Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 155 (2019) 686–691

Procedia Computer Science

www.elsevier.com/locate/procedia

2nd International Workshop on "Recent advances on Internet of Things : Technology and Application Approaches (IoT-T&A 2019)
August 19-21, 2019, Halifax, Canada

A survey on game theoretic approaches for privacy preservation in data mining and network security

Authors: Hitarth Shah[a], Vishruti Kakkad[a], Reema Patel[a], Nishant Doshi [a,*]
[a] *Pandit Deendayal Petroleum University, Gandhinagar, India*

Abstract

Game theory has found widespread use in various fields like Economics, Biology, Political Science, etc. and forms an aegis for logical decision making in these areas. In computer science, due to advancing technologies, there has been a pressing need to use game theory in various problems due to the lack of scalability of traditional solutions. There has been ongoing research in various fields of computer science like security, machine learning, cloud computing, etc. where game theoretic approaches are extensively used. In this paper, we present a review on game theoretical approaches to various fields in computer science such as privacy preservation, network security and intrusion detection and resource optimization. In the end, this paper provides a comparative study of various game models used in different applications in a tabular format.

*Keywords:*Game Theory, Privacy Preservation, Network Security, Intrusion Detection and Resource Optimization

1. Introduction

Game Theory is described as the study of strategic interactions among rational individuals. One of the major assumptions of Game Theory is that all individuals are "rational". This means that each individual has well-defined

*Corresponding author. Tel:+917923275458
Email address: Nishant.Doshi@sot.pdpu.ac.in

10.1016/j.procs.2019.08.098

objectives and each individual will implement the best available strategy to achieve them. Generally, Game Theory includes:

- Identification of the players and analysing the payoff for each player.
- Using various equilibrium strategies to make descriptive or prescriptive predictions.[8][9]

Thus, it can be said that Game Theory attempts at finding solutions that are in the best interests of all the players. Game Theory is used to develop strategies to maintain an equilibrium between data utility and preservation of privacy of data in Privacy Preservation. Game Theory in Network Security and Intrusion Detection is used to provide a mathematical framework of the strategies between the Network Attackers and System Defenders.

The layout of the paper is as follows: Section 2 consists of the basic terminologies required to understand Game Theory and its various components. Section 3 describes the applications of Game Theory in Computer Science. This is followed by a comparative table that provides the details of the various types of Game Models used to solve different problems in Computer Science, which is present in Section 4 of the paper. Section 5 provides the conclusion of the survey.

2. Basic Terminologies

In this section, all those terminologies related to game theory are explained.

- Players: A player makes decisions and performs action in the game whilst interacting with each other with some strategies constrained with payoffs associated with each player. [5]
- Payoffs: It is the measure of satisfaction obtained by each player during any decision made by the player. The payoff can be negative or positive based on the decision made by the player. [5]
- Strategy: It is the action plan of a player throughout the game on the basis of which the player makes decisions and plans the next action. This is decided on the basis of the past actions of the opponent. [5]
- Nash Equilibrium: Nash Equilibrium can be defined as an action profile such that no player can achieve better results by performing a different action, provided that all other players adhere to the action profile. It is the intersection of the best responses of all the players involved in the game. [7]
- Static Games: Static games are those games where all the players perform their actions at the same time. [5]
- Stochastic Games: These type of games consist of a sequence of states and transitions in which each state has a probabilistic transition associated with it. The action performed by each player depends upon the payoff of the present state of the game. [5]
- Dynamic Games: Dynamic games are different from static games in the sense that dynamic games include multiple actions taking place between the attackers and defenders.
- Sequential Games: Sequential games are those games which fulfil the following conditions:
    - Each player chooses their action before the other player chooses theirs.
    - Each player should have knowledge of the actions chosen by previous players.[7]

3. Applications of Game Theory in Various fields of Computer Science

The application of Game Theory in Privacy Preservation and Network Security are described as follows:

3.1 Privacy Preservation

There have been noteworthy advances in the field of data mining since a few years. This has led to a consequential increase in the amount and variety of data collected. However, with this increase in data, there has been an increasing concern about the privacy preservation of the sensitive information of individuals. Thus, in order to perform Privacy Preservation in Data Mining (PPDM), one of the approaches employed is game theory. [1]The primary aim of Game theory models used in PPDM is to obtain the Nash Equilibrium between the players involved. The different game models for privacy preservation are as follows:

In [1], Baby et al. present the technique of Privacy Preserving Distributed association rule mining (PPRADM) using Game theory to prevent the problem of collusion. Collusion is the situation where more than one player combines together in order to expose the private information of other players. Thus, using PPRADM, the Nash

Equilibrium for the game can be achieved.

In [2], Xu et al. present the discussion of the trade-off between data utility and privacy preservation and how game theory can be used to complete this trade-off. The game described in [2] consists of the following players:

- Data Providers: Data Providers consist of those individuals whose data is made available to the data collectors. The data providers get an incentive from the data collectors in return of providing private information to the data collector.
- Data Collector: The role of the data collector is to collect data from the data providers, apply Privacy Preserving Data Publishing (PPDP) on the obtained data and then finally providing this modified data to the data user.
- Data User: Data user performs the role of data mining on the modified data set. The data user obtains this data set by providing incentives to Data Collector. On the other hand, the Data User gets important information from mining the database, which is the source of income for the Data User.

The trade-off between data utility and privacy preservation can be described as follows:

The profit made by a data user is by performing data mining on the data offered by the data collector. The data user would prefer to have less anonymization so that more relevant data can be extracted. Contrarily, the data providers would want their private data to remain secret. As a result of this, the data providers would want high anonymization of their data. Hence, it is the responsibility of the data collector to decide how much anonymization of the data should take place. In [2], a sequential game model is used to describe the interactions between data user and data collector followed by backward induction in order to attain the Subgame perfect Nash Equilibrium.

In [3], Xu et al. discuss the concept of trade, which is defined as the idea of the data owner willingly giving private information to the data collector in return of money or other incentives, and how game theory can be used to obtain an agreement between the parties involved in this trade. Two different approaches by different groups have been discussed. One research group formalizes this problem in the form of a static game. However, it is observed that in the static game model, there is a possibility of collusion. Thus, in order to prevent the problem of collusion, the concept of cheap talk has been proposed by this research group. Another research group approaches this problem as a sequential game. In this game model, the data collector would anonymize the data so as to preserve data providers' privacy. The game model is used in order to obtain the Nash equilibrium which would help determine the level of anonymization to be applied on the data such that there is a consensus between both the data collector as well as the data user.

In [4], Kumari et al. propose a new model of game theory for privacy game which is based on Cooperative Game Theory. The model presents the idea of Cooperative Privacy (CoP), which is defined as the idea that each player preserves his own privacy, which would eventually contribute to preserving the privacy of other players. Further, this model divides the data set into groups, which are also known as coalitions. Finally, CoP is achieved in terms of Nash Equilibrium after implementing anonymization techniques.

3.2 Network Security

There have been various types of network attacks such as Browser attacks, Denial-of-Service (DDoS) attacks, Worm attacks, and Malware attacks to name a few [5]. In order to tackle such network attacks, several game theory techniques have been suggested. One of the key advantages of approaching the problem with Game theory is that it presents all the scenarios of the problem before deciding the appropriate course of action. [5]

In [11], Wang et al. introduce the concept of Honeypot and how several Bayesian-Nash Equilibria can be used in order to solve the problems caused by Distributed Denial of Service attacks in the Smart Grid using Bayesian Honeypot Game Models.

In [6], Durkota et al. introduce the use of a Stackelberg model for the network hardening problem. In this game-theoretic model, the defender initially adds honeypots in the network. The model further assumes that the attacker has knowledge about the number and types of honeypots added to the network. When the attacker attacks any honeypot, the defender would successfully detect the attacker as a result of which the attack would be terminated. The losses for the defender according to this model include the cost of installation of the honeypots and the total expected loss in the case of an attack. This model aims to achieve a Stackelberg equilibrium which would ensure the minimum loss of the defender under the assumption that the attacks are optimal in nature.

In [5], Liang et al. provide a classification of game models and how each model can be employed to provide better solutions for network security. The authors have classified the game models into two major categories:

- Cooperative Game Models: In the cooperative game model, there is a positive and negative influence matrix. The basic requirement in a coalition is that the positive effect increases and the negative effect decreases between the two divisions in comparison to a situation where there was no coalition. A fascinating inference from the cooperative game is that that to form a new coalition from two coalitions consisting of multiple divisions each, the coalition per unit fraction should be below a previously agreed value. [5]
- Non-Cooperative game Models
    - Static Games: The static games are classified into two categories- static games of complete information and static games of incomplete information. It is to be noted that all static games have imperfect information. Only those situations where attackers interact with defenders are considered in Static Games of Complete Information. The Nash Equilibrium represents the outcome of this model. On the contrary, interactions between regular users and defenders are observed in static games of incomplete information. The solution for such models is obtaining the Bayesian Nash Equilibrium.[5]
    - Dynamic Games: The Stackelberg Network intrusion detection game uses the techniques of Dynamic games of complete and perfect information in order to achieve the Stackelberg Equilibrium. Dynamic Games of Complete and Imperfect Information are used to develop the most favourable policy for the administrator to reduce the potential dangers in a network. The final result obtained after using this game would be the Nash equilibrium for the problem. Dynamic Games of incomplete and perfect information are used to obtain the Perfect Bayesian Equilibrium for Intrusion Detection in Ad Hoc wireless network. A Perfect Bayesian Equilibrium is obtained when Dynamic Games of Incomplete and Imperfect Information is used for Two-player Multi-stage Bayesian game. [5]

3.2.1 Intrusion Detection and Resource Optimization

Intrusion Detection System (IDS) are those systems that help recognize any malicious or anonymous activity. The use of IDS became more prominent as the network attacks became more sophisticated and complex in nature. One of the aspects of IDS is to understand the attacker's strategy. Game Theory provides a framework in order to study such a strategy [10].

In [7], Kiennert et al. describe how different game-theoretic approaches can be used in IDS optimization. The authors have classified IDS optimization into 3 areas - Resource Allocation Optimization, IDS Configuration Optimization, and Countermeasure Optimization.

For the Resource Allocation Optimization, various versions of static games are proposed to solve problems such as optimization of network link sampling, optimization of cluster defence strategy in sensor networks, optimization of resource sharing between nodes, etc. The static game like Zero-sum static game, N-player nonzero-sum static game, and Nonzero-sum static game are proposed. Also, zero-sum stochastic game can be used for resource allocation optimization.

The problems which are approached in IDS configuration Optimization using Game Theory include Configuration of IDS sensitivity, IDS reaction optimization to user interactions and optimization of survivability and attack mitigation in Wireless Sensor Network. These approaches include game theory models like Zero-sum static game, N+M player non-zero sum stochastic game, Zero-sum stochastic game, Nonzero-sum game with dynamic information Incomplete information nonzero-sum stochastic game and Repeated game.

The IDS optimization problems in Countermeasure Optimization include Optimization of Unavailability time of Network Nodes, Computation of the optimal response in multi-stage attacks, computation of optimal countermeasures in a Wireless Sensor Network, etc. The games proposed to solve these problems are MDP, Nonzero-sum stochastic game, incomplete information nonzero-sum sequential game and incomplete information zero-sum sequential game.

4. Comparative Study of Uses of Different Game Theoretic Models

Table 1. provides a concise representation of how various game theoretic models are used in different

problems of privacy preservation and network security.

Table 1. - Summary of game theoretic approaches in Privacy Preservation in Data Mining and Network Security

Application	Model Used	Problem Statement	Solution	Limitations
Privacy Preservation	1.PPRADM Using Game Theory	Collusion Prevention in Privacy Preservation	Nash Equilibrium[1]	Enhancing game theoretic solutions for specific PPDP techniques[2] Formulation of real-world problems using realistic manners so that more practical solutions can be obtained through Game Theory[1]
	2.Sequential Game Model	Trade-off resolution between data utility and privacy preservation	Subgame Perfect Nash Equilibrium [2]	
	3.Static Game	Agreement between parties with respect to the trade of incentives and private information.	Nash Equilibrium [3]	
	4.Sequential Game			
	5. Cooperative Game	Privacy Preservation in Data Publishing using the concept of Cooperative Privacy.	Nash Equilibrium [4]	
Network Security	1. Stackelberg Model	Network Hardening Problem.	Stackelberg equilibrium[6]	Not enough game theoretic models for three or more players[5] The Stochastic Game Model needs to be made more realistic[5]
	2. Bayesian Honeypot Game Model.	Honeypot Game Model For Distributed Denial Of Service in Smart Grid.	Bayesian-Nash Equilibrium[11]	
	3. Static games of complete information	Risk assessment of a network, Information warfare	Nash Equilibrium, Mixed Strategy Nash Equilibrium.[5]	
	4. Static games of incomplete information	Information security game between a rational expert and multiple short-sighted agents.	Nash Equilibrium[5]	
	5.Dynamic Games of complete and perfect information	Stackelberg network intrusion detection game	Stackelberg Equilibrium[5]	
	6.Dynamic Games of complete and imperfect information	To determine the best strategies for the administrator to diffuse the potential dangers in a network.	Nash Equilibrium[5]	
	7 .Dynamic Games of incomplete and perfect information.	Intrusion detection in Ad hoc wireless network	Perfect Bayesian Equilibrium[5]	
	8. Dynamic Games of incomplete and imperfect information.	Two-player Multi-stage Bayesian game	Perfect Bayesian Equilibrium[5]	
	9. Zero-sum static game, N-player nonzero-sum static game, Nonzero-sum static game, Zero-sum stochastic game.	Resource Allocation Optimization	Bayesian Nash Equilibrium, Nash Equilibrium[7]	Better techniques for Integration of the game theoretic models into real world scenarios are required[7]
	10. Zero-sum static game, N+M player non-zero sum stochastic game, Zero-sum stochastic game, Nonzero-sum game	IDS configuration Optimization	Quantal Response Equilibrium, Markov Perfect Equilibrium[7]	

with dynamic information Incomplete information nonzero-sum stochastic game, Repeated game.		
11. MDP, Nonzero-sum stochastic game, Incomplete information nonzero-sum sequential game, Incomplete information zero-sum sequential game.	IDS optimization problems in Countermeasure Optimization	Nash Equilibrium[7]

5. Conclusion and Future Work

Privacy Preservation is one of the most researched problems in recent times. Also, traditional methods for network security have not been largely successful with an increase in complex and sophisticated attacks. As a result, new approaches have to be used in order to find novel solutions to these problems. One of the approaches is using various models of Game Theory. This paper presents various models that are used to solve various problems in Privacy Preservation and Network Security and the solution obtained from these models. As future work, one of the enhancements would be designing game models with mixed strategies rather than pure strategies to encompass the benefits of the various strategies used.

The research gaps observed in game theory is the lack of implementation of mixed game models. The mathematical framework for the same has either not been developed or it is at its infancy. The development of games with mixed strategies would help in rectifying the drawbacks of several game theories with pure strategies. This is because where one particular strategy might fall short, another strategy might be able to address these shortcomings. Thus, the scope for research lies in the development of the mathematical framework of such games.

References

[1] Asha Baby, Anoop Jose and Jisha C.T. (2015) “Analysis of Game Theoretic Approach in Data Mining Security.” *International Journal of Innovations & Advancement in Computer Science (IJIACS).*

[2]L. Xu, C. Jiang, J. Wang, Y. Ren, J. Yuan and M. Guizani. (2015) "Game theoretic data privacy preservation: Equilibrium and pricing." *2015 IEEE International Conference on Communications (ICC)*: 7071-7076.

[3]L. Xu, C. Jiang, Y. Chen, J. Wang and Y. Ren. (2016) "A Framework for Categorizing and Applying Privacy-Preservation Techniques in Big Data Mining." in *Computer*, **49(2)**: 54-62.

[4] Kumari V, Chakravarthy S. (2016) “Cooperative privacy game: a novel strategy for preserving privacy in data publishing.” *Humancentric Comput Inf Sci* **6(1)**:12.

[5]X. Liang and Y. Xiao. (2013) “Game theory for network security.” IEEE Commun. Surveys Tuts. **15(1)**: 472–486.

[6]. K. Durkota et al. (2015) “Optimal Network Security Hardening Using Attack Graph Games.” *Proc. 24th Int’l Joint Conf. Artificial Intelligence*: 526–532.

[7]Christophe Kiennert, Ziad Ismail, Herve Debar, and Jean Leneutre. (2018) “A Survey on Game-Theoretic Approaches for Intrusion Detection and Response Optimization.” *ACM Comput. Surv.* **51(5).**

[8]A.R. Karlin, Y. Peres (2017) “Game Theory, Alive.” *American Mathematical Society*.

[9]M. J. Osborne. (2013) “An Introduction to Game Theory.” *Oxford University Press*.

[10] B. Subba, S. Biswas, and S. Karmakar. (2015) “Intrusion detection in mobile ad-hoc networks: Bayesian game formulation.” *Engineering Science and Technology, an International Journal.*

[11]K. Wang, M. Du, S. Maharjan and Y. Sun. (2017) "Strategic Honeypot Game Model for Distributed Denial of Service Attacks in the Smart Grid." in *IEEE Transactions on Smart Grid* **8(5)**:2474-2482.