on Game

The program of the big data environment between the user and the service acquisition process deem incomplete information game, on the one hand by the rules of the game to build services offer incentives security; on the other hand, users of the trust relationship management services to enhance user efforts to protect its own resources. Specific user data big data environment protection method designed with the following objectives:

(1) The behavior of a big data environment service interactive mode, seen as services and between users of the game, trust is important information for both games, the application rules of the game, encouraging both behavior is credible.

(2) Game organization and management information including quantitative methods and trust service selection methods.

(3) Big data environment, services both responsible for their behavior, if there is fraud or malicious attacks, service interactions will affect future that pay the price, so its services according to certain criteria to obtain, make rational decisions.



**Figure 1.** Work flow chart

### 3.3. Service Selection Method

Under the trust game as a service information interaction between the two sides, and how the trust the same or similar service resources, select one of the most appropriate services to meet the security conditions in the premise of the program, to demand personalized service is controlled one important issue of user data protection solution implemented, this section will give service selection module approach instance programs.

Services option is to evaluate the choice of alternative service problems, the number is limited classified as attribute decision making, you can choose to study theory. Known for service attribute weights, weights based on decision analysis and decision-making matrix, sort, the more common methods are: linear distribution method. Because individual needs and preferences trust users, using the linear assignment.

Linear distribution method is an attribute decision making problems to sort methods. The law does not require the given decision matrix, only need to know the various services of the priorities for the

election of each property, use this order to solve the corresponding binary linear programming, so that the corresponding right and reaches the maximum, in order to determine the most likely location solutions. The need to construct a weight matrix method to estimate the location of each program aligned with the  $n$  and  $m$  for selected service attributes, the general steps are as follows:

Step 1: single objective sorting special case. The program may appear side by side, the virtual target method, which set the goal  $g_k$  for the corresponding weight  $w_k$  for the  $g_k$  decomposition  $h$  virtual target  $g_{k1}, g_{k2}, \dots, g_{kh}$  and empower heavy  $w_k / h$ , do the equivalent processing;

Step 2: The program sort by a single goal;

Step 3: Given weights and program line up for each target, construct weight matrix;

Step 4: Construct and solve linear programming problems, the definition of binary permutation matrix  $P$ , whose elements  $P_{ik}$  is 0 or 1. According to the principle of linear distribution method, linear structure following questions:

$$\max \sum_{i=1}^n \sum_{k=1}^n w_{ik} P_{ik} \quad s.t. \sum_{k=1}^n P_{ik} = 1, i = 1, \dots, n \quad (1)$$

Solving this problem is to find a set of values  $P_{ik} = 1$  is the right and the corresponding maximum.

In a big data environment, due to the demand for users personalized service, as the service selection policy makers consider the preference structure to describe the need for trust in services, from spatial and temporal properties and behavior of the user feedback and other aspects of the trust describe the service trust attribute (the attribute weights and trust services are described in detail in the article quantified). When users access to services attribute specific weight, based on the right lineament analysis matrix, selection of service plans will be sorted. Here is an examples service selected linear assignment:

Assume elected service  $(X_1, X_2, X_3, X_4)$ , known weight of each attribute  $W^T = (0.232, 0.142, 0.51, 0.116)$ , according to the preference of service users, as shown in Table 2 single attribute sort.

**Table 2.** Service single attribute sorting

Attributes Order	Time	Space	Users trust	Historical behavior
1	$X_3$	$X_1$	$X_2$	$X_3$
2	$X_4$	$X_2$	$X_1$	$X_4$
3	$X_2$	$X_4$	$X_4$	$X_1$
4	$X_1$	$X_3$	$X_3$	$X_2$

Weight matrix structure  $\overline{W}$ :

$$\overline{W} = \begin{bmatrix} 0.142 & 0.51 & 0.116 & 0.232 \\ 0.51 & 0.142 & 0.232 & 0.116 \\ 0.348 & 0 & 0 & 0.652 \\ 0 & 0.348 & 0.652 & 0 \end{bmatrix}$$

Weight matrix for service by sorting, sorting method is as follows:

If  $\overline{W}$  all elements of the  $i$ th row of the first one in the maximum value, the service  $X_i$  in the first row, and so on, should have multiple services on the same ranking, you need to consider the following order of precedence values. The priority of the following four services:  $X_2 > X_1 > X_4 > X_3$ ,

corresponding rights  $w_{21} + w_{12} + w_{43} + w_{34} = 0.51 + 0.51 + 0.652 + 0.652 = 2.324$ . The results achieve  $\overline{W}$  the corresponding right to maximum.

#### 4. Conclusion

In this paper, we proposed a user privacy protection method in the big data environment. Firstly, from the perspective of game theory, We describe the service interaction in the big data environments as the behavioral game between the service and the user, the trust and decision (service selection) as the game Information. Applying game constraints, the method motivated both parties to implement credible behavior, the user's service selection makes decisions through a trust mechanism to provide customized protection for user privacy data.

#### 5. Acknowledgments

This work is supported by the National Natural Science Foundation of China (under Grant 61802161, 51679116), Science and Technology Fund Project of Liaoning Province (under Grant 20170540448), the Innovative Talents Support Plan of Colleges and Universities of Liaoning Province (under Grant LR2017068).

#### 6. References

- [1] Privacy-preserving publishing of opinion polls[J] . Ahmed Abdalaal, Mehmet Ercan Nergiz, Yucel Saygin. Computers & Security. 2013
- [2] Parallel search over encrypted data under attribute based encryption on the Cloud Computing[J] . Thouraya Bouabana-Tebibel, Abdellah Kaci. Computers & Security. 2015
- [3] PPTD: Preserving personalized privacy in trajectory data publishing by sensitive attribute generalization and trajectory local suppression[J] . Elahe Ghasemi Komishani, Mahdi Abadi, Fatemeh Deldar. Knowledge-Based Systems . 2015
- [4] The new casper: query processing for location services without compromising privacy. M.F. Mokbel, C.Y. Chow, W. G. Aref. Proc of the 32nd international conference on Very Large Data Bases (VLDB) . 2006
- [5] Building access control policy model for privacy preserving and testing policy conflicting problems[J] . Hua Wang, Lili Sun, Elisa Bertino. Journal of Computer and System Sciences. 2014
- [6] Priv Bayes [J] . Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, Xiaokui Xiao. ACM Transactions on Database Systems (TODS) . 2017 (4)
- [7] Tools for privacy preserving distributed data mining[J] . Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, Michael Y. Zhu. ACM SIGKDD Explorations Newsletter . 2002 (2)
- [8] PHOABE: Securely outsourcing multi-authority attribute based encryption with policy hidden for cloud assisted IoT[J] . Sana Belguith, Nesrine Kaaniche, Maryline Laurent, Abderrazak Jemai, Rabah Attia. Computer Networks . 2018