

# Indirect Reciprocity Game Modelling for Cooperation Stimulation in Cognitive Networks

Yan Chen and K. J. Ray Liu

**Abstract**—In cognitive networks, since nodes generally belong to different authorities and pursue different goals, they will not cooperate with others unless cooperation can improve their own performance. Thus, how to stimulate cooperation among nodes in cognitive networks is very important. However, most of existing game-theoretic cooperation stimulation approaches rely on the assumption that the interactions between any pair of players are long-lasting. When this assumption is not true, according to the well-known Prisoner's Dilemma and the backward induction principle, the unique Nash equilibrium (NE) is to always play non-cooperatively. In this paper, we propose a cooperation stimulation scheme for the scenario where the number of interactions between any pair of players are finite. The proposed algorithm is based on indirect reciprocity game modelling where the key concept is “*I help you not because you have helped me but because you have helped others*”. We formulate the problem of finding the optimal action rule as a Markov Decision Process (MDP) and propose a modified value iteration algorithm to find the optimal action rule. Using the packet forwarding game as an example, we show that with an appropriate cost-to-gain ratio, the strategy of forwarding the number of packets that is equal to the reputation level of the receiver is an evolutionarily stable strategy (ESS). Finally, simulations are shown to verify the efficiency and effectiveness of the proposed algorithm.

**Index Terms**—Indirect reciprocity, cognitive network, game theory, evolutionarily stable strategy, cooperation, markov decision process.

## I. INTRODUCTION

A COGNITIVE network is a network composed of elements that can dynamically adapt to varying network conditions to optimize end-to-end performance through learning and reasoning [1]. In such a network, nodes are intelligent and have the ability to observe, learn, and act to optimize their performance. Since nodes generally belong to different authorities and pursue different goals, fully cooperative behaviors, such as unconditionally forwarding packets for each other, cannot be taken for granted. Instead, nodes will only cooperate with others when cooperation can improve their own performance. We regard the nodes with such behaviors as selfish nodes. Therefore, a key problem in cognitive networks is how to stimulate cooperation among selfish nodes.

In the literature, many schemes have been proposed to stimulate node cooperation for different cognitive networks, such as [2][3] for ad hoc networks and [4][5] for peer-to-peer networks. One way to stimulate cooperation among selfish

nodes is to use payment based methods [6][7]. Although these schemes can achieve promising cooperation stimulation results, the requirement of tamper-proof hardware or central billing services greatly limits their potential applications.

Another way to stimulate cooperation among selfish nodes is to use reputation-based methods with necessary monitoring [8][9][10]. Marti et. al [11] propose a mechanism, called “watchdog”, to identify the misbehaving nodes and another mechanism, called “pathrater”, to deflect the traffics around them. The major drawback of their method is that misbehaving nodes are not punished. Therefore, there is no incentive for the nodes to cooperate. To overcome this problem, Buchegger and Boudec [12] as well as Michiardi and Molva [13] propose reputation-based mechanisms to enforce node cooperation. In both approaches, nodes observe the behavior of each other, store this information locally, and distribute this information in reputation reports. According to their observations, nodes isolate the misbehaving nodes by denying forwarding packets to them. However, there is no theoretical justification about the optimality of such approaches.

Recently, efforts have been made to mathematically analyzing cooperation in cognitive networks using game theory [14][15][16][17]. Srinivasan et al. [18] propose to use generous TIT-FOR-TAT strategy while Urpi et al. [19] propose to use Bayesian games. In [20], Felegyhazi et al. investigate equilibrium conditions of packet forwarding strategies based on game theory and graph theory by taking into account the network topology. In [21], Yu and Liu propose a game theoretic framework to jointly analyze cooperation stimulation and security in autonomous mobile ad hoc networks. Their results show that, for a two-player packet forwarding game, the unique cheat-proof Nash equilibrium for every node is not to help the opponent more than the opponent has helped him/her.

However, most of the existing game theoretical frameworks rely on the assumption that the game between a pair of players is directly played for infinite times. In reality, due to mobility or changes of environment, nodes will periodically update their partners to achieve better performance, which means that any pair of players are supposed to play for only finite times with the termination time are either known or can be estimated by both players. Note that every player can experience infinite times with many players but never always with the same partner. In such a case, according to the well known Prisoner's Dilemma and backward induction principle [22], the only optimal strategy is to always play non-cooperatively. The major reason causing such a non-cooperative optimal strategy is the implicit assumption of direct reciprocity in most games, where the action of a player taking towards his/her opponent

Paper approved by B. Sikdar, the Editor for Wireless Packet Access and Cross-Layer Design of the IEEE Communications Society. Manuscript received March 12, 2010; revised July 13, 2010.

The authors are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA (e-mail: {yan, kjrlu}@umd.edu).

Digital Object Identifier 10.1109/TCOMM.2010.110310.100143

is purely determined by the history of how the opponent treats him/her. Obviously, under such a scenario, all players have no incentive to play cooperatively since their behaviors will not be evaluated by other players except their opponents.

To stimulate the players' incentive to play cooperatively, not only the evaluations from the opponents but also the evaluations from other observers should be taken into account, which leads to the notion of "indirect reciprocity". Indirect reciprocity is a key mechanism for the evolution of human cooperation and has recently drawn a lot of attentions in the area of social science and evolutionary biology [23][24]. The key concept of indirect reciprocity is "*I help you not because you have helped me but because you have helped others*". In this paper, we propose to use the indirect reciprocity game modelling to stimulate cooperation among selfish nodes for the scenario where the number of interactions between any pair of players are finite. The main contributions of this paper are summarized as follows.

- We propose a cooperation stimulation scheme to stimulate cooperation among selfish users in cognitive networks using indirect reciprocity game modelling. Different from the existing game-theoretic approaches, our proposed scheme does not rely on the assumption that the number of interactions between a pair of players are infinite.
- In the proposed scheme, we first develop the concept of reputation distribution to capture not only the mean behavior of the transmitter's reputation but also all likelihoods of the transmitter's reputation that may be. Then, we develop a reputation updating policy for the receiver and observers to update the transmitter's reputation distribution based on the transmitter's previous reputation distribution and his/her action toward the receiver. We also propose a gradient descent algorithm to find the stationary reputation distribution of the whole population for any given optimal action rule.
- In the proposed scheme, we formulate the problem of finding the optimal action rule as a Markov Decision Process (MDP) and proposed a modified value iteration algorithm to find the optimal action rule.
- We show that with an appropriate cost-to-gain ratio, the strategy of forwarding the number of packets that is equal to the reputation level of the receiver is an evolutionarily stable strategy (ESS). We also show that even with only 60 percentage of population adopting the optimal action rule at the beginning, by natural selection, the optimal action rule will quickly spread over the whole population. And once the whole population use the optimal action rule, no one will deviate. Moreover, we find that such an ESS will lead to a "good" society with more than 90 percentage of the population have good reputation.

The rest of this paper is organized as follows. In Section II, we describe the problem formulation and introduce basic components in our system model. Then, we show in details how to find the optimal action rule in Section III. In Section IV, we describe two action spreading algorithms due to natural selection. Finally, we show the simulation results in Section V and draw conclusions in Section VI.

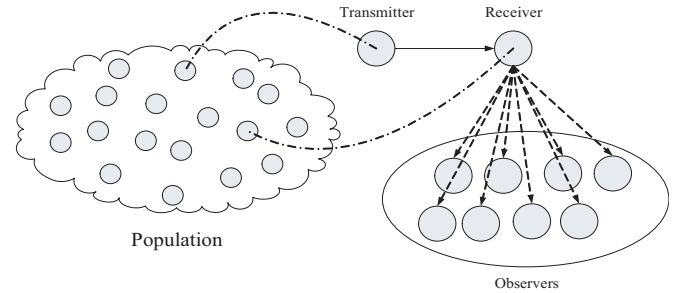


Fig. 1. System model. Within every interaction, a pair of transmitter and receiver is randomly sampled from the population. Then, the transmitter will forward a certain amount of packets to the receiver according to the receiver's and his/her own reputations. After the transmission, the transmitter's reputation will be updated by the receiver and the observers. Finally, the transmitter's reputation is propagated to the whole population from the receiver and the observers through a noisy gossip channel.

## II. THE SYSTEM MODEL

As shown in Figure 1, let us consider a cognitive network with sufficiently large population of nodes. Due to mobility and/or changes of environment, short interactions rather than long-lasting associations between anonymous partners are dominant. At each time slot, a fraction of players is chosen from the population to form pairs to forward packets. Within each pair, one player acts as a transmitter and the other player as a receiver. Let  $\mathbf{A} = \{0, 1, \dots, L\}$  stand for the action set that the transmitter may choose, where the action  $i \in \mathbf{A}$  stands for the transmitter forwards  $i$  packets to the receiver.

In the simplest model with  $L = 1$ , the receiver can obtain a gain  $g$  at a cost  $c$  to the transmitter. We should always assume that the gain  $g$  is greater than the cost  $c$ . Otherwise, no transmission will occur. In such a case, if both players cooperate with each other and forward one packet to the other player, both players receive  $g - c$ , which is better than what they would obtain by both defecting, namely 0. However, a unilateral defector would earn  $g$ , which is the highest payoff, and the exploited cooperator would pay the cost  $c$  without receiving any benefit. The payoff structure yields an instance of the well-known Prisoner's Dilemma game and the unique Nash equilibrium (NE) is defecting, i.e. both players will not forward the packet to the other player. Moreover, with backward deduction, the NE remains the same even the game is played a finite number of times. Such a non-cooperative optimal strategy is mainly because of the use of direct reciprocity, where the action of a transmitter taking towards a receiver is purely determined by the history of how the receiver treats him/her. Obviously, under such a scenario, all transmitters have no incentive to forward packets since their behaviors will not be evaluated by other players except their corresponding receivers.

To stimulate the cooperation under such a scenario, we use the indirect reciprocity game modelling, where the essential concept is: "*I help you not because you have helped me but because you have helped others*". Therefore, a key concept in indirect reciprocity game is the establishment of the notion of reputation, which is the evaluation of the history of the players' action. In this paper, to simplify the analysis, we assume that the reputation is quantized to  $L+1$  levels with "0"

being the worst reputation and “ $L$ ” being the best reputation, i.e., the reputation set can be represented as  $\mathbf{T} = \{0, 1, \dots, L\}$ . However, the results can be easily extended to the case that the reputation set has different size from the action set. Here, we also assume that everyone agrees on the reputation of an individual and no private opinions are allowed. However, errors in assigning reputation are possible. During each interaction, the transmitter determines his action, i.e. how many packets to forward to the receiver, based on the receiver’s and his/her own reputations. After each interaction, the reputation of the receiver remains the same, while the reputation of the transmitter is first updated by the receiver and the observers, and propagated to the whole population through a noisy gossip channel. Then, each participant (including both the transmitter and receiver) goes back to the population with probability  $\delta$  or leaves the population with probability  $1 - \delta$ . The parameter  $\delta$  can be treated as a discounting factor of the future. For every player who leaves the population, a new individual enters with an initial reputation randomly chosen from the reputation set with equal probability  $\frac{1}{L+1}$ .

### A. Social Norms

A social norm,  $\mathbf{Q}$ , is a matrix used for updating the immediate reputation of players, where the immediate reputation is the reputation that a transmitter can immediately obtain by taking an action. Each element  $Q_{i,j}$  in the social norm stands for the immediate reputation assigned to a transmitter who has taken the action  $i$  toward a receiver whose reputation is  $j$ . Without loss of generality, we assume that all players in the population share the same norm. Although the immediate reputation is only determined by the action of the transmitter and the reputation of the receiver, we can see from the later discussion, the final reputation updating rule also involves the reputation of the transmitter.

Since both the cardinalities of the action set and the reputation set are  $L+1$ , there are  $(L+1)^{(L+1) \times (L+1)}$  possible social norms. Based on the intuition that forwarding packets to the receiver with good reputation or denying forwarding packets to the receiver with bad reputation should receive good reputation, in this paper, we define the immediate reputation  $Q_{i,j}$  as follows

$$Q_{i,j} = L - |i - j|, \quad (1)$$

which means that the social norm is

$$\mathbf{Q} = \begin{pmatrix} L & L-1 & \dots & 1 & 0 \\ L-1 & L & \dots & \vdots & 1 \\ \vdots & L-1 & \ddots & L-1 & \vdots \\ 1 & \vdots & \dots & L & L-1 \\ 0 & 1 & \dots & L-1 & L \end{pmatrix}. \quad (2)$$

For the special case when  $L = 1$ , the  $2 \times 2$  social norm can be written as

$$\mathbf{Q}^{2 \times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (3)$$

where “1” stands for good reputation and “0” stands for bad reputation.

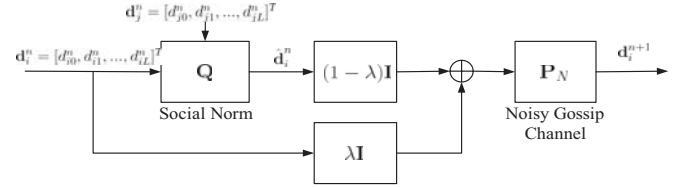


Fig. 2. Reputation updating policy.

With such a social norm shown in (3), we can see that the transmitter can obtain a good immediate reputation by either forwarding packets to the receiver with good reputation or denying forwarding packets to the receiver with bad reputation. On the other hand, the transmitter will obtain a bad immediate reputation if he/she either denies forwarding packets to the receiver with good reputation or forwards packets to the receiver with bad reputation.

### B. Action Rules

An action rule,  $\mathbf{a}$ , is an action table of the transmitter, where the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column element  $a_{i,j}$  stands for the action of the transmitter based on his/her own reputation  $i$  and the corresponding receiver’s reputation  $j$ . Since both the cardinalities of the action set and the reputation set are  $L+1$ , there are  $(L+1)^{(L+1) \times (L+1)}$  possible action rules. The optimal action rule,  $\mathbf{a}^*$ , should be the one that maximizes the payoff function as discussed later.

## III. OPTIMAL ACTION RULE

### A. Reputation Updating Policy

A key concept in indirect reciprocity game is reputation [24]. There is a similar notion of trust [25], however, which is mostly based on direct reciprocity. Players monitor the social interactions within their group and help others establish the reputation of being a helpful player. Therefore, one important step in indirect reciprocity game modelling is how to update reputation based on players’ actions. In this subsection, we develop a reputation updating policy based on the action of the transmitter, the reputation of the transmitter and the reputation of the receiver.

To capture not only the mean behavior of the transmitter’s reputation but also all likelihoods of the transmitter’s reputation that may be, we assign a reputation distribution for each player. Let  $\mathbf{d} = [d_0, d_1, \dots, d_L]^T$  be a reputation distribution for a specific player. Then  $d_i$  stands for the likelihood of the player being assigned with reputation  $i$ .

The proposed reputation updating policy is shown in Fig. 2. Suppose, at time index  $n$ , a transmitter with a reputation distribution  $\mathbf{d}_i^n$  is matched with a receiver with a reputation distribution  $\mathbf{d}_j^n$ . By taking a certain action, the transmitter is assigned with an immediate reputation  $\hat{\mathbf{d}}_i^n$  based on the social norm. Then, the receiver and the observers will update the transmitter’s reputation distribution using a linear combination of the transmitter’s original and immediate reputations, where the weight  $\lambda$  can be treated as a discounting factor of the past reputation. Finally, the transmitter’s reputation is propagated among the population by the receiver and observers through a noisy gossip channel.

In a simple example, we assume that the transmitter's reputation distribution is  $\mathbf{d}_i^n = \mathbf{e}_i$  and the receiver's reputation distribution is  $\mathbf{d}_j^n = \mathbf{e}_j$ , where  $\mathbf{e}_i$  and  $\mathbf{e}_j$  are the standard basis vectors. Let  $a_{i,j}$  be the action the transmitter takes towards the receiver. Then, the immediate reputation of the transmitter is  $\mathbf{e}_{Q_{a_{i,j},j}}$ . According to the reputation updating policy in Fig. 2, after the transmission, the transmitter's reputation distribution becomes

$$\mathbf{d}_i^{n+1} = \mathbf{P}_N \left( \lambda \mathbf{e}_i + (1 - \lambda) \mathbf{e}_{Q_{a_{i,j},j}} \right), \quad (4)$$

where  $\mathbf{P}_N$  is the transition matrix of the noisy channel. Without loss of generality<sup>1</sup>, we define  $\mathbf{P}_N$  as follows

$$\mathbf{P}_N = \begin{pmatrix} 1 - \mu & \mu/L & \dots & \mu/L & \mu/L \\ \mu/L & 1 - \mu & \dots & \vdots & \mu/L \\ \vdots & \mu/L & \ddots & \mu/L & \vdots \\ \mu/L & \vdots & \dots & 1 - \mu & \mu/L \\ \mu/L & \mu/L & \dots & \mu/L & 1 - \mu \end{pmatrix}, \quad (5)$$

with  $\mu \in [0, 0.5]$  being a constant.

The  $\mathbf{d}_i^{n+1}$  in (4) is the updated reputation distribution of the transmitter after the transmitter with an original reputation  $\mathbf{e}_i$  takes an action  $a_{i,j}$  towards the receiver with a reputation  $\mathbf{e}_j$ . Since this updated reputation distribution will be used later in the analysis for finding the optimal action rule, we use a specific symbol  $\tilde{\mathbf{d}}_{i \rightarrow j}$  to denote it, i.e.,

$$\tilde{\mathbf{d}}_{i \rightarrow j} = \mathbf{P}_N \left( \lambda \mathbf{e}_i + (1 - \lambda) \mathbf{e}_{Q_{a_{i,j},j}} \right). \quad (6)$$

For the general case that  $\mathbf{d}_i^n \neq \mathbf{e}_i$  and/or  $\mathbf{d}_j^n \neq \mathbf{e}_j$ , the transmitter's updated reputation distribution cannot be simply expressed using (4) since, given an action rule, different combinations of the transmitter's and receiver's reputations may lead to the same immediate reputation. In such a case, we need to first find the immediate reputation using

$$\hat{\mathbf{d}}_i^n(k) = \sum_p \sum_{q: Q_{a_{p,q},q}=k} \mathbf{d}_i^n(p) \mathbf{d}_j^n(q). \quad (7)$$

Then, according to Fig. 2, the transmitter's updated reputation distribution can be computed by

$$\mathbf{d}_i^{n+1} = \mathbf{P}_N \left( \lambda \mathbf{d}_i^n + (1 - \lambda) \hat{\mathbf{d}}_i^n \right). \quad (8)$$

### B. Stationary Reputation Distribution

Let  $\mathbf{x} = [x_0, x_1, \dots, x_L]^T$  stand for the reputation distribution of the entire population, where  $x_i$  is the portion of the population that have the reputation  $i$ . Since every pair of transmitter and receiver is chosen from the population, given the transmitter with reputation  $i$ , the probability of matching with the receiver with reputation  $k$  is  $x_k$ . After the transmission, the reputation of the transmitter is updated using the policy shown in Fig.2. Therefore, the evolution of  $\mathbf{x}$  can be described by the following differential equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{x}^{new} - \mathbf{x}, \quad (9)$$

<sup>1</sup>Note that the analysis in this paper are also applicable to the  $\mathbf{P}_N$  with other forms.

### Algorithm 1 : Finding Stationary Reputation Distribution Using Gradient Descent

1. Given the optimal action  $a_{i,j}^*$ ,  $\forall i, \forall j$ , the tolerance  $\eta_0 = 0.01$ , the index  $t = 0$ , and step size  $\alpha = 0.1$ , initialize  $\mathbf{x} = [x_0, x_1, \dots, x_L]^T$  with  $\mathbf{x}^0 = [x_0^0, x_1^0, \dots, x_L^0]^T$ , set  $\epsilon = 1$ , and let  $\mathbf{F}(\mathbf{x}) = \mathbf{P}_N (\lambda \mathbf{I} + (1 - \lambda) \mathbf{P}_T(\mathbf{x})) \mathbf{x} - \mathbf{x}$ .
2. while  $\epsilon > \eta_0$ 
  - Compute the updating vector  $\Delta \mathbf{x}^{t+1}$  using  $\Delta \mathbf{x}^{t+1} = -\alpha \times \nabla \mathbf{F}(\mathbf{x}^t) \times \mathbf{F}(\mathbf{x}^t)$ .
  - Update  $\mathbf{x}^{t+1}$  by  $\mathbf{x}^{t+1} = \mathbf{x}^t + \Delta \mathbf{x}^{t+1}$ .
  - Normalize  $\mathbf{x}^{t+1}$  using  $\mathbf{x}^{t+1} = \frac{\mathbf{x}^{t+1}}{\|\mathbf{x}^{t+1}\|_2}$ .
  - Update the parameter  $\epsilon$  by  $\epsilon = \|\mathbf{x}^{t+1} - \mathbf{x}^t\|_2$ .
  - Update the index  $t = t + 1$ .
- End
3. The stationary reputation distribution is  $\mathbf{x}^* = \mathbf{x}^t$ .

where  $\mathbf{x}^{new}$  is the new reputation distribution of the entire population and can be computed by

$$\mathbf{x}^{new} = \mathbf{P}_N (\lambda \mathbf{I} + (1 - \lambda) \mathbf{P}_T) \mathbf{x}, \quad (10)$$

with the  $i^{th}$  row and  $j^{th}$  column element of the matrix  $\mathbf{P}_T$  being defined as

$$\mathbf{P}_T(j, i) = \sum_{k: Q_{a_{i,k},k}=j} x_k. \quad (11)$$

According to (9), (10), and (11), the stationary reputation distribution  $\mathbf{x}^*$  is the solution to the following equation

$$\mathbf{P}_N (\lambda \mathbf{I} + (1 - \lambda) \mathbf{P}_T) \mathbf{x}^* = \mathbf{x}^*. \quad (12)$$

From (11) and (12), we can see that, given the optimal action  $\mathbf{a}^*$ , the stationary reputation distribution can be found by solving the nonlinear equations in (12). In Algorithm 1, we propose a gradient descent algorithm for finding the stationary reputation distribution given the optimal action rule.

### C. Payoff Function

Suppose that the cost of forwarding a packet is a constant,  $c$ , the total cost of the transmitter with reputation  $i$  taking action  $a_{i,j}$  towards a receiver with reputation  $j$  is given by

$$C(a_{i,j}) = a_{i,j}c. \quad (13)$$

Similarly, if the gain of receiving a packet is a constant,  $g$ , the total gain of the receiver with reputation  $i$  can be computed by

$$G(a_{j,i}) = a_{j,i}g, \quad (14)$$

where  $a_{j,i}$  is the action of the corresponding transmitter with reputation  $j$ .

Let  $W_{i,j}$  denote the maximum payoff that a player, currently having reputation  $i$  and being matched with a player with reputation  $j$ , can gain from this interaction to future. Obviously, if the player with reputation  $i$  serves as a transmitter, then the long-term expected payoff that he/she can obtain by taking action  $a_{i,j}$  would be

$$f_1(a_{i,j}) = -a_{i,j}c + \delta \sum_k \sum_l \tilde{\mathbf{d}}_{i \rightarrow j}(k) \mathbf{x}^*(l) W_{k,l}, \quad (15)$$

where the first term  $a_{i,j}c$  is the immediate cost the transmitter incurred by taking action  $a_{i,j}$ , and the second term

---

**Algorithm 2 : Modified Value Iteration For Optimal Action Selection Given Stationary Reputation Distribution**


---

1. Given the stationary reputation  $\mathbf{x}^*$ , tolerance  $\eta_0 = 0.01$ , initialize  $a_{i,j}^*$  with  $a_{i,j}^0 \forall i \forall j$ , set  $\epsilon_1 = 1$  and  $\epsilon_2 = 1$ .
  2. while  $\epsilon_1 > \eta_0$ 
    - Set  $\epsilon_2 = 1$ .
    - Initialize  $W_{i,j} = 0 \forall i \forall j$ .
    - while  $\epsilon_2 > \eta_0$ 
      - Compute  $\tilde{\mathbf{d}}_{i \rightarrow j}$  using  $\tilde{\mathbf{d}}_{i \rightarrow j} = \mathbf{P}_N \left( \lambda \mathbf{e}_i + (1 - \lambda) \mathbf{e}_{Q_{a_{i,j},j}} \right)$ .
      - Compute  $\hat{W}_{i,j}$  using
$$\hat{W}_{i,j} = \max_{a_{i,j}} \left[ \frac{1}{2} \left( -a_{i,j}c + \delta \sum_k \sum_l \tilde{\mathbf{d}}_{i \rightarrow j}(k) \mathbf{x}^*(l) W_{k,l} \right) + \frac{1}{2} \left( a_{j,i}^*g + \delta \sum_l \mathbf{x}^*(l) W_{i,l} \right) \right].$$
      - Compute  $\hat{a}_{i,j}$  using
$$\hat{a}_{i,j} = \arg \max_{a_{i,j}} \left[ \frac{1}{2} \left( -a_{i,j}c + \delta \sum_k \sum_l \tilde{\mathbf{d}}_{i \rightarrow j}(k) \mathbf{x}^*(l) W_{k,l} \right) \right].$$
      - Update the parameter  $\epsilon_2$  by  $\epsilon_2 = \|\hat{\mathbf{W}} - \mathbf{W}\|_2$ .
      - Update  $\mathbf{W}$  by  $\mathbf{W} = \hat{\mathbf{W}}$ .
      - End
      - Update the parameter  $\epsilon_1$  by  $\epsilon_1 = \|\hat{\mathbf{a}} - \mathbf{a}^*\|_2$ .
      - Update  $\mathbf{a}^*$  by  $\mathbf{a}^* = \hat{\mathbf{a}}$ .
  - End
  3. The optimal action is  $\mathbf{a}^*$ .
- 

$\sum_k \sum_l \tilde{\mathbf{d}}_{i \rightarrow j}(k) \mathbf{x}^*(l) W_{k,l}$  stands for the benefit he gains in the future with a discounting factor  $\delta$ . According to (6), by taking action  $a_{i,j}$ , the reputation distribution of the transmitter will change from  $\mathbf{e}_i$  to  $\tilde{\mathbf{d}}_{i \rightarrow j}$ . Since his opponent in the next round is randomly sampled from the population with a stationary reputation distribution  $\mathbf{x}^*$ , with probability  $\tilde{\mathbf{d}}_{i \rightarrow j}(k) \mathbf{x}^*(l)$ , the transmitter's reputation becomes  $k$  and his opponent's reputation is  $l$ .

On the other hand, if the player with reputation  $i$  serves as a receiver, the long-term expected payoff that he/she can obtain is

$$f_2 = a_{j,i}^*g + \delta \sum_l x_l^* W_{i,l}, \quad (16)$$

where the first term  $a_{j,i}^*g$  is the immediate gain he/she can obtain when the transmitter takes the optimal action  $a_{j,i}^*$ , and the second term  $\sum_l x_l^* W_{i,l}$  stands for the benefit he gains in the future with a discounting factor  $\delta$ . As a receiver, the reputation will not change after the transmission. Since his opponent in the next round is randomly sampled from the population with a stationary reputation distribution  $\mathbf{x}^*$ , with probability  $\mathbf{x}^*(l)$ , the receiver's reputation is  $i$  and his opponent's reputation is  $l$ .

With each interaction, the play acts either as a transmitter or as a receiver with equal probability  $\frac{1}{2}$ . Therefore, the Bellman equation of  $W_{i,j}$  can be written as

$$W_{i,j} = \max_{a_{i,j}} \left[ \frac{1}{2} \left( -a_{i,j}c + \delta \sum_k \sum_l \tilde{\mathbf{d}}_{i \rightarrow j}(k) \mathbf{x}^*(l) W_{k,l} \right) + \frac{1}{2} \left( a_{j,i}^*g + \delta \sum_l \mathbf{x}^*(l) W_{i,l} \right) \right], \quad (17)$$

---

**Algorithm 3 : An Alternative Algorithm For Finding Stationary Reputation Distribution And Optimal Action**


---

1. Given the tolerance  $\eta_0 = 0.01$ , initialize  $\mathbf{a}^*$  with  $\mathbf{a}^0$  and set  $\epsilon = 1$ .
  2. while  $\epsilon > \eta_0$ 
    - Given the optimal action  $\mathbf{a}^*$ , finding the stationary reputation distribution  $\mathbf{x}^*$  using Algorithm 1.
    - Given the stationary reputation distribution  $\mathbf{x}^*$ , finding the optimal action  $\hat{\mathbf{a}}^*$  using Algorithm 2.
    - Update the parameter  $\epsilon$  by  $\epsilon = \|\hat{\mathbf{a}}^* - \mathbf{a}^*\|_2$ .
    - Update  $\mathbf{a}^*$  by  $\mathbf{a}^* = \hat{\mathbf{a}}^*$ .
  - End
  3. The stationary reputation distribution is  $\mathbf{x}^*$  and the optimal action is  $\mathbf{a}^*$ .
- 

and the optimal action  $a_{i,j}^*$  can be computed by

$$\begin{aligned} a_{i,j}^* &= \arg \max_{a_{i,j}} W_{i,j} \\ &= \arg \max_{a_{i,j}} \left[ \frac{1}{2} \left( -a_{i,j}c + \delta \sum_k \sum_l \tilde{\mathbf{d}}_{i \rightarrow j}(k) \mathbf{x}^*(l) W_{k,l} \right) \right]. \end{aligned} \quad (18)$$

From (17) and (18), we can see that the problem of finding the optimal action rule is a Markov Decision Process (MDP), where the state is the reputation pair  $(i, j)$ , the action is  $a_{i,j}$ , the transition probability is determined by  $\tilde{\mathbf{d}}_{i \rightarrow j}$  and  $\mathbf{x}^*$ , and the reward is determined by  $c$  and  $g$ . Therefore, given the stationary reputation distribution, the optimal action can be found by solving (18) using dynamic programming. In this paper, we propose a modified value iteration algorithm to find the optimal action given stationary reputation distribution, which is shown in Algorithm 2.

#### D. Optimal Action Using An Alternative Algorithm

From the previous two subsections, we can see that given the optimal action, the stationary reputation distribution can be found using Algorithm 1, and given the stationary reputation distribution, the optimal action can be found using Algorithm 2. Therefore, we can obtain the optimal action and the stationary reputation distribution alternatively by iteratively fixing one and solving the other. The detailed processes are summarized in Algorithm 3. Note that the convergence speed of Algorithm 3 is highly determined by the initial action rule  $\mathbf{a}^0$ . Nevertheless, it will converge since the number of the possible action rules is finite. Moreover, Algorithm 3 can also be used to test the evolutionary stability of any action rule. The idea is to set the tested action rule as the initial action rule and see whether it can converge in one iteration. The details will be discussed in section V.

#### IV. ACTION SPREADING DUE TO NATURAL SELECTION

Based on Algorithm 3, we can find the optimal action rule and the stationary reputation distribution. However, during the above analysis, we do not include the perturbation effect, where players may take non-optimal action rule due to uncertainty of the system and/or the incorrect (noisy) parameters. Taking the perturbation effect into account, we need to evaluate the stability of the optimal action rule. Here, we adopt the concept of evolutionarily stable strategy (ESS) [26], which is “a strategy such that, if all members of the

population adopt it, then no mutant strategy could invade the population under the influence of natural selection". In the following subsections, we first discuss, by natural selection, how the action rules spread over the population. Specifically, we discuss two action spreading algorithms: one is action spreading algorithm using Wright-Fisher model [27] and the other is action spreading algorithm using replicator dynamic equation [26]. Then, we examine, in Section V, the stability of the optimal action rule derived by Algorithm 3 by the simulations.

Let  $M$  be the number of action rules,  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M$ , used in the population. Let  $p_i^t$  be the percentage of the population that uses action rule  $\mathbf{a}_i$  at time  $t$ . Then, we have  $\sum_{i=1}^M p_i^t = 1$ . Let  $U_i^t$  be the average payoff using action rule  $\mathbf{a}_i$  at time  $t$ .

#### A. Action Spreading Algorithm Using Wright Fisher Model

The Wright-Fisher model is by far the most popular stochastic model for reproduction in population genetics [27]. It is based on the assumption that the probability of an individual adopting a certain strategy is proportional to the expected payoff of the population using that strategy. Due to its simplicity and capability of capturing the essence of the biology involved, we use the Wright-Fisher model here to characterize how action rules spread over the population.

Let  $y_i$  be the probability of an individual using action  $\mathbf{a}_i$ . Then, we have  $\sum_{i=1}^M y_i = 1$ . With the Wright-Fisher Model, we assume that  $y_i$  is proportional to the total payoff of the users using  $\mathbf{a}_i$ . Therefore,  $y_i$  can be computed by

$$y_i = \frac{p_i^t U_i^t}{\sum_{j=1}^M p_j^t U_j^t}, \quad (19)$$

where the numerator  $p_i^t U_i^t$  is the total payoff of the users using action  $\mathbf{a}_i$ , and the denominator  $\sum_{j=1}^M p_j^t U_j^t$  is the total payoff of the whole population, which is the normalization term that ensures  $\sum_{i=1}^M y_i = 1$ .

Based on the assumption that the population size is sufficiently large, the percentage of the population using action  $\mathbf{a}_i$  is equal to the probability of an individual using  $\mathbf{a}_i$ . Therefore, the action spreading equation can be written as

$$p_i^{t+1} = \frac{p_i^t U_i^t}{\sum_{j=1}^M p_j^t U_j^t}. \quad (20)$$

#### B. Action Spreading Algorithm Using Replicator Dynamic Equation

Replicator dynamic equation is widely used to characterize the population evolution in evolutionary game theory [26]. It is based on the following intuition: if a certain strategy results in a higher payoff than the average level, the population share using that strategy will grow with the growth rate proportional to the difference between the expected payoff of the population using that strategy and the expected payoff of the entire population. In this subsection, we use the replicator dynamic equation to model the evolution of the percentage of the population using a certain action rule, which means that the evolution of  $p_i$  is given by the following equation

$$\frac{dp_i}{dt} = \eta \left( U_i - \sum_{j=1}^M p_j U_j \right) p_i, \quad (21)$$

where  $\eta$  is a scale factor controlling the speed of the evolution.

By discretizing the replicator dynamic equation in (21), we have the action spreading equation

$$\begin{aligned} p_i^{t+1} &= p_i^t + \eta \left( U_i^t - \sum_{j=1}^M p_j^t U_j^t \right) p_i^t \\ &= p_i^t \left[ 1 + \eta \left( U_i^t - \sum_{j=1}^M p_j^t U_j^t \right) \right]. \end{aligned} \quad (22)$$

### V. EVOLUTIONARILY STABLE STRATEGY AND SIMULATIONS

To verify the proposed algorithm, we simulate the packet forwarding game. We study a fixed-size population,  $N = 1000$ . Each new player receives an initial reputation, which is randomly chosen from  $\{0, 1, \dots, L\}$  with equal probability  $\frac{1}{L+1}$ . Each player uses one of  $(L+1)^{(L+1) \times (L+1)}$  possible action rules. All players in the population share the fixed social norm defined in (2). Before any one elementary step of action updating, each individual has exactly 20 interactions with other randomly chosen individuals. Individuals act as transmitter and receiver on average 10 times each. After each interaction, the reputation of the transmitter is updated according to the reputation updating policy shown in Fig. 2. We assume that every player in the population agrees on the reputation generated by the reputation updating policy. No private lists of reputation are considered. After all 20 interactions have taken place, each participant including both the transmitter and receiver goes back to the population with probability  $\delta$  or leaves the population with probability  $1 - \delta$ . For every player who leaves, a new individual enters the population to keep the total population size constant. The initial reputation of the new coming is randomly chosen from  $\{0, 1, \dots, L\}$  with equal probability  $\frac{1}{L+1}$ . Then, the players in the population, including the old players who stay in the population and the new players who enter the population, choose their new action rules according to previous payoff history of the whole population. There are two possible action spreading algorithms as shown in the previous section. One is the action spreading algorithm using Wright Fisher Model, which is denoted as "WFM", and the other one is the action spreading algorithm using Replicator Dynamic Equation, which is denoted as "RDE". After updating the action rule, the payoffs of all players are reset to zero. Therefore, older players do not accumulate their payoffs. In all the following simulations, the parameters  $\lambda$ ,  $\delta$ , and  $\mu$  are set to be 0.5, 0.9, and 0.95 respectively. The parameter  $\eta$  that controls the speed of the evolution in RDE is set to be 0.1.

#### A. Binary Reputation Scenario

To give more insights into the proposed algorithm, we first evaluate the binary reputation scenario where  $L = 1$ . We assume that the gain per unit is 1 and the cost per unit is 0.1, i.e.  $g = 1$  and  $c = 0.1$ . According to Algorithm 3, with different initial conditions, we can find three pairs of stationary reputation distribution  $\mathbf{x}^*$  and the optimal action rule  $\mathbf{a}^*$ , which

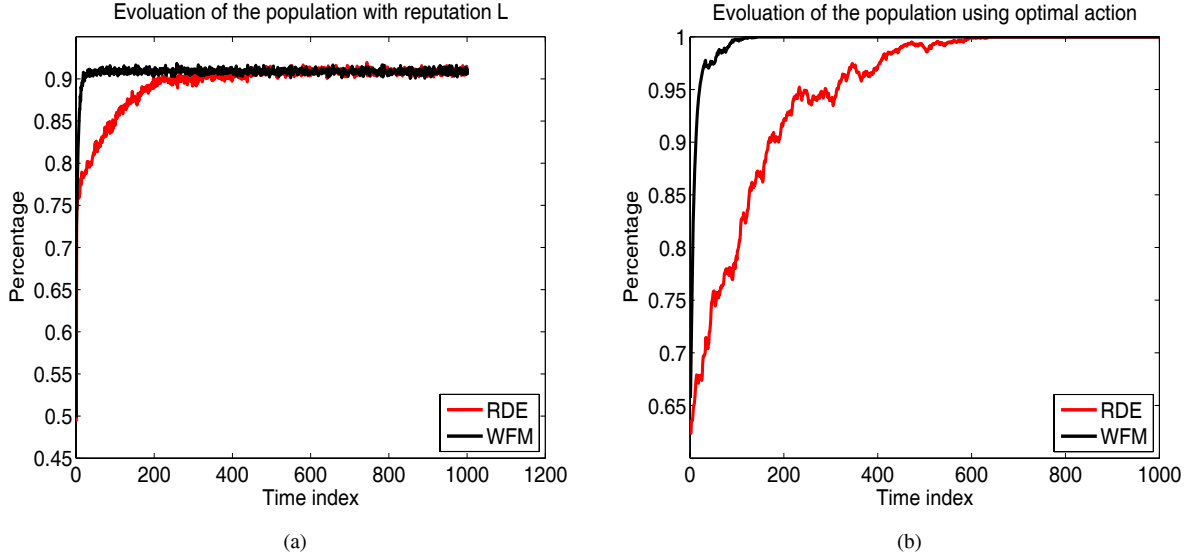


Fig. 3. The population evolution when  $L = 1$ ,  $g = 1$  and  $c = 0.1$ : (a) the percentage of the population with reputation  $L = 1$ ; (b) the percentage of the population using optimal action shown in (24).

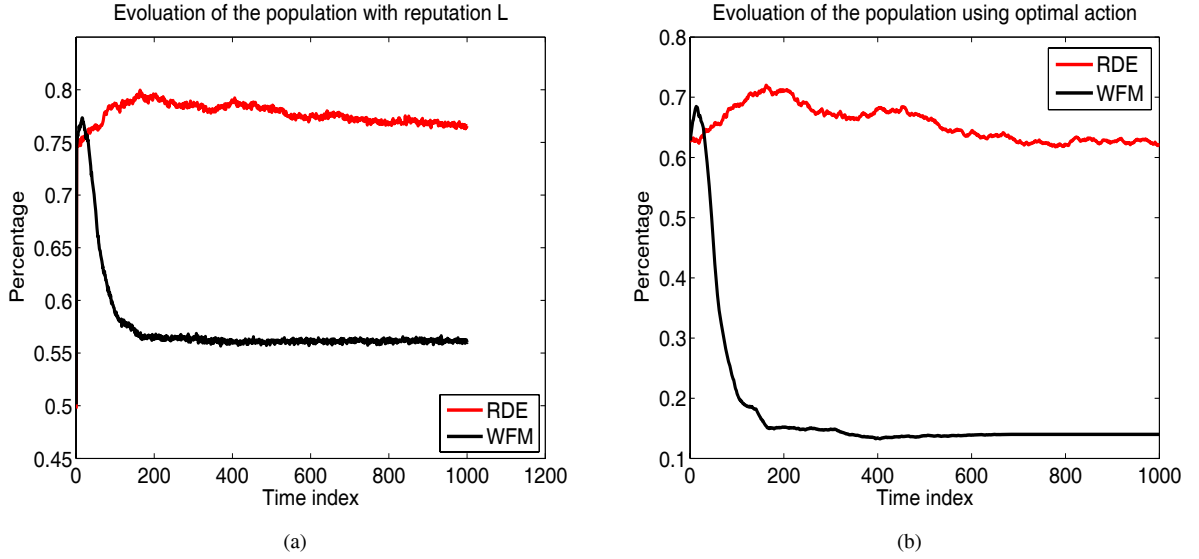


Fig. 5. The population evolution when  $L = 1$ ,  $g = 1$  and  $c = 0.6$ : (a) the percentage of the population with reputation  $L = 1$ ; (b) the percentage of the population using optimal action shown in (24).

are

$$\mathbf{x}_1^* = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \quad \mathbf{a}_1^* = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \quad (23)$$

$$\mathbf{x}_2^* = \begin{pmatrix} 0.0909 \\ 0.9091 \end{pmatrix}, \quad \mathbf{a}_2^* = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}. \quad (24)$$

$$\mathbf{x}_3^* = \begin{pmatrix} 0.9091 \\ 0.0909 \end{pmatrix}, \quad \mathbf{a}_3^* = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}. \quad (25)$$

With  $(\mathbf{x}_1^*, \mathbf{a}_1^*)$ , the transmitter will not forward any packet to the receiver regardless his/her own reputation and the corresponding receiver's reputation. Obviously, it is a bad strategy since, with such a strategy, there is no cooperation and the payoff of every player is zero. The pairs  $(\mathbf{x}_2^*, \mathbf{a}_2^*)$  and  $(\mathbf{x}_3^*, \mathbf{a}_3^*)$  are symmetric where with the former pair, the transmitter will always forward packets to the receiver who has good

reputation, and with the latter pair, the transmitter will always forward packets to the receiver who has bad reputation. We can also find that the pair  $(\mathbf{x}_2^*, \mathbf{a}_2^*)$  leads to a population with more than 90 percentage of the players are good reputation while  $(\mathbf{x}_3^*, \mathbf{a}_3^*)$  leads to a population with more than 90 percentage of the players are bad reputation. Here, we prefer  $(\mathbf{x}_2^*, \mathbf{a}_2^*)$  since it leads to a "good" society with more than 90 percentage of the population are good reputation.

Then, we evaluate the evolutionary stability of  $(\mathbf{x}_2^*, \mathbf{a}_2^*)$ . In the simulation, the initial frequency of the optimal action rule  $\mathbf{a}^*$  shown in (24) is set to be 0.6. The initial frequencies of the other action rules are randomly chosen. The initial reputation of new players is randomly chosen from  $\{0, 1\}$  with equal probability  $\frac{1}{2}$ . In Fig. 3 (a), we show the evolutionary results of the percentage of the population with reputation level  $L = 1$ . From Fig. 3 (a), we can see that for both **WFM** and **RDE**, the reputation distribution converges to the

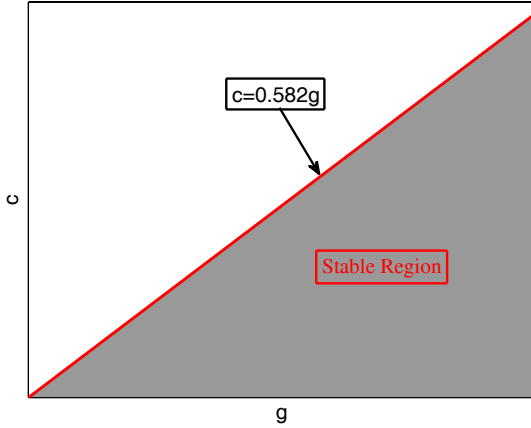


Fig. 4. The stable region for the optimal action rule shown in (24) when  $L = 1$ .

stationary reputation distribution  $\mathbf{x}_2^*$ . Compared with **WFM**, the convergence speed of **RDE** is a bit slower since a small speed controlling parameter  $\eta = 0.1$  is used in **RDE**.

The evolutionary results of the percentage of the population using the action rule  $\mathbf{a}_2^*$  are shown in Fig. 3 (b). From Fig. 3 (b), we can see that for both **WFM** and **RDE**, the action rule  $\mathbf{a}_2^*$  will spread over the whole population. And once the whole population adopt  $\mathbf{a}_2^*$ , no one will deviate. Therefore, the action rule  $\mathbf{a}_2^*$  is an evolutionarily stable strategy (ESS) [26] in this case.

From (17), we can see that the optimal action rule is determined by the values of  $g$  and  $c$ . Intuitively, if  $g \gg c$ , every player is willing to cooperate with other players since in such a scenario, the potential cooperation gain will be greater than the immediate cooperation cost. On the other hand, if  $c \gg g$ , every player tends not to cooperate with other players since the potential cooperation gain will be smaller than the immediate cooperation cost in such a scenario. Based on the intuition, there should exist a critical cost-to-gain ratio  $\gamma$  such that the optimal action rule  $\mathbf{a}_2^*$  is stable if  $c < \gamma g$  and is not stable otherwise.

By setting  $\mathbf{a}_2^*$  as the initial action rule  $\mathbf{a}^0$  in Algorithm 3 and varying  $g$  and  $c$ , we find that if  $\frac{c}{g} \leq 0.582$ , the optimal action rule found by Algorithm 3 is  $\mathbf{a}_2^*$ . On the other hand, if  $\frac{c}{g} > 0.582$ , the optimal action rule changes to be  $\mathbf{a}_1^*$ . Therefore, the critical cost-to-gain ratio  $\gamma$  is equal to 0.582 in this case, which means that the stable region for  $\mathbf{a}_2^*$  is the shadow region shown in Fig. 4.

We verify the above statement by evaluating the stability of  $\mathbf{a}_2^*$  when  $g = 1$  and  $c = 0.6$ . The corresponding evolutionary results are shown in Fig. 5. From Fig. 5 (b), we can see that when  $\frac{c}{g} = 0.6 > 0.582$ , the percentage of the population using action rule  $\mathbf{a}_2^*$  does not converge to 1 for both **WFM** and **RDE**. Therefore,  $\mathbf{a}_2^*$  is not stable in this case. Correspondingly, we can also see from Fig. 5 (a) that the reputation distribution does not converge to  $\mathbf{x}_2^*$  in this case.

### B. Multi-Level Reputation Scenario

For the multi-level reputation scenario where  $L \geq 2$ , due to the large dimension of the action space  $((L+1)^{(L+1) \times (L+1)})$ ,

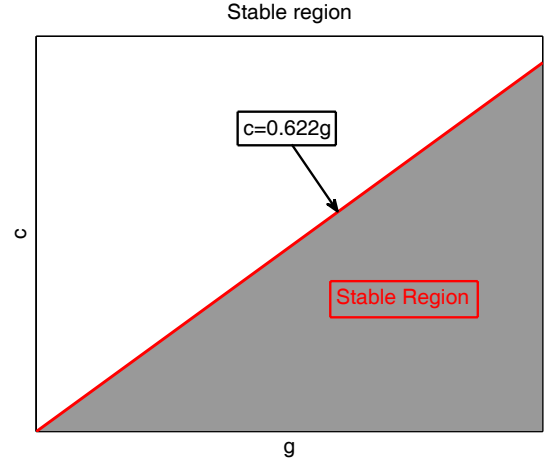


Fig. 6. The stable region for the optimal action rule shown in (26) when  $L = 4$ .

it is difficult to find all the possible pairs of stationary reputation distribution  $\mathbf{x}^*$  and optimal action rule  $\mathbf{a}^*$ . However, based on the results in the binary reputation scenario, we can infer that one possible optimal action rule  $\mathbf{a}_0^*$  is to forward  $i$  packets to the receiver with reputation  $i$ , i.e.  $\mathbf{a}_0^*$  can be written as

$$\mathbf{a}_0^* = \begin{pmatrix} 0 & 1 & \dots & L \\ 0 & 1 & \dots & L \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & L \end{pmatrix}. \quad (26)$$

According to Algorithm 1, we can find the corresponding stationary reputation distribution  $\mathbf{x}_0^*$ . For the special case with  $L = 4$ ,  $\mathbf{x}_0^*$  is

$$\mathbf{x}_0^* = (0.0235 \quad 0.0235 \quad 0.0235 \quad 0.0235 \quad 0.906)^T. \quad (27)$$

Then, similar to the binary reputation scenario, we obtain the stable region for the optimal action rule  $\mathbf{a}_0^*$ . By setting  $\mathbf{a}_0^*$  as the initial action rule  $\mathbf{a}^0$  in Algorithm 3 and varying  $g$  and  $c$ , we find that if  $\frac{c}{g} \leq 0.622$ , the optimal action rule found by Algorithm 3 is still  $\mathbf{a}_0^*$ . On the other hand, if  $\frac{c}{g} > 0.622$ , the optimal action rule changes. Therefore, the critical cost-to-gain ratio  $\gamma$  in this case is equal to 0.622, which means that the stable region for  $\mathbf{a}_0^*$  is the shadow region shown in Fig. 6.

We then verify the above statement by simulating the packet forwarding game with two different cost-to-gain ratio settings. One is  $g = 1$  and  $c = 0.5$ , i.e.  $\frac{c}{g} = 0.5 < 0.622$ , and the other is  $g = 1$  and  $c = 0.7$ , i.e.  $\frac{c}{g} = 0.7 > 0.622$ . The evolutionary results for the former setting are shown in Fig. 7. From Fig. 7, we can see that when the cost-to-gain ratio is set to be  $\frac{c}{g} = 0.5 < 0.622$ , the reputation distribution converges to  $\mathbf{x}_0^*$  and the optimal action rule  $\mathbf{a}_0^*$  spreads over the whole population for both **WFM** and **RDE**, which verifies that  $\mathbf{a}_0^*$  is an ESS in this case.

The evolutionary results for the latter cost-to-gain ratio setting are different and shown in Fig. 8. From Fig. 8, we can see that when the cost-to-gain ratio is set to be  $\frac{c}{g} = 0.7 > 0.622$ , for both **WFM** and **RDE**, the action rule  $\mathbf{a}_0^*$  does not spread over the whole population and the reputation distribution does not converge to  $\mathbf{x}_0^*$ . Therefore,  $\mathbf{a}_0^*$  is not stable in this case.



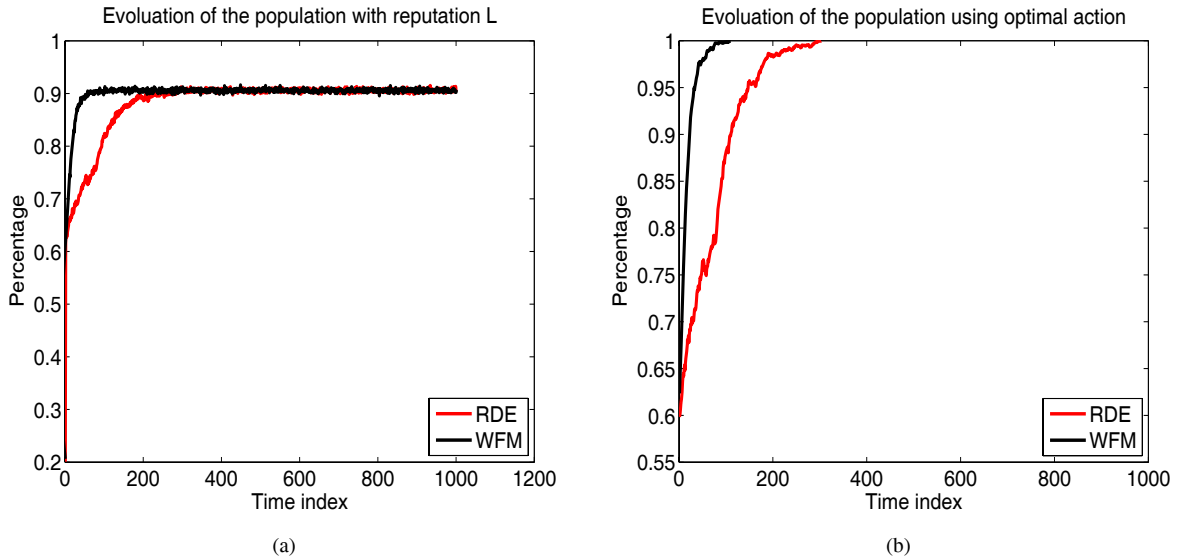


Fig. 7. The population evolution when  $L = 4$ ,  $g = 1$  and  $c = 0.5$ : (a) the percentage of the population with reputation  $L = 4$ ; (b) the percentage of the population using optimal action shown in (26).

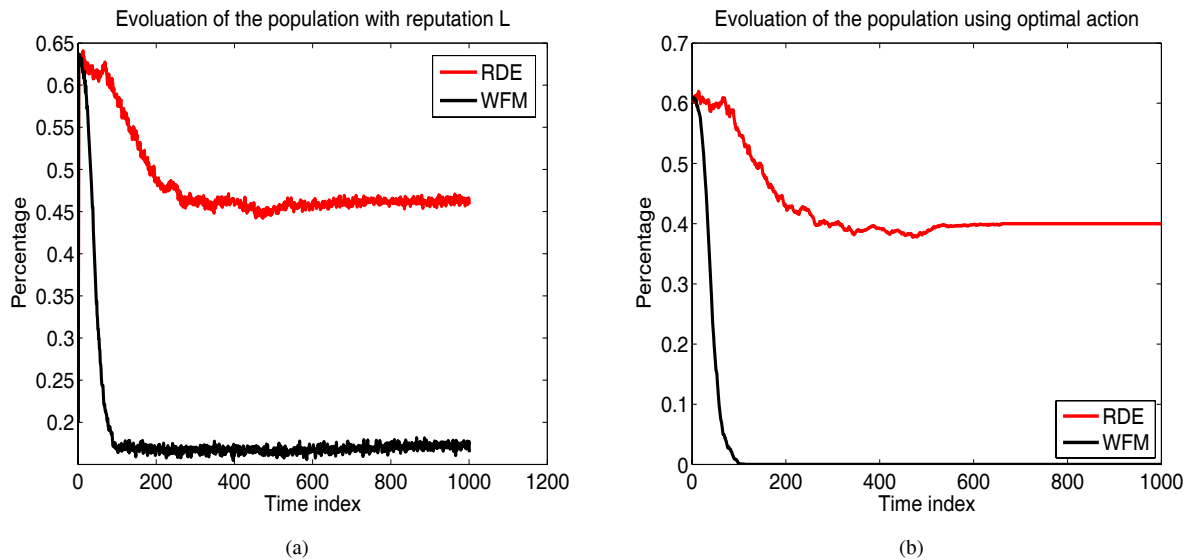


Fig. 8. The population evolution when  $L = 4$ ,  $g = 1$  and  $c = 0.7$ : (a) the percentage of the population with reputation  $L = 4$ ; (b) the percentage of the population using optimal action shown in (26).

## VI. CONCLUSIONS

In this paper, we propose a cooperation stimulation scheme for cognitive networks using indirect reciprocity game modelling. Different from the existing game theoretic approaches, our proposed scheme does not rely on the assumption that the number of interactions between a pair of players are infinite. From the simulation results, we can see that with a proper cost-to-gain ratio, the action rule of forwarding  $i$  packets to the receiver with reputation level  $i$  is an ESS. Even starting with only 60 percentage of population adopting the optimal action rule, the optimal action rule will quickly spread over the whole population by natural selection. And once the whole population use the optimal action rule, no one will deviate. Moreover, such an ESS will lead to a “good” society where more than 90 percentage of the population have good reputation.

## REFERENCES

- [1] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, “Cognitive networks: adaptation and learning to achieve end-to-end performance objectives,” *IEEE Commun. Mag.*, vol. 44, pp. 51-57, Dec. 2006.
- [2] L. Buttyan and J. P. Hubaux, “Enforcing service availability in mobile ad-hoc networks,” in *Proc. ACM MobiHoc*, 2000.
- [3] S. Zhong, J. Chen, and Y. R. Yang, “Sprite: a simple, cheat-proof, credit-based system for mobile ad-hoc networks,” in *Proc. IEEE INFOCOM*, 2003.
- [4] V. Vishumurthy, S. Chandrakumar, and E. Sirer, “Karma: a secure economic framework for peer-to-peer resource sharing,” in *Proc. 2003 Workshop Economics Peer-to-Peer Syst.*, 2003.
- [5] P. Golle, K. Leyton-Brown, and I. Mironov, “Incentive for sharing in peer-to-peer networks,” in *Proc. ACM Conf. Electronic Commerce*, 2001.
- [6] S. Zhong, L. Li, Y. G. Liu, and Y. R. Yang, “On designing incentive-compatible routing and forwarding protocols in wireless ad-hoc networks—an integrated approach using game theoretical and cryptographic techniques,” in *Proc. ACM MobiCom*, 2005.
- [7] L. Anderegg and S. Eidenbenz, “Ad hoc-vcg: a truthful and cost-efficient

- routing protocol for mobile ad hoc networks with selfish agents," in *Proc. ACM MobiCom*, 2003.
- [8] W. Yu and K. J. R. Liu, "Attack-resistant cooperation stimulation in autonomous ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 2260-2271, Dec. 2005.
- [9] S. Marti and H. Garcia-Molina, "Limited reputation sharing in p2p systems," in *Proc. 5th ACM Conf. Electronic Commerce*, 2004.
- [10] M. Gupta, P. Judge, and M. Ammar, "A reputation system for peer-to-peer networks," in *Proc. ACM 13th International Workshop Netw. Operating Syst. Support Digital Audio Video*, 2003.
- [11] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," in *Proc. ACM MobiCom*, 2000.
- [12] S. Buchegger and J. Y. L. Boudec, "Performance analysis of the confidant protocol," in *Proc. ACM MobiHoc*, 2002.
- [13] P. Michiardi and R. Molva, "Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks," in *Proc. IFIP-Comm. Multimedia Security Conf*, 2002.
- [14] —, "A game theoretical approach to evaluate cooperation enforcement mechanisms in mobile ad hoc networks," in *Proc. Workshop Modeling Optimization Mobile, Ad Hoc Wireless Netw. (WiOPT'03)*, 2003.
- [15] J. Crowcroft, R. Gibbens, F. Kelly, and S. Ostring, "Modeling incentives for collaboration in mobile ad hoc networks," in *Proc. Workshop Modeling Optimization Mobile, Ad Hoc Wireless Netw. (WiOPT'03)*, 2003.
- [16] W. S. Lin, H. V. Zhao, and K. J. R. Liu, "Incentive cooperation strategies for peer-to-peer live multimedia streaming social networks," *IEEE Trans. Multimedia*, vol. 11, pp. 396-412, 2009.
- [17] M. Xiao and D. Xiao, "Understanding peer behavior and designing incentive mechanism in peer-to-peer networks: an analytical model based on game theory," in *Proc. International Conf. Algorithms Architectures Parallel Process. (ICA3PP)*, 2007.
- [18] V. Srinivasan, P. Nuggehalli, C. F. Chiasserini, and R. R. Rao, "Cooperation in wireless ad hoc networks," in *Proc. IEEE INFOCOM*, 2003.
- [19] A. Urpi, M. Bonuccelli, and S. Giordano, "Modeling cooperation in mobile ad hoc networks: a formal description of selfishness," in *Proc. Workshop Modeling Optimization Mobile, Ad Hoc Wireless Netw. (WiOPT'03)*, 2003.
- [20] M. Felegyhazi, J. P. Hubaux, and L. Buttyan, "Nash equilibria of packet forwarding strategies in wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 5, pp. 463-476, May 2006.
- [21] W. Yu and K. J. R. Liu, "Game theoretic analysis of cooperation stimulation and security in autonomous mobile ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 6, pp. 459-473, May 2007.
- [22] M. J. Osborne and A. Rubinsten, *A Course in Game Theory*. The MIT Press, 1994.
- [23] M. A. Nowak and K. Sigmund, "Evolution of indirect reciprocity," *Nature*, vol. 437, pp. 1291-1298, 2005.
- [24] H. Ohtsuki, Y. Iwasa, and M. A. Nowak, "Indirect reciprocity provides only a narrow margin for efficiency for costly punishment," *Nature*, vol. 457, pp. 79-82, 2009.
- [25] Y. Sun, W. Yu, Z. Han, and K. J. R. Liu, "Information theoretic framework of trust modelling and evaluation for ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 305-317, 2006.
- [26] J. M. Smith, *Evolutionary and the Theory of Games*. Cambridge University Press, 1982.
- [27] R. Fisher, *The Genetical Theory of Natural Selection*. Clarendon Press, 1930.



forensic. He received the University of Maryland Future Faculty Fellowship in 2010.



**Yan Chen** (S'06) received the Bachelor's degree from University of Science and Technology of China (USTC) in 2004 and the M.Phil degree from Hong Kong University of Science and Technology (HKUST) in 2007. Now he is a Ph.D. student in the Department of Electrical and Computer Engineering at University of Maryland, College Park.

His current research interests are in game theoretical modelling for multimedia social networks, multimedia signal processing, cooperative multimedia communication and networking, and multimedia

**K. J. Ray Liu** (F'03) is named a Distinguished Scholar-Teacher of University of Maryland, College Park, in 2007. He is Associate Chair of Graduate Studies and Research of Electrical and Computer Engineering Department and leads the Maryland Signals and Information Group conducting research encompassing broad aspects of wireless communications and networking, information forensics and security, multimedia signal processing, and biomedical engineering.

Dr. Liu is the recipient of numerous honors and awards including IEEE Signal Processing Society Technical Achievement Award and Distinguished Lecturer. He also received various teaching and research recognitions from University of Maryland including university-level Invention of the Year Award; and Poole and Kent Senior Faculty Teaching Award and Outstanding Faculty Research Award, both from A. James Clark School of Engineering. Dr. Liu is a Fellow of IEEE and AAAS.

Dr. Liu is President-Elect and was Vice President - Publications of IEEE Signal Processing Society. He was the Editor-in-Chief of *IEEE Signal Processing Magazine* and the founding Editor-in-Chief of *EURASIP Journal on Advances in Signal Processing*.

His recent books include *Cognitive Radio Networking and Security: A Game Theoretical View* (Cambridge University Press, 2010); *Cooperative Communications and Networking* (Cambridge University Press, 2008); *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications* (Cambridge University Press, 2008); *Ultra-Wideband Communication Systems: The Multiband OFDM Approach* (IEEE-Wiley, 2007); *Network-Aware Security for Group Communications* (Springer, 2007); *Multimedia Fingerprinting Forensics for Traitor Tracing* (Hindawi, 2005); *Handbook on Array Processing and Sensor Networks* (IEEE-Wiley, 2009).