A review of iterated prisoner's dilemma strategies

Marko Jurišić, Dragutin Kermek and Mladen Konecki
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{mjurisic, dkermek, mlkoneck}@foi.hr

***Abstract*—The iterated prisoner's dilemma game is a widely used tool for modelling and formalization of complex interactions within groups. Every player tries to find the best strategy which would maximize long-term payoffs. Tournaments were organized to determine whether there is a single best stable strategy.**

This paper presents a summary of tournaments held in 1980, 2004 and 2005, reviews strategies which were presented during the last 30 years, both in tournaments and in scientific literature and outlines current issues and trends.

I. Introduction

Game theory and more specifically prisoner's dilemma allow the formalization of the interaction within the group. The prisoner's dilemma is an elegant way of modelling the problem of cooperation between participants who have opposing interests. Modelling the interaction with the prisoners dilemma is applied to many fields, in politics [1], biology [2], economy [3] etc. There are applications in computer systems in determining the best route for packets that traverse the network [4], or better allocation of network resources with BitTorrent protocol [5].

The basic form of the game takes place in Chicago, where the district attorney knows that the two gangsters are guilty of some major crime, but he can not convict anybody without a confession. Both individuals are arrested and brought to interrogation separately and offered the following deal: each has a possibility to confess the crime to obtain a minor sentence, or accuse the other gangster to get away without a sentence [6]:

> If the first suspect accuses the other prisoner and the other suspect confesses his guilt, the first suspect is free (his reward is usually denoted by T - the temptation) and the other suspect gets the maximum sentence (marked with S - suckers payoff). In case that both suspects confess their guilt, they both get minor prison sentence (marked with R - reward for cooperation). If both suspects accuse each other, then both receive a major prison sentence (marked with P - punishment - the punishment for mutual defection).

The game is defined by $T > R > P > S$ and $R < (S + T)/2$ so that the continuous cooperation is better than the alternating cooperation and defection [7].

In a single prisoner's dilemma game, the dominant strategy is defection (accusing the other) because it carries the highest reward regardless of the opponent's strategy [6]. In games that have multiple iterations and where is a high probability that opponents will meet again, cooperative strategies achieve better results, as proved through competitions that were organized by Axelrod (1979) based on strategies of experts in game theory from the field of economics, sociology, political science and mathematics [7]. Similar results were gathered on competitions that were held in years 2004 and 2005.

This paper analyses the strategies that have emerged in competitions. Chapter 2 gives a brief overview of the first competition and chapter 3 summarizes the results of the first competition. Chapter 4 outlines the important strategies that appeared in between the competitions (1979-2004). Chapter 5 summarizes the results of the competitions in years 2004 and 2005. Chapter 6 analyses the open questions on the subject and chapter 7 concludes the paper.

II. First competition

The first competition was organized by Axelrod in the year 1979. It was based on the strategies by experts in game theory from the field of economics, sociology, political science and mathematics. It was a contest of mean 200 rounds with the following values: $S = 0, P = 1, R = 3, T = 5$, as shown in Table 1.

TABLE I Result matrix

Player A \ Player B	Cooperation	Defection
Cooperation	$R = 3$ Reward for mutual cooperation	$S = 0$ Suckers payoff
Defection	$T = 5$ Temptation for defection	$P = 1$ Punishment for mutual defection

The surprising result was that the simplest strategy won the competition. It was sent by professor Anatol Rappoport, called Tit for Tat (TFT) which is based on attempt of cooperation and later on reciprocity - copying the opponent's last move. In the second round of the first competition conducted after the announcement of the results of the first round, the same strategy won again between 63 other strategies including random strategy. The third round was simulating the evolution in a way that successful strategies passed descendants to the next generation based on their success in the previous generation. The same strategy won again.

The main characteristic of strategies that have achieved good results is the tendency to cooperate - nice strategies are the ones that never defect first, or at least not until the last few rounds. The results of the first competition showed that the first 8 strategies were nice strategies and none of the others were nice and there was even a gap in the score between nice strategies and the others [8].

Axelrod [8] analyses the strategies and recognizes the importance of "kingmaker" strategies, which did not have good results themselves but because of them nice strategies had better results and other strategies were punished for defecting (this will be especially important in later competitions with possibility of sending multiple strategies). Two such strategies are DOWNING and GRAASKAMP (in the original paper all strategies except TFT are called by the authors).

DOWNING begins with two defections and later at every step computes the probability of opponent's cooperation after its cooperation and also the probability of opponent's cooperation after its defection. After each move those estimates are updated and strategy chooses the choice that will bring the best long term payoff. If the probability of cooperation and defection is close enough, the strategy concludes that the opponent does not respond and the strategy defects afterwards. The downside is that the first two moves are defections so it automatically loses the possibility of further cooperation with some strategies that severely retaliate on any defection.

Another similar strategy is GRAASKAMP which plays TFT the first 50 moves and then it defects once. The following 5 moves it continues with the TFT, and then it analyses the current development of the game. Defection at 51st round serves to identify whether the strategy is playing against its own twin, TFT or any other known strategy and then it adjusts its moves depending on the opponent. The problem with this strategy is that when it meets an unknown strategy, it assumes that this is a RANDOM strategy and it defects for the rest of the game. This strategy was not nice itself, it did not have good results but it was lowering the results of other strategies that were not nice and TFT had very good results with this strategy.

RANDOM strategy that randomly selects an action had even better results than nice strategies in the individual games, but in the end, nice strategies prevailed. If some strategy wants to get good results against RANDOM, it needs to start with complete defection early on, but the question is whether it really is a RANDOM strategy or is it some unknown strategy that it can not recognize. Complete defection in the beginning of the game may prevent establishment of cooperation later on.

An extreme example of this strategy is the strategy called FRIEDMAN (also known as the GRIM [6], grudger [2], spite [9], spiteful[10]) that cooperates until the first defection and after that it defects until the end of the game. This strategy is very good against RANDOM strategy and against all nice strategies but it had bad results against strategies that defected, especially early on, because there is no way to re-establish cooperation after a defection.

III. Conclusions of the first competition

The first competitions brought a few surprises [8]:

1) Tit for Tat as the simplest rule won the competition
2) Most strategies that tried to improve on TFT had worse results because they attempted to slip occasional unexpected defections which then resulted in chains of mutual defections: "they were too clever"
3) The clearest success factor was a "niceness" - the first 8 strategies were nice which means that they were never the first to defect.
4) It pays to forgive: a strategies that had some mechanism of forgiving defection and restoration of cooperation had better results. Those that forgave earlier and tried to cooperate had better results.
5) Existence of "kingmaker" strategies - the ultimate success of the best strategies is the result of their cooperation with "kingmaker" strategies.
6) Despite that the complicated strategies did not have better results than TFT, it is easy to find a better strategies, for example TFTT - retaliate after two defections or the use of artificial intelligence methods.

Axelrod gives a detailed analysis of the strategies and sets out basic rules for successful strategies [11]:

1) The strategy should not be jealous, it must not attempt to gain more points than the opponent by defection
2) The strategy should not be the first to defect.
3) The strategy must clearly and *quickly* respond on cooperation or defection in a way that consequences are clear and that opponent strategy can adapt
4) The strategy must be simple and clear - it should not be too smart and make assumptions too soon about its opponent. If the strategy is "too smart" it can prematurely declare that the opponent is a

RANDOM strategy and move on with defections that lower the scores of both strategies.

Competitions have given new perspectives on the social and biological developments, especially the results that are related to the evolutionary experiments which showed that cooperation is the only stable option (interpretation of the relationship between figs and wasps or between hydras and algae [7], or the interpretation of sudden and spontaneous truces during the World War I [2]). One of the conclusions is that there is no single evolutionary stable strategy which is proven in [12], stated in [9] - nice strategies always cooperate with each other so that for instance, strategy that only cooperates (ALLC) can penetrate into population that is consisted only of TFT and then become easy prey for any defecting strategy.

Some of the strategies that have emerged in the early competitions were taken as standard and they continue to appear in the literature as a measure of the success of new strategies and those new strategies are often based on existing ones, especially TFT, with modifications to facilitate the establishment of cooperation or exploitation of RANDOM or ALLC opponents.

IV. Strategies that have emerged in between competitions

One of the most famous and later most upgraded strategies that had better results than the TFT is Pavlov [13] or win-stay, lose-shift strategy. Strategy cooperates if and only if both strategies in the previous round had the same action. The name comes from the fact that the strategy exhibits almost a reflex-like adjustment to the result of previous round - repeat previous move if it was rewarded with R or T points and change behaviour if the result was just P or S points. Deficiency of strategy is that is has bad results with the strategies that constantly do defections because every second move tries to cooperate but it has much better results with the strategies that do only cooperation. There is no problem with constant defection if it turns out that the opponent does not reciprocate.

Beaufils and some other researchers reject the last conclusion that Axelrod gave in the book - simplicity. They provide a view where complexity plays a major role [9]. They bring strategy *gradual* which at first behaves like TFT but after each defection reciprocates with another defection: after the first defection it will respond with one defection, after the second defection it will respond with two defections etc. By using genetic algorithms they reached even better strategy that has even better results against predetermined strategies and strategies from the first competition but in the subsequent competitions they didn not show such good results as in their own laboratory [14].

Other strategies are mostly based on aforementioned approaches - the base is always TFT or Pavlov. Examples are adaptable strategy [15], forgiving strategy [10] and winning strategies in subsequent competitions.

There is also interesting research of teams of strategies that collude to increase one member's score. One strategy is set as master and others as slaves, who then work towards the unconditional cooperation with master strategy and unconditional defection towards other strategies so they lower results of other strategies and in that way help their master [16], [17], cited in [18].

Some authors see the simplicity of the original prisoner's dilemma as a limitation and extend the payoff matrix with additional fields [19]. There is also the concept of noise in the communication channel which means that the opponent's response can only be interpreted with certain probability, altering the experiment results and expected behaviour of strategies [20]. In order to take into account these ideas, competitions were repeated in 2004 and 2005. Those competitions included several variants of competition. In the literature, the prisoner's dilemma that was studied by Axelrod in called classical iterated prisoner's dilemma [9] or traditional iterated prisoner's dilemma [21], [22], [23].

V. 2004 and 2005 competitions

Competitions had four variants [24]:

1) Repetition of original Axelrod experiment, to determine whether the TFT is still dominating strategy or there has been found a new better strategy during the period of 25 years in between the competitions.
2) The competition is identical to the previous except the addition of the noise - there is a small probability that the cooperation or defection might be misinterpreted.
3) The competition that allows sending strategies with more players and more choices.
4) The competition that copies the original Axelrod experiment but has the following additional rules:
   - Only one strategy is permitted per competitor, collaboration is prohibited
   - Organizers of the competition will add only a RANDOM strategy as a default strategy
   - Each strategy also competes against its own copy
   - Each competition consists of the same number of moves that is unknown to the authors of the strategies - every event is performed five times using different number of moves
   - Payoff matrix is the same as in the original events (shown in Table 1)
   - Each strategy has to play a move in less than two seconds (the difference from the original competition, added to avoid infinite loops)
   - Each strategy must be accompanied by a description and source code in order to reduce the possibility of collusion

In these competitions, strategies were received via e-mail. It was possible to send a strategy written in the Java programming language using provided ipdlx package, or one could determine the behaviour of the strategy through the web interface. Strategies could use external resources over the internet but they had a time limit to give response which disabled some more complicated approaches such as using genetic algorithms.

The competition in year 2004 consisted of 223 strategies, including 9 default strategies shown in Table 2 (RAND - a random selection, NEG - denies the last action of the opponent, ALLC - always cooperate, ALLD - always defect, TFT, STFT - TFT beginning with defection, TFTT - retaliates only after two defections, GRIM and Pavlov). In the end the winners were the strategies submitted by the University of Southampton who worked in teams. One strategy was team leader and the rest were members of the team. All members cooperated with their leader and they constantly defected to all strategies outside of their team. Strategies that were team leaders had 13% more points than the others, they also won in the competition with the noise because they implemented a way to resolve mistakes [18].

Slany and Kienreich disputed the results of the first competition since from 223 strategies, 112 were sent by the winner and they were not allowed to send as many strategies as they wanted which would make them the winners [25]. The basic approach they had was the same as the winning strategies - team strategies, but the team leader from University of Southampton played defection towards team members and towards others it played standard TFT. Slany and Kienreich modified TFT to deal with the chains of mutual defections which is the main problem of TFT, especially in cases where the noise simulation was added. They also added the ability to recognize and exploit the RANDOM strategy in a way that after an opponent strategy crosses a certain randomness threshold they conclude that the opponent is a RANDOM strategy and change the behavour to act as ALLD. They called their strategy OmegaTFT [25].

TABLE II DEFAULT TYPES OF STRATEGIES

Designation	Description
ALLC	Strategy always plays cooperation
ALLD	Strategy always plays defection
RAND	Strategy has a 50% probability to play cooperation or defection
GRIM	It starts with cooperation, but after the first defection of its opponent continues with defection
TFT	It starts with cooperation and then it copies the moves of the opponent
TFTT	As TFT but defects after two consecutive defections
STFT	As TFT but starts with defection
TTFT	As TFT but for each defection retaliates with two defections
Pavlov	Action results are divided into 2 groups, positive actions are T and R and negative actions are P and S - if the result of previous action belonged to the first group, action is repeated and if the result was in the second group, then the action was changed, it is also called win-stay, lose shift

Competition held in 2005 took into account the results and criticisms and introduced additional restrictions so that each institution could send only one team of 20 strategies in the first two contests. This time there were many more teams and first four places came from different institutions. The winners of the first competition were Slany and Kienreich with their Cosa Nostra Godfather / hitman strategy. In the second competition, which included noise in the data transfer, the winner was again team from University of Southampton with their team of Greek gods [26].

The rules of fourth competition allowed only one strategy per participant and the winning strategy was Adaptive Pavlov (APavlov) that was sent by Jiawei Li. It is based on Pavlov strategy with addition of recognition of the opponent and its classification into one of the default strategy types. Then, the strategy responds to each opponent in an optimal way. If a strategy is unknown to APavlov, it is classified as a RANDOM strategy and it behaves accordingly - it responds with constant defection [27].

VI. CONCLUSIONS OF 2004. AND 2005. COMPETITIONS

Since it has been shown that the collaboration of strategies that work as a team has a key role in the success of individual strategies, one of the major new issues is how to detect and prevent collusion of strategies. Slany and Kienreich showed that it is possible to deceive the organizers. They invented a person, wrote a mail with broken English and submitted strategies that were matched with their main strategies. The results showed that some other completely unrelated strategies accidentally met the conditions so that their main strategy recognized them as their own team member strategies and began to exploit them with unconditional defections [25].

Team from the University of Southampton had focused on communication and coordination between agents in the environment with noise. They implemented their strategies by using Hamming codes that are usually used in information theory in order to eliminate errors, for example, to send teletext via analogue signal [28]. Their strategies have been very successful in competition with noise, while their good result in year 2004 was probably based on great number of strategies that they had sent to work in their favour.

APavlov heuristic strategy that won the 2005 competition, presented by Li in his paper, also had some advantages and disadvantages. It is always a problem if some strategy is too early classified as RANDOM strategy, because there is no way of establishing cooperation later on. It is necessary to constantly develop new rules for new strategies that will appear. The crucial problem is to balance between the need to recognize any possible rival strategy and the need that the strategy remains as simple as possible [27].

VII. Conclusion

Prisoner's dilemma is still a current research area with nearly 15000 papers during the past two years (Source: Google Scholar). New strategies are developed and old ones are reused in new areas. But basic rules for cooperation that were recognized by Axelrod in the first competition are still valid: kindness, provocability, forgiveness and simplicity. Most of new successful strategies are based on principles that were set up 20 or 30 years ago (from 223 strategies at the competition 1 in the year 2004 there were 73 based on TFT principle).

New approaches upgrade known ideas through genetic algorithms and heuristic approaches to successfully recognize opponents, to anticipate their moves and try to achieve better results. But there is always a problem of possibility to misjudge opponent which will bring worse results in the end. However, the information carries the key role in any sort of intelligent activities and strategies. Individuals with more information will have advantage in most of situations so the strategies that learn about the opponents and adjust their own behaviour will certainly have an increasingly important role in the future.

References

[1] J. Conybeare, "Public goods, prisoners' dilemmas and the international political economy," *International Studies Quarterly*, pp. 5–22, 1984.

[2] R. Dawkins, *The selfish gene*. Oxford University Press, USA, 2006.

[3] K. Corfman and D. Lehmann, "The prisoner's dilemma and the role of information in setting advertising budgets," *Journal of Advertising*, pp. 35–48, 1994.

[4] U. Shevade, H. Song, L. Qiu, and Y. Zhang, "Incentive-aware routing in dtns," in *Network Protocols, 2008. ICNP 2008. IEEE International Conference on*, pp. 238–247, IEEE, 2008.

[5] B. Cohen, "Incentives build robustness in bittorrent," in *Workshop on Economics of Peer-to-Peer systems*, vol. 6, pp. 68–72, Citeseer, 2003.

[6] K. Binmore, *Playing for real*. Oxford university press, 2007.

[7] R. Axelrod and W. Hamilton, "The evolution of cooperation," *Science*, vol. 211, no. 4489, p. 1390, 1981.

[8] R. Axelrod, "Effective choice in the prisoner's dilemma," *Journal of Conflict Resolution*, vol. 24, no. 1, p. 3, 1980.

[9] B. Beaufils, J. Delahaye, and P. Mathieu, "Our meeting with gradual: A good strategy for the iterated prisoner's dilemma," *Artificial life five*, vol. 5, p. 202, 1997.

[10] C. O'Riordan, "A forgiving strategy for the iterated prisoner's dilemma," *Journal of Artificial Societies and Social Simulation*, vol. 3, no. 4, pp. 56–58, 2000.

[11] R. Axelrod, *The evolution of cooperation*. Basic Books, 1984.

[12] R. Boyd and J. Lorberbaum, "No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game," 1987.

[13] M. Nowak and K. Sigmund, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game," *Nature*, vol. 364, no. 6432, pp. 56–58, 1993.

[14] J. Humble, "Prisoner's dilemma tournament results," June 2004. URL: http://www.prisoners-dilemma.com/results/cec04/ipd_cec04_full_run.html, pristupano 02.11.2011.

[15] E. Tzafestas, "Toward adaptive cooperative behavior," in *Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior, SAB*, vol. 2, pp. 334–340, Citeseer, 2000.

[16] J. P. Delahaye and P. Mathieu, "L'altruisme perfectionné," *Pour La Science (French Edition of Scientific American)*, vol. 187, pp. 102–107, 1993.

[17] J. P. Delahaye, *Logique, informatique et paradoxes*. Paris, France: Éditions Belin, 1995. Bibliothèque Pour La Science.

[18] A. Rogers, R. Dash, S. Ramchurn, P. Vytelingum, and N. Jennings, "Coordinating team players within a noisy iterated prisoner's dilemma tournament," *Theoretical Computer Science*, vol. 377, no. 1-3, pp. 243–259, 2007.

[19] T. To, "More realism in the prisoner's dilemma," *The Journal of Conflict Resolution*, vol. 32, no. 2, pp. 402–408, 1988.

[20] P. Molander, "The optimal level of generosity in a selfish, uncertain environment," *Journal of Conflict Resolution*, pp. 611–618, 1985.

[21] E. Stanley, D. Ashlock, and L. Tesfatsion, "Iterated prisoner's dilemma with choice and refusal of partners," *Staff General Research Papers*, 1994.

[22] S. Chong and X. Yao, "Self-adapting payoff matrices in repeated interactions," in *Computational Intelligence and Games, 2006 IEEE Symposium on*, pp. 103–110, IEEE.

[23] P. Borges, R. Pacheco, R. Barcia, and S. Khator, "A fuzzy approach to the prisoner's dilemma," *Biosystems*, vol. 41, no. 2, pp. 127–137, 1997.

[24] Y. X. Kendall Graham, Darwen Paul, "Prisoner's dilemma tournament rules," 2003. URL: http://www.prisoners-dilemma.com/competition.html, pristupano 02.11.2011.

[25] W. Slany and W. Kienreich, "On some winning strategies for the iterated prisoner's dilemma, or, mr. nice guy and the cosa nostra," *The iterated prisoners' dilemma: 20 years on*, vol. 4, p. 171, 2007.

[26] J. Humble, "Prisoner's dilemma tournament results," June 2005. URL: http://www.prisoners-dilemma.com/results/cig05/comp%202%20run%2001.html, pristupano 02.11.2011.

[27] J. Li, "How to design a strategy to win an ipd tournament," *The iterated prisoners' dilemma: 20 years on*, vol. 4, p. 89, 2007.

[28] J. Chambers, "Broadcast teletext system," June 29 1982. US Patent 4,337,485.