Achieving Sustainable Cooperation in Generalized Prisoner's Dilemma with Observation Errors*

(Extended Abstract)

[1]Fuuki Shigenaka, [1]Shun Yamamoto, [1]Motohide Seki,
[2]Tadashi Sekiguchi, [3]Atsushi Iwasaki, and [1]Makoto Yokoo
1: Kyushu University, {shigenaka@agent., syamamoto@agent., seki@, yokoo@}inf.kyushu-u.ac.jp
2: Kyoto University, sekiguchi@kier.kyoto-u.ac.jp
3: University of Electro-Communications, iwasaki@is.uec.ac.jp

ABSTRACT

A repeated game is a formal model for analyzing cooperation in long-term relationships. The case where each player observes her opponent's action with some observation errors (imperfect private monitoring) is difficult to analyze, and existing works show that cooperative relations can be sustainable only in ideal situations. We deal with a generic problem that can model both the prisoner's dilemma and the team production problem. We examine a situation with an additional action that is dominated by another action. By adding this seemingly irrelevant action, players can achieve sustainable cooperative relations far beyond the ideal situations. Moreover, for a two-player case, the obtained welfare matches a theoretical upper bound.

Keywords

Game theory, Repeated games, Private monitoring, Prisoner's dilemma, Belief-free equilibrium

1. INTRODUCTION

A repeated game, where players repeatedly play the same stage game over an infinite time horizon, is a formal model for analyzing cooperation in long-term relationships and has received considerable attention in MAS and economics literature. The case of perfect monitoring, where each player can observe other players' actions, is now well understood. There is also a large body of literature on the *imperfect monitoring* case, where players' actions are only imperfectly observed through some signals. Such imperfect monitoring cases are further classified into *public* and *private monitoring* cases. If *all* players observe the same set of signals that imperfectly indicate players' actions, we have an *imperfect public monitoring* case. In contrast, suppose that each player observes her opponent's action with some observation errors. Assume that each player chooses cooperation (C) or defection (D), and a signal, which determines a player's outcome, can be either good (g) or bad (b). If the opponent plays C, a player usually observes g, but she may observe b with a small probability. An important feature of this model is that a player's observation is her private information that is not known to the opponent. This is an example of repeated games with *imperfect private monitoring*, where each player privately receives signals about the actions of other players. Then, players do not share a common understanding about whom to punish and when punishment should start (and end). Hence constructing effective punishment, which sustains cooperation and is voluntarily followed by players, is substantially more difficult than in the public monitoring case. Verifying an equilibrium becomes hard since we need to check that no player has an incentive to deviate under any possible belief she might have on the past histories of other players.

On the other hand, a special type of an equilibrium called *belief-free* equilibrium is identified, where checking whether a profile of strategies forms such an equilibrium is more tractable [1, 3]. However, these existing works show that cooperative relations can be sustainable only in ideal cases where the discount factor (δ) is close to 1 and/or the observation error rate (ϵ) is close to 0.

We deal with a generic problem that can model both the repeated Prisoner's Dilemma (PD) game and the team production problem (in which alternating (C, D) and (D, C) maximizes players' welfare). Furthermore, we introduce an additional action that we call C' as well as an associated observation g' for this action. This action is dominated by another action, and playing it decreases the players' total welfare. Thus, adding it is irrelevant in a one-shot game. To our surprise, it turns out that by adding this action, players can achieve sustainable cooperative relations far beyond the ideal cases identified in existing works. More specifically, we identify a class of strategies called one-shot punishment strategy that constitutes a belief-free equilibrium in a wide range of δ and ϵ. Moreover, when the number of players is two, we show that the sum of the discounted average payoffs achieved by the one-shot punishment strategies is actually theoretically *optimal*, i.e., it matches a theoretical upper bound for any belief-free equilibrium.

2. TWO-PLAYER MODEL

Let us explain only the case with two players due to space limitations, though our results have been generalized to cases

*This work was partially supported by JSPS KAKENHI Grant Number 24220003 and 26280081.

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

Table 1: Stage game payoff

	$a_2 = C$	$a_2 = D$	$a_2 = C'$
$a_1 = C$	1	$-y$	$1-\alpha$
$a_1 = D$	$1+x$	0	$1+x-\alpha$
$a_1 = C'$	1	$-y$	$1-\alpha$

with an arbitrary number of players. There exists a set of players $N = \{1, 2\}$. Two players repeatedly play the same stage game over an infinite horizon $t = 0, 1, 2, \ldots$. In each period, player i takes an action $a_i \in A = \{C, D, C'\}$, and her expected payoff in that period is given by stage game payoff function $u_i(\boldsymbol{a})$, where $\boldsymbol{a} = (a_1, a_2) \in A^2$ is an action profile in that period.

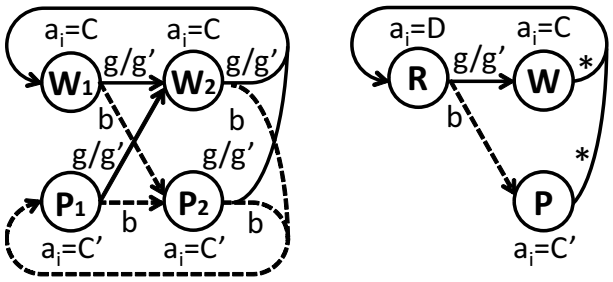
The stage game payoff is shown in Table 1. Here, we only show player 1's payoff since the game is symmetric. We assume x, y, and $\alpha > 0$. Here, the payoff for playing C' is identical to C for the player who plays it. On the other hand, the player can hurt the other player (by α) by playing C' instead of C. Since action C' is dominated by D, adding it is irrelevant when the stage game is played only once. When $x < y + 1$, this stage game corresponds to the well-known PD game, where (C, C) is the outcome that maximizes the total payoff of two players. When $x > y+1$, this stage game corresponds to the team production problem [2], where the outcome that maximizes the total payoff of two players is either (C, D) or (D, C).

Within each period, player i observes her private signal $\omega_i \in \Omega = \{g, b, g'\}$ that is related to the opponent's action. Here, g (or b, g') is the "correct" signal for C (or D, C'). Let a_{-i} denote the opponent's action, and $o(\omega_i \mid a_{-i})$ denote the marginal distribution of ω_i given a_{-i}. We set $o(\omega_i \mid a_{-i})$ to $1 - 2\epsilon$ when ω_i is the correct signal for a_{-i}, and otherwise to ϵ. We assume $0 < \epsilon < 1/3$, i.e., a correct signal is more likely to be observed. Player i's *realized* payoff is denoted as $\pi_i(a_i, \omega_i)$. Hence, her expected payoff is given by $\sum_{\omega_i \in \Omega} \pi_i(a_i, \omega_i) \cdot o(\omega_i \mid a_{-i})$. This formulation ensures that realized payoff π_i conveys no more information than a_i and ω_i. Player i's expected discounted payoff from a sequence of action profiles $\boldsymbol{a}^0, \boldsymbol{a}^1, \ldots$ is $\sum_{t=0}^{\infty} \delta^t u_i(\boldsymbol{a}^t)$, with discount factor $\delta \in (0, 1)$. The (expected) discounted *average payoff* (payoff per period) is defined as $(1-\delta)\sum_{t=0}^{\infty} \delta^t u_i(\boldsymbol{a}^t)$.

3. ONE-SHOT PUNISHMENT STRATEGY

We define *one-shot punishment* strategies. The strategy profile $\boldsymbol{\sigma}^{L^*}$ is defined by an FSA (Finite-State Automaton). Here, $L^* \in \{1, 2\}$ represents the number of players who are supposed to play C in each period. Each player basically follows a prescribed cycle of two states. When $L^* = 2$, the actions of both states are C. When $L^* = 1$, the action of one prescribed state is C, and the action of the other state is D.

Figure 1(a) shows an FSA where $L^* = 2$. Player 1 starts from W_1 and player 2 starts from W_2. Here, the upper-side cycle of W_1 and W_2 is the prescribed cycle. When player 1 is at W_1, player 2 should be at W_2 (or P_2) and should play C (or C'). If player 1 observes b, she punishes player 2 in the next period by moving to P_2. Then player 2 should be at W_1 (or P_1). If player 1 observes b again, she punishes player 2 again in the next period by moving to P_1. Otherwise, she returns to the prescribed cycle, i.e., moves to W_1.



(a) $L^* = 2$ (b) $L^* = 1$

Figure 1: One-shot punishment strategy

Figure 1(b) shows an FSA where $L^* = 1$. Here, the prescribed cycle has two states: W and R. Therefore, only one punishment state P exists.

THEOREM 1. *A one-shot punishment strategy profile $\boldsymbol{\sigma}^{L^*}$ forms a belief-free equilibrium if and only if the following Inequality* (1) *holds:*

$$\alpha \geq w/[\delta(1 - 3\epsilon)]. \tag{1}$$

Where $w = x$ for $L^* = 2$, and $w = y$ for $L^* = 1$.

Note that we can extend the one-shot punishment strategy for cases with an arbitrary number of players and have successfully identified the equilibrium condition.

Let $E_N(\boldsymbol{s})$ denote the sum of all the players' discounted average payoffs given by a strategy profile $\boldsymbol{s}$. Let $E_N^*(\boldsymbol{\sigma}^{L^*})$ denote the maximum of E_N by varying α in the range where Inequality (1) is satisfied. This value is actually optimal for any belief-free equilibria; the following theorem holds.

THEOREM 2. *For any given δ, x, y, and ϵ, if a profile of strategies $\boldsymbol{s}$ forms a belief-free equilibrium, then the following Inequality* (2) *holds:*

$$E_N(\boldsymbol{s}) \leq \max(E_N^*(\boldsymbol{\sigma}^2), E_N^*(\boldsymbol{\sigma}^1), 0). \tag{2}$$

4. CONCLUSIONS

We introduced a generic problem that can model both the repeated PD game and the team production problem, and examined a situation where seemingly irrelevant action C' is added. We identified one-shot punishment strategy, which can constitute a belief-free equilibrium in a wide range of parameters. Moreover, when the number of players is two, we showed that the obtained welfare of this equilibrium matches the theoretical upper bound.

Our future works include applying a similar idea to different settings, e.g., C' requires an additional cost, a player can choose the level of punishment, and so on.

REFERENCES

[1] J. C. Ely and J. Välimäki. A robust folk theorem for the prisoner's dilemma. *Journal of Economic Theory*, 102(1):84–105, 2002.

[2] H. Kobayashi, K. Ohta, and T. Sekiguchi. Repeated partnerships with decreasing returns. Public Economics Seminar, Keio University, October 2014.

[3] M. Piccione. The repeated prisoner's dilemma with imperfect private monitoring. *Journal of Economic Theory*, 102(1):70–83, 2002.