Non-Cooperative Power Control for Wireless Ad Hoc Networks with Repeated Games

Chengnian Long, *Member, IEEE*, Qian Zhang, *Senior Member, IEEE*, Bo Li, *Senior Member, IEEE*, Huilong Yang, and Xinping Guan, *Senior Member, IEEE*

Abstract— One of the distinctive features in a wireless ad hoc network is lack of any central controller or single point of authority, in which each node/link then makes its own decisions independently. Therefore, fully cooperative behaviors, such as cooperation for increasing system capacity, mitigating interference for each other, or honestly revealing private information, might not be directly applied. It has been shown that power control is an efficient approach to achieve quality of service (QoS) requirement in ad hoc networks. However, the existing work has largely relied on cooperation among different nodes/links or a pricing mechanism that often needs a third-party involvement. In this paper, we aim to design a non-cooperative power control algorithm without pricing mechanism for ad hoc networks. We view the interaction among the users' decision for power level as a repeated game. With the theory of stochastic fictitious play (SFP), we propose a reinforcement learning algorithm to schedule each user's power level. There are three distinctive features in our proposed scheme. First, the user's decision at each stage is self-incentive with myopic best response correspondence. Second, the dynamics arising from our proposed algorithm eventually converges to pure Nash Equilibrium (NE). Third, our scheme does not need any information exchange or to observe the opponents' private information. Therefore, this proposed algorithm can safely run in a fully selfish environment without any additional pricing and secure mechanism. Simulation study demonstrates the effectiveness of our proposed scheme.

Index Terms— Wireless Ad Hoc Networks, Power Control, Non-cooperative, Repeated Games, Stochastic Fictitious Play, Reinforcement Learning, Nash Equilibrium.

I. INTRODUCTION

THE POWER control in a wireless ad hoc network deals with the selection of proper transmission power for each packet at each link in a distributed fashion. This is a complex and intriguing problem since the selection of the power level fundamentally affects many aspects of the operation of the network and its resulting performance; for instance the quality of the signal received at the receiver, the interference it creates for the other receivers and energy consumption at each node. In traditional emergency or military situations, the nodes in an ad hoc network usually belong to the same authority with certain common objectives. To maximize the overall system performance, nodes usually work in a fully cooperative way and can unconditionally follow the designated algorithm with a system optimization method [20], [37], [40]. Recently, emerging applications of ad hoc networks are envisioned in civilian and commercial usage, where nodes typically do not belong to a single authority and may not pursue a common goal. The usual assumption of spontaneous willingness to cooperate is unrealistic for autonomous users, especially when there are conflicting interests among the users [3], [28], [32]. The autonomous nature of this evidently affects many aspects in the design of ad hoc networks including the selection power level, which is the focus of study in this paper.

In the power control scheme of ad hoc networks, there are usually two fundamentally conflicting objects. On the one hand, the higher the Signal-to-Interference plus Noise Ratio (SINR), the better the service a receiving node can expect; on the other hand, this higher SINR is achieved at the expense of increased drainage on the battery consumption and higher interference to the signals of other users. If each user demonstrates totally selfish behavior, the network performance may significantly degrade, even leading to interrupted service [9], which is called *tragedy of the commons* in economics. Therefore, to deploy an ad hoc network successfully in a self-organized manner, the issue of cooperation stimulation should be solved. Most of the existing works rely on a pricing mechanism as an incentive to a certain degree to facilitate cooperation among self-interested users in power control for cellular systems [2], [3], [18], [25], [30], [38], and ad hoc networks [19], [22]. However, such a pricing mechanism is often difficult to configure and implement in ad hoc networks, since there lacks a central controller or/and there is no single *creditable* node in wireless ad hoc networks. There have been fine attempts using cryptographic techniques [39] or mechanism design [31] to implement the distributed and reliable price-based incentive scheme, yet these can incur significant overheads in the overall protocol design. For example, to bring about incentive-, communication-, and algorithm compatibility in the mechanism design, these need to be a combination of payments, redundancy, problem partitioning, and cryptographic techniques [31]. Each user may perform redundant computation as a *checker*, which creates the opportunity for a

C. N. Long, H. L. Yang, and X. P. Guan are with the Centre for Networking Control and Bioinformatics (CNCB), Department of Electrical Engineering, Yanshan University, Qinhuangdao, 066004, Hebei Province, P.R. China (e-mail:{dylcn,xpguan}@ysu.edu.cn).

Q. Zhang, and B. Li are with Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China (e-mail:{qianzh, bli}@cse.ust.hk).

catch-and-punish scheme that provides incentives for rational nodes to be faithful. It should be noted that there still needs to be a bank to detect manipulation and discuss appropriate reactions to detected manipulations.

Our work is motivated by the demand of a new power control with low complexity for wireless ad hoc networks, in which there are no *creditable* agent(s) and no infrastructure support. In this paper, we formulate the power control in ad hoc networks with non-cooperative games. To avoid the complexity incurred by a pricing mechanism, the power control decision of each user requires a rational *self-incentive* property. The *self-incentive* property means that the designated dynamics should largely comply with each user's self-interests, that is, the user runs spontaneously with the given update rule. Specifically, we answer the following two questions: *Does there exist a control rule with a self-incentive property for non-cooperative power control game? How to design a rational self-incentive dynamics to approximate the Nash Equilibrium (NE) of the non-cooperative power control game?*

To formulate the non-cooperative power control analytically, we first model the self-interest property for power control in ad hoc networks. In general, the concept of utility, a term in microeconomics, is commonly used [10], [30]. The utility refers to the level of satisfaction the decision-taker receives as a result of its actions. In this paper, we construct a utility model with the consideration of both the QoS requirement and the energy-efficiency property. Based on the utility model, we then formulate the power control problem with the theory of non-cooperative games.

Power control in wireless ad hoc networks is inherently a repeated process. It is natural to model the interaction among users with repeated games [14]. Our argument is that every rational user has the motivation to improve its performance during the interaction even if they are selfish by nature. Based on this observation, we can set up the rational dynamics in the stage game. With this dynamics, each user will be endowed with sufficient *intelligent behavior* to make a correct decision. Consequently, we introduce the learning theory in games to complete the stage game design [15]. Of the many existing learning models in games, one of the best known is *fictitious play (FP)*. In *FP* and its variant, each user chooses the best response to its beliefs about its opponents, which is given by the time average of past play. Interestingly, this myopic learning process is consistent with the desirable *self-incentive* property in non-cooperative power control design. To cope with observing the opponents' private information problem, we propose a utility based reinforcement learning algorithm with *SFP* framework, which makes a *myopically optimal decision (MOD)* for the next stage under current observation. Here, *MOD* means that users update their strategies each using the best response to the strategies of the others. This learning process only needs to observe its own action and the derived utility, which avoids the complex security and mechanism implementations incurred by information exchange or observation.

The challenging problem in this scheme is how to guarantee the convergence of the learning dynamics in the repeated game. To this end, we first show that the non-cooperative power control with utility function is a supermodular game.

The supermodularity reflects the interdependence among the users' actions. Furthermore, it provides a strong evidence that the user has the incentive to learn the environment for improving its own utility. According to the continuous dynamics arising from *SFP*, we show that the behavior of *MOD* can almost surely converge to a bounded set of *NE*. To avoid observing the opponents' private information completely, we introduce a utility-based reinforcement learning algorithm. By exploiting the utility structure, we can significantly speed up the convergence. It is an inherently two-time-scale dynamics in the proposed learning algorithm. With the stochastic approximation theory, it can be proved that the dynamics of the utility-based reinforcement learning algorithm is almost surely an asymptotic pseudotrajectory of the flow defined by the *SFP* dynamics. Furthermore, it can be shown that, in the supermodular games, the mixed *NE* of *SFP* learning dynamics is unstable. We can therefore present the important result that the dynamics of our utility based *SFP* learning algorithm almost surely converges to a *pure Nash Equilibrium (PNE)*. It is a very interesting property for the non-cooperative power control problem of wireless ad hoc networks.

Our work is the first step toward exploring the learning theory in repeated games to design non-cooperative power control, which can eliminate the performance degradation incurred by selfish behavior. We provide the following two important insights:

- Firstly, for non-cooperative power control with required QoS in wireless ad hoc networks, we show that the self-incentive dynamics using the *MOD* exist in the learning process.
- Secondly, the self-incentive learning dynamics can converge to the set of *PNE*. The simulation results show that the required QoS performance can always be guaranteed with a feasible assumption. For the energy-efficiency utility, the average performance loss in the steady state is about $13\% - 20\%$ compared with the system optimum.

The remaining sections of the paper are organized as follows. In Section II we introduce the preliminary knowledge on repeated games, learning theory and supermodular games. In Section III we formulate the non-cooperative power control problem formally with the non-cooperative power control game. We also present the design objective and challenges. In Section IV we propose the utility based stimulus-response learning algorithm within the *SFP* framework. To get a high convergence rate, we further propose an improved algorithm by exploiting the utility structure. In Section V the convergence of the learning dynamics is investigated. The simulation study is included in Section VI to show the validity and efficiency of our proposed algorithm. In Section VII, we summarize some related work on power control and non-cooperative behavior in wireless networks. Finally, we present in Section VIII an overview of this paper.

II. NOTATION AND PRELIMINARIES

A. Notation

To make readers easy to follow, we begin by briefly summarizing notation and definitions in the paper, which is shown in Table I.

TABLE I
THE SUMMARIZATION OF THE NOTATION

Notation	Definition
N	the number of active links (users) at a certain time slot.
$\mathcal{N}$	the set of all users $\{1, \ldots, N\}$. So $|\mathcal{N}| = N$.
$\mathcal{A}_i$	the finite action set of user i, $i \in \mathcal{N}$.
$\boldsymbol{\pi}_i$	the mixed strategies for user i, $i \in \mathcal{N}$.
$\mathcal{S}_i$	the set of mixed strategies for user i, $i \in \mathcal{N}$.
$\mathcal{U}_i$	the utility function of user i, $i \in \mathcal{N}$.
p_i	the power level of user i, $i \in \mathcal{N}$.
γ_i	the SINR of user i, $i \in \mathcal{N}$.
G	the *non-cooperative power control (NPC)* game $[\mathcal{N}, \{\mathcal{P}_i\}, \{\mathcal{U}_i(\cdot)\}]$.
$\overline{\mathcal{BR}}_i(\cdot)$	smooth best response function.
$\dot{\boldsymbol{\pi}}_i$	the perturbed best response dynamic $\dot{\boldsymbol{\pi}}_i = \overline{\mathcal{BR}}_i(\boldsymbol{\pi}_{-i}) - \boldsymbol{\pi}_i$
$F(x)$	game differential equation $\frac{dx}{dt} = F(x)$
$\aleph$	the set of fixed points $\{x \in X : F(x) = 0\}$ of game differential equation
$\mathcal{L}(x)$	the limit set of state x : $\{z \in X : \lim_{t \to \infty} \Psi_t x = z\}$
$\Psi_t(Q)$	invariant set $Q \subset \Re^n$ for F

B. Preliminaries

The process of *fictitious play (FP)* is one of the most widely used learning models [14], [15]. In this process, users behave as if they think they are facing a stationary, but unknown, distribution of opponents' strategies. At each stage, *FP* makes *MOD* based on its observed environment. Therefore, it is a suitable learning method among self-interested users. In this Section, we introduce the basic concept of *FP* and its variant, *stochastic fictitious play (SFP),* in the repeated games.

Supermodular games are those characterized by "*strategic complementarities*" – basically, this means that when one player takes a higher action, the others want to do the same. In power control, the conflict of the QoS requirement from users has the property of "*strategic complementarities*", that is, once one user increases its own power level for improving the QoS, the other users may increase those power levels to guarantee their own QoS requirements. Supermodular games are particularly interesting because they tend to be analytically appealing – they have nice comparative statics properties and behave well under various learning rules [29].

1) Learning Theory in Repeated Games: A repeated game is the repeated play of a particular *stage game*. Non-cooperative power control is inherently a repeated process. Each user collects local information about the network environment and the decision about power level is then made locally. New information (perhaps dependent on actions in the prior periods) are emergent, the decision process repeats itself. In this paper, we model the users' interaction in power control with a repeated game.

a) Fictitious Play: In this subsection, we define the process of *fictitious play (FP)*. We first introduce some notations to describe normal form games. A N player[1] normal form game G is defined by a collection of finite action sets $\mathcal{A}_1, \ldots, \mathcal{A}_N$ and a collection of utility functions $\mathcal{U}_1, \ldots, \mathcal{U}_N$. We assume that player i has a fixed finite set $\mathcal{A}_i = \{1, \ldots, |\mathcal{A}_i|\}$ of pure strategies, called player i's *action*

---

[1]In this paper, we use the terms "user", "player" and "link" interchangeably.

---

space. Player i's utility function $\mathcal{U}_i$ is a map from the set of action profiles of $\mathcal{A} = \Pi_j \mathcal{A}_j$ to the real line. Finally, $\mathcal{A}_{-i} = \Pi_{j \neq i} \mathcal{A}_j$ denotes the set of action profiles of player i's opponents.

The set $\mathcal{S}_i$ of the mixed strategies for player i is the unit simplex $\Delta^{n_i - 1} \subset \Re^{n_i}$ of the dimension $n_i - 1$

$$\mathcal{S}_i = \left\{ \boldsymbol{\pi}_i \in \Re_+^{n_i} : \sum_{k \in \mathcal{A}_i} \pi_i^k = 1 \text{ and } \pi_i^k \geq 0 \text{ for all } k \in \mathcal{A}_i \right\}.$$

We identify $\mathcal{S}_i$ with the set of probability measures on $\mathcal{A}_i$. If $a_i \in \mathcal{A}_i$, $\mathbf{a}_i \in \mathcal{S}_i$ denotes the corresponding vertex (i.e. the lth component of $\mathbf{a}_i$ is 1 for $a_i = l$ and 0 otherwise). The game's state space is the following compact convex polyhedron

$$\boldsymbol{\Sigma} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_N \subset \Re_+^{n_1} \times \ldots \times \Re_+^{n_N}.$$

The payoffs to player i are determined by its utility function $\mathcal{U}_i : \mathcal{A}_i \times \mathcal{A}_{-i} \to \Re$.

The model of *FP* supposes that players choose their actions in each period to maximize that period's expected payoff given the opponents' *empirical frequencies* of actions. One can think of the *FP* as a mechanism employed by the users to select their actions. At each time step, $t = 0, 1, \ldots$, users propose action vector

$$\boldsymbol{a}(t) := (a_1(t), \ldots, a_N(t)),$$

where $a_i(t) \in \mathcal{A}_i$ is the label of the action proposed by user i. The objective is to construct a mechanism so that the proposed actions, $\boldsymbol{a}(t)$, ultimately converge at a large enough t. In the *FP*, the scheduled action at step t is a function of past proposed actions over the interval $[0, t-1]$ as follows. First, enumerate the actions available to user i as $\mathcal{A}_i$. For any $j \in [1, |\mathcal{A}_i|]$, let $\kappa_i^j(t)$ denote the total number of times the user i proposed action A_i^j up to stage t. We define the empirical frequency vector, $\rho_i(t) \in \Re^{|\mathcal{A}_i|}$, of the user i as follows:

$$\rho_i(t) = \left(\frac{\kappa_i^1(t-1)}{t-1} \quad \cdots \quad \frac{\kappa_i^{|\mathcal{A}_i|}(t-1)}{t-1} \right).$$

Now we define the *FP* process. At stage game t, user i selects action $a_i(t) \in \mathcal{A}_i$ in accordance with maximizing its expected utility as though all the other users make an independent random selection of their actions, $\Pi_{j \neq i} \rho_j(t)$, i.e. $a_i(t) \in \arg\max_{\alpha \in \mathcal{A}_i} \mathbf{E}_{\boldsymbol{a}_{-i}}[\mathcal{U}_i(\alpha, \boldsymbol{a}_{-i})]$. Noticing that this is an "inclusion", since the best response correspondence for these beliefs is not necessarily single-valued. The discontinuity inherent in *determinate FP* is troubling descriptively and can lead to poor long-term performance.

b) Stochastic Fictitious Play: Since best responses are generically pure, a player's choices under *determinate FP* are often sensitive to the exact value of its beliefs. Small changes in beliefs can lead to discrete changes in behavior. Even when beliefs converge to *NE*, actual behavior may not be the case. In particular, the behavior can never converge to the mixed equilibrium of a game. To contend with these issues, Fudenberg and Levine [15] introduced *SFP*. In this model, each player i chooses a strategy $\boldsymbol{\pi}_i$ to maximize the perturbed utility

$$\mathcal{U}_i(\boldsymbol{\pi}_i, \boldsymbol{\pi}_{-i}) + \lambda \eta_i(\boldsymbol{\pi}_i), \tag{1}$$

where $\lambda > 0$ is a temperature parameter that controls the level of randomization, $\eta_i : \mathcal{S}_i \to \Re$ is a player-dependent smoothing function, which is a smooth, strictly differentiable concave function such that as $\boldsymbol{\pi}_i$ approaches the boundary of $\mathcal{S}_i$, the slope of η_i becomes infinite. The conditions on η_i mean that, for the fixed $\boldsymbol{\pi}_{-i}$, there is a unique maximizing $\boldsymbol{\pi}_i$, so we can define the *smooth best response function*

$$\overline{\mathcal{BR}}_i(\boldsymbol{\pi}_i) = \arg\max_{\boldsymbol{\pi}_i}\{\mathcal{U}_i(\boldsymbol{\pi}_i, \boldsymbol{\pi}_{-i}) + \lambda\eta_i(\boldsymbol{\pi}_i)\}. \quad (2)$$

A typical example of perturbation functions that satisfy these conditions is the entropy function $\mathcal{H}(\cdot)$:

$$\mathcal{H}(\boldsymbol{\pi}_i) = -\boldsymbol{\pi}_i^T \log(\boldsymbol{\pi}_i) = \sum_{a_i \in \mathcal{A}_i} -\pi_i(a_i)\log\pi_i(a_i). \quad (3)$$

For any $\lambda > 0$, we can explicitly solve for $\overline{\mathcal{BR}}_i$ as

$$\overline{\mathcal{BR}}_i(\boldsymbol{\pi}_{-i})[a_i] = \frac{\exp\left((1/\lambda)\mathcal{U}_i(a_i, \boldsymbol{\pi}_{-i})\right)}{\sum_{r_i \in \mathcal{A}_i} \exp\left((1/\lambda)\mathcal{U}_i(r_i, \boldsymbol{\pi}_{-i})\right)}. \quad (4)$$

We are interested in small values of $\lambda > 0$ because for such values, $\overline{\mathcal{BR}}_i(\boldsymbol{\pi}_{-i})$ approximately maximizes user i's unperturbed utility.

2) Supermodular Games: S-modular games are developed in [34] for games where the strategy space S_i of player i is a compact sublattice of $\Re^m$. By sublattice, we mean that it has the property that for any two elements a and b which are contained in S_i, also the component-wise minimum $\min(a, b)$ (denoted by $a \wedge b$) and component-wise maximum $\max(a, b)$ (denoted by $a \vee b$) are contained there. In particular, a compact sublattice has component-wise smallest and largest elements. In the following, we present the main results for the case of $m = 1$. Consider N players, and the utility of player i corresponding to the N-dimensional vector of strategies $\boldsymbol{x}$ is denoted by $\mathcal{U}_i(\boldsymbol{x})$. Let $\varkappa$ denote the space of all strategies. We present the property of *increasing differences (supermodularity)* with the following definition [4].

Definition 1: The utility function $\mathcal{U}_i$ for player i is *super-modular* if and only if for all $x, y \in \varkappa$

$$\mathcal{U}_i(x \wedge y) + \mathcal{U}_i(x \vee y) \geq \mathcal{U}_i(x) + \mathcal{U}_i(y).$$

Consider the case that $\mathcal{U}_i$ is a twice differentiable function. It is *supermodular* in $\boldsymbol{x} = (x_1, \ldots, x_N)$ if it has the *increasing differences* property, i.e., $\frac{\partial^2 \mathcal{U}(x)}{\partial x_i \partial x_j} \geq 0$ for all $\boldsymbol{x} \in \varkappa$ and $j \neq i$. For one-dimension discrete strategy space, the *increasing differences* property can be stated that the difference $\mathcal{U}_i(\boldsymbol{a}) - \mathcal{U}_i(\boldsymbol{a}')$ is (strictly) increasing in a_j, i.e.

$$\mathcal{U}_i(a_i, a_{-i}) - \mathcal{U}_i(a_i', a_{-i}) \geq \mathcal{U}_i(a_i, a_{-i}') - \mathcal{U}_i(a_i', a_{-i}') \quad (5)$$

for $a_i \geq a_i'$ and $a_{-i} \geq a_{-i}'$.

Remark 1: It should be noted that the condition (5) for *increasing differences* can be extended to the mixed strategy $\boldsymbol{\pi}_i$.

Proposition 1: A game $G = [\mathcal{N}, \{\mathcal{P}_i\}, \{\mathcal{U}_i(\cdot)\}]$ is *super-modular* if for each player $i \in N$,
(i) the strategy $\mathcal{P}_i$ is a nonempty and compact sublattic,
(ii) the utility function $\mathcal{U}_i$ is continuous in all players' strategies, is supermodular in player i's own strategy, and has increasing differences between any component of player i's strategy and any component of any other player's strategy.

III. NON-COOPERATIVE POWER CONTROL GAME

In this paper, we propose a non-cooperative power control with *self-incentive* dynamic property in self-interested ad hoc networks, which does not need a pricing mechanism and can avoid implementational complexity. We emphasize that *self-incentive* dynamic property, where the user is willing to obey the designated dynamics with a myopic view of its selfish interest. A key component for describing the selfish interest is utility function. In this section, we present a utility model for power control, which considers both the QoS requirement and the energy-efficiency property. With the utility model, we formalize the power control problem with non-cooperative games. Finally, we state the design objective and highlight the challenging problems.

A. Utility Model

The goal in power control in wireless networks is to ensure that no user's SINR γ_i falls below its required threshold γ_i^* chosen to ensure adequate QoS, i.e., to maintain

$$\gamma_i \geq \gamma_i^*, \forall i \quad (6)$$

where the subscript i is the set of users.

Considering a wireless ad hoc network with a set $\mathcal{N} = \{1, \ldots, N\}$ of distinct node pairs. Each link consists of one dedicated transmitter and one dedicated receiver. We designate the transmitted power and SINR for the ith user by p_i ($p_i^{\min} \leq p_i \leq p_i^{\max}$) and γ_i, respectively. The users' transmit power vector is denoted by $\boldsymbol{p} = (p_1, \ldots, p_N)$. We denote the background (receiver) noise within the user's bandwidth by σ_i. In the deterministic formulation of the power control problem for wireless networks, the noise power σ_i is dealt with a constant. We use a "snapshot" model with the assumption that link gains evolute slowly with respect to the SINR evolution. In this problem formulation, the SINR of the ith user is

$$\gamma_i(\boldsymbol{p}) = \frac{p_i h_{ii}}{\sigma_i + \sum_{j \neq i} p_j h_{ji}}, \quad (7)$$

where h_{ii} is the link gain from the ith user's transmitter T_i to its intended receiver R_i, and h_{ji} is the link gain from the jth user's transmitter T_j to the ith user's receiver R_i. We denote the interference plus noise of user i by

$$\mathcal{I}_i(\boldsymbol{p}_{-i}) = \sigma_i + \sum_{j \neq i} p_j h_{ji}. \quad (8)$$

To pose the power control problem as a non-cooperative game, we need to define a utility function suitable for data applications. Most data applications are sensitive to error but tolerant to delay. It is clear that a higher SINR level at the output of the receiver will generally result in a lower bit error rate and hence higher throughput. However, achieving a high SINR level requires the user terminal to transmit at a high power, which in turn results in low battery life. Moreover, a user with high power will increase the magnitude of the interference it creates for other receivers. This trade-off can be quantified by defining the utility function of the average amount of data received correctly per unit of energy consumption

$$u_i(\boldsymbol{p}) = \frac{T_i(\boldsymbol{p})}{p_i}, \quad (9)$$

where $T_i(\boldsymbol{p})$ is the effective throughput (goodput). We assume that transmitters use variable-rate M-QAM, with a bounded probability of symbol error and trellis coding with a nominal coding gain. Thus, the effective throughput $T_i(\boldsymbol{p})$ can be well approximated [35] by

$$T_i(\boldsymbol{p}) = W \log_2\left(1 + \frac{\gamma_i(\boldsymbol{p})}{\Omega}\right), \tag{10}$$

which is a function of global power vector $\boldsymbol{p}$ and channel conditions. Here, W is the bandwidth of the channel, and $\Omega \geq 1$ is the gap between uncoded M-QAM and the capacity, minus the coding gain.

Considering both the QoS requirement (6) and energy efficient utility (9), we express the utility of user i formally as follows

$$\mathcal{U}_i(\boldsymbol{p}_i, \boldsymbol{p}_{-i}) = \begin{cases} \frac{W \log_2(1 + \gamma_i(\boldsymbol{p})/\Omega)}{p_i} & \text{if } \gamma_i(\boldsymbol{p}) \geq \gamma_i^*, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

B. Non-cooperative Power Control Game

In this subsection, we formulate the users' selfish behavior with a non-cooperative game framework. Let $G = [\mathcal{N}, \{\mathcal{P}_i\}, \{\mathcal{U}_i(\cdot)\}]$ denote the *non-cooperative power control (NPC) game* where $\mathcal{N} = \{1, \ldots, N\}$ is the index set for active users currently in an ad hoc network, $\mathcal{P}_i$ is the strategy set, and $\mathcal{U}_i(\cdot)$ is the utility function of user i. Each users selects a power level p_i such that $p_i \in \mathcal{P}_i$. Let the power vector $\boldsymbol{p} = (p_1, \ldots, p_N) \in \mathcal{P}$ denote the outcome of the game in terms of the selected power levels of all the users, where $\mathcal{P}$ is the set of all power vectors. The utility function (11) demonstrates the strategic interdependence among users. The level of utility each user gets depends on its own power level and also on the choice of other players' strategies, through the SINR of that user. We assume that each users' strategy is rational, that is, each user maximizes its own utility in a distributed fashion. Formally, the *NPC* game G is expressed as

$$\max_{p_i \in \mathcal{P}_i} \mathcal{U}_i(p_i, p_{-i}), \text{ for all } i \in \mathcal{N}, \tag{12}$$

where $\mathcal{U}_i$ is given in (11) and $\mathcal{P}_i = [p_i^{\min}, p_i^{\max}]$ is the strategy space of user i. In this game $\boldsymbol{p}$ is the strategy profile, and the strategy profile of user i's opponents is defined to be $p_{-i} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_N)$, so that $\boldsymbol{p} = (p_i, p_{-i})$. A similar notation will be used for other quantities.

User i's *best response* is $\mathcal{BR}_i(p_{-i}) = \arg\max_{p_i \in \mathcal{P}_i} \mathcal{U}_i(p_i, p_{-i})$, i.e., the p_i that maximizes $\mathcal{U}_i(p_i, p_{-i})$ given a fixed p_{-i}. With the best response concept, we can present the following definition for the *Nash Equilibrium (NE)* of *NPC* game G [16].

Definition 2: A strategy profile p^* is a Nash Equilibrium (NE) of NPC game G if it is a fixed point of the best response, i.e. $\mathcal{U}_i(p_i^*, p_{-i}^*) \geq \mathcal{U}_i(p_i', p_{-i}^*)$ for any $p_i' \in \mathcal{P}_i$ and any user i.

The *NE* concept offers a predictable, stable outcome of a game where multiple agents with conflicting interests compete through self-optimization and reach a point where no player wishes to deviate. However, such a point does not necessarily exist. In general, we have the following result of the sufficient condition for the existence of an *NE* in a non-cooperative game [30].

Proposition 2: An *NE* exists in *NPC* game $G = [\mathcal{N}, \{\mathcal{P}_i\}, \{\mathcal{U}_i(\cdot)\}]$ if, for all $i = 1, \ldots, N$:
(i) The strategy profile $\mathcal{P}_i$ is a nonempty, convex, and compact subset of some Euclidean space $\Re^n$.
(ii) $\mathcal{U}_i(\cdot)$ is continuous in $\boldsymbol{p}$ and quasi-concave in p_i.

To be compatible with the *SFP* framework, we discretize the continuous power profile $p_i \in \mathcal{P}_i = [p_i^{\min}, p_i^{\max}]$ as the following

$$p_i(a_i) = \left(1 - \frac{a_i}{M_i}\right) p_i^{\min} + \frac{a_i}{M_i} p_i^{\max}, a_i = 0, \ldots, M_i. \tag{13}$$

Then the action of user i is denoted by an integer

$$a_i \in \mathcal{A}_i = \{0, \ldots, M_i\}. \tag{14}$$

The *NPC* game $G = [\mathcal{N}, \{\mathcal{P}_i\}, \{\mathcal{U}_i(\cdot)\}]$ can be converted to the discrete form $G_d = [N, \{\mathcal{A}_i\}, \{\mathcal{U}_i\}]$, i.e., each user chooses its strategy to maximize its own utility

$$\max_{\boldsymbol{\pi}_i \in \mathcal{S}_i} \boldsymbol{\pi}_i^T \mathcal{U}_i(a_i, \boldsymbol{\pi}_{-i}), \forall i. \tag{15}$$

The concept of *NE* in discrete *NPC* game G_d is similar to that of *NPC* game G.

C. Design Objective and Challenging Problems

In this subsection, we present the design objective and challenging problems in *NPC* game. First, we give a precondition of our design. We define the feasible power vector set

$$\mathbf{p}^{feasible} = \{\boldsymbol{p} \in \mathcal{P}, \gamma_i(\boldsymbol{p}) \geq \gamma_i^*, \forall i\}. \tag{16}$$

As in [20], we give the following assumption:

Assumption: Given the routing and MAC protocol, N concurrent users can transmit data successfully and satisfy the QoS constraint (6), i.e., there is feasible power vector set $\mathbf{p}^{feasible}$ at a particular schedule.

In the discrete strategy space $\mathcal{A}_i$ with feasible power vector set $\mathbf{p}^{feasible}$, we can observe that the utility function (11) does not have the quasi-concave property for pure action a_i or mixed strategy $\boldsymbol{\pi}_i$, which is different from the utility of the energy-effiency power control design in [30]. To investigate the existence of *NE* in *NPC* game G_d, we should refer to the graceful property of supermodular games.

Our aim is to design non-cooperative power control in wireless ad hoc networks with self-interested users. We apply the repeated games to model the interaction among the users' power decisions. Compared with the related work [13], the fundamental difficulty is that the *NE* of a repeated game with utility (11) is unknown. Therefore, the simple tit-for-tat (TFT) punishment mechanism in repeated games can not be directly applied in this problem. We, hence, introduce the reinforcement learning method to find each user's best decision over its interest. The problem is challenging in several aspects due to the following constraints:

- Each user can observe its own private information, such as the historical and current decision and derived utilities. However, no user can observe its opponents' private information. Otherwise, this would result in implementational complexity.

- Each user's power level decision should be self-incentive. Otherwise, the user would not be willing to obey the designated dynamics.

The first constraint shows that *SFP* cannot solve the power control problem directly because *SFP* should trace its opponents' historical decisions. In *SFP*, each user selects an action that maximizes its immediate expected payoff under the assumption that opponents will play a mixed strategy by the historical frequencies of past plays. The second constraint shows that the dynamic update rule should be *myopic best response* from the user's view for the environment. However, the user's view may be wrong in a certain time period. Therefore, the challenging problems are: *how to design a non-cooperative power control with the myopic best response dynamics over private and incomplete information? How to guarantee the convergence and a satisfactory convergence speed?*

IV. NON-COOPERATIVE POWER CONTROL WITH UTILITY BASED *SFP* APPROACH

In this section, we propose a non-cooperative power control with the myopic best response dynamics over private and incomplete information. The main disadvantage of *SFP* is its requirement to account for the opponents' historical joint behavior distribution. In non-cooperative power control, however, the users only know what payoff they are getting from their statues quo action. To make *SFP* sense in private information environment, it is clear that the users need to evaluate their utilities more directly without using the empirical frequencies. This motivates utility based *SFP*. Furthermore, we propose an improved algorithm by exploiting the utility structure. This will increase the convergence speed in the learning process.

A. Utility Based SFP

To avoid observing opponents' private information, we should propose an algorithm only with its own information in the dynamics. In stage game of power control, user i selects its action, $a_i(t)$, according to a probability distribution $\pi_i(t-1)$. User i neither knows the opponents' action distribution $\pi_{-i}(t-1)$ nor the utility $\mathcal{U}_i(a_i(t))$ before running its own action $a_i(t)$. But the user i can compute the attainable utility $\mathcal{U}_i(a_i)$ with the feedback information from receiver. It should be noted that the only use of the opponent joint strategy π_{-i} is in the assessment of the action values $\mathcal{U}_i(a_i, \pi_{-i})$ in *SFP*. Accordingly, user i compute an estimate $\bar{\mathcal{U}}_i^{a_i}(t)$ using the following recursion

$$\bar{\mathcal{U}}_i^{a_i}(t) = \begin{cases} \frac{\mathcal{U}_i^{a_i}(t) - \bar{\mathcal{U}}_i^{a_i}(t-1)}{\pi_i^{a_i}(t)t} + \bar{\mathcal{U}}_i^{a_i}(t-1) & \text{if } a_i(t) = a_i, \\ \bar{\mathcal{U}}_i^{a_i}(t-1) & \text{otherwise.} \end{cases}$$
(17)

where $\mathcal{U}_i^{a_i}(t)$ is the utility obtained by user i at time t.

The distinction between action observation and utility based *SFP* is that the users predict their utilities during the stage game based on the actual utilities corresponding to the previous estimated utilities. At any time step, $t \geq 1$, user i is always assumed to know its own proposed actions, $a_i(1), \ldots, a_i(t-1)$, and the probabilities,

$\pi_i(1), \ldots, \pi_i(t-1)$, with which its own actions were selected. The only additional information required for user i to estimate its average utility at the time $t \geq 1$ is the utilities $\bar{\mathcal{U}}_i^{a_i(1)}(1), \ldots, \bar{\mathcal{U}}_i^{a_i(t-1)}(t-1)$. In particular, the users do not need to compute the empirical frequencies of the past actions made by any user and compute their expected utilities based on the empirical frequencies. This significantly alleviates the implementation and computational bottleneck of *SFP*.

Once user i computes its estimated utility, $\bar{\mathcal{U}}_i(t)$, it proposes an action $a_i(t+1)$ for next time according to a probability distribution, $\pi_i(t) \in \mathcal{S}_i$, that maximizes its perturbed utility

$$\pi_i^T(t)\bar{\mathcal{U}}_i(t) + \lambda\mathcal{H}(\pi_i),$$
(18)

where $\mathcal{H}(\cdot)$ is the entropy function given in Eq. (3).

B. Utility based SFP with Exploiting the Utility Structure

A user does not need to know its own utility structure in utility based *SFP* approach. Therefore the recursion (17) of utility estimate only updates the value of current selected action. It may result in slow convergence and cannot fit in the application for power control. Actually, user knows its utility structure (11) in power control game. With exploiting utility structure and the measured interference and noise $\widehat{\mathcal{I}}_i$, the user can get its attainable utility of each action, $\hat{\mathcal{U}}_i^j(t)$, $j = 0, \ldots, M_i$

$$\hat{\mathcal{U}}_i^j(t) = \begin{cases} \frac{W}{\log 2} \frac{\log\left(1 + h_{ii}p_i^j/\widehat{\mathcal{I}}_i(t)\right)}{p_i^j} & \text{if } \frac{h_{ii}p_i^j}{\widehat{\mathcal{I}}_i(t)} \geq \gamma_i^*, \\ 0 & \text{otherwise.} \end{cases}$$
(19)

Herein,

$$\widehat{\mathcal{I}}_i(t) = \frac{h_{ii}p_i^{a_i(t)}(t)}{\gamma_i(t)},$$
(20)

where $a_i(t)$ is the selected action of user i at stage t. Therefore, the average utility estimate of every action can be updated based on current observation as follows

$$\bar{\mathcal{U}}_i^j(t) = \begin{cases} \bar{\mathcal{U}}_i^j(t-1) + \frac{\hat{\mathcal{U}}_i^j(t) - \bar{\mathcal{U}}_i^j(t-1)}{t\pi_i^j(t-1)} & \text{if } j = l, \\ \bar{\mathcal{U}}_i^j(t-1) + \frac{\alpha\left(\hat{\mathcal{U}}_i^j(t) - \bar{\mathcal{U}}_i^j(t-1)\right)}{t+\alpha} & \text{otherwise,} \end{cases}$$
(21)

where α is the filter parameter. Since $t\pi_i^j(t)$ will be approximately equal to the number of times user i selected the action j until time t, we can explain that the utility estimate $\bar{\mathcal{U}}_i^j(t)$ as an approximation to user i's average utility for action j corresponding to those past actions where user i has selected. Note that $\frac{\alpha}{t+\alpha}$ is decreasing with time evolution and

$$\frac{\alpha}{t+\alpha} \begin{cases} \geq 0.5 & \text{if } t \leq \alpha, \\ < 0.5 & \text{otherwise.} \end{cases}$$

We can thus explain that α is a believable parameter for the historical learning process. With this manner, we can increase the learning speed and make it sense for power control.

In the following, we can update the probability distribution $\pi_i^j(t)$ with maximizing the perturbed utility (18). To guarantee the convergence, we propose a weighted filter algorithm as follows:

$$\pi_i^j(t) = \frac{\alpha^2}{t^2 + \alpha^2} \frac{\exp\left((1/\lambda)\bar{\mathcal{U}}_i^j(t)\right)}{\sum_{k=0}^{M_i} \exp\left((1/\lambda)\bar{\mathcal{U}}_i^k(t)\right)} + \frac{t^2\pi_i^j(t-1)}{t^2 + \alpha^2}.$$
(22)

Using these update rules, the USFP_EUS algorithm is given in Algorithm 1.

Remark 2: Given the different time-varying filter parameters in (21) and (22), the algorithm inherently a two-time-scale dynamics with average utility estimate dynamics (fast dynamics) and probability distribution dynamics (slow dynamics).

Remark 3: USFP_EUS algorithm is based on the *SFP* framework. At each stage, the user's decision is *MOD*. Therefore, USFP_EUS algorithm has the self-incentive property.

Remark 4: Eq. (19) shows that each link only needs to measure its own SINR, which does not need to any message passing among the links. Eq. (21) shows that the filter algorithm only records the latest average utility estimate. Therefore, USFP_EUS algorithm is a fully decentralized algorithm with small memory unit and light computation overhead.

Algorithm 1: **USFP_EUS Algorithm**

1. For $t = 0$, initialization:

(i) Draw $a_i(0) = \text{rand}(0, M_i)$, compute $p_i(a_i(0)) = \left(1 - \frac{a_i(0)}{M_i}\right) p_i^{\min} + \frac{a_i(0)}{M_i} p_i^{\max}$.

(ii) Measure the SINR $\gamma_i(0)$ with the feedback information of the intended receiver, then derive the estimated interference and noise $\widehat{\mathcal{I}}_i(0) = h_{ii} p_i^{a_i(0)}(0) / \gamma_i(0)$.

(iii) Compute the attainable utility $\hat{\mathcal{U}}_i^j(0)$ with Eq. (19) for all the action j based on the measured $\widehat{\mathcal{I}}_i(0)$.

(iv) Initialize the average utility $\bar{\mathcal{U}}_i^j(0) = \hat{\mathcal{U}}_i^j(0)$, and the probability $\pi_i^j(0)$ for the action $j = 0, \ldots, M_i$

$$\pi_i^j(0) = \frac{\exp\left((1/\lambda)\bar{\mathcal{U}}_i^j(0)\right)}{\sum_{k=0}^{M_i} \exp\left((1/\lambda)\bar{\mathcal{U}}_i^j(0)\right)}.$$

2. For time $t \geq 1$

(i) Draw $a_i(t) = l, l \in \{0, \ldots, M_i\}$ randomly accordingly to the probabilities $\boldsymbol{\pi}_i(t-1)$, compute $p_i^l(t) = \left(1 - \frac{l}{M_i}\right) p_i^{\min} + \frac{l}{M_i} p_i^{\max}$.

(ii) Measure the SINR $\gamma_i(t)$ with the feedback information of the intended receiver, then derive the estimated interference and noise $\widehat{\mathcal{I}}_i(t)$ with Eq. (20).

(iii) Compute the attainable utility $\hat{\mathcal{U}}_i^j(t)$ with Eq. (19) for all the action j based on the measured $\widehat{\mathcal{I}}_i(t)$.

(iv) Compute the average utility estimate with Eq. (21) for $j = 0, \ldots, M_i$.

(v) Update $\pi_i^j(t)$ with Eq. (22) for $j = 0, \ldots, M_i$.

V. CONVERGENCE ANALYSIS

In this Section, we present a formally analysis on our proposed learning algorithm. First, we show that the supermodularity of the non-cooperative power control game. Second, we show the limit set of continuous time dynamics arising from *SFP* in supermodular game converges to the set of *NE* of the non-cooperative game. Third, we show that the dynamics of USFP_EUS algorithm is almost surely an asymptotic pseudotrajectory of the flow defined by the smooth best response dynamics (*SFP* dynamics). Finally, with the above results, we can derive our main conclusion: *USFP_EUS algorithm is almost surely to converge the set of pure NE*. The detail proof for Theorem 1 and Theorem 3 can be found in [26].

A. NPC Supermodular Games

We first show that *NPC* game G and G_d are supermodular game. It is an important property for our design.

Theorem 1: The continuous *NPC* with the energy efficient utility function (9) in continuous strategy space is supermodular game. Moreover, in the set of feasible power vector $\mathbf{p}^{feasible}$, the *NPC* with the utility function (11) are supermodular games both in discrete pure strategy space and in mixed strategy space.

The property of increasing differences shows that supermodular games are games in which each player's strategy set is partially ordered, the marginal returns to increasing one's strategy rise with increases in the competitors' strategies. Supermodular games are of particular interest since they have Nash Equilibria. The simplicity of supermodular games makes convexity and differentiability assumptions unnecessary [30]. The following proposition summarizes several important properties of supermodular games [20].

Proposition 3: In a supermodular game $G = [\mathcal{N}, \{\mathcal{P}_i\}, \{\mathcal{U}_i(\cdot)\}]$.

(i) The set of NEs is a nonempty and compact sublattice and so there is a component-wise smallest and largest NE.

(ii) If the users' best responses are single-valued, and each user uses the myopic best response updates starting from the smallest (largest) elements of its strategy space, then the strategies monotonically converge to the smallest (largest) NE.

(iii) If each user starts from any feasible strategy and uses myopic best response updates, the strategies will eventually lie in the set bounded component-wise by the smallest and largest NE. If the NE is unique, the myopic best response updates globally converge to that NE from any initial strategies.

B. SFP Dynamics in Supermodular Games

In this subsection, we present some results on *SFP* dynamics in supermodular games with stochastic approximation theory [6], [7], [21].

1) Continuous Time Dynamics Arising From SFP: In *SFP*, the empirical frequency vector $\boldsymbol{\pi}_i(t) \in \mathcal{S}_i$ of player i after the first $t \geq 1$ games is the vector

$$\boldsymbol{\pi}_i(t) = \frac{1}{t} \sum_{j=1}^{t} \mathbf{a}_i(t).$$

The lth component $(\boldsymbol{\pi}_i(t))_l$ of $\boldsymbol{\pi}_i(t)$ is the proportion of times in first t games that player i has played action $l \in \mathcal{A}_i$. We can obtain a recursive definition of $\boldsymbol{\pi}_i(t)$

$$\boldsymbol{\pi}_i(t+1) = \frac{1}{t+1} (t\boldsymbol{\pi}_i(t) + a_i(t+1)).$$

We can then compute the expected increments of $\boldsymbol{\pi}_i(t)$

$$\mathbf{E}(\boldsymbol{\pi}_i(t+1) - \boldsymbol{\pi}_i(t) | \boldsymbol{\pi}(t) = \pi)$$
$$= \frac{1}{t+1} [E(a_i(t+1) | \boldsymbol{\pi}(t) = \pi) - \pi]$$
$$= \frac{1}{t+1} \left(\overline{\mathcal{BR}}_i(\boldsymbol{\pi}_{-i}(t)) - \boldsymbol{\pi}_i(t)\right).$$

Thus, we see that after a reparameterization of time, expected changes in the time average $\pi(t)$ are governed by the perturbed best response dynamic

$$\dot{\pi}_i = \overline{\mathcal{BR}}_i (\pi_{-i}) - \pi_i. \tag{23}$$

This dynamics is defined on the space of mixed strategy profiles Σ.

Our analysis of the *SFP* process will relay on a close connection between the asymptotic behavior of sample paths of such a stochastic process $\{\pi(t)\}$ and the the following game differential equation

$$\frac{dx}{dt} = F(x). \tag{24}$$

2) Limit Sets and Chain Recurrence: To understand the dynamics (23) of *SFP*, we introduce the concept of limit sets and chain recurrence.

Consider a dynamic (24) that generates a semiflow $\Psi : \Re_+ \times X \to X$ on the compact set $X \subset \Re^n$. The set of fixed points of (24) can be defined as

$$\aleph = \{x \in X : F(x) = 0\} = \{x \in X : \Psi_t x = x \text{ for all } t \geq 0\} \tag{25}$$

The limit set of state x is the set of limit points of the solution trajectory $\Psi_t x$ starting at x

$$\mathcal{L}(x) = \left\{ z \in X : \lim_{t \to \infty} \Psi_t x = z \right\}$$

An invariant set for F is a set $Q \subset \Re^n$ such that $\Psi_t(Q) = Q$ for all t. For any invariant set Q we denote by $\Psi|Q$ the restriction of the flow Ψ to Q. Let Q denote a compact invariant set. A subset K of Q is called an attractor for $\Psi|Q$ provided by K is nonempty, compact and invariant, and there is neighborhood $\Upsilon \subset Q$ of K with the property that $\lim_{t \to \infty} \text{dist}(\Psi_t x, K) = 0$ uniformly for $x \in \Upsilon$. Here $\text{dist}(\Psi_t x, K)$ means the distance from $\Psi_t x$ to the nearest point of K. We call Q attractor-free if Q is nonempty compact invariant set that contains no proper attractor, which means the attractor of Q is either asymptotically stable limit cycle and equilibrium or the whole space Q. With the above notion, we give the limit set theorem from Theorem 3.3 of [7].

Proposition 4: Consider a dynamic (24). With probability one, the state limit set $\mathcal{L}(x), x \in X$ has the following properties:

(i) $\mathcal{L}(x), x \in X$ is an invariant set for the flow of the game vector field F.

(ii) $\mathcal{L}(x), x \in X$ is compact, connected and attractor-free.

Another important concept is chain recurrence. Call a sequence $\{x = x_0, x_1, \ldots, x_k = y\}$ an ε-chain from x to y if for each $i \in \{1, \ldots, k\}$, there is a $t_i > 1$ such that $|\Psi_{t_i} x_{i-1} - x_i| < \varepsilon$. The ε-chain specifies $k+1$ segments of solution trajectories to (24). We call the state x chain recurrent if there is an ε-chain from x to itself for all $\varepsilon > 0$. We let $\Im$ denote the set of chain recurrent points of (24). If every point of $x \in X$ is chain recurrent then Ψ is a chain recurrent semiflow. If there is ε-chain for all $x, y \in X$, we say that flow Ψ is chain transitive. Let $\Lambda \subset X$ be a nonempty invariant set. Ψ is called chain recurrent on Λ if every point $q \in \Lambda$ is a chain recurrent point for $\Psi|\Lambda$. A compact invariant set on which Ψ is chain recurrent (or chain transitive) is called an internally chain recurrent (or internally chain transitive) set. The proposition 5.3 in [6] makes precise the relation between the difference notions we have introduced.

Proposition 5: Let $\Lambda \subset X$. The following claims are equivalent.

(i) Λ is internally chain-transitive.

(ii) Λ is connected and internally chain-recurrent.

(iii) Λ is compact invariant set and $\Psi|\Lambda$ admits no proper attractor.

3) Convergence of SFP in Supermodular Games: The following Theorem states that the *SFP* dynamics (24) may converge to the bounded set of the fixed point almost surely.

Theorem 2: Consider *SFP* π_t starting from arbitrary initial conditions. Suppose that G is a N player supermodular game. Then

$$\Pr \left(\mathcal{L}\{\pi_t\} \subset \aleph \text{ or } \mathcal{L}\{\pi_t\} \subset \varrho_i \cap [\underline{\pi}, \overline{\pi}] \text{ for some } i \right) = 1.$$

where ϱ_i is a Lipschitz submanifold, and every persistent non-convergent trajectory is asymptotic to a trajectory in an ϱ_i, $[\underline{\pi}, \overline{\pi}]$ is the interval of the fixed point set $\aleph$. In particular, if the set of fixed points $\aleph$ is $\{\pi^*\}$, then $\Pr \left(\lim_{t \to \infty} \pi_t = \pi^* \right) = 1$.

The proof for the Theorem 2 can be based on the Proposition 4, Proposition 5 in this paper and Corollary 5.5 in [21]. The detailed proof procedure is omitted here.

C. Convergence of USFP_EUS Algorithm

First, we present a result on the two-time-scale dynamics property of USFP_EUS algorithm.

Theorem 3: For the USFP_EUS algorithm, each player i

$$\bar{\mathcal{U}}_i(t) = \bar{\mathcal{U}}_i(t-1) + \frac{1}{\mathbf{\Gamma}_i(t)} \left(\hat{\mathcal{U}}_i(t) - \bar{\mathcal{U}}_i(t-1) \right),$$

$$\pi_i(t) = \frac{\alpha^2}{t^2 + \alpha^2} \overline{\mathcal{BR}}_i \left(\bar{\mathcal{U}}_i(t) \right) + \frac{t^2}{t^2 + \alpha^2} \pi_i(t-1),$$

where $\mathbf{\Gamma}_i(t) = \left(\Gamma_i^0(t) \quad \ldots \quad \Gamma_i^{M_i}(t) \right)$

$$\Gamma_i^j(t) = \begin{cases} \frac{1}{t \pi_i^j(t)} & \text{if } j = a_i(t), \\ \frac{\alpha}{t + \alpha} & \text{otherwise}, \end{cases}$$

we have

$$\left\| \bar{\mathcal{U}}_i^{a_i}(t) - \mathcal{U}_i(a_i, \pi_{-i}(t)) \right\| \to 0 \text{ as } t \to \infty \text{ a.s.}$$

and a suitable interpolation of the $\pi_i(t)$ processes will almost surely be an asymptotic pseudotrajectory of the flow defined by the smooth best response dynamics

$$\dot{\pi}_i = \overline{\mathcal{BR}}_i (\pi_{-i}) - \pi_i.$$

The proof of Theorem 3 is based on Bokan's result on the asymptotic behavior analysis for two-time-scale coupled stochastic dynamic systems [8] and asymptotic pseudotrajectory relating the *SFP* dynamic with that of deterministic dynamic systems [6].

With the above Theorem 2, Theorem 3, Theorem 1 in [11] and the monotone property for its own action in utility function (11), we can state our main result in Theorem 4.

Theorem 4: The dynamics arising from the USFP_EUS algorithm is almost surely to converge the set of *pure NE*.
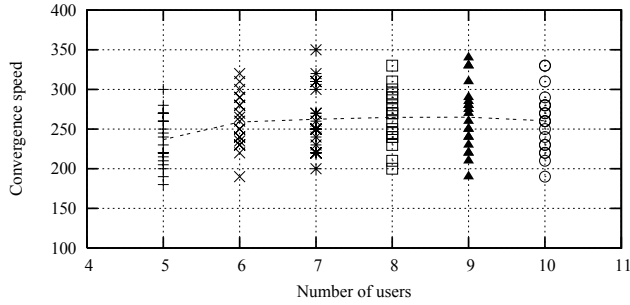
Fig. 1. The relation between the convergence speed and network scale.

Fig. 2. Utility performance ratio at different network scales.

VI. NUMERICAL EXAMPLES

We simulate a network contained in a 300m×300m square area. There are 100 nodes in the square area in a random placing manner. Two nodes can communication directly if their distance is no more than 50m. We choose N links randomly from the square area.

Each user's utility is Eq. (11). The bandwidth $W = 10^6$Hz, the AWGN noise $\sigma_i = 1.0 \times 10^{-10}$ and $\Omega = 1$. The transmitter powers are 50mW$\le p_i \le$ 100mW. The power gains are given by $h_{ij} = KS_{ij} (d_0 / d_{ij})^\beta$, where d_{ij} is the distance between the nodes, K and d_0 are normalization constants set to $K = 10^{-6}$ and $d_0 = 10$m, respectively, the path loss exponent $\beta = 4$, and the shadowing factor S_{ij} are random, independent and identically generated from a lognormal distribution with a mean of 0 dB and variance $\delta = 8$dB (so $S_{ij} = 10^{N_{ij}/10}$ and N_{ij} is Gaussian with expectation $E[N_{ij}] = 0$ and standard deviation $\delta_{N_{ij}} = 8$). In the USFP_EUS algorithm, $\alpha = 100, M_i = 49, i = 1, \ldots, N$. To overcome the overflow of the exponential operation, we set

$$\lambda_i (t) = \max \left(\frac{\min_j \left(\bar{\mathcal{U}}_i^j (t) > 0 \right)}{200}, \frac{\max_j \left(\bar{\mathcal{U}}_i^j (t) \right)}{650} \right)$$

where $\min_j \left(\bar{\mathcal{U}}_i^j (t) > 0 \right)$ is the minimum utility of all the actions with positive utilities at time t.

A. Convergence Speed

The major concern for USFP_EUS algorithm is the convergence speed of the repeated dynamics. In this simulation, we study the relation between convergence speed and network scale. We run it 20 times for each N users' scenario, $N = 5, \ldots, 10$. At each simulation, the USFP_EUS algorithm is executed with the derived power gains matrix $H = [h_{ij}]$. The relation between the convergence speed and network scale is shown in Fig. 1. Each column represents the results for a group of simulation with the same number of users, and each data point is the convergence speed of a single simulation. Fig. 1 shows that the number of iterations is no more than 350 in all the simulations and the average convergence speed is about 265. The graph shows that the convergence speed of the USFP_EUS algorithm is not affected by the network scale.

B. Performance Loss

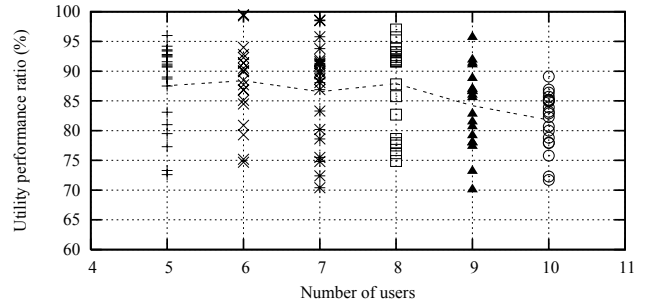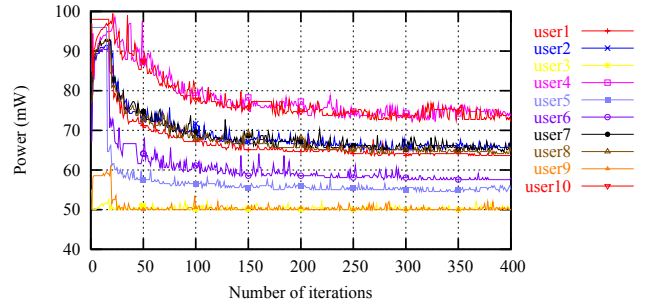Another important concern for the distributed algorithm with selfish utility is the performance

Fig. 3. Utility dynamics with $N = 10$.

loss compared with the system's optimum $\mathcal{U}^* = \max_{\boldsymbol{p}^{\min} \preceq \boldsymbol{p} \preceq \boldsymbol{p}^{\max}} \sum_{i=1}^N \mathcal{U}_i (p_i, \boldsymbol{p}_{-i})$ .we compute the utility performance ratio ς at each simulation with $\varsigma = \sum_{i=1}^N \mathcal{U}_i (j) / \mathcal{U}^*$, where $\mathcal{U}_i (j)$ is the user i's utility at steady action j. The utility performance ratios at different network scales are shown in Fig. 2. This figure illustrates that the average performance losses are from $13\% - 20\%$ at different network scales. The worst case is no more than 30%.

Finally, we depict the power dynamics in Fig. 3 and utility dynamics in Fig. 4 using one of simulation results with $N = 10$. They validate the convergence of the USFP_EUS algorithm.

VII. RELATED WORK

A. Power Control in Wireless Networks

The power control in CDMA/TDMA cellular networks has been extensively studied in the literature. Most of the work is on designing a power control algorithm to minimize transmitting power subject to guarantee the SINR of ongoing connections with the assumption of *cooperation* among the users [17], [37]. There is conflicting interest in power control. Hence, it is appropriate to address power control of cellular networks within a non-cooperative game-theoretic framework [3], [23], [25], [30], [38]. To give incentive for a certain degree of cooperation among the users, they, in general, introduced a *price-based mechanism* to constrain the totally selfish utility. Obviously, the price mechanism has a significant effect on the users' behavior, *NE* and system performance. In cellular network systems, the price can be managed by base station [30].

Compared with the cellular networks, little work has been done on the selfish behavior of power control in wireless
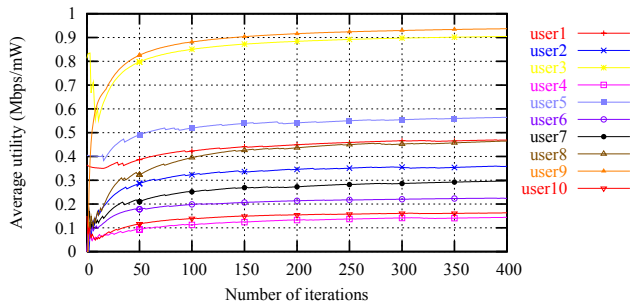
Fig. 4. Power dynamics with $N = 10$.

ad hoc networks. In [20], Huang et al. proposed a price-based power control framework for wireless ad hoc networks. Assuming that the users voluntarily cooperate by exchanging interference information, users announce prices to reflect their sensitivities to the current interference levels, and then adjust their power to maximize their surplus. They introduced fictitious non-cooperative games[2] as the convergence proof technique, whereas the actual users in the network are assumed to be cooperative, i.e., the price dynamics are forced to obey the system's optimal conditions. They pointed out that the power control for non-cooperative users in wireless ad hoc networks is still an open problem. Our work is motivated by this problem. However, we do not follow with the pricing mechanism. We propose a new reinforcement learning algorithm for non-cooperative power control, where the pricing mechanism is not required. In [22], Ileri et al. studied the network-geometric dependence of incentivized cooperation in wireless ad hoc networks with energy-efficient utility function. They designed a pricing-based joint user-and-network centric incentive mechanism that induces forwarding among selfish users by compensating the real and opportunity costs of the forwarders. However, the work is inherently to need a central node (access node) for computing the price as in wireless cellular systems.

In wireless ad hoc networks, the choice of the power level fundamentally affects the performance of multiple protocol layers. Recently, there has been much work on formulating the power control problem with cross layer design. The interested reader is referred to [10] and cited reference therein. However, most work assumes that the users are cooperative. Thus, the cross layer design problem can be converted to the system's optimal design. Our future work is to study the effect selfish behavior has on the cross layer design and non-cooperative power control with cross-layer design using the proposed reinforcement learning algorithm from this paper.

B. Non-cooperative Behavior in Wireless Networks

The problem of non-cooperative nodes/links in wireless networks has been widely addressed in the network layer [5], [12], [13], [32], [39], MAC layer [9], [27], and application layer [28], [36], whereas little work has been done on the power control. Emerging research in game theory is applied to

analyze the non-cooperative behaviors at different layers. They show much promise in helping to understand the complex interactions between nodes/links in this highly dynamic and distributed environment [33].

A one-shot game among self-interested users may result in extraordinary low utility for each user [9], [13]. However, much of the users' selfish behavior at different layers may overlap many times, which can be modeled with repeated games. The application of repeated games to model the interaction among the selfish users is still at the nascent stage [1], [13], [24]. In [24] and [1], they studied the price incentive mechanism with dynamic repeated games for Internet routing and packet forwarding in autonomous mobile ad hoc networks, respectively. Félegyházi et al. [13] studied the Nash Equilibria of packet forwarding strategies in wireless ad hoc networks with the TFT punishment strategy in repeated games. They have proven that cooperation solely based on the self-interest of the nodes can *in theory* exist. However, the conditions of such cooperation are virtually never satisfied *in practice*.

In this paper, we consider a multiuser competing wireless resource viewed as a non-cooperative game, i.e. maximizing the energy-efficiency performance with the required QoS constraint of each transceiver, regardless of what all the other users do. Compared to the related work, our work has the following unique characteristics. First, our work does not follow the general pricing mechanism design framework. We present a reinforcement algorithm with *self-incentive* dynamics, where the update rule is *MOD* based on the user's own information. Second, our work is the first step in exploring the combination function between *SFP* learning in repeated games and the strategic complementary properties in self-interested environments for designing non-cooperative power control. We state that the dynamics arising from the learning dynamics can eventually converge to a steady state with a satisfactory performance.

VIII. CONCLUSION

In this paper we have developed a non-cooperative power control algorithm with repeated games that captures the notion of repetition. Without complex price and secure mechanisms, our proposed USFP_EUS algorithm can be executed safely in a self-interest environment, which is vital for many practical applications. We provide the important insight that a felicitous intelligent learning behavior with self-incentive dynamics can eventually converge to steady state with a satisfactory system performance. This result may provide an alternative tool to design a simple protocol for a self-interest environment.

REFERENCES

[1] M. Afergan, "Using repeated games to design incentive-based routing systems," in *Proc. IEEE INFOCOM*, Barcelona, April 2006.
[2] T. Alpcan , T. Basar, R. Srikant and E. Altman, "CDMA uplink power control as a noncooperative game," *IEEE Conference on Decision and Control*, Orlando, Florida, U.S.A., Dec. 2001
[3] T. Alpcan, T. Basar, and S. Dey, "A power control game based on outage probabilities for multicell wireless data networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 890 - 899, April 2006.
[4] E. Altman, Z. Altman, "S-modular games and power control in wireless networks," *IEEE Trans. Autom. Control*, vol. 48, no. 5, pp. 839–842, May 2003.

---

[2] It should be noted that the concept of noncooperative fictitious play in [20] is different from that of the fictitious play in classic learning theory, which is introduced in this paper.

[5] L. Anderegg, S. Eidenbenz, "Ad hoc-VCG: a truthful and cost efficient routing protocol for mobile ad hoc networks with selfish agents," in *Proceedings of the 9th ACM annual international conference on Mobile computing and networking (MobiCom)*, 2003, pp. 245–259.

[6] M. Benaïm, "Dynamics of stochastic approximation algorithms," In *Séminaire de Probabilités XXXIII*, Lecture Notes in Mathematics 1709, ed. by J. Azéma et. al., Berlin: Springer, 1-68, 1999.

[7] M. Benaïm, M. W. Hirsch, "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games," *Games Econ. Behav.*, vol. 29, pp. 36–72, 1999.

[8] V. S. Borkar, "Stochastic approximation with two time scales," *Syst. & Contr. Lett.*, vol. 29, pp. 291-294, 1997.

[9] M. Čagalj, S. Ganeriwal, I. Aad, and J.-P. Hubaux, "On selfish behavior in CSMA/CA networks," in *Proc. IEEE INFOCOM*, Miami, USA, volume 4, Mar. 2005, pp. 2513-2524.

[10] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.

[11] F. Echenique, A. Edlin, "Mixed equilibria are unstable in games of strategic complements," *Journal of Economic Theory*, vol. 118, pp. 61-79, 2004.

[12] S. Eidenbenz, V. S. Anil Kumar, S. Zust, "Equilibria in Topology Control Games for Ad Hoc Networks," *Mobile Networks and Applications*, vol. 11, pp. 143–159, 2006.

[13] M. Félegyházi, J.-P. Hubaux, and L. Buttyán, "Nash equilibria of packet forwarding strategies in wireless ad hoc networks," *IEEE Trans. Mobile Computing*, vol. 5, no.4, pp. 463-476, April 2006.

[14] D. Fudenberg and D. Kreps, "Learning mixed equilibria," *Games Econ. Behav.*, vol. 5, pp. 320–367, 1993.

[15] D. Fudenberg, D. K. Levine, *The Theory of Learning in Games*. Cambridge, MA: MIT Press, 1998.

[16] D. Fudenberg, J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[17] G. J. Foschini, Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 641–646, Nov. 1993.

[18] Z. Han and K. J. Ray Liu, "Non-Cooperative Power Control Game and Throughput Game over Wireless Networks," *IEEE Trans. Commun.*, vol.53, no.10, pp. 1625-1629, Oct. 2005.

[19] T. Heikkinen, "A potential game approach to distributed power control and scheduling," *Comput. Networks,* vol. 50, no. 13, pp. 2295-2311, Sep. 2006.

[20] J. Huang, R. Berry and M. L. Honig, "Distributed interference compensation for wireless networks", *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074-1084, May 2006.

[21] J. Hofbauer, W. H. Sandholm, "On the global convergence of stochastic fictitious play," *Econometrica*, vol. 70, no. 6, pp. 2265-2294, Nov. 2002.

[22] O. Ileri, S.-C. Mau, N. B. Mandayam, "Pricing for enabling forwarding in self-configuring ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 151-162, Jan. 2005.

[23] R. A. Iltis, S.-J. Kim, and D.A. Hoang, "Noncooperative iterative MMSE beamforming algorithms for ad hoc networks," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 748- 759, 2006.

[24] Z. Ji, W. Yu, and K. J. R. Liu, "An optimal dynamic pricing framework for autonomous mobile ad hoc networks," in *Proc. IEEE INFOCOM*, Barcelona, April 2006.

[25] S. Koskie, Z. Gajic, "A nash game algorithm for SIR-based power control in 3G wireless CDMA networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 5, pp. 1017-1026, Oct. 2005.

[26] C. Long, Q. Zhang, B. Li, H. Yang and X. Guan, "Non-cooperative power control for wireless ad hoc networks with repeated games," *Technique Report*, HKUST, Sept. 2006.

[27] A. B. MacKenzie, S. B. Wicker, "Stability of multipacket slotted Aloha with selfish users and perfect information," in *Proceedings of IEEE INFOCOM*, San Francisco, CA, Mar.-Apr. 2003, vol. 3, pp. 1583-1590.

[28] P. Marbach, Y. Qiu, "Cooperation in wireless ad hoc networks: A market-based approach," *IEEE/ACM Trans. Netw.*, vol. 13, no. 6, pp. 1325–1338, Dec. 2005.

[29] P. Milgrom, J. Roberts, "Rationalizability and learning in games with strategic complementarities," *Econometrica*, vol. 58, pp. 1255-1278, 1990.

[30] C. U. Saraydar, N. B. Mandayam and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 291-303, Feb. 2002.

[31] J. Shneidman, D. C. Parkes, "Overcoming rational manipulation in mechanism implementations," 2004 [Online]. Available: http://www.gsb.stanford.edu/facseminars/events/economics/pdfs/ccac.pdf

[32] V. Srinivasan, P. Nuggehalli, C.F. Chiasserini, and R. R. Rao, "Cooperation in wireless ad hoc networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar.-Apr. 2003, vol. 2, pp. 808-817.

[33] V. Srivastava, J. Neel, A. B. Mackenzie, R. Menon, L. A. Dasilva, J. E. Hicks, J. H. Reed, R. P. Gilles, "Using game theory to analyze wireless ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 7, no. 4, pp. 46-56, Fourth Quarter 2005.

[34] D. M. Topkis, "Equilibrium points in nonzero-sum n-person submodular games," *SIAM J. Control Optim.*, vol. 17, pp. 773–778, 1979.

[35] S. Toumpis, A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp 736-748, July 2003.

[36] W. Wang, B. Li, "Market-driven bandwidth allocation in selfish overlay networks," in *Proc. IEEE INFOCOM*, Miami, USA, Mar. 2005, vol. 4, pp. 2578-2589.

[37] C. Wu, D. P. Bertsekas, "Distributed power control algorithms for wireless networks," *IEEE Trans. Veh. Technol.*, vol. 50, no. 2, pp. 504–514, Mar. 2001.

[38] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power control scheme in wireless cellular systems," *IEEE/ACM Trans. Netw.*, vol. 11, pp. 210–221, Mar 2003.

[39] S. Zhong, L. Li, Y. G. Liu, and Y. R. Yang, "On designing incentive-compatible routing and forwarding protocols in wireless ad-hoc networks – an integrated approach using game theoretical and cryptographic techniques," in *Proc. Mobicom*, Cologne, Germany, Aug.-Sep., 2005, pp. 117-131.

[40] Y. Wu, P. Chou, Q. Zhang, K. Jian, W. Zhu, and S.Y. Kung, "Network Planning in Wireless Ad hoc Networks: A Cross-layer Approach," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, Jan. 2005, pp. 136–150.

Chengnian Long (M'07) received the BS, MS, and PhD degrees from Yanshan University, China, in 1999, 2001, and 2004, respectively, all in control theory and engineering. He is an associate professor at Yanshan University. He visited Department of Computer Science and Engineering, Hongkong University of Science and Technology in 2006. Now he is a Killam postdoctoral fellow at Department of Electrical and Computer Engineering, University of Alberta. His current research interests are in the area of noncooperative behavior and incentive mechanism design in wireless multi-hop networks, energy-efficiency protocol design in wireless sensor networking, and pricing mechanism in Internet.

Qian Zhang (M'00-SM'04) received the BS, MS, and PhD degrees from Wuhan University, China, in 1994, 1996, and 1999, respectively, all in computer science. She joined the Hong Kong University of Science and Technology in September 2005 as an associate professor. Before that, she was at Microsoft Research Asia, Beijing, China, from July 1999, where she was the research manager of the Wireless and Networking Group. She has published more than 150 refereed papers in international leading journals and key conferences in the areas of wireless/Internet multimedia networking, wireless communications and networking, and overlay networking. She is the inventor of about 30 pending patents. Her current research interests are in the areas of wireless communications, IP networking, multimedia, P2P overlay, and wireless security. She has also participated many activities in the IETF ROHC (Robust Header Compression) WG group for TCP/IP header compression.

Dr. Zhang is the associate editor for the IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technologies, IEEE Transactions on Multimedia, Computer Networks and Computer Communications. She has also served as guest editor for the IEEE Wireless Communications, IEEE Journal on Selected Areas in Communications, ACM/Springer Journal of Mobile Networks and Applications (MONET), and Computer Networks. Dr. Zhang received TR 100 (MIT Technology Review) world's top young innovator award in 2004, the Best Asia Pacific (AP) Young Researcher Award elected by the IEEE Communication Society in 2004, and the Best Paper Award by the Multimedia Technical Committee (MMTC) of the IEEE Communications Society and Best Paper Award in QShine 2006. She received the Oversea Young Investigator Award from the National Natural Science Foundation of China (NSFC) in 2006. Dr. Zhang is Vice-Chair and also the Award Committee Chair of the Multimedia Communication Technical Committee of the IEEE Communications Society. She is also a member of the Visual Signal Processing and Communication Technical Committee and the Multimedia System and Application Technical Committee of the IEEE Circuits and Systems Society.

Bo Li (S'89-M'92-SM'99) received his B. Eng. (summa cum laude) and M. Eng. degrees in the Computer Science from Tsinghua University, Beijing in 1987 and 1989, respectively, and the Ph.D. degree in the Electrical and Computer Engineering from University of Massachusetts at Amherst in 1993. Between 1993 and 1996, he worked on high performance routers and ATM switches in IBM Networking System Division, Research Triangle Park, North Carolina. Since 1996, he has been with the Department of Computer Science, Hong Kong University of Science and Technology. Since 1999, he has also held an adjunct researcher position at the Microsoft Research Asia (MSRA), Beijing, China. His current research interests are on adaptive video multicast, peer-to-peer streaming, resource management in mobile wireless systems, across layer design in multihop wireless networks, content distribution and replication. He has published 90 journal papers and held several patents. He received the Outstanding Young Investigator Award by the National Natural Science Foundation of China (NSFC) in 2004. He is currently a Distinguished Lecturer in IEEE Communications Society.

He has been on editorial board for IEEE Transactions on Wireless Communications, IEEE Transactions on Mobile Computing, IEEE Transactions on Vehicular Technology, ACM/Kluwer Journal of Wireless Networks (WINET), IEEE Journal of Selected Areas in Communications (JSAC) - Wireless Communications Series, ACM Mobile Computing and Communications Review (MC2R), Elsevier Ad Hoc Networks, SPIE/Kluwer Optical Networking Magazine (ONM), KICS/IEEE Journal of Communications and Networks (JCN). He served as a guest editor for IEEE Communications Magazine Special Issue on Active, Programmable, and Mobile Code Networking (April 2000), ACM Performance Evaluation Review Special Issue on Mobile Computing (December 2000), and SPIE/Kluwer Optical Networks Magazine Special Issue in Wavelength Routed Networks: Architecture, Protocols and Experiments (January/February 2002), IEEE Journal on Selected Areas in Communications Special Issue on Protocols for Next Generation Optical WDM Networks (October 2000), Special Issue on Recent Advances in Service-Overlay Network (January 2004), and Special Issue on Quality of Service Delivery in Variable Topology Networks (September 2004), ACM/Kluwer Mobile Networks and Applications (MONET) Special Issue on Energy Constraints and Lifetime Performance in Wireless Sensor Networks (2nd Quarter of 2005), and ACM/Springer Mobile Networks and Applications (MONET) Special Issue on Advances in Wireless Mesh Networks, to appear in 2007. In addition, He has been involved in organizing over 40 conferences, esp. IEEE Infocom since 1996. He was the Co-TPC Chair for IEEE Infocom 2004.

Huilong Yang received the BS, and MS degrees in Control Theory and Engineering from Yanshan University, China, in 2005 and 2007, repectively. His research interesting are on Internet congestion control for high-speed networks, resource allocation in wireless networks.

Xinping Guan (M'02-SM'04) received the B.S. degree in mathematics from Harbin Normal University, Harbin, China, and the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering, both from Harbin Institute of Technology, in 1986, 1991, and 1999, respectively. He is currently a Professor and Dean of the Institute of Electrical Engineering, Yanshan University, Qinhuangdao, China. He is the (co)author of more than 100 papers in mathematical, technical journals, and conferences. As (a)an (co)-investigator, he has finished more than 19 projects supported by National Natural Science Foundation of China (NSFC), the National Education Committee Foundation of China, and other important foundations. His current research interests include wireless sensor networks, congestion control of networks, robust control and intelligent control for nonlinear systems, chaos control and synchronization.

Dr. Guan is serving as an Associate Editor of IEEE Transaction on Systems, Man and Cybernetics (Part C), a Reviewer of Mathematic Review of America, a Member of the Council of Chinese Artificial Intelligence Committee, and Vice-Chairman of Automation Society of Hebei Province, China. Dr. Guan is "Cheung Kong Scholars Programme" Special appointment professor, Ministry of Education, 2005, and National Science Fund for Distinguished Young Scholars of China, 2005.