Study of Agent Cooperation Incentive Strategy Based on Game Theory in Multi-Agent System

Shuyin Wang[1](✉) and Limei Jiang[2]

[1] Department of Computer Science and Technology, Heilongjiang Institute of Technology, Harbin 150001, China
shuyinwang616@sina.com

[2] Department of Computer, North China Electric Power University, Baoding 071003, China
qhdllmh@163.com

Abstract. The "Folk Theorem" in Game Theory proves that in the process of one-stage game, the game player with selfish behavior may produce cooperative behavior in the process of repeated game. This paper mainly studies the cooperative behavior of Agents in Multi-Agent System from the perspective of Game Theory, against the problem of the "Prisoner's Dilemma" caused by malicious Agents in game, this paper analyzes the influence of three different incentive strategies on trust cooperation behavior of Agents in the process of repeated game, and then gives the equilibrium boundary conditions and corresponding inferences under the three incentive strategies which make the game equilibrium happen. It is found that the incentive strategy can motivate malicious Agents to participate in cooperation in the process of infinite repeated games. Finally, an experiment is carried out to verify.

Keywords: Multi-Agent System · Cooperation game · "Prisoner's Dilemma" Incentive strategies · TFT Strategy

1 Introduction

In the Multi-Agent System, the choice of strategy when Agents cooperate with each other is not only a simple decision process, but also a complex strategy game. Because the inherent rationality of Agents is to maximize their utility, in the absence of effective incentive mechanism, malicious entities in the system tend to be more selfish, and choose the "non-cooperate" strategy as the preferred strategy. The inherent rationality of malicious entities seriously affects the overall revenue of the Multi-Agent System. In order to improve the overall revenue of the system, it is necessary to introduce an effective incentive mechanism to motivate the malicious entities to choose the "co-operate" strategy as the preferred strategy in the process of cooperation, which makes the entities more active. This paper mainly studies the cooperative behavior of Agents in Multi-Agent System from the perspective of Game Theory: Firstly the model of "Prisoner's Dilemma" is presented when the Agents are in one-stage cooperation game, and then according to the "Folk Theorem", the incentive strategies are put forward during infinitely repeated "Prisoner's Dilemma", and thirdly each incentive strategy

Q. Liang et al. (Eds.): CSPS 2017, LNEE 463, pp. 1871–1878, 2019.
https://doi.org/10.1007/978-981-10-6571-2_227

and corresponding equilibrium boundary conditions are analyzed and concluded, and then verified by experiments.

2 "Prisoner's Dilemma" in One-Stage Cooperation Game

Definition 2.1. The cooperation game model for Agents in Multi-Agent System can be defined as a quadruple $B = (I, T, A, u)$

1. the player in game are all the Agents in Multi-Agent System, expressed as $I = \{i, j\}$;
2. the type space of player i is $T_i = \left\{t_i^n, t_i^s\right\}$, t_i^n expresses that player i is a good behavior server provider; t_i^s expresses that player i is a malicious server provider; the type space of player j is $T_j = \left\{t_j^n, t_j^s\right\}$, t_j^n expresses that player j is a good behavior service requester; t_j^s expresses that player j is a malicious service requester.
3. the action set of players based on its type can be expressed as $A = \left\{a_i, a_j\right\}$, including $a_i(t_i^n)$, $a_i(t_i^s)$, $a_j(t_j^n)$, $a_j(t_j^s)$. Here, $a_i(t_i^n) = \left\{a_i(t_i^n)_1, a_i(t_i^n)_2\right\}$ $a_i(t_i^n)_1 = \{\text{CO(cooperate)}\}$, $a_i(t_i^n)_2 = \{\text{NO(non-cooperate)}\}$; Similarly $a_i(t_i^s)_1 = \{\text{CO}\}$, $a_i(t_i^s)_2 = \{\text{NO}\}$; $a_j(t_j^n)_1 = \{\text{CO}\}$, $a_j(t_j^n)_2 = \{\text{NO}\}$; $a_j(t_j^s)_1 = \{\text{CO}\}$, $a_j(t_j^s)_2 = \{\text{NO}\}$.
4. the pay-off set of players based on its type can be expressed as $u = \left\{u_i, u_j\right\}$, including $u_i(a_i(t_i^n))$, $u_i(a_i(t_i^s))$, $u_j(a_j(t_j^n))$, $u_j(a_j(t_j^s))$.

For a good behavior game player i (No matter it is a service requester or service provider), if its opponent j in the game is a malicious Agent, the pay-off matrix of both sides in one-stage game can be expressed as in Fig. 1.

		malicious player j	
		CO	NO
player i	CO	$h_i(\eta), h_j(\eta)$	$-g_i(\eta), -C_j(\Delta T_j)+\rho_j$
	NO	$-C_i(\Delta T_i), -g_j(\eta)$	$-C_i(\Delta T_i), -C_j(\Delta T_j)$

Fig. 1. The pay-off of Agents in one-stage game

For the malicious player j, $u_j(a_i(t_i^n)_1,\ a_j(t_j^s)_2) > u_j(a_i(t_i^n)_1,\ a_j(t_j^s)_1)$ && $u_j(a_i(t_i^n)_2,\ a_j(t_j^s)_2) > u_j(a_i(t_i^n)_2,\ a_j(t_j^s)_1)$, to choose "non-cooperate" is better for player j, and then player i get a better choice "non-cooperate", this game model has only one Nash equilibrium: $(a_i(t_i^n)^*,\ a_i(t_j^s)^*) = (\text{NO, NO})$. Although both sides choose “cooperate” strategy at the same time is better than other strategy combination to obtain the pay-off, the rationality of players makes both sides choose “non-cooperate” strategy as the Nash equilibrium in one-stage game. This is a typical “Prisoner’s Dilemma” problem.

For the classic “Prisoner’s Dilemma” model, there is only one Nash equilibrium, that is, both sides choose “non-cooperate” strategy at the same time. But if the game infinitely repeated, both sides will adopt “cooperate” strategy, and the “cooperate” strategy becomes the only Nash equilibrium.

3 Incentive Strategy in Infinitely Repeated “Prisoner’s Dilemma”

In the case of infinite repeated games, the cooperation phenomenon may arise as long as the one-stage game has a Nash equilibrium. This theorem is called “Folk Theorem”. The “Folk Theorem” shows that because the opportunity for retaliation is always present in an infinitely repeated game, the short-term profits resulting from a one-stage “non-cooperate” will be negligible in an infinite repeated game. Therefore, every participant will not seek short-term benefits of “non-cooperate” in order to achieve long-term benefits of “cooperate”, so as to achieve successful cooperation. The reasonable choice for both is to choose the trigger strategy, but not to adhere to the Nash equilibrium of original game, so the Nash equilibrium of “cooperate” will appear in the “Prisoner’s Dilemma”.

4 Incentive Strategy and Corresponding Equilibrium Boundary Conditions

In the Multi-Agent System, the cooperation process of Agents can be regarded as infinite repeated games. We can motivate malicious Agents to participate in cooperation through some appropriate trigger strategies. Based on infinitely repeated game between the Agents, three typical incentive strategies in infinitely repeated “Prisoner’s Dilemma” are expounded and analyzed in this section, and corresponding equilibrium boundary conditions and inference are given.

In order to analysis the incentive strategy, we use the repeated probability game model, assuming the probability to choose “cooperate” strategy for a Agent within a time frame is s. $s = 0$ means the Agent does not participate in any cooperation; $s = 1$ represents the Agent participates in all cooperation. For a malicious player j, the benefits of the game within a t time frame can be expressed as $u_j^t(s_i^t, s_j^t)$: and the total benefits of the repeated game can be expressed as:

$$u_j = \sum_{t=1}^{\infty} \delta^{t-1} u_j^t(s_i^t, s_j^t) \tag{4.1}$$

Here, δ is discount factor.

If all Agents have always chose "cooperate" strategy, the total benefits of the repeated game for malicious player j can be expressed as:

$$u_j(1,1) = h_j * \sum_{t=1}^{\infty} \delta^{t-1} \tag{4.2}$$

Under the incentive strategies of both sides, if the total benefits for malicious Agent j to choose "cooperate" strategy is superior to choosing other strategies during the repeated game, Agent j will select the "cooperate" strategy.

4.1 "Tit for Tat" Strategy

"Tit for Tat" is an effective strategy for repeated "Prisoner's Dilemma". The steps are as follows: the player chooses "cooperate" strategy in the first stage; then in next stage, makes his choice according to the choice of opponent from previous stage, if the opponent choose "non-cooperate", this stage the player will betray; if the opponent choose "cooperate", the cooperation will continue in this stage. In repeated games, the TFT Strategy is as follows:

$$s_i^t = \begin{cases} 1, & t = 1 \\ s_j^{t-1}, & t > 1 \end{cases} \tag{4.3}$$

In the repeated probability game model, if the Agent j unilaterally reduces the probability $s_j^1 = p$ of its "cooperate" strategy at the first time frame, the Agent i will cooperate with Agent j at the same probability $s_i^2 = p$ in the second time frame, and so on. The following stage strategy is shown in Table 1

Table 1. The game strategy under TFT

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	...
s_i^t	1	p	1	p	...
s_j^t	p	1	p	1	...

To facilitate simplification, the benefits of a one-stage cooperation game between Agents in Sect. 2 are expressed as:

$$u_j(a_i(t_i^n)_1, a_j(t_j^s)_1) = h_j = a; u_j(a_i(t_i^n)_1, a_j(t_j^s)_2) = -C_j + \rho_j = b;$$
$$u_j(a_i(t_i^n)_2, a_j(t_j^s)_1) = -g_j = -c; u_j(a_i(t_i^n)_2, a_j(t_j^s)_2) = -C_j = -d.$$

Here, $b > a > |-c| > |-d|$. Under the TFT Strategy, the total benefits of the malicious player j when it deviates from the cooperative behavior is:

$$u_j(1,p) = (p \cdot a + (1-p)b) + (p \cdot a + (1-p)(-c))\delta + (p \cdot a + (1-p)b)\delta^2 +$$
$$(p \cdot a + (1-p)(-c))\delta^3 + \cdots \quad = \frac{(a-b)p + b + \delta((a+c)p - c)}{1-\delta^2} \tag{4.4}$$

Under the TFT Strategy, the condition for the game to reach a cooperative equilibrium state is:

$$E_{TFT} = u_j(1,1) - u_j(1,p) = a * \sum_{t=1}^{\infty} \delta^{t-1} - \frac{(a-b)p + b + \delta((a+c)p - c)}{1-\delta^2} > 0 \\ \Rightarrow \frac{a+c}{b-a} > 1/\delta, 0 < p < 1 \tag{4.5}$$

Corollary 4.1. If the Agents adopt TFT Strategy, and the equilibrium boundary condition $\frac{a+c}{b-a} > 1/\delta$ is established, the malicious Agent j to choose "cooperate" with probability $p = 1$ can obtain more benefits than to choose a lower probability of "cooperate". So the "cooperate" strategy is the best choice for Agent j.

4.2 GT Strategy

The "Grim Trigger" is an incentive strategy which adopts trigger mechanism in repeated game. In GT Strategy, the player i chooses "cooperate" strategy in the first stage, once the opponent j takes the "non-cooperate" strategy so as to touch the trigger condition, the player i will always choose "non-cooperate" strategy in the future to revenge player j. In repeated games, the GT Strategy is as follows:

$$s_i^t = \begin{cases} 1, t = 1 \\ 1, \forall m < t : s_j^m \geq d \ \&\ t > 1 \\ 0, else \end{cases} \tag{4.6}$$

Here, d is the trigger threshold, $0 < d \leq 1$. If in any stage $s_j < d$, the permanent penalty strategy is triggered. $s_j^1 = p$ represents the probability of "cooperate" strategy for player j. (1) if $p \geq d$, player i will choose "cooperate" strategy; (2) if $p < d$, the player i will always choose "non-cooperate" strategy in the future to revenge player j. This procedure is shown in Table 2.

Table 2. The game strategy under GT

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	...
s_i^t	1	0	0	p	...
s_j^t	p	1	0	0	...

Under the GT Strategy, the total benefits of the malicious Agent j when it deviates from the cooperative behavior is:

$$u_j(1,p) = (p \cdot a + (1-p)b) + (-c) \cdot \delta + (-d)\delta^2 + (-d)\delta^3 + \cdots \tag{4.7}$$

Under the GT Strategy, the condition for the game to reach a cooperative equilibrium state is:

$$E_{GT} = u_j(1,1) - u_j(1,p) = a * \sum_{t=1}^{\infty} \delta^{t-1} - (ap - bp + b - c\delta - \tfrac{d\delta^2}{1-\delta}) > 0$$
$$\Rightarrow (a-b)(p-1)(\delta-1) + (a+c)\delta + (c+d)\delta^2 > 0 \qquad (4.8)$$

Corollary 4.2. If the Agents adopt GT Strategy, and the equilibrium boundary condition $(a-b)(p-1)(\delta-1)+(a+c)\delta+(c+d)\delta^2 > 0$ is established, the malicious Agent j to choose "cooperate" with probability $p = 1$ or $p \geq d$ can obtain more benefits than to choose a lower probability $p < d$. So the "cooperate" strategy is the best choice for Agent j.

4.3 OT Strategy

The OT Strategy also adopts trigger mechanism to punish the malicious Agent in repeated game. But the OT Strategy only punishes the opponent at one stage, leaving the possibility of reconstruction cooperation for later stages. Therefore, it is a relatively mild punishment mechanism. In repeated games, the OT Strategy is as follows:

$$s_i^t = \begin{cases} 1, t = 1 \\ 1, s_j^{t-1} \geq d \ \& \ t > 1 \\ 0, else \end{cases} \qquad (4.9)$$

Here, d is the trigger threshold, $0 < d \leq 1$. If in any stage $s_j < d$, the penalty strategy is triggered. $s_j^1 = p$ represents the probability of "cooperate" strategy for player j. (1) if $p \geq d$, player i will choose "cooperate" strategy; (2) if $p < d$, the player i will choose "non-cooperate" strategy in the next stage to revenge player j. But the penalty only lasts some stages, and then restores the cooperation after the penalty. This procedure is shown in Table 3.

Table 3. The game strategy under OT

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	...
s_i^t	1	0	1	0	...
s_j^t	p	1	0	1	...

Under the OT Strategy, the total benefits of the malicious Agent j when it deviates from the cooperative behavior is:

$$u_j(1,p) = (p \cdot a + (1-p)b) + (-c) \cdot \delta + b\delta^2 + (-c)\delta^3 + b\delta^4 + (-c)\delta^5 \cdots \qquad (4.10)$$

Under the OT Strategy, the condition for the game to reach a cooperative equilibrium state is:

$$E_{GT} = u_j(1,1) - u_j(1,p) = a * \sum_{t=1}^{\infty} \delta^{t-1} - (ap - bp + b + (b\delta^2 - c\delta)\frac{1}{1-\delta^2}) > 0$$
$$\Rightarrow (ap - bp + c)\delta^2 + (a + b - c)\delta + (a - b)(1 - p) > 0 \tag{4.11}$$

Corollary 4.3. If the Agents adopt OT Strategy, and the equilibrium boundary condition $(ap - bp + c)\delta^2 + (a + b - c)\delta + (a - b)(1 - p) > 0$ is established, the malicious Agent j to choose "cooperate" with probability $p = 1$ or $p \geq d$ can obtain more benefits than to choose a lower probability $p < d$. So the "cooperate" strategy is the best choice for Agent j.

5 Experimental Verification

In order to verify the influence of the three incentive strategies, this paper uses Matlab to simulate the network simulation environment. 500 entities are set up in the network, and the experimental indexes are the cooperative success rate of nodes, and the values of the related parameters in the given payoff matrix are $h_i = h_j = a = 6$, $g_i = g_j = c = 4$, $C_i = C_j = d = 2$, $-C_j + \rho_j = b = 8$, $\delta = 0.6$, $p = 0.5$. Figure 2 shows the simulation results of the cooperative success rate with the change amount of malicious entities under the three strategies. When the incentive strategy is adopted, the malicious entity will choose "cooperate" strategy under the pressure of future punishment. As can be seen, TFT Strategy is the best to restrain malicious entities, followed by OT Strategy, while the GT Strategy is worst to restrain malicious entities.

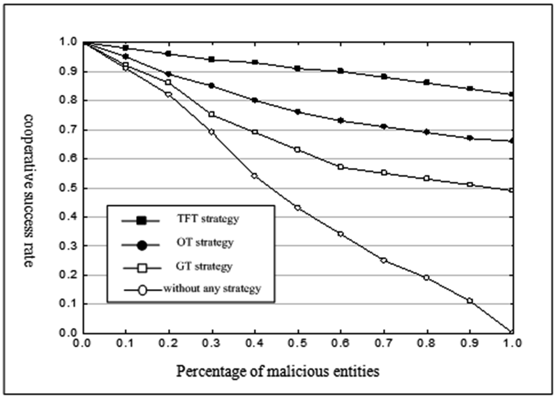

Fig. 2. The change of the success rate of cooperation under three strategies

6 Conclusion

Using appropriate incentive strategies can motivate malicious Agents to participate in cooperation in the process of infinite repeated games, which can improve the cooperative success rate of the overall system. Experiments show that the TFT Strategy has the best effect to restrain malicious entities, and can effectively promote the cooperation among entities.

References

1. Mailath, G., Samuelson, L.: Repeated Games and Reputations: Long-run Relationships. Oxford University Press, New York (2006)
2. Sim, K.-B., Kim, J.-Y., Lee, D.-W.: Game theory based coevolutionary algorithm: a new computational coevolutionary approach. Int. J. Control Autom. Syst. **2**(4), 463–474 (2004)
3. Tisue, S., Wilensky, U.: NetLogo: design and implementation of a multi-agent modeling environment. In: Wolff, K.E., Pfeiffer, H.D. (eds.) Proceeding of the 12th International Conference on Conceptual Structures. LNCS, vol. 3127, pp. 16–21. Springer (2004)
4. Axelrod, R., Hamilton, W.D.: The evolution of cooperation. Science **211**, 1390–1396 (1981)