# A Cooperation Stimulation Strategy in Wireless Multicast Networks

Binglai Niu, H. Vicky Zhao, *Member, IEEE*, and Hai Jiang, *Member, IEEE*

***Abstract*—Cooperative multicast is an effective technique to provide reliable multicast service in wireless networks. However, selfish nodes may act as free riders to maximize their payoffs, and an incentive mechanism is needed to stimulate cooperation. In this paper, we employ a game theoretic approach to analyze the interactions among selfish mobile nodes in wireless multicast networks. The cooperation process is modeled as an infinite repeated game and the desired cooperation state that achieves absolute fairness and Pareto optimality is derived. A Worst Behavior Tit-for-Tat (WBTFT) incentive strategy is proposed to stimulate cooperation at the desired cooperation state. In the proposed scheme, a node monitors others' behaviors and makes decisions according to the worst behaved node. With perfect monitoring, we analyze the conditions for the proposed strategy to be subgame perfect. To address the issue of imperfect monitoring, an interval based estimation method is proposed and the subgame perfect equilibrium conditions are derived under the assumption that nodes are bounded rational. Simulation results show that the proposed strategy can efficiently enforce cooperation even with imperfect monitoring, and its performance is close to that when all nodes fully cooperate with each other and when they can perfectly monitor each other's behavior without errors.**

*Index Terms* – wireless multicast, cooperation, repeated game, incentive mechanism, bounded rational.

## I. Introduction

In the past decade, multimedia broadcast/multicast applications over wireless networks have become popular, where multimedia data are delivered to a group of nodes simultaneously. Examples include the Internet Protocol Television (IPTV) over WiMax [1] and multimedia broadcast/multicast service (MBMS) in 3G networks [2], [3]. To combat channel fading, cooperative multicast has been shown as an effective technique to provide reliable multicast service in wireless networks [4]–[11], where successful nodes are selected as relays to rebroadcast the data. In most of the existing cooperation strategies, it is assumed that nodes will cooperate unconditionally, and always forward data when selected as relays. However, in many applications, nodes are selfish and prefer receiving data from others but would not contribute at all, since relaying data costs extra energy. Therefore, it is critical to design an incentive mechanism to stimulate cooperation among selfish nodes.

In the literature, a number of incentive mechanisms have been proposed to stimulate cooperation in wireless networks, which can be briefly classified into three categories: payment based mechanisms [12]-[14], reputation based mechanisms [15]-[22] and punishment based mechanisms [24]-[25]. Payment based mechanisms introduce credits (or virtual currency) as payment for cooperative services, where nodes who get services should be charged and those who help should be remunerated. Examples include the "nuglets" mechanism in [12] and the "Sprite" system in [13]. However, these mechanisms require temper-proof hardware or central banking service to coordinate the credit exchange, which greatly limits their applications. In reputation based mechanisms, nodes monitor each other's behavior and cooperate with those who maintain good reputation. Based on this idea, CORE [16] and CONFIDENT [17] are proposed to enforce cooperation, which can efficiently detect and isolate misbehaved nodes. In [20] and [21], an attack-resist cooperation stimulation mechanism is proposed with consideration of noisy channels and the imperfect monitoring process. The main idea in punishment based mechanisms is to punish non-cooperative nodes by employing some punishment strategy. In [23], a "punish-and-forgive" strategy is proposed to enforce cooperative spectrum sharing in cognitive radio networks, where a punishment period will be triggered whenever a deviating behavior is detected. A similar strategy is employed in [24], and a cartel maintenance framework is proposed to stimulate cooperation among selfish nodes in a distributed manner.

Most of the existing incentive mechanisms are proposed for unicast scenarios, where nodes communicate with each other through point-to-point links. The special features of wireless multicast networks pose great challenges for cooperation stimulation. First, in multicast networks, communication is point-to-multiple points. The broadcast nature in wireless channels and the heterogeneous behaviors of nodes in the multicast group make it difficult to design payment based mechanisms or reputation based mechanisms (i.e.,


Manuscript received August 6, 2010; revised December 1, 2010 and January 31, 2011; accepted Feb 2, 2011. The review of this paper was coordinated by Dr. Huaiyu Dai.

This work was supported by the Natural Science and Engineering Research Council (NSERC) of Canada under a Strategic Project Grant, and by the Alberta Innovates - Technology Futures, Alberta, Canada under an Alberta Ingenuity New Faculty Award. This work was completed when B. Niu was a Master's student at the University of Alberta.

B. Niu is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4 (email: bniu@ece.ubc.ca). H. V. Zhao and H. Jiang are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2V4 (email: {vzhao, hai.jiang}@ece.ualberta.ca).

it is hard to decide who should pay for the cooperative services, and a node may not have the choice to only cooperate with those good-reputation nodes due to the broadcast nature). Second, due to the existence of noise and fading in wireless channels, packets may be dropped during transmission. Therefore, a cooperative behavior may not provide the expected service, and hence may not get any reward. Moreover, in real applications with distributed nodes, the monitoring process is usually imperfect. Since a multicast node may need to monitor all others' behaviors, its monitoring results may have large errors, which may result in frequent undesired decisions and may discourage node cooperation.

In this work, we design an incentive mechanism to stimulate cooperation among selfish nodes in wireless multicast networks. We consider a two-phase cooperative multicast scenario where a BS multicasts data packets to a group of nodes in the first phase, and nodes cooperate with each other to distribute the data in the second phase. We consider the scenario where nodes are selfish and can decide their transmission power when selected as relays. To address the aforementioned challenges, we employ game theory to analyze the cooperation process and propose a punishment based mechanism, where nodes decrease their transmission power (or even do not cooperate) when a misbehavior is detected. The major contributions of this work can be summarized as follows:

- We provide game theoretic analysis of the cooperative multicast process. We model the cooperative multicast process as an infinite repeated game, and find the desired cooperation power level that satisfies the absolute fairness and the maximum payoff criteria. We also demonstrate that the desired power level is Pareto optimal. To our best knowledge, this is the first work to study cooperation stimulation in wireless multicast networks using game theoretic approaches.
- We propose a punishment based incentive strategy to stimulate cooperation for multicast applications. To address the challenge of heterogeneous behaviors in multicast networks, we introduce a *Worst Behavior Tit-for-Tat* incentive strategy. The main idea is that each node monitors others' behaviors within a certain period, and makes a similar move to the worst behavior it has observed. Through theoretical analysis, we show that by employing this strategy, nodes are encouraged to perform cooperatively, since any deviation from the desired power level leads to a low-cooperative or even non-cooperative phase, which lowers their payoffs.
- We develop a novel interval based estimation method to address the issue of imperfect monitoring. We consider the realistic scenario that the monitoring process is imperfect. In this scenario, nodes make decisions based on unreliable information and may not cooperate with each other at the desired level. To address this issue, we propose an interval based estimation method that gives more reliable estimation results when nodes choose transmission power from a discrete set. Through simulations, we show that the proposed strategy with interval based estimation can enforce cooperation efficiently even with imperfect monitoring, at a level that is close to the desired cooperation state.

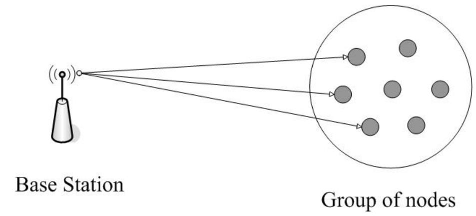


Fig. 1. System model.

The rest of the paper is organized as follows. Section II describes the system model and formulates the repeated game. Equilibrium of the game is studied in Section III, and the cooperation stimulation strategy is proposed and analyzed in Section IV. Performance evaluation is provided in Section V, and conclusion is drawn in Section VI.

## II. System Model and Game Formulation

Consider a single hop multicast network as shown in Fig. 1, where a base station (BS) multicasts data packets to a group of $N$ mobile nodes that are randomly deployed within a circular area. The distance between the BS and a node is much larger than that between any pair of nodes. To improve the network performance, a cooperative multicast strategy is employed, which divides a time slot into two subslots. In the first subslot, denoted as phase 1, the BS broadcasts a packet and then randomly selects a relay based on the feedbacks from nodes who correctly receive the packet. Here, we say a node succeeds in phase 1 if it receives the packet correctly from the BS. In the second subslot, denoted as phase 2, the selected relay rebroadcasts the packet to those who do not succeed in phase 1. The wireless channel between the BS and a node experiences Rayleigh fading and AWGN noise (with complex Gaussian distribution with zero mean and variance $N_0$, i.e., $\mathcal{CN}(0, N_0)$), and the channel gain is assumed to be independent and identically distributed (i.i.d.) with complex Gaussian distribution $\mathcal{CN}(0, \sigma_B^2)$[1]. Similar to [26], the probability that a node correctly receives a packet from the BS is $p_B = \exp\left(\frac{N_0(1-2^{\gamma_0})}{P_0\sigma_B^2}\right)$, where $P_0$ is the BS's transmission power and $\gamma_0$ is the BS's transmission rate. Similarly, the channel gain between node

[1]In this work, we only consider path loss and Rayleigh fading channels, and the i.i.d. assumption is reasonable when the BS is far away from the node group while the nodes are close to each other. Note that the i.i.d. assumption is only for analysis. In the simulations in Section V, the channel from the BS to a node is based on the real distance between them and Rayleigh fading.

$i$ and node $j$ is modeled as $\mathcal{CN}(0,\sigma_{ij}^2)$, and the probability that node $j$ successfully receives a packet from node $i$ is $p_{ij}=\exp\left(\frac{N_0(1-2^{\gamma_1})}{P_i\sigma_{ij}^2}\right)$,[2] where $P_i$ is node $i$'s transmission power and $\gamma_1$ is node $i$'s transmission rate[3]. Since nodes are mobile, it can be seen that the success probability $p_{ij}$ changes from time to time. To simplify the analysis, we use the following approximation: for any pair of mobile nodes $i$ and $j$, we can calculate the statistical mean of $\frac{N_0(2^{\gamma_1}-1)}{\sigma_{ij}^2}$ according to the mobility model and the statistics of the channel gain, which is denoted as $D$, and the overall probability that node $j$ successfully receives a packet from node $i$ is approximated as $p_{ij} \approx p_r(P_i)\stackrel{\triangle}{=}\exp\left(\frac{-D}{P_i}\right)$, $\forall\, i,j\in\{1,2,...N\}, i\neq j$ [4].

In the above system model, it can be seen that forwarding data benefits unsuccessful nodes, but incurs some cost to the relay, such as extra energy consumption. Therefore, a selfish node would expect others to cooperate but would not cooperate itself. Consider the scenario where nodes are all selfish and aim to maximize their own profits. Then, the cooperative multicast process can be modeled as a game, where each node is a player. When selected as a relay, a node can decide the transmission power to maximize its *payoff*, which is defined as the reward for receiving packets (either from the BS or from relays) minus the cost to forward packets. Since the multicast session usually lasts for a long time, nodes may perform the decision-making process repeatedly. Therefore, the interactions between nodes can be formulated as a repeated game. Denote the reward for correctly receiving a packet as $r_0>0$ and the cost for transmitting a packet with unit power as $c_0>0$, where $r_0$ and $c_0$ are application dependent constants. Consider $M$ time slots as a stage, in which we assume that a node does not change its transmission power when selected as a relay. Then the cooperative multicast game can be defined as follows:

*Definition 1: The Cooperative Multicast Game is the game* $G=\langle \mathcal{N}, \boldsymbol{P}^k, \boldsymbol{\pi}^k\rangle$, *where*

- $\mathcal{N}=\{1,2,...,N\}$ *is the set of nodes (players);*
- $\boldsymbol{P}^k=[P_1^k,P_2^k,...,P_N^k]$. $P_i^k\in[0,P_{\max}]$ *is node $i$'s action in stage $k$, which means that, if node $i$ is selected as a relay, it transmits the packet with power $P_i^k$.*[5] $P_{\max}$ *is the maximum power that a node is willing to use for cooperation during a long-period multicast program;*[6]
- $\boldsymbol{\pi}^k=[\pi_1^k,\pi_2^k,...,\pi_N^k]$. $\pi_i^k=r_0(n_{Bi}^k+n_{Ri}^k)-c_0n_i^kP_i^k$ *is node $i$'s expected payoff in stage $k$, where $n_{Bi}^k$ is the expected number of packets correctly received from the BS in stage $k$, $n_{Ri}^k$ is the expected number of packets not successfully received from the BS but successfully received from relays in stage $k$, and $n_i^k$ is the expected number of packets that node $i$ should forward to others in stage $k$.*

We consider the scenario where nodes stay in the multicast session for a long time and no one knows exactly when others will leave the multicast service and when the game will end. Then the cooperative multicast process can be viewed as an $N$-player infinite repeated game. In such a repeated game, at the beginning of stage $k$, node $i$ makes a decision of $P_i^k$ based on other nodes' past behaviors according to strategy $s_i$. We assume that any node $i$ is rational [27] and intelligent,[7] and it makes decisions to maximize its long term expected payoff

$$V_i=\sum_{k=0}^{\infty}(\delta)^k\pi_i^k, \tag{1}$$

where $\pi_i^0$ is node $i$'s expected payoff in the current stage, and $\delta\in(0,1)$ is a common discount factor that characterizes how much nodes care about their future payoffs. From the above game model, it can be seen that, in order to motivate nodes to work cooperatively, a proper incentive strategy should be designed such that nodes will get more benefit by following the strategy rather than deviating from it.

## III. Equilibrium Analysis of Cooperative Multicast Game

In the previous section, we have modeled the cooperative multicast process as an $N$-player infinite repeated game. Next, before designing the incentive strategy, it is important to analyze the Nash Equilibrium (NE) points of the game. NE is a steady state of the game where players do not change their actions (or strategies). Obviously, the incentive strategy should enforce cooperation at an NE from which no one has intention to deviate. According to the Folk theorem [28], any *feasible* and *enforceable* action profile of a stage is possible to be an NE of the corresponding infinite repeated game. In our model, *feasible* means the action profile can be realized (where $P_i^k\in[0,P_{\max}]$, $\forall i\in\mathcal{N}$), and *enforceable* means each node's expected payoff should be no worse than that without cooperation. Then the question is which NE is the desired one that should be enforced. Several issues need to be considered when choosing the NE. First, in many applications, nodes should be treated fairly. Second, nodes tend to maximize their expected payoffs. Based on these, we choose the desired NE according to the following criteria: (i) absolute fairness, which requires that the expected payoffs of all nodes are the same, and (ii) maximum payoff, which

[2] Throughout this manuscript, when there are two subscripts for a symbol, e.g. $p_{ij}$, if not explicitly specified, the first one means the transmitter and the second means the receiver.

[3] In this work, we assume the transmission rate of a relay is fixed, which is determined by the packet size and length of the time slot.

[4] In the simulations in Section V, $p_{ij}$ is calculated based on the real distance between node $i$ and node $j$ and the Rayleigh fading model. The simulation results show that our strategy designed with approximation can work efficiently in real scenarios.

[5] If $P_i^k=0$, it means node $i$ does not relay any packet within stage $k$.

[6] In this work, we consider the homogeneous scenario where all nodes have the same $P_{\max}$.

[7] *Rational* means a node makes decisions consistently in pursuit of its own profit, and *intelligent* means a node knows everything about others' rationality and takes into account others' performance when making its decision.

means the desired NE achieves the maximum payoff profile under the absolute fairness constraint.

Next, we find the desired NE action profile according to the above criteria. Denote $\boldsymbol{P} = [P_1, P_2, ..., P_N]$ as the action profile in a stage and $\boldsymbol{\pi}(\boldsymbol{P}) = [\pi_1(\boldsymbol{P}), \pi_2(\boldsymbol{P}), ..., \pi_N(\boldsymbol{P})]$ as the corresponding expected payoff vector. Then finding the desired NE is equivalent to solving the following optimization problem:

$$\max_{\boldsymbol{P}} \pi_i(\boldsymbol{P}) \text{ subject to } \pi_i(\boldsymbol{P}) = \pi_j(\boldsymbol{P}),\ \forall\ i, j \in \mathscr{N}, i \neq j. \tag{2}$$

Since the channel gain between the BS and a node is i.i.d. and the relay is randomly selected, each node has the same chance to be selected as a relay, and that probability is

$$\begin{aligned} q_1 &= \sum_{t=1}^{N-1} \Pr\{\text{node } i \text{ is selected}|t \text{ nodes succeed} \\ &\quad \text{in phase 1}\} \Pr\{t \text{ nodes succeed in phase 1}\} \\ &= \sum_{t=1}^{N-1} \frac{1}{N} \binom{N}{t} p_B^t (1-p_B)^{N-t} \\ &= \frac{1}{N}\left(1 - p_B^N - (1-p_B)^N\right). \end{aligned} \tag{3}$$

Moreover, the probability that node $i$ does not receive the packet correctly from the BS while node $j$ is selected as a relay is

$$\begin{aligned} q_2 &= \sum_{t=1}^{N-1} \Pr\{\text{node } j \text{ is selected}|\text{node } i \text{ does not succeed} \\ &\quad \text{in phase 1 and } t \text{ nodes succeed in phase 1}\} \\ &\quad \times \Pr\{\text{node } i \text{ does not succeed in phase 1} \\ &\quad \text{and } t \text{ nodes succeed in phase 1}\} \\ &= \sum_{t=1}^{N-1} \frac{1}{N-1} \binom{N-1}{t} p_B^t (1-p_B)^{N-t} \\ &= \frac{1}{N-1}\left(1 - p_B - (1-p_B)^N\right). \end{aligned} \tag{4}$$

In a stage, the expected number of packets that node $i$ correctly receives from the BS is $M \cdot p_B$, the expected number of packets received from relay node $j$ when node $i$ does not receive correctly from the BS is either $M \cdot q_2 \cdot p_r(P_j)$ (if $P_j > 0$) or 0 (if $P_j = 0$), and the expected number of packets that node $i$ should transmit is $M \cdot q_1$. In a stage, let $r(x)$ denote the reward that a node gets from another node whose transmission power is $x$, and $c(x)$ denote the cost of a node if it uses power $x$ to cooperate. Then, we have

$$r(x) = \begin{cases} 0 & \text{if } x = 0, \\ M q_2 r_0 p_r(x) & \text{otherwise,} \end{cases} \quad \text{and} \quad c(x) = M q_1 c_0 x. \tag{5}$$

The expected payoff of node $i$ in one stage is

$$\pi_i(\boldsymbol{P}) = M p_B r_0 + \sum_{j \in \mathscr{N}, j \neq i} r(P_j) - c(P_i). \tag{6}$$

Note that the channel between the BS and a node is assumed to be i.i.d., then the packet delivery success probability from the BS to any node ($p_B$) is the same. According to (6), it can be easily shown that to achieve absolute fairness (where $\pi_1(\boldsymbol{P}) = \pi_2(\boldsymbol{P}) = ... = \pi_N(\boldsymbol{P})$), the transmission power of all nodes should be the same. That is $P_1 = P_2 = ... = P_N$. Thus, the expected payoff for node $i$ under the fairness constraint can be represented as:

$$\begin{aligned} \pi_i(P_i) &= M p_B r_0 + (N-1) r(P_i) - c(P_i) \\ &= M p_B r_0 + g(P_i), \end{aligned} \tag{7}$$

where $g(x) \triangleq (N-1)r(x) - c(x)$. Note that, in order to enforce cooperation, the expected payoff of each node at the desired NE should be larger than that when nodes do not cooperate. That is, $\pi_i(P_i) > M p_B r_0$, or equivalently, $g(P_i) > 0$. Assume there exists a $P_i \in (0, P_{\max}]$ such that $g(P_i) > 0$. Then solving (2) is equivalent to solving

$$\max_x g(x), \quad \text{subject to} \quad x \in (0, P_{\max}] \text{ and } g(x) > 0. \tag{8}$$

The solution to (8) is provided in Appendix A, and the desired power for all relays is

$$P^* = \begin{cases} \arg\max_{x>0} g(x) & \text{if } P_{\max} \geq \arg\max_{x>0} g(x); \\ P_{\max} & \text{if } P_{\max} < \arg\max_{x>0} g(x). \end{cases} \tag{9}$$

Therefore, the desired action profile that satisfies the absolute fairness and the maximum payoff criteria is $\boldsymbol{P}^* = [P^*, ..., P^*]$. Note that the system throughput reaches its maximum when everyone uses the maximum power ($P_{\max}$) to cooperate. Interestingly, the solution in (9) shows that $P_{\max}$ is exactly the desired power if it is less than a threshold value. Moreover, the desired NE also has the following property.

*Lemma 1:* The NE with $\boldsymbol{P}^* = [P^*, ..., P^*]$ is Pareto optimal, that is, no one can increase its expected payoff without decreasing any other node's expected payoff by taking a different action.

The proof is provided in Appendix B. Thus, we have found the NE that can achieve absolute fairness and Pareto optimality. Recall that $P_{\max}$ is the maximum power that a node is willing to offer during cooperation for the long-period multicast process. In the following, for simplicity of presentation, we consider only the scenario that $P^* = P_{\max}$, that is, $P_{\max} < \arg\max_{x>0} g(x)$ in (9), and design an incentive strategy to stimulate cooperation at the desired NE point $\boldsymbol{P}^* = [P_{\max}, ..., P_{\max}]$. Note that when $P^* < P_{\max}$ (that is, $P_{\max} \geq \arg\max_{x>0} g(x)$ in (9)), the desired cooperation power is $P^*$, and selfish nodes would not offer more power than $P^*$ for cooperation. The proposed cooperation stimulation strategy in Section IV can still work after replacing $P_{\max}$ with $P^*$.

## IV. Cooperation Stimulation Strategy

### A. *Worst Behavior Tit-for-Tat Incentive Strategy*

In this section, we first propose a *Worst Behavior Tit-for-Tat* incentive strategy and then analyze its effectiveness in perfect and imperfect monitoring scenarios. Before introducing the strategy, we first define a node's behavior as follows:

*Definition 2:* The behavior of node $i$ observed by node $j$ in stage $k$, denoted as $b_{ij}^k$, is defined as the probability that node $j$ successfully receives a packet from node $i$. Mathematically, if node $i$ takes the action $P_i^k$, we have $b_{ij}^k = p_r(P_i^k) = \exp(-D/P_i^k)$.

According to this definition, the observed behavior is a function of a node's action. Note that under our approximation of $p_{ij}$ as $p_r(P_i)$ in Section II, for any node $i$, its behavior observed by all other nodes is the same, which can be further simplified as $b_i^k = p_r(P_i^k)$. Although the action of a node is private information, the node's behavior can be estimated and monitored based on the number of successfully delivered packets from the node. During the multicast process, a node can monitor others' behaviors and adjust its decision accordingly. Then the proposed *Worst Behavior Tit-for-Tat* strategy $s^*$ is as follows:

(i) At the beginning of the multicast session, all nodes cooperate with action $P_{\max}$.

(ii) In each stage, all nodes monitor others' behaviors.

(iii) In stage $k$, if the worst observed behavior among node $i$'s observations in the previous stage, denoted as $b^\dagger$, is greater than a threshold $b_T$, then node $i$ takes an action that gives the same behavior as $b^\dagger$. Otherwise, node $i$ does not cooperate. Mathematically, from the definition of behavior, we have

$$P_i^k = \begin{cases} \frac{-D}{\ln b^\dagger}, & \text{if } b^\dagger > b_T; \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

(iv) If node $i$'s observations of all other nodes' behaviors are the same as its own behavior in the previous $L$ consecutive stages, then node $i$ should use $P_{\max}$ to resume cooperation in stage $k$.

The main idea of this strategy is that if a node deviates from the desired action ($P_{\max}$), all other nodes would behave the same as the deviating node, or even do not cooperate when the deviating behavior is below a threshold. Note that in this strategy, nodes make decisions mainly based on the one-stage observations, and a deviating behavior will result in a reduced reward immediately in the next stage. Then, intuitively, a selfish and rational node will not deviate if the benefit by deviating is less than the corresponding decrease of reward. If a node happens to make a mistake (e.g., use a lower transmission power) due to imperfect monitoring, other nodes take the misbehavior as a deviation and the desired cooperation state may not be achieved. To address this issue, the proposed strategy allows nodes to resume cooperation at the desired level when everyone takes the same behavior for $L$ consecutive stages.

### *B. Equilibrium Analysis of WBTFT Strategy with Perfect Monitoring*

According to the proposed strategy, nodes make decisions based on their monitoring results. Thus, their actions are greatly affected by the accuracy of the monitoring technique. To gain some insight, we first analyze the proposed strategy under the ideal case, where the monitoring process is perfect and everyone knows others' actual behaviors in the previous stage. We analyze the proposed strategy from two aspects: whether the desired NE can be achieved if everyone follows the proposed strategy, and whether the strategy is a *subgame perfect equilibrium strategy* (which means a node cannot get more benefits by deviating to any other strategy from any stage if other nodes follow the proposed strategy).

First, consider the scenario that all nodes follow the proposed strategy from the initial stage. Since the monitoring process is perfect, all nodes' observations are the same (as the behavior by taking action $P_{\max}$) in the first stage. According to the proposed strategy, nodes will cooperate with the desired power $P_{\max}$ in the second stage. Following the same idea, it can be seen that in the subsequent stages, nodes will always cooperate with $P_{\max}$. Hence, the desired cooperation state can be achieved if all nodes follow the proposed strategy.

The next question is whether the proposed strategy is a subgame perfect equilibrium strategy. In the literature, *One-Shot Deviation Principle* is used to analyze the subgame perfection of a strategy [29], which states that a strategy is subgame perfect if a player cannot get more benefit by deviating for one stage and then coming back to follow the strategy again. Based on this principle, the WBTFT strategy can be analyzed as follows.

Denote $P^\dagger$ as the desired action for the current stage according to the proposed strategy $s^*$. Without loss of generality, we assume $P^\dagger$ has been used by all nodes in the previous $K$ stages, where $K < L$. Then if everyone follows $s^*$, node $i$'s long term expected payoff is given by

$$V_i(s^*) = \sum_{k=0}^{L-K-1} (\delta)^k \pi_i(\boldsymbol{P^k}) + \sum_{k=L-K}^{\infty} (\delta)^k \pi_i(\boldsymbol{P^*}), \tag{11}$$

where $P_j^k = P^\dagger$ for $k < L-K$, $\forall j \in \mathscr{N}$. Note that $P^\dagger$ is the action in the current stage according to the incentive strategy (which may or may not be the same as $P_{\max}$). Since we assume that the action $P^\dagger$ has been used by all nodes for $K$ stages, then after another $L-K$ stages, all users will resume cooperation with the desired power level $P_{\max}$ according to the proposed strategy. Therefore, in (11), the first term denotes the summation of payoffs before resuming cooperation, while the second term denotes the summation of payoffs after everyone resumes cooperation at the desired power level. Note that with perfect monitoring, any one-shot deviation behavior of node $i$ will be recognized by other nodes, who will react accordingly in the next $L$ stages. Assume node $i$ employs a one-shot deviation strategy $s'$ in the current stage that gives $P_i'^0 = P_i' < P^\dagger$. The corresponding long term expected payoff is given by

$$V_i(s') = \pi_i(\boldsymbol{P}'^0) + \sum_{k=1}^{L} (\delta)^k \pi_i(\boldsymbol{P}'^k) + \sum_{k=L+1}^{\infty} (\delta)^k \pi_i(\boldsymbol{P}^*), \tag{12}$$

where $P_j'^0 = P^\dagger$ ($\forall j \neq i$) and $P_j'^k = P_i'$ ($\forall j \in \mathcal{N}$, $0 < k < L+1$) according to the WBTFT strategy $s^*$. According to the *One-Shot Deviation Principle*, $s^*$ is a subgame perfect equilibrium strategy if $V_i(s^*) \geq V_i(s')$. It is shown in Appendix C that this inequality holds under certain conditions, which are summarized in the following proposition.

*Proposition 1: In the cooperative multicast game with perfect monitoring, the WBTFT strategy is a subgame perfect equilibrium strategy if the following conditions are satisfied:*
*(a)* $G(x) \triangleq \frac{\sum_{k=1}^{L} \delta^k}{\sum_{k=0}^{L} \delta^k}(N-1)r(x) - c(x) > 0$, $\forall x \in (-D/\ln b_T, P_{\max}]$, *and (b)* $P_{\max} \leq \arg\max_{x>0} G(x)$.

Given that $\delta$ is close to 1, we can select $b_T$ and $L$ properly such that the above two conditions are satisfied. Hence, cooperation can be stimulated using the proposed WBTFT strategy.

### *C. Refined Strategy and Equilibrium Analysis with Imperfect Monitoring*

The previous analysis is based on the assumption that the monitoring process is perfect. Here, we extend the analysis to a more realistic case, where nodes' monitoring results may be erroneous. In a distributed scenario, a node can estimate others' behaviors based on the number of successfully received packets from them. However, due to packet loss and bit errors, the monitoring results may be erroneous, and therefore undesired actions may be carried out. To address this issue, we propose an interval based estimation method, and derive the equilibrium conditions for the proposed strategy with imperfect monitoring.

Typically, estimation of another user's behavior is based on some observed information. In this game model, since the behavior is defined in terms of probability, then any node $j$'s observed information of another node $i$ can be defined as the proportion of packets that node $j$ correctly receives from node $i$ among all the packets that node $i$ should transmit. Denote $y_{ij}^k = \frac{n_{ij}^k}{\overline{n}}$ as the information that node $j$ observes from node $i$ in stage $k$, where $n_{ij}^k$ is the number of packets that node $j$ receives correctly from node $i$, and $\overline{n}$ is the average number of packets that a node should transmit in a stage. Note that if node $i$ transmits $\overline{n}$ packets with a certain power that results in a behavior $b_i^k$, then statistically $n_{ij}^k$ follows a binomial distribution $B(\overline{n}, b_i^k)$, and therefore the mean of the observed information $y_{ij}^k$ is $b_i^k$. Based on this fact, we propose an interval based estimation method as follows. We divide [0,1] into $m+1$ intervals, $[0,\beta_1], (\beta_1,\beta_2], ...(\beta_m, 1]$, and within the $l$th ($0 \leq l \leq m$) interval, we select a behavior $\hat{b}_l$ to represent the behavior level. If node $j$'s observed information of node $i$, $y_{ij}^k$, falls in the $l$th interval, then behavior level $\hat{b}_l$ is selected as node $j$'s estimation result of node $i$'s behavior in stage $k$, i.e., $b_{ij}^k = \hat{b}_l$. The interval boundaries ($\{\beta_l\}$) and behavior levels ($\{\hat{b}_l\}$) are designed according to a parameter $\epsilon$ such that if node $i$ transmits $\overline{n}$ packets with behavior $\hat{b}_l$, then any other node's estimation error probability (i.e., the probability that the estimation of node $i$'s behavior is not $\hat{b}_l$) is no larger than a small value $2\epsilon$. The design procedure is shown in Algorithm 1 for $1 \leq l \leq m$. Note that $\hat{b}_0 = 0$.

**Algorithm 1** Determination of $\hat{b}_l$ and $\beta_l$ ($l = 1, 2, ..., m$).

1: Select a index value $m = 100$ and let $\hat{b}_m = b^* \triangleq \exp(-D/P_{\max})$;
2: Find $\beta_m \in \{\frac{1}{\overline{n}}, \frac{2}{\overline{n}}, ..., \frac{\overline{n}-1}{\overline{n}}, 1\}$ such that $\Pr\{y_{ij}^k \leq \beta_m | b_i^k = \hat{b}_m, n_i^k = \overline{n}\} = F(\overline{n}\beta_m; \overline{n}, \hat{b}_m) = \epsilon$;
3: Set $l = m$;
4: **while** $\beta_l > b_T$ **do**
5: **for** $\beta_{l-1} = \beta_l - \frac{1}{\overline{n}}$ to 0 with step $\frac{1}{\overline{n}}$ **do**
6: Find $\hat{b}_{l-1}$ that satisfies $F(\overline{n}\beta_{l-1}; \overline{n}, \hat{b}_{l-1}) = \epsilon$;
7: **if** $1 - F(\overline{n}\beta_l; \overline{n}, \hat{b}_{l-1}) \leq \epsilon$ **then**
8: break and go to step 11;
9: **end if**
10: **end for**
11: Set $l = l - 1$;
12: **end while**
13: Adjust the index values $l \to 1$, $l+1 \to 2$,..., $m \to m - l + 1$.

In this algorithm, $F(x; n, p)$ is the cumulative distribution function of a binomial random variable $X \sim B(n, p)$, and $\beta_l$ is selected from the set $\{\frac{1}{\overline{n}}, \frac{2}{\overline{n}}, ..., \frac{\overline{n}-1}{\overline{n}}, 1\}$. Note that the number of intervals $m+1$ is determined by the system parameters (i.e., $P_{\max}$, $D$, $\overline{n}$, $b_T$ and $\epsilon$), and its value is unknown until all the intervals are determined. Since we start searching the intervals from the highest level, we initialize $m$ using an arbitrarily chosen large number in the first step, and in this work, we use $m$=100 as an example. Then, in the last step, after all intervals are determined, we adjust $m$ and the interval indices accordingly based on the number of determined intervals. The details of Algorithm 1 are explained as follows.

In step 1, we set the desired cooperation behavior $b^*$ (where nodes cooperate with $P_{\max}$) as the highest behavior level $\hat{b}_m$. In step 2, we determine $\beta_m$ according to the design parameter $\epsilon$, such that if node $i$ takes behavior $\hat{b}_m$, the probability that $y_{ij}^k$ falls in $[0, \beta_m]$ is $\epsilon$.[8] Then from step 4 to step 12, the boundary points and the corresponding behavior levels for other intervals are found one by one until $\beta_l$ is smaller than the threshold $b_T$. We use the $(m-1)$th interval as an example. Its right boundary $\beta_m$ is already determined

[8]Note that $\epsilon$ is selected from the set $\{F(1; \overline{n}, b^*), F(2; \overline{n}, b^*), ..., F(\overline{n}-1; \overline{n}, b^*), 1\}$ so that there exists such a $\beta_m$ in step 2 of Algorithm 1.

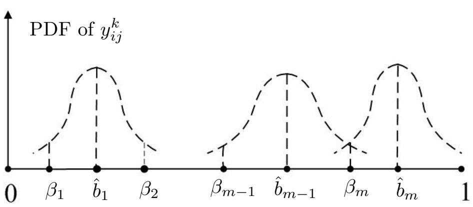


Fig. 2. Probability density function (PDF) of $y_{ij}^k$.

when the $m$th interval is dealt with. So we need to determine its left boundary $\beta_{m-1}$. We search the left boundary $\beta_{m-1}$ from $\beta_m - \frac{1}{\overline{n}}$ to 0 with a step size $\frac{1}{\overline{n}}$. For each possible value of $\beta_{m-1}$, (in step 6) we find the corresponding $\hat{b}_{m-1}$ such that if node $i$ takes behavior $\hat{b}_{m-1}$, the probability that $y_{ij}^k$ falls in $[0, \beta_{m-1}]$ is $\Pr\{y_{ij}^k \in [0, \beta_{m-1}] | b_i^k = \hat{b}_{m-1}, n_i^k = \overline{n}\} = F(\overline{n}\beta_{m-1}; \overline{n}, \hat{b}_{m-1}) = \epsilon$ (i.e., the probability in the left tail in the PDF figure of $y_{ij}^k$, as shown in Fig. 2, is $\epsilon$). The desired searching result is the largest value of $\beta_{m-1}$ such that its corresponding $\hat{b}_{m-1}$ and itself satisfy $\Pr\{y_{ij}^k \in (\beta_m, 1] | b_i^k = \hat{b}_{m-1}, n_i^k = \overline{n}\} = 1 - F(\overline{n}\beta_m; \overline{n}, \hat{b}_{m-1}) \le \epsilon$ (i.e., the probability in the right tail of the PDF figure of $y_{ij}^k$ is no larger than $\epsilon$).[9]

Using the above estimation method, if nodes take behaviors from the set $\mathscr{B} = \{0, \hat{b}_1, ..., \hat{b}_m\}$, the monitoring results will be more accurate when $\epsilon$ approaches zero. Denote $\mathscr{P} = \{0, \hat{P}_1, ..., \hat{P}_m\}$ as the set of power levels that are associated with the behavior levels in $\mathscr{B}$, where $\hat{P}_l = -D/\ln \hat{b}_l$, ($\hat{b}_l \in \mathscr{B}$, $0 < l \le m$). Then, the WBTFT strategy in Section IV-A can be refined as follows. In step (ii), nodes employ the above interval based estimation to estimate each other's behavior level. In addition, a communication period is added at the end of each stage where nodes exchange the worst observed behavior level index.[10] In step (iii), at the beginning of a stage, nodes choose their transmission power from the set $\mathscr{P}$ according to the smallest behavior level index obtained at the communication period in the previous stage. Note that when all nodes follow the proposed strategy, estimation errors may occur (even with small probabilities), which will lead the cooperation to a lower power level (e.g., the worst behavior is estimated as $b^\dagger < b^*$ and then all nodes will cooperate with $P^\dagger = -D/\ln b^\dagger$). According to step (iv) of the proposed strategy, if there is no estimation error from then on, after $L$ stages, nodes will resume cooperation at the desired power level $P_{\max}$.

The next question is whether the refined strategy is a subgame perfect equilibrium strategy when there are estimation errors. With imperfect monitoring, a node is uncertain about others' behaviors in the previous stage, and it is also not sure how others will react in the next stage if it takes a certain action. Therefore, it is difficult to find a closed-form formulation of the long term expected payoff due to the infinite number of possible future paths, which makes it difficult to check the equilibrium condition. Note that in reality, a node may be *bounded rational* [30] when making a decision, that is, it only cares those outcomes with large probabilities rather than considering all possible cases. Based on this fact, we analyze the strategy under the bounded rational assumption.

We assume a bounded rational node $i$ has the following characteristics. When node $i$ estimates other nodes' behaviors and its own expected payoff, if node $t$ takes an action that gives the behavior $b_t^k$ in stage $k$, node $i$ will consider the scenario where another node $j$'s observed information $y_{tj}^k$ falls in the interval $(u(b_t^k), v(b_t^k))$, which satisfies $\Pr\{y_{tj}^k \le u(b_t^k) | b_t^k\} \le \eta$ and $\Pr\{y_{tj}^k \ge v(b_t^k) | b_t^k\} \le \eta$, where $\eta$ is a small value close to zero; and node $i$ will ignore the small-probability scenarios where $y_{tj}^k$ falls outside the interval $(u(b_t^k), v(b_t^k))$. Note that when analyzing the equilibrium of the proposed strategy, a node always assumes that others will follow this strategy. Therefore, a node can predict others' behaviors in future stages and estimate its long term payoff accordingly. It can be seen that when estimating the long term payoff, instead of considering all possible outcomes, the bounded rational nodes will focus on the monitoring results (predicted for future stages) within a range with a large coverage probability (equal to or greater than $1 - 2\eta$). Intuitively, when $\eta$ becomes smaller, the estimated payoff is closer to the expected payoff. On the other hand, a smaller $\eta$ may result in higher computational complexity that is difficult to handle. Note that when all nodes follow the refined strategy, they will choose behaviors in the set $\mathscr{B}$. In this case, according to the interval based estimation, it is easy to check that a bounded rational node $i$ with $\eta \ge \epsilon$ will only consider the scenario that its observed information from another node $j$ with behavior $b_j^k = \hat{b}_l$ falls in the $l$th interval ($\beta_l \le u(b_j^k) < v(b_j^k) \le \beta_{l+1}$), and node $i$'s estimation result of node $j$'s behavior is exactly $\widetilde{b}_{ji}^k = \hat{b}_l (= b_j^k)$. Similarly, when estimating its long term payoff, node $i$ will only consider the scenario that all nodes' estimations in future stages are correct, which greatly simplifies the calculation of the estimated payoff. In this work, we use $\eta = \epsilon$ when analyzing the equilibrium.

Similar to the analysis of the scenario with perfect monitoring, the *One-Shot Deviation Principle* is also employed to study the equilibrium condition for the refined strategy.

[9]Note that $\beta_l$ is selected from a discrete set, and we may not guarantee that both $\Pr\{y_{ij}^k \in [0, \beta_l] | b_i^k = \hat{b}_l, n_i^k = \overline{n}\}$ and $\Pr\{y_{ij}^k \in (\beta_{l+1}, 1] | b_i^k = \hat{b}_l, n_i^k = \overline{n}\}$ are equal to $\epsilon$. In this work, we let $\Pr\{y_{ij}^k \in [0, \beta_l] | b_i^k = \hat{b}_l, n_i^k = \overline{n}\} = \epsilon$ and $\Pr\{y_{ij}^k \in (\beta_{l+1}, 1] | b_i^k = \hat{b}_l, n_i^k = \overline{n}\} \le \epsilon$ for $0 < l < m$.

[10]We assume the information exchange is perfect, which can be achieved by introducing some trust management scheme. The discussion of trust management is beyond the scope of this paper, which will be considered in future work. With the information exchange, all nodes will make the same decisions by following the refined strategy, so that cooperation can be resumed when undesired decisions are carried out.

Assume that all nodes follow the refined strategy in the past stages. Denote the worst behavior level index in the previous stage as $l^\dagger$. Without loss of generality, we assume that if all nodes follow the strategy, they will cooperate with $P_{l^\dagger}$ for $L$ consecutive stages and then resume cooperation at $P_{\max}$. Then according to the bounded rational assumption and interval based estimation, node $i$ will consider that all nodes' estimations are correct, and the estimated long term payoff can be represented as

$$\widetilde{V}_i^* = \sum_{k=0}^{L-1}(\delta)^k \pi_i(P_{l^\dagger}) + \sum_{k=L}^{\infty}(\delta)^k \pi_i(P_{\max}). \tag{13}$$

If node $i$ decides to take a one-shot deviation in the current stage with $P_i' < P_{l^\dagger}$ that gives behavior $b_i^0 \in (\hat{b}_l, \hat{b}_{l+1}]$, where $b_T < \hat{b}_{l+1} \leq \hat{b}_{l^\dagger}$. Based on the bounded rational assumption with $\eta = \epsilon$, node $i$ will believe that any other node $j$'s observation $y_{ij}^0$ falls in the interval $(u(b_i^0), v(b_i^0))$, where $\Pr\{y_{ij}^0 \leq u(b_i^0)|b_i^0\} \leq \epsilon$ and $\Pr\{y_{ij}^0 \geq v(b_i^0)|b_i^0\} \leq \epsilon$. From the interval based estimation, it can be shown that $u(b_i^0) \geq \beta_l$ and $v(b_i^0) \leq \beta_{l+2}$. Therefore, node $i$ will believe that node $j$'s estimated behavior $\widetilde{b}_{ij}^0$ is either $\hat{b}_l$ or $\hat{b}_{l+1}$. Denote $w_{ij}(\hat{b}_{l+1}|b_i^0)$ as node $i$'s estimated probability[11] that node $j$'s estimation is $\widetilde{b}_{ij}^0 = \hat{b}_{l+1}$ given $b_i^0 \in (\hat{b}_l, \hat{b}_{l+1}]$, which can be calculated as

$$w_{ij}(\hat{b}_{l+1}|b_i^0) = \frac{\Pr\{y_{ij}^0 \in (\beta_{l+1}, v(b_i^0))|b_i^0\}}{\Pr\{y_{ij}^0 \in (u(b_i^0), v(b_i^0))|b_i^0\}}. \tag{14}$$

Since nodes exchange their estimated worst behavior levels at the end of each stage, node $i$'s estimated probability that the worst behavior level in the current stage is $\hat{b}_{l+1}$, denoted as $w_i(\hat{b}_{l+1}|b_i^0)$, is given by

$$w_i(\hat{b}_{l+1}|b_i^0) = \prod_{j\in\mathcal{N}, j\neq i} w_{ij}(\hat{b}_{l+1}|b_i^0). \tag{15}$$

Then node $i$'s estimated probability that the worst behavior level is $\hat{b}_l$ is $w_i(\hat{b}_l|b_i^0) = 1 - w_i(\hat{b}_{l+1}|b_i^0)$, and its estimated long term payoff can be represented as

$$\begin{aligned} & \widetilde{V}_i' \\ = \; & M p_B r_0 + (N-1) r(\hat{P}_{l^\dagger}) - c(P_i') \\ & + \sum_{k=1}^{L}(\delta)^k \left\{\pi_i(\hat{P}_l) w_i(\hat{b}_l|b_i^0) + \pi_i(\hat{P}_{l+1}) w_i(\hat{b}_{l+1}|b_i^0)\right\} \\ & + \sum_{k=L+1}^{\infty}(\delta)^k \pi_i(P_{\max}). \end{aligned} \tag{16}$$

Following analysis similar to that in Section IV-B, the conditions for $\widetilde{V}_i^* \geq \widetilde{V}_i'$ can be derived, which are in the following proposition.

*Proposition 2: In the cooperative multicast game, the refined WBTFT strategy is a subgame perfect equilibrium strategy with imperfect monitoring under the bounded rational assumption if the following conditions are satisfied:*
$(a)$ *Conditions in Proposition 1 are satisfied.*
$(b)$ $\widetilde{G}(x,l) > 0 \;\; \forall\, 0 \leq l < m$ *and* $x \in (\hat{b}_l, \hat{b}_{l+1})$,
*where*

$$\widetilde{G}(x,l) = c\left(\frac{-D}{\ln x}\right) - \left[w_i(\hat{b}_l|x)c(\hat{P}_l) + w_i(\hat{b}_{l+1}|x)c(\hat{P}_{l+1})\right].$$

The proof is in Appendix D. The above conditions can be satisfied by appropriately choosing the parameters $\epsilon$ and $L$, and hence the refined strategy maintains an equilibrium even with the imperfect monitoring.

### D. WBTFT with Optimal Power Allocation Scheme

Our previous analysis is based on the assumption that nodes use the same transmission power for all time slots during one stage. However, nodes may employ different power levels for different time slots or only transmit a portion of packets when being selected as a relay in a stage. Here, we discuss the power allocation scheme within a stage and show that WBTFT strategy actually achieves the optimal power allocation when the parameter $b_T$ is no less than $1/e$.

First, we consider the scenario where nodes vary the transmission power within a stage. Suppose node $i$ transmits $k$ packets in a stage, the power allocated for the $t$th packet is denoted as $P_t$, $t = 1, 2, ..., k$, and the total cost in this stage is $c_i = \sum_{t=1}^{k} c_0 P_t$. We further denote the average behavior of node $i$ observed by any other one as $\overline{b}_i = \frac{1}{k}\sum_{t=1}^{k} p_r(P_t)$. Obviously, a node prefers to provide a large $\overline{b}_i$ with low cost. Then given a desired behavior $\overline{b}_i$, the optimal power allocation scheme should achieve the lowest cost. Based on this idea, the following lemma can be proved.

*Lemma 2:* To obtain the same $\overline{b}_i$, where $\overline{b}_i > \frac{1}{e}$, the optimal power allocation scheme that gives the lowest cost $c_i$ is $P_t = \overline{P} = \frac{-D}{\ln(\overline{b}_i)}$, $\forall\, t = 1, 2, ..., k$.

The proof is provided in Appendix E. Lemma 2 implies that given the desired observed information by others, using the same power to transmit all packets in a stage gives the smallest cost. According to the WBTFT strategy, nodes make decisions based on the observed information. Therefore, to gain more profit, they will keep the transmission power unchanged within a stage.

The next question is whether a node will transmit all the time when being selected as a relay in a stage. Suppose node $i$ should transmit $n_i$ packets in a stage, and its target behavior is denoted as $b^\dagger$, which is assumed to be greater than $1/e$. Lemma 2 implies that node $i$ will transmit packets with the same power. However, it can decide to transmit with a certain probability (or equivalently, only transmit a portion of $n_i$ packets). Denote $\alpha_i$ as the probability (or the proportion of $n_i$) that node $i$ decides to transmit with fixed power $\overline{P}_i$ when being selected as a relay. Then the expected number of packets that any other node $j$ can successfully receive is $n_{ij} = n_i \alpha_i p_r(\overline{P}_i)$. Equivalently, the average behavior that any other node should observe in the stage

[11] Here node $i$'s estimated probability of an event means the probability that node $i$ estimates the event happens.

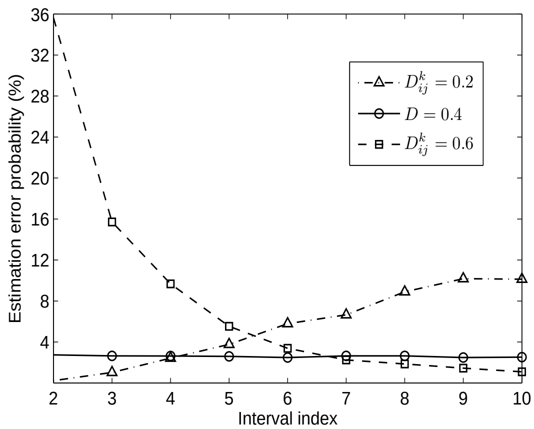


Fig. 3. Estimation error probability versus interval index

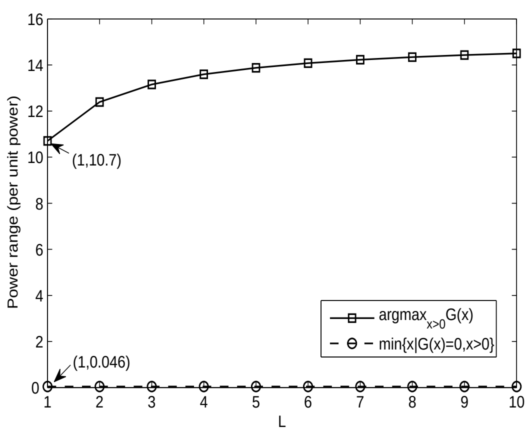


Fig. 4. Conditions in Proposition 1.

can be represented as $\overline{b}_i = \alpha_i p_r(\overline{P}_i)$, and the corresponding cost in the stage is given by $c_i = c_0 n_i \alpha_i \overline{P}_i$. Since nodes are selfish and tend to minimize their cost, determining the optimal values of $\alpha_i$ and $\overline{P}_i$ is equivalent to solving the following optimization problem:

$$\min \ \alpha_i \overline{P}_i \quad \text{subject to} \quad \overline{b}_i = \alpha_i p_r(\overline{P}_i) = b^\dagger. \tag{17}$$

The solution to (17) is given in Appendix H, which is

$$\begin{cases} \alpha_i = e b^\dagger, \overline{P}_i = D, & \text{if } b^\dagger < \frac{1}{e}; \\ \alpha_i = 1, \overline{P}_i = \frac{-D}{\ln b^\dagger}, & \text{if } b^\dagger \geq \frac{1}{e}. \end{cases} \tag{18}$$

It can be seen that if the target behavior $b^\dagger$ is no less than $1/e$, a node will always transmit with fixed power when being selected as a relay. In addition, if $b_T \geq 1/e$, then the refined strategy achieves the optimal power allocation within a stage.

### *E. Further Discussion of Designing Issues*

In the previous discussion, we have refined the WBTFT strategy with an interval based estimation method to address the issue of imperfect monitoring. However, the design of intervals is based on the approximation that the packet delivery success probability between node $i$ and node $j$ at each stage is $p_{ij} \approx \exp(\frac{-D}{P_i})$, where $D$ depends on the long term average channel condition between node $i$ and node $j$. In real applications, given the mobility model, the average channel condition between two mobile nodes may fluctuate from stage to stage, which may result in large estimation errors when using the intervals designed according to $D$. To address this problem, we make the following adjustment to the interval based estimation scheme. First, we design a general behavior level set $\mathscr{B} = \{0, \hat{b}_1, ..., \hat{b}_m\}$ (with interval boundaries $\{\beta_l\}$) according to Algorithm 1 based on $D$, and find its corresponding power set $\mathscr{P}$. Note that in the refined strategy, nodes choose transmission power from the same power set $\mathscr{P}$ based on the worst behavior level index. Next, in stage $k$, node $j$ estimates the current average channel condition between node $i$ and itself (either by exchanging topology information or using pilot signals), and calculates the average value of $\frac{N_0(1-2^{\gamma_1})}{\sigma_{ij}^2}$, denoted as $D_{ij}^k$. Then, node $j$ adjusts its estimation of node $i$'s behavior level as follows. In stage $k$, given $\mathscr{P}$ and $D_{ij}^k$, if node $i$ chooses power $\hat{P}_l \in \mathscr{P}$ ($1 \leq l \leq m$), node $j$'s observed behavior of node $i$ is $\hat{b}_{ijl}^k = \exp\left(-D_{ij}^k/\hat{P}_l\right) = (\hat{b}_l)^{D_{ij}^k/D}$, and we update node $j$'s observed behavior level set of node $i$ as $\mathscr{B}_{ij}^k = \{0, \hat{b}_{ij1}^k, \hat{b}_{ij2}^k, ..., \hat{b}_{ijm}^k\}$. We then update the interval boundaries for each behavior level in $\mathscr{B}_{ij}^k$, by using $\beta_{ijl}^k = (\beta_l)^{D_{ij}^k/D}$. Then node $j$ estimates node $i$'s behavior level according to the new intervals at the end of stage $k$. That is, if node $j$'s observed information of node $i$, $y_{ij}^k$, falls in the $l$th interval $(\beta_{ijl}^k, \beta_{ij(l+1)}^k]$, then node $j$'s estimation result of node $i$'s behavior level is $\hat{b}_{ijl}^k$. Note that for the original behavior levels in $\mathscr{B}$, the intervals are designed such that the estimation error probability for each behavior level is no larger than $2\epsilon$ (i.e., $\Pr\left\{y_{ij}^k \in [0, \beta_l] \cup y_{ij}^k \in (\beta_{l+1}, 1] | b_i^k = \hat{b}_l, n_i^k = \overline{n}\right\} \leq 2\epsilon$). Unfortunately, this feature cannot be maintained for the adjusted intervals. The estimation error probability decreases with the behavior level index when $D_{ij}^k > D$ and increases when $D_{ij}^k < D$. An example is shown in Fig. 3. To guarantee the accuracy of estimation, we set a threshold value $\epsilon_T$. In stage $k$, if node $j$'s observed information from node $i$, $y_{ij}^k$, falls in the $l$th adjusted interval, then node $j$ first calculates $p_e = \Pr\left\{y_{ij}^k \in [0, \beta_{ijl}^k] \cup y_{ij}^k \in (\beta_{ij(l+1)}^k, 1] | b_i^k = \hat{b}_{ijl}^k, n_i^k = \overline{n}\right\}$ and compares $p_e$ with $\epsilon_T$. If $p_e > \epsilon_T$, then the estimation result is unreliable and will be discarded. At the end of each stage, nodes only exchange the worst behavior level index from their reliable estimation results and make decisions accordingly.

## V. Performance Evaluation

In this section, we evaluate the performance of the proposed strategy using simulations. Before showing the simulation results, it is worth investigating how to select the design parameters to satisfy the equilibrium conditions.

First, consider the conditions in Proposition 1. As shown in the proof of Proposition 1, $G(x)$ has the same property as $g(x)$. That is, if there exists an $x$ such that $G(x) > 0$, then $G(x)$ monotonically increases in the interval $(\min\{x|G(x)=0, x>0\}, \arg\max_{x>0} G(x)]$. Then, the conditions in Proposition 1 become $\min\{x|G(x) = 0, x > 0\} \leq -D/\ln(b_T)$ and $P_{\max} \leq \arg\max_{x>0} G(x)$. Consider a system setup with $N = 10$, $p_B = 0.4$, $D = 0.4$, $\delta = 0.99$ and $r_0/c_0 = 100$. Fig. 4 shows $\min\{x|G(x) = 0, x > 0\}$ and $\arg\max_{x>0} G(x)$ with different $L$. It can be seen that as $L$ increases, $\arg\max_{x>0} G(x)$ becomes larger, while $\min\{x|G(x) = 0, x > 0\}$ is close to zero, which implies the conditions are easier to satisfy when $L$ becomes larger. Note that to achieve optimal power allocation in a stage, $b_T$ should be no less than $1/e$. Therefore, we can select $b_T$ such that $b_T \geq \max\{\exp(-D/\min\{x|G(x) = 0, x > 0\}), 1/e\}$. In this system setup, if the unit power is $P_{\text{UNIT}} = 10$ mW, and $P_{\max}$ is less than 107 mW, then selecting $L = 1$ and $b_T = 1/e$ can satisfy the required conditions.

Next, consider the conditions in Proposition 2. It can be seen that condition $(b)$ of Proposition 2 is related to the behavior intervals, which are determined based on the parameter $\epsilon$. According to the interval based estimation, a small $\epsilon$ is preferred because the estimation result becomes more and more accurate when $\epsilon$ approaches zero. However, a small $\epsilon$ may violate the equilibrium conditions. Thus, given a network setup, the smallest $\epsilon$ that satisfies the conditions in Proposition 2 can be found via numerical methods. Some values of the minimum $\epsilon$ under different $P_{\max}$ are given in Table I. Here, the design parameter $\overline{n}$ is 1000, and other parameters are the same as that in Fig. 4. It can be seen that the minimum $\epsilon$ values are all smaller than 0.02, which is acceptable.

We use simulations to demonstrate the effectiveness of the proposed incentive strategy. The simulated multicast network consists 10 mobile nodes who are randomly deployed within a circular area with radius 25m, and a BS located 250m away from the center of the circular area. Each node is moving according to the random waypoint model: a node randomly chooses a destination within the circle and moves forward to the destination at a velocity uniformly chosen in [0.5m/s, 2.5m/s]. When arriving at the destination, the node will choose a new location and a new speed to move on. The BS broadcasts a packet every 10 ms with rate 1 Mbps over a wireless channel with 1MHz bandwidth. The maximum power that a node offers for cooperation is 40 mW. When selected as a relay, a node can choose the power between 0 and 40 mW to rebroadcast the packet with rate 1 Mbps. All the wireless channels undergo path loss with exponent 2 and Rayleigh fading, and the average received SNR from the BS to a node with distance 250m is 0 dB. The average received SNR between a pair of nodes with distance 25m and unit transmission power 8 mW is 2 dB. The packet delivery success probability (either from BS to a node or between two nodes) is calculated is calculated based on the real distance (either between the BS and a node or between two nodes) and the Rayleigh fading model. The design parameter $D$ is calculated by averaging results from simulations of the above mobility model in $10^7$ time slots and is approximated by $D = 0.40$. Other simulation parameters are listed in Table II.

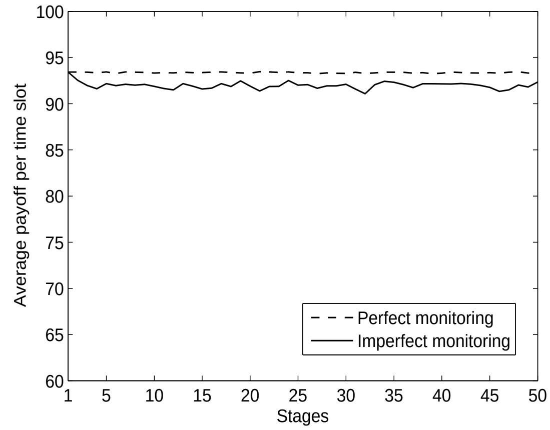


Fig. 5. Average payoff with perfect and imperfect monitoring

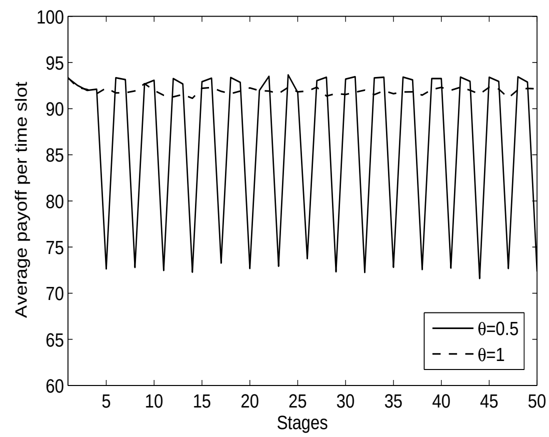


Fig. 6. Average payoff with deviating node

Fig. 5 shows the average payoff of all nodes in different stages when everyone follows the proposed strategy. The intervals are designed according to $\overline{n} = 1000$ and $\epsilon = 0.0158$. The corresponding stage length is $M = 10103$ time slots. Other parameters are $L = 1$, $b_T = 1/e$ and $\epsilon_T = 0.1$. This result is based on 50 simulation runs, each with 50 stages, and the average payoff of all nodes is plotted. The results with perfect monitoring process where all nodes cooperate at the desired power level are also plotted as a benchmark. It can be seen that the WBTFT strategy with interval based estimation can achieve cooperation efficiently

TABLE I
MINIMUM $\epsilon$ VERSUS $P_{\max}/P_{\text{UNIT}}$.

| $P_{\max}/P_{\text{UNIT}}$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Minimum $\epsilon$ | 0.0119 | 0.0113 | 0.0106 | 0.0126 | 0.0158 |
| $P_{\max}/P_{\text{UNIT}}$ | 6 | 7 | 8 | 9 | 10 |
| Minimum $\epsilon$ | 0.0090 | 0.0121 | 0.0132 | 0.0181 | 0.0133 |

TABLE II
SIMULATION PARAMETERS.

| Parameter | Value | Parameter | Value | Parameter | Value |
|---|---|---|---|---|---|
| Distance from BS to the center of the circular area | 250 m | $\gamma_0$ | 1 Mbps | $D$ | 0.4 |
| Radius of the circular area | 25 m | $\gamma_1$ | 1 Mbps | $c_0$ | 2 |
| SNR (from BS to a node with distance 250 m) | 0 dB | $P_{\max}$ | 40 mW | $r_0$ | 100 |
| SNR (between two nodes with distance 25 m and power $P_{\text{UNIT}}$) | 2 dB | $P_{\text{UNIT}}$ | 8 mW | $\delta$ | 0.99 |

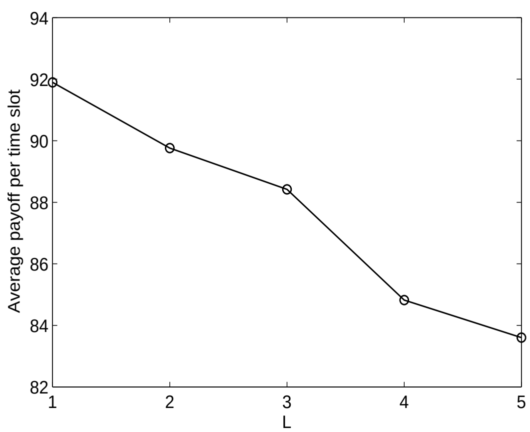


Fig. 7. Average payoff under different $L$

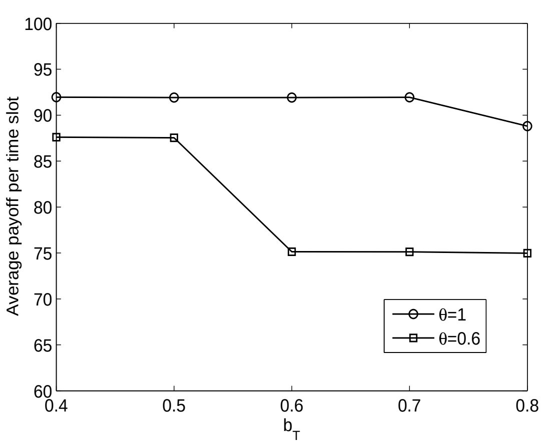


Fig. 8. Average payoff under different $b_T$

with an average payoff of 92.93 per time slot, which is close to the desired cooperation state (with perfect monitoring) with an average payoff of 94.33 per time slot.

In Fig. 6, the payoffs of a node with two different strategies, the proposed strategy and a deviation strategy, are compared. We define the deviating ratio $\theta$ as the ratio between the deviating behavior and the desired behavior generated from the proposed strategy. Specifically, if the desired behavior in stage $k$ is $b^k$ for all nodes, then a node with a deviating ratio $\theta$ will take an action $P_i^k$ such that $b_i^k = \exp(-D/P_i^k) = \theta \cdot b^k$. In this simulation, node 1 deviates with $\theta = 0.5$ in stage 4, stage 7, stage 10,..., stage 49, and in other stages, it follows the proposed WBTFT strategy. Other parameters are the same as those in Fig. 5. It can be seen that whenever node 1 deviates, its misbehavior will be identified and the corresponding payoff is reduced in the next stage. Therefore, the proposed strategy is able to punish the deviating node. The average payoff with deviation is 87.40 per time slot, which is smaller than that without deviation. Hence the proposed strategy can motivate cooperation efficiently.

Fig. 7 shows the impact of $L$ on the performance. Other parameters are the same as the previous figures and there is no deviating node. It can be seen that as $L$ increases, the performance slightly degrades. This is mainly because when estimation error occurs, a smaller $L$ can help resume cooperation faster. Therefore, when designing the incentive strategy for real applications, a small $L$ is preferred so that cooperation can be easily recovered when undesired decisions are carried out.

Fig. 8 shows the performance with different $b_T$. Other parameters are the same as those in Fig. 5. Note that $\theta = 1$ means no one deviates. It can be seen that changing of $b_T$ has more impact on the performance with deviating nodes. The reason is that when $b_T$ becomes larger, the punishment becomes more severe, since no one will cooperate when a misbehavior that falls below $b_T$ is detected. Although a large $b_T$ is good to provide incentives, it may also degrades the performance even everyone follows the proposed strategy, since the cooperative behavior might be estimated as a deviation if it falls below $b_T$ due to estimation errors.

Next, we evaluate the performance of the proposed strategy under different designing parameter $\overline{n}$, and the result is shown in Fig. 9. The incentive strategy is designed according to $\overline{n} = 100$, 500 and 1000 (such that $M = 1010$,

TABLE III
BEHAVIOR LEVELS FOR $\overline{n} = 100, 500, 1000$.

| $l$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Behavior for $\overline{n} = 100$ | 0.382 | 0.596 | 0.782 | 0.923 | | | | | | |
| Behavior for $\overline{n} = 500$ | 0.339 | 0.432 | 0.526 | 0.621 | 0.710 | 0.792 | 0.864 | 0.923 | | |
| Behavior for $\overline{n} = 1000$ | 0.390 | 0.457 | 0.526 | 0.593 | 0.659 | 0.722 | 0.781 | 0.835 | 0.883 | 0.923 |

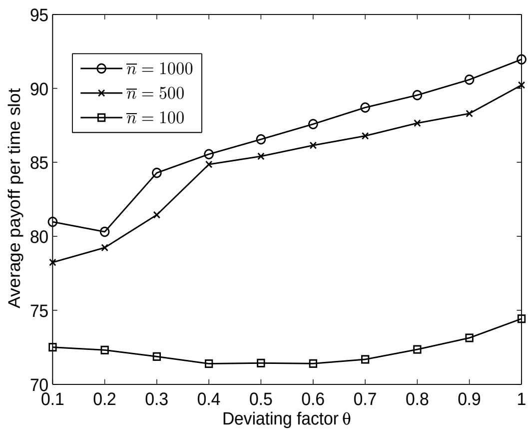


Fig. 9. Average payoff under different $\overline{n}$

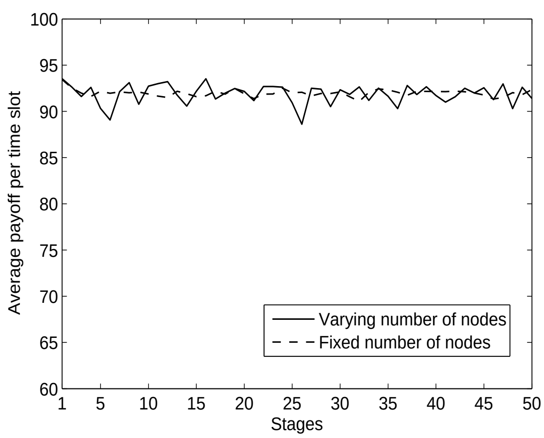


Fig. 10. Average payoff with fixed and dynamic membership

5052 and 10103), and the corresponding minimum $\epsilon$ is $0.0210, 0.0173,$ and $0.0158$, respectively. Other parameters are the same as those in Fig. 5. It can be seen that among the three designs, $\overline{n} = 1000$ and $\overline{n} = 500$ achieve relatively close performance, with $\overline{n} = 1000$ being slightly better, and both are better than $\overline{n} = 100$. This can be explained as follows. First, a larger $\overline{n}$ means a longer stage, in which nodes can collect more information to estimate each other's behavior. Moreover, the approximation of $p_{ij}$ as $p_r(P_i)$ in the system model is more accurate when the stage is longer, which reduces the estimation errors. Second, according to Algorithm 1, a larger $\overline{n}$ gives more behavior levels and smaller interval ranges for each level. The behavior levels for $\overline{n} = 100, 500, 1000$ are listed in Table III. For $\overline{n} = 1000$ and $\overline{n} = 500$, the adjacent behavior levels are close to each other, while for $\overline{n} = 100$, the difference between two adjacent behavior levels is larger. Therefore, estimation errors have greater impact on the performance when $\overline{n} = 100$. Based on the above discussion, when designing the incentive strategy, a larger $\overline{n}$ is preferred.

Finally, we evaluate the performance of the proposed strategy with dynamic membership in each stage. The design parameters are the same as those in Fig. 5. In the initial stage, there are 10 mobile nodes. At the beginning of each stage, one node randomly joins or leaves the group (where the total number is bounded between 8 and 12). To cope with the dynamic membership, we fix the average number of packets that a node should transmit in one stage ($\overline{n}$) so that the design parameters for the incentive strategy do not change. This can be achieved by adjusting the length of a stage ($M$) as follows: in the current stage with a total of $N'$ nodes, we use (3) to calculate $q_1$, the probability that a node is selected as a relay, and select the length of the current stage as $M' = \lceil \overline{n}/q_1 \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Then, nodes follow the proposed incentive strategy in Section IV to make decisions. We simulate both the fixed member scenario with 10 mobile nodes and the aforementioned dynamic membership scenario, and the result is shown in Fig. 10. It can be seen that in the scenario with dynamic members, the proposed strategy can achieve an average payoff of 91.89 per time slot, which is close to that of 91.98 per time slot in the fixed member scenario with 10 nodes.

## VI. CONCLUSION

In this work, we study cooperation stimulation among selfish mobile nodes in wireless multicast networks. We formulate the cooperative multicast process as an infinite repeated game and find the desired cooperation state that achieves absolute fairness and Pareto optimality. To enforce cooperation, we propose a *Worst Behavior Tit-for-Tat* incentive strategy, and find the subgame perfect equilibrium conditions in the scenario when nodes can accurately monitor each other's behavior. To address the issue of imperfect monitoring due to packet loss and bit errors in wireless networks, we develop an interval based estimation method, and derive the subgame perfect equilibrium conditions under the assumption that nodes are bounded rational. Simulation results show that even with imperfect monitoring, the proposed strategy can efficiently enforce cooperation, and its performance is close to that of the desired cooperation state where all nodes fully cooperate with perfect monitoring.

## Appendix A: Solution to (8)

To solve (8), we first analyze the property of $g(x)$. Let $R \stackrel{\triangle}{=} (N-1)Mq_2r_0$ and $C \stackrel{\triangle}{=} Mq_1c_0$, Then $g(x) = (N-1)r(x) - c(x) = Rp_r(x) - Cx$. We have

$$\begin{aligned} \frac{dg(x)}{dx} &= R\left\{\exp\left(-\frac{D}{x}\right)\cdot Dx^{-2}\right\} - C \\ &= RD\frac{p_r(x)}{x^2} - C \\ \text{and} \quad \frac{d^2g(x)}{dx^2} &= RD\frac{p_r(x)}{x^4}(D-2x). \end{aligned} \tag{19}$$

From (19), it can be seen that $\frac{dg(x)}{dx}$ monotonically increases with $x$ in $(0, D/2)$ and monotonically decreases with $x$ in $(D/2, +\infty)$. Moreover, it is easy to prove that $\lim_{x\to 0^+}\frac{dg(x)}{dx} = -C < 0$ and $\lim_{x\to 0^+} g(x) = 0$, which implies that $g(x)$ first decreases and takes negative values when $x$ increases from 0. Based on the above property, if there exists an $x$ such that $g(x) > 0$, then there exists at least one $x$ that satisfies $\frac{dg(x)}{dx} > 0$. Since $\frac{dg(x)}{dx}$ achieves the maximum value at $x = D/2$, we have $\frac{dg(x)}{dx}|_{x=D/2} > 0$. Therefore, $\frac{dg(x)}{dx} = 0$ has two roots $x_{g1}$ and $x_{g2} (0 < x_{g1} < x_{g2})$, where $\frac{dg(x)}{dx} > 0$ for $x \in (x_{g1}, x_{g2})$ and $\frac{dg(x)}{dx} < 0$ for $x \in (0, x_{g1}) \cup (x_{g2}, +\infty)$. It implies that $g(x)$ monotonically increases in $[x_{g1}, x_{g2}]$ and decreases in other intervals with $x_{g2} = \arg\max_{x>0} g(x)$. Then, we have $g(x_{g1}) < 0$ and $g(x_{g2}) > 0$. Based on the increasing property of $g(x)$ in $(x_{g1}, x_{g2})$, there exists a point $x_g^* \in (x_{g1}, x_{g2})$ such that $g(x_g^*) = 0$ and $g(x) > 0$ in $(x_g^*, x_{g2})$. Since there exists $P_i \in [0, P_{\max}]$ such that $g(P_i) > 0$, then $P_{\max} > x_g^*$. Then the solution to (8) is

$$x^{opt} = \begin{cases} x_{g2} = \arg\max_{x>0} g(x) & \text{if } P_{\max} \geq x_{g2}; \\ P_{\max} & \text{if } P_{\max} < x_{g2}. \end{cases} \tag{20}$$

## Appendix B: Proof of lemma 1

*Proof:* Take the summation of all nodes' expected payoffs in a stage, we have

$$\begin{aligned} \sum_{i=1}^{N} \pi_i(\boldsymbol{P}) &= NMp_Br_0 + \sum_{i=1}^{N}\sum_{j\in\mathcal{N}, j\neq i} r(P_j) - \sum_{i=1}^{N} c(P_i) \\ &= NMp_Br_0 + \sum_{i=1}^{N} g(P_i). \end{aligned} \tag{21}$$

By using the action profile $\boldsymbol{P}^* = [P^*, ..., P^*]$, we have

$$\sum_{i=1}^{N} \pi_i(\boldsymbol{P}^*) = NMp_Br_0 + Ng(P^*). \tag{22}$$

Since $P^*$ is the solution to (8), according to (21), it is obvious that $\max\left\{\sum_{i=1}^{N}\pi_i(\boldsymbol{P})\right\} = \sum_{i=1}^{N}\pi_i(\boldsymbol{P}^*)$. For any other action profile $\boldsymbol{P}'$, if $\pi_i(\boldsymbol{P}') > \pi_i(\boldsymbol{P}^*) > 0$, then we have

$$\begin{aligned} \sum_{j\in\mathcal{N}, j\neq i} \pi_j(\boldsymbol{P}') &= \sum_{i=1}^{N}\pi_i(\boldsymbol{P}') - \pi_i(\boldsymbol{P}') \\ &< \sum_{i=1}^{N}\pi_i(\boldsymbol{P}^*) - \pi_i(\boldsymbol{P}^*) \\ &= \sum_{j\in\mathcal{N}, j\neq i} \pi_j(\boldsymbol{P}^*). \end{aligned} \tag{23}$$

This indicates that a node cannot increase its expected payoff without decreasing any other node's expected payoff. Thus the NE point $\boldsymbol{P}^*$ is Pareto optimal. This completes the proof. ■

## Appendix C: Proof of Proposition 1

*Proof:* First, note that for the one-shot deviating node $i$, it will not deviate to an action $P_i' > 0$ in the current stage such that $0 < b_i^0 < b_T$, because in this case all other nodes who follow the proposed strategy will not cooperate during the next $L$ stages, which is the same as choosing $P_i' = 0$. Therefore, the deviating behavior $b_i^0$ of node $i$ will either be 0 or $b_i^0 > b_T$, and all other nodes will employ the same action during the next $L$ stages. Then from (11) and (12) we have

$$\begin{aligned} V_i(s^*) &= \sum_{k=0}^{L-K-1}(\delta)^k\{Mp_Br_0 + g(P^\dagger)\} \\ &\quad + \sum_{k=L-K}^{\infty}(\delta)^k\{Mp_Br_0 + g(P_{\max})\} \end{aligned} \tag{24}$$

where $g(x) = (N-1)r(x) - c(x)$, and

$$\begin{aligned} V_i(s') &= Mp_Br_0 + (N-1)r(P^\dagger) - c(P_i') \\ &\quad + \sum_{k=1}^{L}(\delta)^k\{Mp_Br_0 + g(P_i')\} \\ &\quad + \sum_{k=L+1}^{\infty}(\delta)^k\{Mp_Br_0 + g(P_{\max})\}. \end{aligned} \tag{25}$$

Since $P_{\max}$ maximizes $g(x)$ in $(0, P_{\max}]$, then $g(P_{\max}) \geq g(P^\dagger)$, we have

$$\begin{aligned} & V_i(s^*) - V_i(s') \\ = & \sum_{k=1}^{L-K-1}(\delta)^k g(P^\dagger) - c(P^\dagger) + \sum_{k=L-K}^{L}(\delta)^k g(P_{\max}) \\ & - \left\{\sum_{k=1}^{L}(\delta)^k g(P_i') - c(P_i')\right\} \\ \geq & \sum_{k=1}^{L}(\delta)^k g(P^\dagger) - c(P^\dagger) - \left\{\sum_{k=1}^{L}(\delta)^k g(P_i') - c(P_i')\right\} \\ = & \sum_{k=0}^{L}(\delta)^k\{G(P^\dagger) - G(P_i')\} \end{aligned} \tag{26}$$

where $G(x) \stackrel{\triangle}{=} \left(\sum_{k=1}^{L}(\delta)^k / \sum_{k=0}^{L}(\delta)^k\right)(N-1)r(x) - c(x) = \left(1/\sum_{k=0}^{L}(\delta)^k\right)\{\sum_{k=1}^{L}(\delta)^k g(x) - c(x)\}$. The inequality in (26) is obtained by replacing $g(P_{\max})$ with $g(P^{\dagger})$ in the summation term $\sum_{k=L-K}^{L}(\delta)^k g(P_{\max})$. Note that $G(x)$ has the same structure as $g(x)$ except for the constant factor $\left(\sum_{k=1}^{L}(\delta)^k / \sum_{k=0}^{L}(\delta)^k\right)$. From Appendix A, it can be seen that the constant factor does not affect the analysis. Similarly, we can prove that if there exists an $x \in (0, P_{\max}]$ such that $G(x) > 0$, then, $G(x)$ monotonically increases in the interval $(\min\{x|G(x)=0, x>0\}, \arg\max_{x>0} G(x)]$. From the above property, it can be seen that if condition $(a)$ holds and $P_{\max} \le \arg\max_{x>0} G(x)$, then $G(x)$ monotonically increases in $(-D/\ln(b_T), P_{\max}]$, which gives $V_i(s^*) - V_i(s^{'}) = \sum_{k=0}^{L}(\delta)^k\{G(P^{\dagger}) - G(P_i')\} > 0$. In summary, node $i$ cannot get more benefit by taking one shot deviation from the proposed strategy. This completes the proof. ■

## Appendix D: Proof of Proposition 2

*Proof:* First, we show that node $i$ cannot get more benefit by deviating to a behavior other than those in $\mathscr{B}$ under conditions in Proposition 2. Consider node $i$ deviates to the behavior $b_i^0 \in (\hat{b}_l, \hat{b}_{l+1}]$ in the current stage, if $b_i^0 = \hat{b}_{l+1}$, then node $i$ believes that everyone's estimation results are correct and the worst behavior level index is $l+1$ at the end of the current stage. Similar to (25), the estimated long term payoff is given by

$$\widetilde{V}_i'(\hat{b}_{l+1}) = \sum_{k=0}^{\infty}(\delta)^k M p_B r_0 + (N-1)r(\hat{P}_{l^{\dagger}}) + \sum_{k=0}^{L}(\delta)^k G(\hat{P}_{l+1}) + \sum_{k=L+1}^{\infty}(\delta)^k g(P_{\max}). \quad (27)$$

If $b_i^0 \in (\hat{b}_l, \hat{b}_{l+1})$, from (16), we have

$$\begin{aligned} &\widetilde{V}_i'(b_i^0) \\ = &\sum_{k=0}^{\infty}(\delta)^k M p_B r_0 + (N-1)r(\hat{P}_{l^{\dagger}}) - c(P_i') \\ &+ \sum_{k=1}^{L}(\delta)^k[w_i(\hat{b}_l|b_i^0)g(\hat{P}_l) + w_i(\hat{b}_{l+1}|b_i^0)g(\hat{P}_{l+1})] \\ &+ \sum_{k=L+1}^{+\infty}(\delta)^k g(P_{\max}) \\ = &\sum_{k=0}^{\infty}(\delta)^k M p_B r_0 + (N-1)r(\hat{P}_{l^{\dagger}}) - c(P_i') \\ &+ w_i(\hat{b}_l|b_i^0)c(\hat{P}_l) + w_i(\hat{b}_{l+1}|b_i^0)c(\hat{P}_{l+1}) \\ &+ \sum_{k=0}^{L}(\delta)^k[w_i(\hat{b}_l|b_i^0)G(\hat{P}_l) + w_i(\hat{b}_{l+1}|b_i^0)G(\hat{P}_{l+1})] \\ &+ \sum_{k=L+1}^{+\infty}(\delta)^k g(P_{\max}) \end{aligned} \quad (28)$$

According to Appendix C, $G(x)$ is an increasing function in $(-D/\ln(b_T), P_{\max}]$ if conditions in Proposition 1 hold. Since $w_i(\hat{b}_l|b_i^0) + w_i(\hat{b}_{l+1}|b_i^0) = 1$, we have

$$\begin{aligned} \widetilde{V}_i'(b_i^0) \le &\sum_{k=0}^{\infty}(\delta)^k M p_B r_0 + (N-1)r(\hat{P}_{l^{\dagger}}) - \widetilde{G}(b_i^0, l) \\ &+ \sum_{k=0}^{L}(\delta)^k G(\hat{P}_{l+1}) + \sum_{k=L+1}^{+\infty}(\delta)^k g(P_{\max}) \end{aligned} \quad (29)$$

where $\widetilde{G}(x,l) = c(\frac{-D}{\ln x}) - w_i(\hat{b}_l|x)c(\hat{P}_l) + w_i(\hat{b}_{l+1}|x)c(\hat{P}_{l+1})$. From condition $(b)$, we have $\widetilde{G}(b_i^0, l) > 0$ if $b_i^0 \in (b_l, \hat{b}_{l+1})$. Then from (27) and (29), we have $\widetilde{V}_i(b_i^0) < \widetilde{V}_i(\hat{b}_{l+1})$. Therefore, node $i$ will not deviate to behaviors other than those in $\mathscr{B}$. On the other hand, if node $i$ takes behaviors from $\mathscr{B}$, according to the bounded rational assumption and interval based estimation, it will believe that everyone's monitoring results are correct in the future stages. Then the problem boils down to analyzing equilibrium with perfect monitoring, and the conditions in Proposition 1 are sufficient for the refined strategy being a subgame perfect equilibrium strategy. This completes the proof. ■

## Appendix E: Proof of Lemma 2

*Proof:* Denote $s^{\ddagger} = [P_1, P_2, ...P_k]^T$ as a power allocation scheme which is different with the equal power allocation scheme $\overline{s} = [\overline{P}, ..., \overline{P}]$. Let $b_t = \exp(-D/P_t)$ for $P_t \in s^{\ddagger}$ and $\overline{b} = \exp(-D/\overline{P})$, assume $s^{\ddagger}$ and $\overline{s}$ obtain the same average behavior $\overline{b}_i$, we have $\overline{b}_i = \frac{1}{k}\sum_{t=1}^{k} b_t = \overline{b}$. The corresponding costs of $s^{\ddagger}$ and $\overline{s}$ are

$$c_i(s^{\ddagger}) = \sum_{t=1}^{k} c_0 \frac{-D}{\ln b_t}, \quad \text{and} \quad c_i(\overline{s}) = k c_0 \frac{-D}{\ln \overline{b}}, \quad (30)$$

respectively. Define $a(x) = -1/\ln x$, then for $\overline{b} > 1/e$, it is proved in Appendix F that $a(x)$ satisfies

$$\begin{aligned} a(\overline{b}) &= a\left(\frac{b_1 + b_2 + ... + b_k}{k}\right) \\ &\le \frac{a(b_1) + a(b_2) + ... + a(b_k)}{k}, \end{aligned} \quad (31)$$

and the equality holds when $b_1 = ... = b_k = \overline{b}$. From (30) and (31), we can conclude that $c_i(\overline{s}) \le c_i(s^{\ddagger})$, this completes the proof. ■

## Appendix F: Proof of (31)

*Proof:* First, taking the second derivative of $a(x)$, we have $\frac{d^2 a(x)}{dx^2} = -\frac{1}{x^2(\ln x)^2}\left(\frac{2}{\ln x} + 1\right)$. Then, it is easy to show that $\frac{d^2 a(x)}{dx^2} > 0$ for $x > 1/e^2$, that is, $a(x)$ is convex in $(1/e^2, +\infty)$. Therefore, according to Jensen's inequality, (31) holds if $b_t > 1/e^2, \ \forall\, t = 1, 2, ..., k$. Next we consider the case that $b_t < 1/e^2$ for some $t$, and use induction to prove (31). From Appendix G, we have

$$a(\lambda x_1 + (1-\lambda)x_2) < \lambda a(x_1) + (1-\lambda)a(x_2),$$

$$\text{for} \quad x_1 < \frac{1}{e} < x_2, 0 < \lambda < 1, \lambda x_1 + (1-\lambda) x_2 > \frac{1}{e}. \quad (32)$$

Then for $k = 2$, with $x_1 = b_1 < 1/e^2 < 1/e$, $x_2 = b_2 > 1/e$, and $\lambda = 1/2$, we have $\lambda x_1 + (1-\lambda)x_2 = \frac{1}{2}(b_1 + b_2) = \bar{b} > 1/e$. From (32), we have $a(\bar{b}) = a\left(\frac{b_1+b_2}{2}\right) < \frac{a(b_1)+a(b_2)}{2}$. Assume that for $k = n$, we also have $a\left(\frac{\sum_{t=1}^n b_t}{n}\right) < \frac{\sum_{t=1}^n a(b_t)}{n}$. Then for $k = n+1$, let $b_{n+1} < 1/e^2 < 1/e$, we have

$$\begin{aligned} a(\bar{b}) &= a\left(\frac{\sum_{t=1}^{n+1} b_t}{n+1}\right) \\ &= a\left(\frac{1}{n+1} b_{n+1} + \frac{n}{n+1} \cdot \frac{\sum_{t=1}^n b_t}{n}\right). \end{aligned} \quad (33)$$

Since $\bar{b} = \bar{b}_i > 1/e$, then $(n+1)\bar{b} = \sum_{t=1}^{n+1} b_t > n/e + b_{n+1}$, which gives $\sum_{t=1}^n b_t / n > 1/e$. Let $x_1 = b_{n+1}$, $x_2 = \sum_{t=1}^n \hat{b}_t / n$, and $\lambda = 1/(n+1)$, from (32), we have

$$\begin{aligned} & a\left(\frac{1}{n+1} b_{n+1} + \frac{n}{n+1} \cdot \frac{\sum_{t=1}^n b_t}{n}\right) \\ & < \frac{1}{n+1} a(b_{n+1}) + \frac{n}{n+1} a\left(\frac{\sum_{t=1}^n b_t}{n}\right). \end{aligned} \quad (34)$$

Since $a\left(\frac{\sum_{t=1}^n b_t}{n}\right) < \frac{\sum_{t=1}^n a(b_t)}{n}$, we conclude that

$$\begin{aligned} a(\bar{b}) &= a\left(\frac{\sum_{t=1}^{n+1} b_t}{n+1}\right) \\ &< \frac{1}{n+1} a(b_{n+1}) + \frac{n}{n+1} \cdot \frac{\sum_{t=1}^n a(b_t)}{n} \\ &= \frac{\sum_{t=1}^{n+1} a(b_t)}{n+1}. \end{aligned} \quad (35)$$

Thus, (31) also holds for $k = n+1$. Based on induction, we can conclude that (31) holds for all $k > 0$. This completes the proof. ∎

## APPENDIX G: PROOF OF (32)

*Proof:* Let $x_1 \in (0, \frac{1}{e})$, $x_2 \in (\frac{1}{e}, 1)$ and $x_\lambda = \lambda x_1 + (1-\lambda) x_2$ where $\lambda \in (0,1)$. It can be seen that if $x_\lambda > \frac{1}{e}$, then $x_\lambda \in (\frac{1}{e}, x_2)$. Denote the line determined by the two points $(x_1, a(x_1))$ and $(x_2, a(x_2))$ as $Z_1(x)$, where $Z_1(x_1) = a(x_1)$ and $Z_1(x_2) = a(x_2)$. Then we have

$$\frac{Z_1(x_\lambda) - Z_1(x_1)}{Z_1(x_\lambda) - Z_1(x_2)} = \frac{x_\lambda - x_1}{x_\lambda - x_2} = \frac{\lambda - 1}{\lambda}. \quad (36)$$

From (36), we get $Z_1(x_\lambda) = \lambda Z_1(x_1) + (1-\lambda) Z_1(x_2) = \lambda a(x_1) + (1-\lambda) a(x_2)$. To prove (32), it is equivalent to prove $a(x) < Z_1(x)$ for all $x \in (\frac{1}{e}, x_2)$. We can also show that $a(x)$ has the following properties:

- For $x \in (0,1)$, $\frac{da(x)}{dx} = \frac{1}{x(\ln x)^2} > 0$, which indicates that $a(x)$ is monotonically increasing;
- $a(x)$ is concave in $(0, \frac{1}{e^2})$ and is convex in $(\frac{1}{e^2}, 1)$;
- $a(\frac{1}{e}) = \frac{-1}{\ln x}|_{x=\frac{1}{e}} = 1$ and $\frac{da(x)}{dx}|_{x=1/e} = e$, which indicates the tangent of $a(x)$ at the point $(\frac{1}{e}, 1)$ is $Z_2(x) = ex$; and
- $\lim_{x\to 0} a(x) = \lim_{x \to 0} \frac{-1}{\ln x} = 0$.

Then, based on the above properties, it is easy to show that $Z_1(x_1) = a(x_1) > Z_2(x_1)$ and $Z_1(x_2) = a(x_2) > Z_2(x_2)$, which indicates line $Z_1(x)$ is above line $Z_2(x)$ in $(x_1, x_2)$. Since $x_1 < \frac{1}{e} < x_2$, it is obvious that $Z_1(\frac{1}{e}) > Z_2(\frac{1}{e}) = a(\frac{1}{e})$. According to the monotonically increasing property of $a(x)$, we have $Z_1(x_2) = a(x_2) > a(x_1) = Z_1(x_1)$, then $Z_1(x)$ is also an increasing function. Since $a(x)$ is a convex and increasing function in $(\frac{1}{e^2}, 1)$, from $Z_1(\frac{1}{e}) > a(\frac{1}{e})$ and $Z_1(x_2) = a(x_2)$, based on the monotonically increasing property of $Z_1(x)$, we can conclude that $a(x) < Z_1(x)$ within $(\frac{1}{e}, x_2)$. This completes the proof. ∎

## APPENDIX H: SOLUTION TO (17)

Let $x$ represent $\overline{P}_i$, according to the constraint, we have $\alpha_i = b^\dagger / p_r(x)$. Denote $f(x) = \alpha_i \overline{P}_i = b^\dagger x / p_r(x)$, then we have $\frac{df(x)}{dx} = \frac{b^\dagger (1 - D/x)}{p_r(x)}$. It can be seen that $f(x)$ has the following property:

$$\begin{cases} \frac{df(x)}{dx} < 0 & \text{if } 0 < x < D; \\ \frac{df(x)}{dx} = 0 & \text{if } x = D; \\ \frac{df(x)}{dx} > 0 & \text{if } x > D. \end{cases}$$

Note that $\alpha_i \in (0, 1]$. Thus we have $p_r(x) \geq b^\dagger$ which leads to $x \geq \frac{-D}{\ln b^\dagger}$. It can be seen that if $\frac{-D}{\ln b^\dagger} < D$ which is $b^\dagger < \frac{1}{e}$ then $\min\{f(x)\} = f(D)$; if $\frac{-D}{\ln b^\dagger} \geq D$, then $\min\{f(x)\} = f(\frac{-D}{\ln b^\dagger})$. Therefore, the solution can be summarized as

$$\begin{cases} \alpha_i = e b^\dagger, \overline{P}_i = D, & \text{if } b^\dagger < \frac{1}{e}; \\ \alpha_i = 1, \overline{P}_i = \frac{-D}{\ln b^\dagger}, & \text{if } b^\dagger \geq \frac{1}{e}. \end{cases} \quad (37)$$

## ACKNOWLEDGMENTS

The authors would like to thank the Associate Editor Dr. Huaiyu Dai and the anonymous reviewers for their constructive comments and suggestions which helped to improve the quality of the manuscript.

## REFERENCES

[1] J. She, P. Ho, and L. Xie, "IPTV over WiMax: Key success factors, challenges, and solutions," *IEEE Commun. Mag.*, vol. 45, no. 8, pp. 87–93, Aug. 2007.

[2] S. Kota, Y. Qian, E. Hossain, and R. Ganesh, "Advances in mobile multimedia networking and QoS," *IEEE Commun. Mag.*, vol. 45, no. 8, pp. 52–53, Aug. 2007.

[3] K. Kang and T. Kim, "Improved error control for real-time video broadcasting over CDMA2000 networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 188–197, Jan. 2009.

[4] I. Güvenc, U. C. Kozat, M.-R. Jeong, F. Watanabe, and C.-C. Chong,"Reliable multicast and broadcast services in relay-based emergency communications," *IEEE Wireless Commun.*, vol. 15, no. 3, pp. 40–47, June 2008.

[5] S. Raza, D. Li, C.-N. Chuah, and G. Cheung, "Cooperative peer-to-peer repair for wireless multimedia broadcast," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1075–1078, 2007.

[6] P. Fan, C. Zhi, C. Wei, and K. B. Letaief, "Reliable relay assisted wireless multicast using network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 749–762, June 2009.

[7] A. del Coso, O. Simeone, Y. Bar-ness, and C. Ibars, "Outage capacity of two-phase space-time coded cooperative multicasting," in *Proc. 40th Asilomar Conf. Signals, Systems and Computers (ACSSC)*, pp. 666–670, 2006.

[8] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, and P. V. Krishnamurthy, "On broadcasting with cooperative diversity in multi-hop wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 484–496, Feb. 2007.

[9] B. Sirkeci-Mergen and A. Scaglione, "On the power efficiency of cooperative broadcast in dense wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 497–507, Feb. 2007.

[10] I. Maric and R. D. Yates, "Cooperative multi-hop broadcast for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1080–88, Aug. 2004.

[11] O. Alay, T. Korakis, Y. Wang, E. Erkip, and S. Panwar, "Layered wireless video multicast using omni-directional relays," in *Proc. IEEE International Conference on Acoustics, Speach and Signal Processing (ICASSP)*, pp. 2149–2152, 2008.

[12] L. Buttyán and J.-P. Hubaux, "Stimulating cooperation in self-organizing mobile ad hoc networks," *ACM/Kluwer Mobile Networks and Applications*, vol. 8, no. 5, pp. 579–592, Oct. 2003.

[13] 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'03*, pp. 1987–1997, 2003.

[14] 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. 9th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pp. 245–259, 2003.

[15] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," in *Proc. 6th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pp. 255–265, 2000.

[16] P. Michiardi and R. Molva, "CORE: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks," in *Proc. 6th IFIP TC6/TC11 Joint Working Conference on Communications Multimedia Security*, pp. 107–121, 2002.

[17] S. Buchegger and J.-Y. L. Boudec, "Performance analysis of the CONFIDANT protocol," in *Proc. 3rd ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 226–236, 2002.

[18] V. Srinivasan, P. Nuggehalli, C. F. Chiasserini, and R. R. Rao, "Cooperation in wireless ad hoc networks," in *Proc. IEEE INFOCOM'03*, pp. 808–817, 2003.

[19] P. Michiardi and R. Molva, "A game theoretical approach to evaluate cooperation enforcement mechanisms in mobile ad hoc networks," in *Proc. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2003.

[20] W. Yu and K. J. R. Liu, "Game theoretic analysis of cooperation and security in autonomous mobile ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 6, no. 5, pp. 459–473, May 2007.

[21] W. Yu and K. J. R. Liu, "Secure cooperation in autonomous mobile ad-hoc networks under noise and imperfect monitoring: A game-theoretic approach," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 2, pp. 317–330, June 2007.

[22] W. Lin, H. V. Zhao, and K. J. R. Liu, "Incentive cooperation strategies for peer-to-peer live streaming social networks," *IEEE Trans. Multimedia*, vol. 11, no. 3, pp. 396–412, Apr. 2009.

[23] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Repeated open spectrum sharing game with cheat-proof strategies," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1922–1933, Apr. 2009.

[24] Z. Han, Z. Ji, and K. J. R. Liu, "A cartel maintenance framework to enforce cooperation in wireless networks with selfish users," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1889–1899, May 2008.

[25] C. Song and Q. Zhang, "Achieving cooperative spectrum sensing in wireless cognitive radio networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 13, no. 2, pp. 14–25, Apr. 2009.

[26] B. Niu, H. Jiang, and H. V. Zhao, "A cooperative multicast strategy in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3136–3143, July 2010.

[27] A. R. Mele and P. Rawling, *The Oxford Handbook of Rationality*. Oxford University Press, 2004.

[28] M.J. Osborne and A.Rubinste, *A Course in Game Theory*. The MIT Press,1994.

[29] G. J. Mailath and L. Samuelson, *Repeated Games and Reputations: Long-run Relationships*. Oxford University Press, 2006.

[30] P. Jehiel, "Bounded rationality and imperfect learning: Game theory vs AI," in *Transfer of Knowledge in Economic Decision Making*, D. J. Zizzo, Ed. Palgrave Macmillan, 2005.

**Binglai Niu** received the B.S. degree in electronics engineering from Fudan University, Shanghai, China, in 2008 and the M.Sc. degree in electrical engineering from the University of Alberta, Edmonton, Alberta, Canada, in 2010. He is currently working towards his Ph.D. degree at the University of British Columbia, Vancouver, British Columbia, Canada. His research interests include optimization, game theory, resource management in wireless networks, and incentive mechanism design for cooperative communications.

**H. Vicky Zhao** (M'05) received the B.S. and M.S. degree from Tsinghua University, China, in 1997 and 1999, respectively, and the Ph. D degree from University of Maryland, College Park, in 2004, all in electrical engineering. She was a Research Associate with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park from Jan. 2005 to July 2006. Since August 2006, she has been an Assistant Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada.

Dr. Zhao's research interests include media-sharing social networks, information security and forensics, digital communications and signal processing. Dr. Zhao received the IEEE Signal Processing Society (SPS) 2008 Young Author Best Paper Award. She is a co-author of "Multimedia Fingerprinting Forensics for Traitor Tracing" (Hindawi, 2005) and "Behavior Dynamics in Media-Sharing Social Network" (Cambridge University Press, to appear 2011). She is the Associate Editor for IEEE SIGNAL PROCESSING LETTERS and ELSEVIER JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION, and a guest editor of special issue on Signal and Information Processing for Social Learning and Networking of IEEE SIGNAL PROCESSING MAGAZINE.

**Hai Jiang** (M'07) received the B.Sc. and M.Sc. degrees in electronics engineering from Peking University, Beijing, China, in 1995 and 1998, respectively, and the Ph.D. degree (with an Outstanding Achievement in Graduate Studies Award) in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2006.

Since July 2007, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. His research interests include radio resource management, cognitive radio networking, and cross-layer design for wireless multimedia communications.

Dr. Jiang is an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He served as a Co-Chair for the General Symposium at the International Wireless Communications and Mobile Computing Conference (IWCMC) in 2007, the Communications and Networking Symposium at the Canadian Conference on Electrical and Computer Engineering (CCECE) in 2009, and the Wireless and Mobile Networking Symposium at the IEEE International Conference on Communications (ICC) in 2010. He received an Alberta Ingenuity New Faculty Award in 2008 and a Best Paper Award from the IEEE Global Communications Conference (GLOBECOM) in 2008.