Robust Additively Coupled Games in the Presence of Bounded Uncertainty in Communication Networks

Saeedeh Parsaeefard, *Student Member, IEEE*, Ahmad R. Sharafat, *Senior Member, IEEE*, and Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—We propose a novel scheme for robust communications in scenarios where channel gains, interference levels, or other measured values are uncertain or erroneously measured due to channel variations, delayed feedback, and users' mobility. When the exact values of such measurements are known, it has been shown in the literature that multiuser wireless interactions can be modeled as additively coupled games (ACGs) in which users converge to a unique Nash equilibrium by following a distributed best-response algorithm. However, in practice, such measurements are uncertain or erroneous, and hence, it is important to analyze how these uncertainties and errors affect the performance of the users playing ACGs. Most importantly, novel adjustment schemes are needed to ensure that the utility of each user is preserved under such uncertainties, i.e., introduce robustness against uncertainties and errors in ACGs. We utilize the worst case robust optimization techniques to analyze the impact of uncertainties on the users' performance and to build robust ACGs (RACGs). We derive sufficient conditions for the existence and uniqueness of their robust equilibrium and compare the outcome of an RACG and an ACG at their respective equilibria in terms of both utilities and the actions taken by the users. To reach the RACG's equilibrium, we propose a novel distributed best-response algorithm and derive sufficient conditions for its convergence. Our analytical results are supported by simulations for power control games in interference channels and for flow control in Jackson networks.**

***Index Terms*—Resource allocation, robust game theory, variational inequalities, worst case robust optimization.**

I. Introduction

DISTRIBUTED designs for multiuser communication networks have been extensively developed during the past decade to implement low-cost and scalable networks with limited overhead in terms of message passing between transmitter–receiver pairs. In doing so, each transmitter–receiver pair with local observations determines its transmit strategies in an autonomous manner. To deploy such designs, it is essential to know whether they converge to a (preferably unique) equilibrium and to evaluate their performances at the emerging equilibrium/equilibria.

Strategic noncooperative game theory is a framework for analyzing and designing networking schemes in which each user (i.e., a transmitter–receiver pair) is a rational and self-interested player that aims to maximize its own utility by choosing its transmit strategy. The notion of Nash equilibrium (NE), at which no user can attain a higher utility by unilaterally changing its strategy, is frequently used to analyze the performance of a noncooperative game. To derive sufficient conditions for NE's existence and uniqueness, different approaches such as fixed-point theory, contraction mapping, and variational inequalities (VI) [2]–[4] have been applied in both wired and wireless networks, including applications to routing in Jackson networks, and to power control in interference channels [5]–[10].

A game-theoretic approach to setting users' strategies requires utilization of local observations and measurements in the network, whose values are often corrupted by noise or are uncertain. Such uncertainties are attributable to many factors, including random delays in feedback channels, errors in estimated values, and channel variations. Obtaining exact measurements is not practical, and as we will demonstrate in Section VIII-A1, performance of the network (the network's social utility) at its NE, e.g., its throughput or its delay, as the case may be, fluctuates to a great extent, which is highly undesirable. Hence, ensuring robustness against errors is of paramount importance and is essential for the correct operation of any network. Our proposed design in this paper is an effort to address these real-world measurement errors and uncertainties.

To make an NE robust against uncertainties, robust optimization theory has been widely used in the literature [11]–[13], in which each uncertain parameter is a new optimization variable. This involves converting the nominal optimization problem (optimization without considering uncertainty) to its robust counterpart (optimization with uncertainty) via two basic approaches: the Bayesian approach, where the statistics of uncertain parameters are considered and the utility of each user is probabilistically guaranteed; and the worst case approach, where a given closed region, called the uncertainty region, is considered for the distance between the exact and the estimated values of uncertain parameters and the utility of each user is guaranteed for any realization of uncertainty within the uncertainty region [14]–[17].

Both of these approaches have been applied to the power allocation problem in spectrum sharing environments. For example, in [18], the probability distribution function of

Manuscript received July 28, 2012; revised March 7, 2013 and June 14, 2013; accepted September 3, 2013. Date of publication October 3, 2013; date of current version March 14, 2014. This work was supported in part by Tarbiat Modares University, Tehran, Iran, by the Iran Telecommunications Research Center under Ph.D. Research Grant TMU 88-11-124, and by the National Science Foundation under NSF Grant CCF 0830556. This paper was presented in part at the IEEE Global Telecommunications Conference (GLOBECOM), Houston, TX, USA, December 5–9, 2011. The review of this paper was coordinated by Prof. O. B. Akan.

S. Parsaeefard and A. R. Sharafat are with the Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran 14117-13116, Iran (e-mail: sharafat@modares.ac.ir).

M. van der Schaar is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2013.2284344

uncertainty in interference of users on each other is assumed to be uniform, and the Bayesian approach is utilized to study the performance of the proposed robust power allocation scheme. In [19], the interference channel gain is the only uncertain parameter, and the worst case approach is used to study the performance of the proposed robust power allocation game in multicarrier systems. In [20], the dynamic worst case robust power allocation game is proposed for multiple-input–multiple-output systems to tackle uncertainty in interference channel gains. In [16], the worst case robust power allocation game is utilized to maintain the uncertain interference level of each user below a given threshold.

To incorporate robustness in communication networks, there exist multiple challenges, such as 1) how to implement robustness in a wider class of problems, 2) how to derive the sufficient conditions for the existence and uniqueness of the robust NE (RNE), 3) determining the impact of considering uncertainties on the system's performance at its equilibrium compared with that of the case with no uncertainty, and 4) how to design a distributed algorithm for reaching the robust equilibrium.

We consider a general class of games—called additively coupled games (ACGs)—in which each user's utility depends on its action and system parameters and is additively coupled to other users' actions or impacts [6]. ACGs can model any problem in communication networks for which such additive couplings exist, as in the three examples that are covered in this paper, namely, transmit power allocation in interference channels in wireless networks, downlink transmit power allocation in digital subscriber line access multiplexers (DSLAMs), and routing delay minimization in Jackson networks [21].

To introduce robustness in a given ACG, we choose the worst case robust approach, as it can preserve the utility of each user under any condition of error in distributed networks. It has been shown in [17], [22], and [24] that an error can be considered to stay within an uncertainty region specific to the nature of uncertainty. We assume that users' impacts on each other are uncertain, caused by variations in system parameters and in other users' strategies. This is different from the assumption in other works (e.g., in [19], where the only cause of uncertainty is variations in interference channel gains). We assume that uncertainty in the additive impact is modeled by a bounded error and that each user aims to maximize its utility for the worst case condition of error [11]–[13]. As in the nomenclature of robust optimization theory, we refer to an ACG in which uncertainty is not considered and its equilibrium as a nominal ACG (NACG) and a nominal NE, respectively; and refer to an ACG in which uncertainty is considered (and the worst case robust optimization is utilized) and its equilibrium as a robust ACG (RACG) and an RNE, respectively. To derive sufficient conditions for the existence and uniqueness of the RNE, we apply VI [3] and show that when uncertainty is bounded and convex, an RNE always exists. We also show that the RNE is a bounded perturbed version of the nominal NE and derive the condition for RNE's uniqueness based on the condition for the nominal NE's uniqueness.

When the nominal NE is unique, we show that the social utility (the sum of utilities of all users) at the RNE is always less than that at the nominal NE and derive the upper bound for differences between users' actions at the RNE and at the nominal NE. When the nominal NE is not unique, we show that the social utility at the RNE may be higher than that at the respective nominal NE and derive a sufficient condition for this phenomenon. Finally, we use the proximal response map associated with the worst case utility function to propose a distributed algorithm for reaching the RNE and derive sufficient conditions for its convergence.

The rest of this paper is organized as follows. In Section II, the system model of the NACG is summarized, followed by Section III, where we introduce the RACG and its RNE. Section IV covers the sufficient conditions for existence and uniqueness of the RNE, and in Section V, we simplify the results for logarithmic utility functions. In Section VI, we propose distributed algorithms for reaching the RNE, followed by Section VII, where we discuss the effects of robustness for the case of multiple nominal NEs. In Section VIII, simulation results validate our analytical developments for the power allocation problem and for the Jackson networks. Finally, conclusions are drawn in Section IX.

II. System Model

Consider a set of communication resources divided into K orthogonal dimensions denoted by $\mathcal{K} = \{1, \ldots, K\}$, e.g., frequency bands, time slots, or routes, which are shared between a set of users denoted by $\mathcal{N} = \{1, \ldots, N\}$, where each user consists of a transmitter and a receiver. We assume that users do not cooperate with each other and formulate the resource allocation problem as a strategic noncooperative game $\mathcal{G} = \{\mathcal{N}, (v_n)_{n\in\mathcal{N}}, \mathcal{A}\}$, where v_n is the utility of user n, $\mathcal{A} = \prod_{n\in\mathcal{N}} \mathcal{A}_n$ is the joint strategy space of the game, and $\mathcal{A}_n \subseteq \mathbb{R}^K$ is the strategy space of user n, where the strategy of each user in each dimension and the sum of its strategies over all dimensions are bounded [6], [9], i.e.,

$$\mathcal{A}_n = \left\{ \mathbf{a}_n = \left(a_n^1, \ldots, a_n^K\right) | a_n^k \in \left[a_{n,k}^{\min}, a_{n,k}^{\max}\right], \right. \\ \left. \text{and} \sum_{k=1}^{K} a_n^k \le a_n^{\max} \right\} \quad (1)$$

where $a_{n,k}^{\max}$ and $a_{n,k}^{\min}$ are the maximum and the minimum transmit strategies of each user in each dimension, and $a_n^{\max}$ is the upper bound on the sum of strategies of user n over all dimensions. In practice, $a_{n,k}^{\min}$ is much less than $a_{n,k}^{\max}$ and can even be negligible. The function $v_n(\mathbf{a}) : \mathcal{A} \to \mathbb{R}$ is the utility function of user n, whose value depends on the chosen strategy vector of all users $\mathbf{a} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$, where $\mathbf{a}_n \in \mathcal{A}_n$ is the action of user n. The vector of actions of all users except user n is $\mathbf{a}_{-n} \in \mathcal{A}_{-n}$, where $\mathcal{A}_{-n} = \prod_{m\in\mathcal{N}, m\neq n} \mathcal{A}_m$ is the strategy space of all users except user n. In a noncooperative strategic game, each user n aims to maximize its own utility subject to its strategy space via $\max_{\mathbf{a}_n\in\mathcal{A}_n} v_n(\mathbf{a})$. Assume

$$v_n\left(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\right) = \sum_{k=1}^{K} v_n^k\left(a_n^k, f_n^k(\mathbf{a}_{-n}, \mathbf{s}_n)\right)$$

where $\mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n) = [f_n^1(\mathbf{a}_{-n}, \mathbf{s}_n), \ldots, f_n^K(\mathbf{a}_{-n}, \mathbf{s}_n)]$ is the $1 \times K$ vector of the additive impacts of other users' strategies and system parameters on user n, whose elements are $f_n^k(\mathbf{a}_{-n}, \mathbf{s}_n) = \sum_{m\in\mathcal{N}, m\neq n} a_m^k x_{nm}^k + y_n^k$ for all $k \in \mathcal{K}$, where $\mathbf{s}_n = [\mathbf{x}_{n1}, \ldots, \mathbf{x}_{n(n-1)}, \mathbf{x}_{n(n+1)}, \ldots, \mathbf{x}_{nN}, \mathbf{y}_n]$ is a $1 \times (N \times K)$ vector for user n, in which $\mathbf{x}_{nm} = [x_{nm}^1, \ldots, x_{nm}^K]$, and x_{nm}^k represents system parameters specific to the impact of the transmitter of user m on the receiver of user n in dimension k (e.g., interference), and $\mathbf{y}_n = [y_n^1, \ldots, y_n^K]$ is the vector of the system's other parameters for the receiver of user n (e.g., ambient noise). The utility of each user is a function of its action and additive impacts of other users' actions and system parameters. Hence, this class of games is called ACGs [6]. In our setup, the receiver of each user measures the impacts of actions of other users and values of system parameters and sends them back to its transmitter via the feedback channel to solve the given optimization problem.

The given formulation can model a number of practical problems in communication networks. For example, consider power control in interference channels, where each user, consisting of a transmitter-and-receiver pair, competes with other users to maximize its total throughput over K distinct subchannels, and the transmit power of user n over any subchannel is upper bounded to $p_n^{\max}$ in [9]. We denote this upper bound as $a_n^{\max}$ in (1) so that our notations would be applicable in general and not to a specific problem. The strategy of each user is its transmit power in K subchannels. The additive impact of other users on user n is their interference in subchannel k, expressed by $f_n^k(\mathbf{a}_{-n}, \mathbf{s}_n) = \sum_{m\neq n} a_m^k h_{nm}^k + \sigma_n^k$, where h_{nm}^k is the channel gain between user m and user n in subchannel k, a_n^k is the transmit power of user n in subchannel k, and σ_n^k is the channel noise in subchannel k of user n. In this example, interference channel gains between users and noise in each subchannel are considered as system parameters, i.e., $\mathbf{x}_{nm} = \mathbf{h}_{nm}$, where $\mathbf{h}_{nm} = [h_{nm}^1, \ldots, h_{nm}^K]$, and $y_n^k = \sigma_n^k$.

In line with the existing literature, we also assume the following: **A1**) As in [23], the utility function of each user is an increasing, twice differentiable, and strictly concave function with respect to $\mathbf{a}_n$ and has bounded gradients. **A2**) As in [6], the utility function of each user is a decreasing, twice differentiable, and strictly convex function with respect to $\mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$. **A3**) As in [6] and [23], the second-order mixed partial derivatives of utility functions, i.e., $((\partial^2 v_k^n)/(\partial a_k^n \partial f_k^n))$ and $((\partial^2 v_k^n)/(\partial f_k^n \partial a_k^n))$, exist.

For the power control game, the utility of user n is its throughput over all subchannels, i.e.,

$$v_n = \sum_{k=1} \log\left(1 + \frac{a_n^k h_{nn}^k}{\sum_{m\neq n} a_m^k h_{nm}^k + \sigma_n^k}\right) \tag{2}$$

and Assumptions **A1–A3** hold.

Interactions between users are studied at the nominal NE, which corresponds to the strategy profile $\mathbf{a}^* = [\mathbf{a}_1^*, \ldots, \mathbf{a}_N^*]$, such that for any other strategy profile, we have $v_n(\mathbf{a}_n^*, \mathbf{f}_n(\mathbf{a}_{-n}^*, \mathbf{s}_n)) \geq v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}^*, \mathbf{s}_n))$ for all $\mathbf{a}_n \in \mathcal{A}_n$ and $n \in \mathcal{N}$, where $\mathbf{a}_{-n}^* = [\mathbf{a}_1^*, \ldots, \mathbf{a}_{n-1}^*, \mathbf{a}_{n+1}^*, \ldots, \mathbf{a}_N^*]$ [2]. In what follows, the utility of user n and the social utility of ACG (the sum of utilities of all users) at the nominal NE are denoted by v_n^* and $v^* = \sum_{n=1}^N v_n^*$, respectively. We derive the sufficient conditions for the nominal NE's existence and uniqueness via reformulating the nominal NE through VI and show that, this way, the nominal NE and the RNE can be analyzed in a similar manner.

Remark 1: Consider the mapping vector $\mathcal{F}(\mathbf{a}) = (\mathcal{F}_n(\mathbf{a}))_{n=1}^N$, where

$$\mathcal{F}_n(\mathbf{a}) = -\nabla_{\mathbf{a}_n} v_n\left(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\right) \tag{3}$$

in which $\nabla_{\mathbf{a}_n} v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))$ denotes the column gradient vector of $v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))$ with respect to $\mathbf{a}_n$. The NE of $\mathcal{G}$ can be obtained by solving $VI(\mathcal{A}, \mathcal{F})$ (see [3, Proposition 1.4.2]) as $(\mathbf{a} - \mathbf{a}^*)\mathcal{F}(\mathbf{a}^*) \geq 0$ for all $\mathbf{a} \in \mathcal{A}$. Since $v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))$ is a continuous and concave function with respect to $\mathbf{a}_n \in \mathcal{A}_n$, $\mathcal{F}(\mathbf{a})$ is a continuous mapping. From (1), set $\mathcal{A}$ is convex and compact. Therefore, the solution set of $VI(\mathcal{A}, \mathcal{F})$ is nonempty and compact (see [3, Th. 2.2.1]). Consequently, the NE of ACG exists.

Remark 2: For the mapping $\mathcal{F}(\mathbf{a})$, we have

$$\alpha_n(\mathbf{a}) \triangleq \text{smallest eigenvalue of} - \nabla_{\mathbf{a}_n}^2 v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))$$

$$\beta_{nm}(\mathbf{a}) \triangleq \|-\nabla_{\mathbf{a}_n\mathbf{a}_m} v_n\left(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\right)\|_2 \ \forall n \neq m$$

where $\nabla_{\mathbf{a}_n}^2 v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))$ and $\nabla_{\mathbf{a}_n\mathbf{a}_m} v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))$ are the $K \times K$ Jacobian matrices of $\mathcal{F}_n(\mathbf{a})$ with respect to $\mathbf{a}_n$ and $\mathbf{a}_m$, respectively, and $\| - \nabla_{\mathbf{a}_n\mathbf{a}_m} v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))\|_2$ is the l_2-norm of $v_n(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n))$. Let

$$\alpha_n^{\min} \triangleq \inf_{\mathbf{a}\in\mathcal{A}} \alpha_n(\mathbf{a}) \tag{4}$$

$$\beta_{nm}^{\max} \triangleq \sup_{\mathbf{a}\in\mathcal{A}} \beta_{nm}(\mathbf{a}) \tag{5}$$

for all users, and as in [4, Sec. 12], define the $N \times N$ matrix $\mathbf{\Upsilon}$ whose elements are

$$[\Upsilon]_{nm} = \begin{cases} \alpha_n^{\min}, & \text{if } m = n \\ -\beta_{nm}^{\max}, & \text{if } m \neq n. \end{cases} \tag{6}$$

When $\mathbf{\Upsilon}$ is a P-matrix (matrix $\mathbf{\Upsilon}$ is a P-matrix if for any nonzero vector $\mathbf{x}$, we have $x_i(\mathbf{\Upsilon}\mathbf{x})_i > 0$, where x_i is the ith element of $\mathbf{x}$ [3]), the nominal NE is unique (see [4, Th. 12.5 in Sec. 12.4.1]).

III. Robust Games

User n may encounter different sources of uncertainty caused by variations in $\mathbf{a}_{-n}$ and/or $\mathbf{x}_{nm}$, which cause variations in the utility of that user and prevent it from attaining its expected performance. We assume that all uncertainties in $\mathbf{a}_{-n}$ and $\mathbf{x}_{nm}$ for user n can be modeled by variations in $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$, i.e., $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n) = \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n) + \mathbf{e}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$, where $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n) = [\widetilde{f}_n^1(\mathbf{a}_{-n}, \mathbf{s}_n), \ldots, \widetilde{f}_n^K(\mathbf{a}_{-n}, \mathbf{s}_n)]$, $\mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n) = [f_n^1(\mathbf{a}_{-n}, \mathbf{s}_n), \ldots, f_n^K(\mathbf{a}_{-n}, \mathbf{s}_n)]$, and $\mathbf{e}_n(\mathbf{a}_{-n}, \mathbf{s}_n) = [e_n^1(\mathbf{a}_{-n}, \mathbf{s}_n), \ldots, e_n^K(\mathbf{a}_{-n}, \mathbf{s}_n)]$ are the estimated values, the exact values, and the error in the impacts of other users on user n, respectively. In the worst case robust optimization, uncertainties are assumed to be bounded to the uncertainty region, i.e.,

$$\Re_n(\mathbf{a}_{-n}) = \left\{\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n) > \mathbf{0}_K | \ \|\mathbf{e}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\|_2 \leq \varepsilon_n\right\}, \quad \forall\, n \in \mathcal{N} \tag{7}$$

where $\mathbf{0}_K$ is the $1 \times K$ zero vector, $\varepsilon_n \geq 0$ is the bound on uncertainty, and $\|\mathbf{e}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\|_2 = \sqrt[2]{\sum_{k=1}^{K} |e_n^k(\mathbf{a}_{-n}, \mathbf{s}_n)|^2}$ is the l_2-norm of $\mathbf{e}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$. Note that in (7), $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n) > \mathbf{0}_K$ means that each element of $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$ is greater than its corresponding element in $\mathbf{0}_K$. Since in communication networks the ellipsoid region (i.e., l_2-norm) has been commonly used to model uncertainty [17], [24], [25], we also use it in our robust game. The uncertainty region for user n, denoted by $\Re_n(\mathbf{a}_{-n})$, depends on system parameters (e.g., the uncertainty region as defined in [19] for interference channel gains) and on other users' actions.

The effect of uncertainty in $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$ is highlighted by a new vector of variables in the utility function of each user. We denote the utility of user n in the robust game by u_n, which depends on uncertain parameters $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$, and write $u_n(\mathbf{a}_n, \widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)) = \sum_{k=1}^{K} u_n^k(a_n^k, \widetilde{f}_n^k(\mathbf{a}_{-n}, \mathbf{s}_n))$. As in [14] and [17], we have

$$v_n\left(\mathbf{a}_n, \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\right) = u_n\left(\mathbf{a}_n, \widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\right)\Big|_{\varepsilon_n = 0} \tag{8}$$

for all $n \in \mathcal{N}$, which means that when the uncertainty region shrinks to zero (i.e., there is no uncertainty), utility functions of the nominal and the robust optimization problems are identical. The objective of the worst case approach is to find the optimal strategy for each user that optimizes its utility under the worst condition of error in the uncertainty region. In this approach, the optimization problem of each user is formulated as [11]

$$\widetilde{u}_n = \max_{\mathbf{a}_n \in \mathcal{A}_n} \min_{\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n) \in \Re_n(\mathbf{a}_{-n})} u_n\left(\mathbf{a}_n, \widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\right) \tag{9}$$

where $\widetilde{u}_n$ is the achieved utility of user n in the worst case approach. In (9), the minimization is due to the fact that we adopt a worst case approach, and the maximization is due to the fact that each user is aiming to maximize its own performance. Note that (9) is obtained via applying the worst case approach and **A2** and means that the utility of each user is preserved under the worst case condition of error [17]. The domain of optimization problem (9) is $\widetilde{\mathcal{A}}_n(\mathbf{a}_{-n}) = \mathcal{A}_n \times \Re_n(\mathbf{a}_{-n})$, which is a function of other users' strategy. We represent the RACG by $\widetilde{\mathcal{G}} = \{\mathcal{N}, (u_n)_{n \in \mathcal{N}}, \widetilde{\mathcal{A}}(\mathbf{a})\}$, where $\widetilde{\mathcal{A}}(\mathbf{a}) = \prod_{n=1}^{N} \widetilde{\mathcal{A}}_n(\mathbf{a}_{-n})$. The strategy of each user is denoted by $\widetilde{\mathcal{A}}_n(\mathbf{a}_{-n})$ to emphasize that it is dependent on strategies of other users, and it is a set-valued mapping. Consequently, $\widetilde{\mathcal{A}}(\mathbf{a})$ is also a set-valued mapping and is a function of $\mathbf{a}$. The solution to (9) for user n is a pair $(\widetilde{\mathbf{a}}_n', \mathbf{f}_n'(\mathbf{a}_{-n}, \mathbf{s}_n)) \in \widetilde{\mathcal{A}}_n(\mathbf{a}_{-n})$ that satisfies [15]

$$\begin{aligned}
&\max_{\mathbf{a}_n \in \mathcal{A}_n} u_n\left(\mathbf{a}_n, \mathbf{f}_n'(\mathbf{a}_{-n}, \mathbf{s}_n)\right) \\
&\quad = u_n\left(\widetilde{\mathbf{a}}_n', \mathbf{f}_n'(\mathbf{a}_{-n}, \mathbf{s}_n)\right) \\
&\quad = \min_{\mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n) \in \Re_n(\mathbf{a}_{-n})} u_n\left(\widetilde{\mathbf{a}}_n', \mathbf{f}_n(\mathbf{a}_{-n}, \mathbf{s}_n)\right)
\end{aligned} \tag{10}$$

which is the saddle point of (9). Using the given formula, the equilibrium of $\widetilde{\mathcal{G}}$ is defined below.

Definition 1: The RNE of RACG corresponds to the strategy profile $\widetilde{\mathbf{a}}^* = [\widetilde{\mathbf{a}}_1^*, \ldots, \widetilde{\mathbf{a}}_N^*]$ if and only if for any other strategy profile $\widetilde{\mathbf{a}}_n$ we have [11]–[13]

$$\begin{aligned}
&\min_{\widetilde{\mathbf{f}}_n(\widetilde{\mathbf{a}}_{-n}^*, \mathbf{s}_n) \in \Re_n(\mathbf{a}_{-n}^*)} u_n\left(\widetilde{\mathbf{a}}_n^*, \widetilde{\mathbf{f}}_n(\widetilde{\mathbf{a}}_{-n}^*, \mathbf{s}_n)\right) \geq \\
&\min_{\widetilde{\mathbf{f}}_n(\widetilde{\mathbf{a}}_{-n}^*, \mathbf{s}_n) \in \Re_n(\mathbf{a}_{-n}^*)} u_n\left(\widetilde{\mathbf{a}}_n, \widetilde{\mathbf{f}}_n(\widetilde{\mathbf{a}}_{-n}^*, \mathbf{s}_n)\right), \quad \forall\, \widetilde{\mathbf{a}}_n \in \mathcal{A}_n
\end{aligned} \tag{11}$$

where $\widetilde{\mathbf{a}}_{-n}^* = [\widetilde{\mathbf{a}}_1^*, \ldots, \widetilde{\mathbf{a}}_{n-1}^*, \widetilde{\mathbf{a}}_{n+1}^*, \ldots, \widetilde{\mathbf{a}}_N^*]$. We denote the achieved utility of user n at the RNE by $\widetilde{u}_n^*$ and the social utility at the RNE by $\widetilde{u}^* = \sum_{n=1}^{N} \widetilde{u}_n^*$. From (11), the RNE is the equilibrium of the game with uncertain parameters, and each user aims to solve its worst case robust optimization problem. At the RNE, each user reaches its maximum utility under the worst condition of error, and no user can reach a higher utility by unilaterally changing its strategy. Note the difference with the nominal game, at whose NE each user aims to maximize its utility by choosing a strategy from its strategy set without considering uncertainty in $\mathbf{f}_n$. When $\varepsilon_n = 0$, the RNE and the NE are identical.

IV. Robust Nash Equilibrium Analysis: Conditions for Robust Nash Equilibrium's Existence and Uniqueness

Now, we derive the characteristics of the RNE in the RACG from the nominal NE in the NACG. For convenience, in what follows, we omit the arguments $\mathbf{a}_{-n}$ and $\mathbf{s}_n$ in $\widetilde{\mathbf{f}}_n(\mathbf{a}_{-n}, \mathbf{s}_n)$.

A. *Existence of the RNE*

In analyzing the existence of the RNE, we encounter two problems. First, by considering uncertainty in the utility of each user, the utility may become nonconvex, and analyzing the RNE may become impossible. Second, the strategy space of user n changes to $\widetilde{\mathcal{A}}_n(\mathbf{a}_{-n}) = \mathcal{A}_n \times \Re_n(\mathbf{a}_{-n})$, which is a set-valued mapping and is a function of other users' actions. Therefore, convexity of the optimization problem of each user is not a sufficient condition for the existence of RNE, and hence, we need to establish the existence of the RNE via other approaches.

Lemma 1:

1) For the uncertainty region in (7), the strategy of each user (i.e., $\widetilde{\mathcal{A}}_n(\mathbf{a}_{-n})$) is a convex, bounded, and closed set.
2) The function $\Psi_n(\mathbf{a}_n, \mathbf{a}_{-n})$ is a concave and continuous differentiable function of $\mathbf{a}_n$ for every $\mathbf{a}_{-n}$, where

$$\Psi_n(\mathbf{a}_n, \mathbf{a}_{-n}) = \min_{\widetilde{\mathbf{f}}_n \in \Re_n(\mathbf{a}_{-n})} u_n(\mathbf{a}_n, \widetilde{\mathbf{f}}_n) = u_n\left(\mathbf{a}_n, \widetilde{\mathbf{f}}_n^*\right) \tag{12}$$

and $\widetilde{\mathbf{f}}_n^* = \mathbf{f}_n - \varepsilon_n \boldsymbol{\vartheta}_n$, where $\widetilde{\mathbf{f}}_n^* = [\widetilde{f}_n^{1*}, \ldots, \widetilde{f}_n^{K*}]$, $\boldsymbol{\vartheta}_n = [\vartheta_n^1, \ldots, \vartheta_n^K]$, and ϑ_n^k is

$$\vartheta_n^k = \frac{\frac{\partial u_n^k(\mathbf{a}_n, \widetilde{\mathbf{f}}_n)}{\partial \widetilde{f}_n^k}}{\sqrt{\sum_{k=1}^{K} \left(\frac{\partial u_n^k(\mathbf{a}_n, \widetilde{\mathbf{f}}_n)}{\partial \widetilde{f}_n^k}\right)^2}}. \tag{13}$$

The robust game is $\widetilde{\mathcal{G}} = \{\mathcal{N}, (\Psi_n)_{n \in \mathcal{N}}, \mathcal{A}\}$.

Proof: See Appendix A. ■

For the robust game in Part 2 of Lemma 1, the mapping for $\widetilde{\mathcal{G}}$ is $\widetilde{\mathcal{F}}(\mathbf{a}) = (\widetilde{\mathcal{F}}_n(\mathbf{a}))_{n=1}^N$, where $\widetilde{\mathcal{F}}_n(\mathbf{a}) = -(\partial\Psi_n(\mathbf{a}_n, \mathbf{a}_{-n}))/(\partial\mathbf{a}_n)$. Since $\widetilde{\mathcal{F}}(\mathbf{a})$ is the set-valued mapping, the RNE can be obtained via generalized VI (GVI) [13], i.e., $\widetilde{\mathbf{a}}^*$ is the RNE if and only if it is a solution to $GVI(\mathcal{A}, \widetilde{\mathcal{F}})$. Now, we study the existence of the RNE.

Theorem 1: For any set of system parameters, users' actions, and the bound on the uncertainty region, there always exists an RNE for $\widetilde{\mathcal{G}}$.

Proof: It is easy to show that all the assumptions in [13, Lemma 3.1 and Th. 3.2] hold, and $GVI(\mathcal{A}, \widetilde{\mathcal{F}})$ has a solution. Hence, $\widetilde{\mathcal{G}}$ has an RNE. ■

B. Uniqueness of the RNE

Since the closed-form solution to (9) cannot be obtained, the fixed-point algorithm and the contraction mapping cannot be applied as in [6] and [9] to derive the sufficient conditions for RNE's uniqueness. However, we will show that the RNE is the bounded perturbed version of the nominal NE of the NACG and that the condition for RNE's uniqueness can be derived without a closed-form solution to (9).

Lemma 2: The mapping $\widetilde{\mathcal{F}}(\mathbf{a})$ is a bounded perturbed version of the mapping $\mathcal{F}(\mathbf{a})$, i.e., there exists a $0 < \wp < \infty$ such that $\|\widetilde{\mathcal{F}}_n(\mathbf{a}) - \mathcal{F}_n(\mathbf{a})\|_2 \le \wp$.

Proof: See Appendix B. ■

From Lemma 1, $\widetilde{\mathcal{A}}(\mathbf{a})$ is a closed and convex set, and from Lemma 2, $\widetilde{\mathcal{F}}(\mathbf{a})$ is a bounded perturbed version of the mapping $\mathcal{F}(\mathbf{a})$. We use this point to derive a sufficient condition for RNE's uniqueness in Theorem 2 as follows.

Theorem 2: When $\boldsymbol{\Upsilon}$ in (6) is a P-matrix, for any bounded $\boldsymbol{\Delta} = [\varepsilon_1, \ldots, \varepsilon_N]$, the following statements hold.

1) $\widetilde{\mathcal{G}}$ has a unique RNE.
2) The social utility at the RNE is always less than or equal to that at the nominal NE, i.e., $\widetilde{u}^* \le v^*$.
3) The distance between the strategy profiles at the RNE and at the nominal NE is

$$\|\mathbf{a}^* - \widetilde{\mathbf{a}}^*\|_2 \le \frac{\|\boldsymbol{\Delta}\|_2}{c_{\mathrm{sm}}(\mathcal{F})} \tag{14}$$

where $c_{\mathrm{sm}} > 0$ is the strong monotonicity constant for the mapping $\mathcal{F}$, which guarantees $(\mathbf{a}_1 - \mathbf{a}_2)(\mathcal{F}(\mathbf{a}_1) - \mathcal{F}(\mathbf{a}_2)) \ge c_{\mathrm{sm}}\|\mathbf{a}_1 - \mathbf{a}_2\|_2^2$ for all $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}$. In such a case, $\mathcal{F}$ is a strongly monotone map [3].

Proof: See Appendix C. ■

Recall that $\boldsymbol{\Upsilon}$ in (6) being a P-matrix establishes strong monotonicity of the mapping $\mathcal{F}(\mathbf{a})$, which is a sufficient condition for the uniqueness of the nominal NE as per [4, Th. 12.5 in Sec. 12.4.1]. When this condition holds and the mapping $\widetilde{\mathcal{F}}(\mathbf{a})$ is a bounded perturbed version of the mapping $\mathcal{F}(\mathbf{a})$, as per Lemma 3 in the proof of Theorem 2, $\widetilde{\mathcal{F}}(\mathbf{a})$ is also strongly monotone, which is a sufficient condition for the uniqueness of the RNE as per [13, Th. 4.3]. Note that there may be other sufficient conditions under which nominal game $\mathcal{G}$ has a unique solution but robust game $\widetilde{\mathcal{G}}$ may not have a unique solution or even any solution.

In addition, from Theorem 2, the difference between the upper bound on users' actions in (14) at the RNE and at the nominal NE can be compared. Moreover, from (14), the social utility's decrement at the RNE (compared with that at the nominal NE) can be approximated. Note that c_{sm} is obtained from $\mathcal{F}$ (see [3, Sec. II] and [29, Sec. VI]). The given discussion means that Theorem 2 provides the means to determine how uncertainty affects the outcome of the robust game and shows how to obtain the RNE from the nominal NE. Consider $\mathbf{W}(\mathbf{a}) = (\mathbf{W}^k(\mathbf{a}))_{k=1}^K$, where $\mathbf{W}^k(\mathbf{a})$ is an $N \times N$ matrix whose elements are

$$W_{nm}^k \equiv \begin{pmatrix} \frac{\partial v_n^k\left(a_n^k, f_n^k\right)}{\partial a_n^k}, & \text{if } m = n \\ \frac{\partial v_n^k\left(a_n^k, f_n^k\right)}{\partial a_m^k} x_{nm}^k, & \text{if } m \ne n \end{pmatrix}, \qquad m, n \in \mathcal{N}. \tag{15}$$

In Appendix D, we show that the difference between social utilities at the RNE and at the nominal NE is

$$\|v^* - \widetilde{u}^*\|_2 \approx \|\mathbf{W}(\mathbf{a}^*)\|_2 \times \frac{\|\boldsymbol{\Delta}\|_2}{c_{\mathrm{sm}}(\mathcal{F})}. \tag{16}$$

In Appendix D, we also show that the gap between the exact value of $\|v^* - \widetilde{u}^*\|_2$ and its approximation in (16) is always less than or equal to $((\|J(\mathcal{F})\|_2\|\boldsymbol{\Delta}\|_2^2)/2)$, where $J(\mathcal{F})$ is the Jacobian matrix of $\mathcal{F}$, and $J(\mathcal{F})$ contains the second derivative of v_n. When the second derivative of v_n is small, (16) is a tight approximation for the difference between social utilities at the RNE and at the nominal NE. When $\boldsymbol{\Upsilon}$ in (6) is a P-matrix, $\mathbf{a}_n^*$ is the attractor for $GVI(\mathcal{A}, \widetilde{\mathcal{F}})$ (see [3, Th. 5.4.4]), i.e., $\lim_{\boldsymbol{\Delta}\to\mathbf{0}_N} \|\mathbf{a}^* - \widetilde{\mathbf{a}}^*\|_2 = 0$, which means that when uncertainty approaches zero, the RNE converges to the nominal NE. From the given discussion, we conclude that when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix, the RNE can be obtained as a variant of the nominal NE from estimated system parameters and the uncertainty bound.

C. Numerical Validation

For the power control game, when the utility of user n is (2), we have

$$-\nabla_{\mathbf{a}_n}^2 v_n(\mathbf{a}_n, \mathbf{f}_n) = \mathrm{diag}\left(\left(\frac{h_{nn}^k}{\sigma_n^k + \sum_{m\in\mathcal{N}} a_m^k h_{nm}^k}\right)^2\right)_{k=1}^K$$

$$-\nabla_{\mathbf{a}_m\mathbf{a}_n} v_n(\mathbf{a}_n, \mathbf{f}_n) = \mathrm{diag}\left(\frac{h_{nm}^k h_{nn}^k}{\left(\sigma_n^k + \sum_{m\in\mathcal{N}} a_m^k h_{nm}^k\right)^2}\right)_{k=1}^K.$$

Consequently, for this game, (4) and (5) are as follows:

$$\alpha_n^{\min} = \min_{k\in\mathcal{K}} \left(\frac{h_{nn}^k}{\sigma_n^k + \sum_{m\in\mathcal{N}} a_{m,k}^{\max} h_{nm}^k}\right)^2$$

$$\beta_{nm}^{\max} = \max_{k\in\mathcal{K}} \frac{h_{nn}^k h_{nm}^k}{\left(\sigma_n^k + \sum_{m\in\mathcal{N}} a_{m,k}^{\min} h_{nm}^k\right)^2}.$$

In addition, for this game, (6) is a P-matrix if and only if

$$\min_{k\in\mathcal{K}} \frac{h_{nn}^k w_n^k}{\left(\sigma_n^k + \sum_{m\in\mathcal{N}} a_{m,k}^{\max} h_{nm}^k\right)^2} > \sum_{m\neq n} \max_{k\in\mathcal{K}} \frac{h_{nm}^k w_m^k}{\left(\sigma_n^k + \sum_{m\in\mathcal{N}} a_{m,k}^{\min} h_{nm}^k\right)^2} \; \forall\, n\in\mathcal{N},\; \forall\, w_n^k \in \mathcal{A}_n \tag{17}$$

which is the condition for the uniqueness of the nominal NE. Otherwise, the problem may have multiple nominal NEs.

To describe the physical meaning of (17), let $I_{n,k}^{\max} = (\sigma_n^k + \sum_{m\in\mathcal{N}} a_{m,k}^{\max} h_{nm}^k)^2$ be the maximum expected interference caused by other users on user n in subchannel k and $I_{n,k}^{\min} = (\sigma_n^k + \sum_{m\in\mathcal{N}} a_{m,k}^{\min} h_{nm}^k)^2$ be the minimum expected interference caused by other users on user n in subchannel k. In addition, let $\min_{k\in\mathcal{K}}((h_{nn}^k w_n^k)/(I_{n,k}^{\max}))$ be the minimum expected signal-to-interference-plus-noise ratio (SINR) of user n over all subchannels and $((h_{nm}^k w_m^k)/(I_{n,k}^{\min}))$ be the expected normalized interference of user m on user n in subchannel k. From (17), when the minimum expected SINRs of all users are greater than the sum of the maximum expected normalized interference levels, the nominal NE is unique. On the other hand, when interference of users on each other is low and channel gains between receivers and transmitters are high, the nominal NE is unique. This is in line with [26, Proposition 5 and Corollary 7].

As a numerical example, consider two users and two subchannels, and assume $a_n^{\max} = a_{n,k}^{\max} = 1$, $a_{n,k}^{\min} = 0.01$, and $\sigma_n^k = 0.001$ for all users in all subchannels. To ensure that (17) holds and (6) is a P-matrix for the power control game, the interference channel gain between users should be much less than the direct channel gain between each transmitter and its receiver, i.e., $h_{nm}^k < 0.01 h_{nn}^k$. In our simulation, the matrix $\boldsymbol{\Upsilon}$ in (6) for the power control game is

$$\boldsymbol{\Upsilon} = \begin{bmatrix} 1.5432 & -0.016 \\ -0.0012 & 1.221 \end{bmatrix}.$$

Since (17) is satisfied for the given matrix, it is a strongly monotone matrix. Fig. 1 shows the surface of the social utility of users 1 and 2 as their strategies in subchannels 1 and 2, respectively. When $\varepsilon_n \le 0.8$ for both users, simulations show the following. 1) Both the NE and the RNE are unique. 2) The social utility at the NE is $v_1^* + v_2^* = 1.6$. The social utility at the RNE is $\widetilde{u}_1^* + \widetilde{u}_2^* = 0.585$, which is less than that at the NE, as expected from Part 1 in Theorem 2. 3) At the NE, the allocated power to users 1 and 2 is $(a_1^{*1} = 0.5, a_1^{*2} = 0.5)$ and $(a_2^{*1} = 0.4, a_2^{*2} = 0.6)$, respectively, and RNEs for users 1 and 2 are $(\widetilde{a}_1^{*1} = 0.4,\ \widetilde{a}_1^{*2} = 0.6)$ and $(\widetilde{a}_2^{*1} = 0.9,\ \widetilde{a}_2^{*2} = 0.1)$, respectively. The upper bound in (14) is 1.3115, and the exact value of $\|\mathbf{a}^* - \widetilde{\mathbf{a}}^*\|_2$ is 0.7211, which is less than its upper bound in (14). Consider $\|\mathbf{W}^k(\mathbf{a})\|_2 = \sqrt{(\lambda_{\max}(\mathbf{W}^{*k}(\mathbf{a})\mathbf{W}^k(\mathbf{a}))}$, where $\mathbf{W}^{*k}(\mathbf{a})$ is the conjugate transpose of $\mathbf{W}^k(\mathbf{a})$, and $\lambda_{\max}$ is the maximum eigenvalue of the matrix. In our simulation, $\lambda_{\max}(\mathbf{W}^{*1}(\mathbf{a})\mathbf{W}^1(\mathbf{a})) = 0.9091$ and $\lambda_{\max}(\mathbf{W}^{*2}(\mathbf{a})\mathbf{W}^2(\mathbf{a})) = 0.59$. From (16), the distance

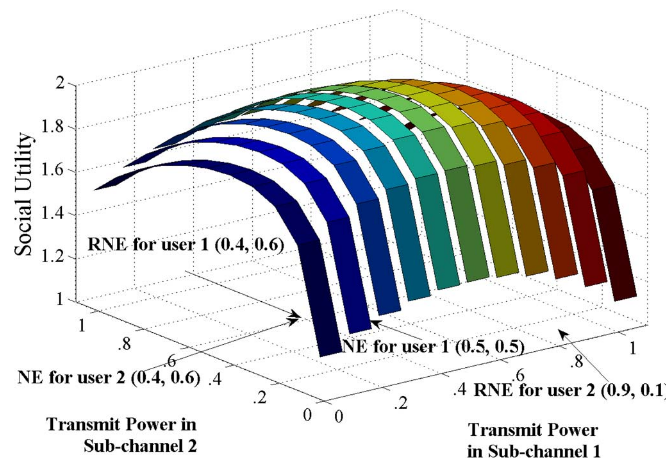


Fig. 1. Social utility for the power control game when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix.

between the social utilities at the RNE and at the nominal NE is 1.02, and in this simulation, this difference is 1.015. This simulation shows that (16) is a tight approximation for the difference between social utilities at the NE and at the RNE for the power control game. The given results numerically illustrate Theorem 2. From the given discussion and Theorem 2, when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix, the social utility at the RNE is less than that at the nominal NE, which is not desirable. However, in case studies in Section VIII, we show that users' utilities are very stable at the RNE as compared with those at the nominal NE. In addition, when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix for the power control game and variations in system parameters exceed the bound on the uncertainty region, the mean of social utility may increase at the RNE as compared with that at the nominal NE.

V. Logarithmic Utility Functions

We now consider logarithmic utility functions and assume that only the system parameters are uncertain. In such cases, VI becomes very simple, and the strong monotonicity requirement is relaxed to the positive definiteness of the affine mapping. Let the utility of each user be [27]

$$v_n^k\left(a_n^k, f_n^k\right) = \begin{cases} \log\left(c_n^k + \frac{a_n^k}{f_n^k}\right), & \text{if } \theta = -1 \\ \frac{\left(c_n^k + \frac{a_n^k}{f_n^k}\right)^{\theta+1}}{\theta+1}, & \text{if } \theta < 0 \text{ and } \theta \neq -1 \end{cases} \tag{18}$$

where c_n^k is the fixed system parameter related to dimension k of user n. Assume that for the uncertain parameter of user n, we have $\widetilde{x}_{nm}^k = x_{nm}^k + \widehat{x}_{nm}^k$, where $\widetilde{x}_{nm}^k$, x_{nm}^k, and $\widehat{x}_{nm}^k$ are the estimated value, the exact value, and the error, respectively. The uncertainty region for each user is $\Re_n^k = \{\sqrt{\sum_{m=1,m\neq n}^N (\widehat{x}_{nm}^k / x_{nn}^k)^2} \le \epsilon_n^k\}$ for all k, where ϵ_n^k is the bound on the uncertainty region for user n in dimension k. The given formulation models the power control game when uncertainty emanates from variations in the channel gain between transmitter m and receiver n. The throughput of each user in (2) corresponds to when in (18), we have $\theta = -1$, and $c_n^k = 1$, and f_n^k is normalized with respect to x_{nn}^k, i.e., the interference on user n in subchannel k is normalized with respect to its direct channel gain. For this case, we also assume that [28]

$$a_n^{\max} < \sum_{k=1}^{K} a_{n,k}^{\max}, \qquad \forall\, n \in \mathcal{N}. \tag{19}$$

Proposition 1: For an ACG with utility function (18), the following statements hold.

1) The nominal NE is the solution to an affine VI (AVI), denoted by $AVI(\mathcal{A}, \mathcal{M})$, where $\mathcal{M}(\mathbf{a}) = (\mathcal{M}_n(\mathbf{a}))_{n=1}^N$, and

$$\mathcal{M}_n(\mathbf{a}) = \boldsymbol{\varpi}_n + \sum_{m=1}^{N} \mathbf{M}_{nm}\mathbf{a}_m^{\mathrm{T}} \tag{20}$$

where $\boldsymbol{\varpi}_n = (\varpi_n^k)_{k=1}^K$, $\varpi_n^k = ((y_n^k + c_n^k)/(x_{nn}^k))$, and $\mathbf{M}_{nm} = \mathrm{diag}(x_{nm}^k/x_{nn}^k)_{k=1}^K$.

2) Consider an $N \times N$ matrix $\mathbf{M}^{\max}$, where $M_{nm}^{\max} = \max_{k\in\mathcal{K}}(x_{nm}^k/x_{nn}^k)$ if $m \neq n$, and $M_{nm}^{\max} = 0$ otherwise. The nominal NE is unique when

$$\max_{n\in\mathcal{N}} \|\mathbf{a}_n\|_2 > \sum_{m\neq n} M_{nm}^{\max}\|\mathbf{a}_m\|_2,\ \forall\, \mathbf{a}_n \in \mathcal{A}_n,\ \forall\, n \in \mathcal{N}. \tag{21}$$

Proof:

1) From (18), the best response of the NACG is $a_n^k = [\mu_n^{(1/\theta)} - ((c_n^k + y_n^k)/x_{nn}^k) - \sum_{m\neq n}(x_{nm}^k/x_{nn}^k)a_m^k]_{a_{nk}^{\min}}^{a_{nk}^{\max}}$, where Lagrange multiplier μ_n for user n is so chosen to satisfy (23). When (19) holds, the best response obtained via AVI is $AVI(\mathcal{A}, \mathcal{M})$, where $\mathcal{M}(\mathbf{a}) = (\mathcal{M}_n(\mathbf{a}))_{n=1}^N$, and $\mathcal{M}_n$ is obtained from (20) [29].

2) For this case, the game has a unique NE when $\mathcal{M}(\mathbf{a})$ is strongly monotone. From [28, Proposition 1], when (21) holds, $\mathcal{M}(\mathbf{a})$ is strongly monotone, and hence, the nominal NE is unique. ■

From Proposition 1, when the impact of users on each other is sufficiently low, the nominal NE is unique. For the example of the power control game, Proposition 1 implies that when interference between users is sufficiently low, the nominal NE of the game is unique [28], [29]. We now derive the RNE's uniqueness condition for such cases.

Theorem 3: For utility function (18), the following statements hold.

1) The AVI mapping of the RNE is $\widetilde{\mathcal{M}}(\mathbf{a}) = (\widetilde{\mathcal{M}}_n(\mathbf{a}))_{n=1}^N$, where

$$\widetilde{\mathcal{M}}_n(\mathbf{a}) \leq \mathcal{M}_n(\mathbf{a}) + \widehat{\mathcal{M}}_n(\mathbf{a}), \qquad \forall\, n \in \mathcal{N} \tag{22}$$

where $\widehat{\mathcal{M}}_n(\mathbf{a}_{-n}) = (\epsilon_n^k\|\mathbf{a}_{-n}^k\|)_{k=1}^K$ and $\mathbf{a}_{-n}^k = [a_1^k, \ldots, a_{n-1}^k, a_{n+1}^k, \ldots, a_N^k]$.

2) When (21) holds, the RNE of $\widetilde{\mathcal{G}}$ is unique for any bounded ϵ_n^k.

3) When (21) holds, the utility of each user at the RNE is always less than or equal to that at the nominal NE, and the upper bound on the strategy profile of each user is $\|\mathbf{a}^* - \tilde{\mathbf{a}}^*\|_2 \leq ((\|\mathbf{E}\|_2)/(\lambda_{\min}(\mathbf{M}^{\max})))$, where $E_{nm} = \|\boldsymbol{\epsilon}_n\|_\infty$ if $m = n$, and $E_{nm} = 0$ otherwise. $\lambda_{\min}$ is the minimum eigenvalue of matrix $\mathbf{M}^{\max}$, $\boldsymbol{\epsilon}_n = [\epsilon_n^1, \ldots, \epsilon_n^K]$, and $\|\cdot\|_\infty$ is the maximum element of the vector.

Proof: See Appendix E. ■

From Theorems 2 and 3, we note that 1) the condition for RNE's uniqueness is not related to the size of the uncertainty region; and 2) for a closed, bounded, and convex uncertainty region, the RNE's uniqueness condition is the same as the NE's uniqueness condition. The given discussion is contingent upon strong monotonicity of $\mathcal{F}$ and $\widetilde{\mathcal{F}}$. Now, one can obtain the impact of uncertainty in the estimated values on the RNE. In brief, the RNE can be easily derived from the NE.

Remark 3: By rearranging AVI of the RNE for utility function (18), the best response of the RACG is $\tilde{a}_n^k = [\mu_n^{(1/\theta)} - \varpi_n^k - \sum_{m\neq n}((x_{nm}^k a_m^k)/x_{nn}^k) - \epsilon_n^k\|\mathbf{a}_{-n}^k\|]_{a_{nk}^{\min}}^{a_{nk}^{\max}}$, where μ_n is the Lagrange multiplier for the strategy space of user n in (1), i.e.,

$$\mu_n \times \left(\sum_{k=1}^{K} a_n^k - a_n^{\max}\right) = 0 \tag{23}$$

which is similar to [19, eq. (13)] for the power control game in spectrum sharing environments. From [19, Th. 2], a sufficient condition for RNE's uniqueness is related to the size of the uncertainty region, whereas when we use AVI to analyze the RNE, we obtain (21) as another sufficient condition for RNE's uniqueness, which is independent of the size of the uncertainty region. These two sufficient conditions do not need to be the same, and in fact, they are different because the approaches for obtaining a sufficient condition for RNE's uniqueness in this paper and in [19] are not the same, and their respective social utilities are not related. Moreover, in [19], uncertainty is confined to channel gains, whereas in this paper, channel gains and impacts of other users are uncertain.

VI. Distributed Algorithms

We now utilize the proximal-point method to propose a distributed and efficient numerical algorithm for obtaining the robust solution to $\widetilde{\mathcal{G}}$. The proximal-point method is a projection method for solving problems that involve set-valued mappings, where a sequence of subproblems is iteratively solved as per [3, Sec. 12] and [4, Sec. 12.6.1]. The main advantages of utilizing the proximal-point method for our robust game are twofold: 1) the optimization problem in the proximal-point method can be decomposed and solved in a distributed and efficient manner as per [4, Sec. 12.6.1]; and 2) the users' objective functions in the game, i.e., Ψ_n, do not need to be strictly or strongly convex for convergence of the distributed algorithm as per [4, Sec. 12.2.4]. We will show that utilizing the proximal-point method for our robust game can lead to a closed-form solution for the problem at hand. For any $\mathbf{b} = [\mathbf{b}_1, \ldots, \mathbf{b}_N] \in \mathcal{A}$, let $\hat{\mathbf{a}}(\mathbf{b}) = [\hat{\mathbf{a}}_1(\mathbf{b}), \ldots, \hat{\mathbf{a}}_N(\mathbf{b})]$ be the solution to the following optimization problem (see [4, Sec. 12.6.1]):

$$\hat{\mathbf{a}}(\mathbf{b}) = \arg\max_{\mathbf{a}\in\mathcal{A}} \left[\sum_{n=1}^{N} \Psi_n(\mathbf{a}_n, \mathbf{b}_{-n}) - \frac{1}{2}\|\mathbf{a} - \mathbf{b}\|_2^2\right] \tag{24}$$

where $\hat{\mathbf{a}}(\mathbf{b})$ is a proximal response map of game $\widetilde{\mathcal{G}}$. From [4, Proposition 12.5 in Sec. 12.2.4], since Ψ_n is concave (Part 2 in Lemma 1), the fixed point of $\hat{\mathbf{a}}(\mathbf{b})$ is the RNE of the robust

TABLE I
DISTRIBUTED ALGORITHM FOR PROXIMAL POINT METHOD

Inputs for Each User
$\mathcal{L} = [1, \cdots, L]$: Users' iterations, ε_n: Uncertainty region for user n, and $0 < \zeta << 1$: Interrupt criteria for all users,
Initialization For $t = 0$,
Set an initial $\mathbf{a}_n(0)$ and a random $\mathbf{f}_n(0)$ for all $n \in \mathcal{N}$,
Iterative Algorithm
For $l = 1, \cdots, L$ and $L \to \infty$
Update $\mathbf{a}_n(l) = \text{argmax}_{\mathbf{a}_n \in \mathcal{A}_n} \Psi_n(\mathbf{a}_n, \mathbf{a}_{-n}(l-1)) - \frac{1}{2}\|\mathbf{a}_n - \mathbf{a}_n(l-1)\|_2^2$ for all users,
Each user transmits the value of $\mathbf{a}_n(l)$ to other users, and measures the aggregate effects of other users (i.e., $\mathbf{f}_n(l)$),
If $\|\mathbf{a}_n(l-1) - \mathbf{a}_n(l)\|_2 \le \zeta$, End. Otherwise $l = l+1$, continue;

game $\widetilde{\mathcal{G}}$. Now, (24) can be decomposed into N subproblems (one for each user) as per [4, Sec. 12.6.1], i.e.,

$$\widehat{\mathbf{a}}_n(\mathbf{b}) = \arg\max_{\mathbf{a}_n \in \mathcal{A}_n} \left[\Psi_n(\mathbf{a}_n, \mathbf{b}_{-n}) - \frac{1}{2}\|\mathbf{a}_n - \mathbf{b}_n\|_2^2\right] \quad (25)$$

for all $n \in \mathcal{N}$. A distributed iterative algorithm is developed as follows. Let $\widehat{\mathbf{a}}_n(\mathbf{b})$ and $\mathbf{b}_n$ be the solutions for user n in its current and previous iterations, respectively. When user n is informed about other users' actions and estimates $\mathbf{f}_n^k$ at its receiver, a solution to (25) can be obtained. The distributed algorithm that is based on the proximal-point method is summarized in Table I. In this algorithm, users update their actions at discrete instances l in $\mathcal{L} = [1, \ldots, L]$, where $\mathbf{a}_n(l)$ is the action of user n at iteration l obtained from (25), and $\mathbf{f}_n(l)$ are the impacts of other users on user n at iteration l, which is observed at the receiver of user n and sent to the respective transmitter.

In Theorem 4, we obtain the sufficient conditions for convergence of the iterative algorithm. Note that the optimization problem in (25) is strongly concave as per [4, Sec. 12.6.1], where it is shown that the regularization term $(1/2)\|\mathbf{a} - \mathbf{b}\|_2^2$ guarantees the strong concavity of each user's optimization problem, and hence, each $\widehat{\mathbf{a}}_n$ can be obtained via efficient convex optimization algorithms. When the distributed algorithm in Table I converges, the regularization term $\|\mathbf{a}_n - \mathbf{b}_n\|_2^2$ tends to zero.

Theorem 4: As $L \to \infty$, the distributed algorithm in Table I converges to the unique RNE from any initial $\mathbf{a}_n(0)$, when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix, and $((\partial^3 v_n^k(a_n^k, f_n^k))/(\partial^2 a_n^k \partial f_n^k)) = ((\partial^3 v_n^k(a_n^k, f_n^k))/(\partial a_n^k \partial^2 f_n^k)) = 0$.

Proof: See Appendix F. ∎

Note that Theorem 4 does not add any new constraint for the power control game. This is because when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix, the condition of Theorem 4 holds for the power control game. In this case, interference in the system is very low, and consequently, the SINR of each user is high, i.e., $((h_{nn}^k a_n^k)/f_n^k) \gg 1$, and the utility function of each user is $v_n(\mathbf{a}_n, \mathbf{f}_n) \approx \sum_{k=1}^K \log((h_{nn}^k a_n^k)/f_n^k)$, which satisfies $((\partial^3 v_n^k(a_n^k, f_n^k))/(\partial^2 a_n^k \partial f_n^k) = 0$. From Theorem 4, the distributed algorithm converges to a unique NE when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix irrespective of the size of the uncertainty region, so long as the uncertainty region is closed and convex.

Remark 4: In solving (25), one has to obtain $\widetilde{\mathcal{F}}_n(\mathbf{a}) = -((\partial\Psi_n(\mathbf{a}_n, \mathbf{a}_{-n}))/(\partial\mathbf{a}_n))$ [3]. When $\widetilde{\mathcal{F}}_n(\mathbf{a})$ is an affine mapping, solving (25) reduces to solving an AVI, which is easy and straightforward. For example, for (18), solving (25) via the proximal-point method at iteration l is similar to solving $VI(\mathcal{A}, \overline{\mathcal{M}}(\mathbf{a}(l-1)))$ (see [3, Sec. 12.3]), where $\overline{\mathcal{M}}_n(\mathbf{a}(l-1)) = (1/2)(\boldsymbol{\varpi}_n + \sum_{m=1}^N \mathbf{M}_{nm}\mathbf{a}_m^{\mathrm{T}}(l-1)) - \mathcal{I}_n$, and $\mathcal{I}_n = (a_n^k(l) - a_n^k(l-1))_{k=1}^K$. The solution to (25) obtained via the proximal-point method is

$$a_n^k(l) = \frac{1}{2}\Bigg[\mu_n^{\frac{1}{\theta}} - \varpi_n^k - \sum_{m \neq n} \frac{x_{nm}^k a_m^k(l-1)}{x_{nn}^k} - \epsilon_n^k \left\|\mathbf{a}_{-n}^k(l-1)\right\| + a_n^k(l-1)\Bigg]_{a_{nk}^{\min}}^{a_{nk}^{\max}} \quad (26)$$

where μ_n satisfies (23). Hence, in the power control game, to solve the waterfilling-like formulation in (26) with few calculations, each user n only needs to know the amount of interference from other users and their actions, as well as its own previous action.

When $\widetilde{\mathcal{F}}_n(\mathbf{a})$ is not an affine mapping, from Lemma 1, $\Psi_n(\mathbf{a}_n, \mathbf{a}_{-n})$ is concave. Hence, the Lagrange function $L_n(\mathbf{a}_n, \mu_n) = u_n(\mathbf{a}_n, \mathbf{f}_n - \varepsilon_n\boldsymbol{\vartheta}_n) - (1/2)\|\mathbf{a}_n(l) - \mathbf{a}_n(l-1)\|_2^2 - \mu_n(\sum_{k=1}^K a_n^k - a_n^{\max})$ is used to iteratively solve (25), where μ_n is the Lagrange multiplier that satisfies (23) for user n.

VII. EFFECTS OF ROBUSTNESS ON SOCIAL UTILITY FOR THE CASE OF MULTIPLE NOMINAL NASH EQUILIBRIA

So far, we obtained the RNE's uniqueness condition from the nominal NE's uniqueness conditions. Now, we study RNEs when the NACG has multiple nominal NEs. In general, this is not easy since the mapping $\mathcal{F}$ for NACG is nonmonotone and nonsmooth [30]–[32]. To compare the case of multiple nominal NEs with that of a unique nominal NE, consider the numerical example in Section IV-C, but when $\boldsymbol{\Upsilon}$ in (6) is not a P-matrix, and $h_{nm}^k > 0.5h_{nn}^k$ for all users in all subchannels. The mapping $\mathcal{F}$ is nonmonotone for both users. As shown in Fig. 2, there are multiple local optima in the utility function, corresponding to multiple nominal NEs. At a nominal NE, the convergence points for users 1 and 2 are ($a_1^{*1} = 0.534$, $a_1^{*2} = 0.463$) and ($a_2^{*1} = 0.417$, $a_2^{*2} = 0.583$), respectively, and $v_1^* + v_2^* = 3.0176$. When uncertainty is $\varepsilon_n < 0.8$, the RNE converges to ($\widetilde{a}_1^{*1} = 0.556$, $\widetilde{a}_1^{*2} = 0.444$) and ($\widetilde{a}_2^{*1} = 0.325$, $\widetilde{a}_2^{*2} = 0.675$), and $\widetilde{u}_1^* + \widetilde{u}_2^* = 3.077$. In this simulation, the initial points in $\mathcal{G}$ and $\widetilde{\mathcal{G}}$ are the same. This example shows that introducing robustness may increase the social utility at the RNE when the NACG has multiple nominal NEs, which is in line with the simulation results in [18] and [19]. We observe that implementing RACG in communication networks may increase the social utility as compared with that of NACG. However,

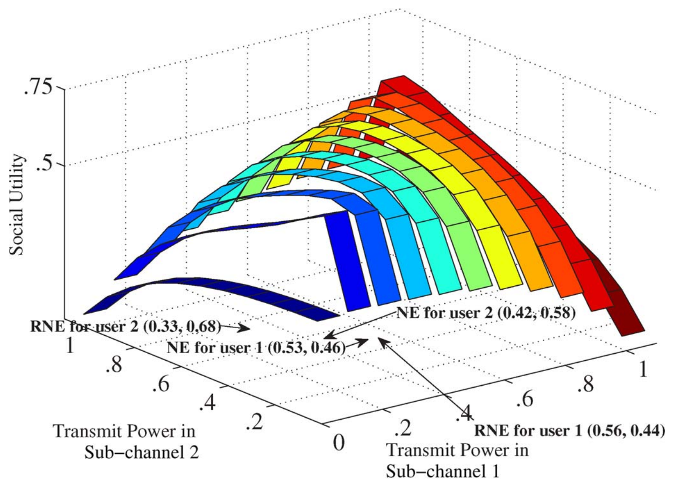


Fig. 2. Social utility for the power control game when $\boldsymbol{\Upsilon}$ in (6) is not a P-matrix.

since the optimal strategy of each user is a nonlinear function of its uncertainty region and other users' strategies, obtaining the sufficient conditions under which the social utility increases is not easy for all cases. Hence, instead of considering the general case, we focus on a special case, in which both $\mathcal{G}$ and $\widetilde{\mathcal{G}}$ start from the same initial point, and the strategy of each user is a decreasing function of the bound on the uncertainty region. Since $\widetilde{\mathcal{F}}$ is a bounded perturbed version of $\mathcal{F}$ from Lemma 2, the converged RNE is in the proximity of the nominal NE.

Proposition 2: Consider $\mathbf{W}(\mathbf{a}) = (\mathbf{W}(\mathbf{a})^k)_{k=1}^K$, where $\mathbf{W}(\mathbf{a})^k$ is defined in (15). If $\mathbf{W}(\mathbf{a})$ is a negative definite matrix and $\nabla_{\varepsilon_n}\mathbf{a}_n < \mathbf{0}_K^{\mathrm{T}}$ for all users (i.e., when the strategy of each user is a decreasing function of the bound on its uncertainty region), the social utility at the RNE is higher than that at the corresponding converged nominal NE.

Proof: Variation in the utility of user n due to variations in the bound on the uncertainty region is

$$\begin{aligned}\lim_{\boldsymbol{\varepsilon}\to 0} &\nabla_{\boldsymbol{\varepsilon}} u_n(\mathbf{a}_n, \widetilde{\mathbf{f}}_n)\\ &= \nabla_{\mathbf{a}_n} v_n(\mathbf{a}_n, \mathbf{f}_n) \times \mathbf{1}_K \times \nabla_{\varepsilon_n}\mathbf{a}_n\\ &\quad + \nabla_{\mathbf{f}_n} v_n(\mathbf{a}_n, \mathbf{f}_n) \times \mathbf{x}_{nm} \times \nabla_{\varepsilon_m}\mathbf{a}_m\\ &\quad + \varepsilon_n \nabla^2_{\mathbf{a}_n\mathbf{f}_n} v_n(\mathbf{a}_n, \mathbf{f}_n) \times \mathbf{X}_{nm} \times \nabla_{\varepsilon_n}\mathbf{a}_n\\ &\quad + \varepsilon_n \mathbf{X}_{nm} \times \nabla^2_{\mathbf{f}_n\mathbf{f}_n} v_n(\mathbf{a}_n, \mathbf{f}_n) \times \nabla_{\varepsilon_m}\mathbf{a}_m + o(\cdot) \end{aligned} \tag{27}$$

where $\mathbf{X}_{nm}$ is the $K \times K$ matrix, $[\mathbf{X}_{nm}]_{kk} = (x_{nm}^k)^2$, $\mathbf{1}_K$ is a $1 \times K$ vector whose elements are equal to one, $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_N]$, and ∇ is the column gradient vector. The last two terms on the right-hand side (RHS) of (27) are always positive because of Assumptions **A1** and **A2** in Section II. The first term on the RHS of (27) is always negative because $v_n(\mathbf{a}_n, \mathbf{f}_n)$ is an increasing function of $\mathbf{a}_n$, and $\nabla_{\varepsilon_n}\mathbf{a}_n < \mathbf{0}_K^{\mathrm{T}}$. The second term on the RHS of (27) is always positive because $v_n(\mathbf{a}_n, \mathbf{f}_n)$ is a decreasing function of $\mathbf{f}_n$, and $\nabla_{\varepsilon_n}\mathbf{a}_m < \mathbf{0}_K^{\mathrm{T}}$. When $\mathbf{W}(\mathbf{a})$ is a negative definite matrix, the negative terms on the RHS of (27) are less than its positive terms, and the social utility of the robust game is higher than that of the nominal game. ■

From Proposition 2, when at the RNE, the reduction in the utility of each user (caused by a reduction in its own strategy, e.g., its own transmit power) is less than the increase in the same user's utility (caused by reductions in other users' strategies), introducing robustness in the game causes a net increase in the social utility. In such cases, the social utility at the RNE is higher than the social utility at the corresponding nominal NE; however, there may be other cases as well, in which this may not be valid.

Remark 5: When the solution of affine VI in Proposition 1, denoted by $AVI(\mathcal{A}, \mathcal{M})$, is a monotone decreasing function of ε_n (i.e., when $\nabla_{\varepsilon_n}\mathbf{a}_n < \mathbf{0}_K^{\mathrm{T}}$ for all users), $\mathbf{M} = (\mathbf{M}_n)_{n=1}^N$ is a negative definite matrix, where $\mathbf{M}_n = \sum_{m=1}^N \mathbf{M}_{nm}$ (see Appendix G). For the power control game, Proposition 2 is simplified to

$$\max_{n\in\mathcal{N}} \|\mathbf{a}_n\|_2 < \sum_{m\neq n} M_{nm}^{\max}\|\mathbf{a}_m\|_2, \qquad \forall\, \mathbf{a}_n \in \mathcal{A}_n \tag{28}$$

for all $n \in \mathcal{N}$. This means that when all interference channel gains are sufficiently greater than direct channels gains, introducing robustness increases the social utility. This is an opportunistic phenomenon when the robust game encounters multiple nominal NEs. To benefit from this (i.e., increase the social utility), we propose a distributed algorithm in Table II. Obviously, checking the conditions of Proposition 2 in a distributed manner is not easy, and the social utility may be increased when some other conditions prevail as well.

By considering these two issues, as shown in Table II, all users first play the game $\mathcal{G}$ by utilizing the conventional distributed algorithm, whose social utility at the nominal NE is $\tilde{u}(0)$. As an example, for the power control game, the simultaneous iterative waterfilling algorithm (IWFA) can be played between users in $\mathcal{G}$ [10], [33]. When $\mathcal{G}$ converges, users play the game $\widetilde{\mathcal{G}}$ by assuming a small arbitrary value for ε_n and calculate their new strategies. Users update their strategies at iterations $l_1 = 1, 2, \ldots, L_1$, where L_1 is the bound on iterations. At the end of iteration l_1, each user measures $\mathbf{f}_n^k$ at its receiver, calculates its achieved utility, and sends it to other users so that all users can obtain the social utility $\tilde{u}(l_1)$. When $\tilde{u}(l_1) > \tilde{u}(l_1 - 1)$, each user's uncertainty region is expanded. Otherwise, the algorithm is terminated, and users' strategies in the previous iteration are applied. In the opportunistic algorithm in Table II, vectors $\mathbf{a}_n(l_1)$ and $\mathbf{f}_n(l_1)$ are the transmit strategies and impacts of other users on user n at iteration l_1, respectively. Note that in the opportunistic algorithm in Table II, each user needs to know other users' utilities in addition to its own strategy, which is not the case for the distributed algorithm in Table I.

VIII. Case Studies

A. Power Control Games

We use simulations for the power control game to provide an insight into the performance of $\widetilde{\mathcal{G}}$ for different bounds on the uncertainty region as compared with that of $\mathcal{G}$ where the utility of each user is defined in (2). In the following simulations,

TABLE II
OPPORTUNISTIC ALGORITHM FOR INCREASING SOCIAL UTILITY

First Stage: All users play $\mathcal{G}$ to reach the NE with utility v_n^*.
Second Stage: When $\mathcal{G}$ converges to NE:
Initialization: Let $\widetilde{u}(0) = \sum_{n=1}^{N} v_n^*$, $0 < \chi < 1$, $\omega(0) = 0$, and $0 < \delta << 1$
Iterative Algorithm
For $l_1 = 1, \cdots, L_1$;
1) Consider the uncertainty region $\varepsilon_n = \omega(l_1)$, where $\omega(l_1) = l_1 \times \chi$ for all n;
2) Update $\mathbf{a}_n(l_1) = \text{argmax}_{\mathbf{a}_n \in \mathcal{A}_n} \Psi_n(\mathbf{a}_n, \mathbf{a}_{-n}(l_1 - 1)) - \frac{1}{2}\|\mathbf{a}_n - \mathbf{a}_n(l_1 - 1)\|_2^2$ for user n;
3) User n transmits with $\mathbf{a}_n(l_1)$, measures $\tilde{\mathbf{f}}_n(l_1)$, and calculates $\widetilde{u}_n(l_1)$;
4) Each user n informs other users of its $\mathbf{a}_n(l_1)$ and $\widetilde{u}_n(l_1)$,
5) The social utility at iteration l_1 is calculated via $\widetilde{u}(l_1) = \sum_{n=1}^{N} \widetilde{u}_n(l_1)$;
6) If $\widetilde{u}(l_1) > \widetilde{u}(l_1 - 1)$ and $\|\widetilde{u}(l_1) - \widetilde{u}(l_1 - 1)\|_2 > \delta$: Continue; Otherwise: End.

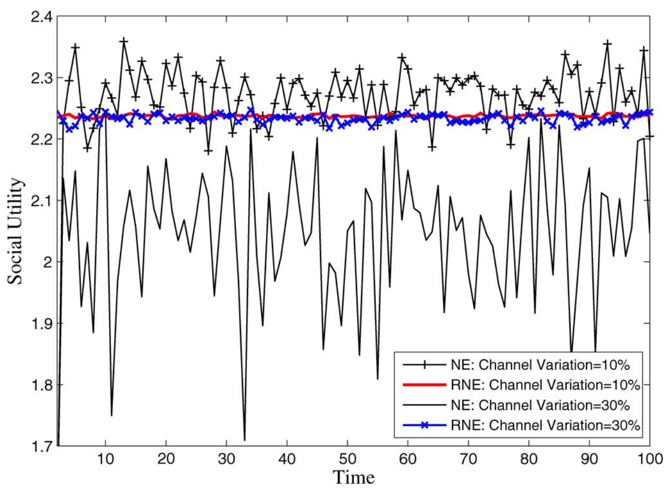


Fig. 3. Impact of channel variations in robust and nonrobust games.

the value of ε_n is normalized to the estimated value of $\mathbf{f}_n$, i.e., $\varepsilon_n = (\|\widetilde{\mathbf{f}}_n - \mathbf{f}_n\|)/(\|\mathbf{f}_n\|)$, and uncertainty for all users is assumed to be the same, denoted by ε. In addition, for the case of multiple nominal NEs, we only consider the cases in which the simulation converges to a local nominal NE. In simulations, we assume $a_{n,k}^{\min} = 0.01$, $a_{n,k}^{\max} = 0.5$, $a_n^{\max} = 1$, and $\sigma_n^k = 0.001$ for all users in all subchannels.

1) Effect of Uncertainty on the Nominal NE: For the power control game, we begin by studying the effect of uncertainty on its performance in both $\mathcal{G}$ and $\widetilde{\mathcal{G}}$ in terms of utility variations at their equilibria. To do so, we consider $N = 3$ users, $K = 16$, and $\varepsilon = 10\%$ at the RNE. After convergence to the RNE and to the nominal NE, channel gains between users are varied up to 10% as well as 30% of their nominal values, which causes variations in the utility of each user at the nominal NE and at the RNE, as shown in Fig. 3 and summarized in Table III. Note that variations in the social utility at the nominal NE are considerable for both cases. In contrast, in the RACG, the social utility at the RNE is stable for both cases. From Table III, the mean of the social utility at the NE is greater than that at the RNE when variations in system parameters are up to 10%; however, its variance is considerably higher than that at the RNE, which means that users experience more variations in their social utility. When variations in system parameters are up to 30%, the mean of the social utility is less than that at the RNE, and its variance is greater than that at the nominal NE. In addition, as can be seen in Table III, when variations in channel gain are up to 70%, the average social utility at the nominal NE is progressively reduced, and its variance is increased, which means that users on the average experience undesirable fluctuations in their respective quality of service. In contrast, the social utility at the RNE is more stable, and the reduction in the social utility at the RNE (caused by taking into account uncertainties and variations in parameters' values) is less than that at the nominal NE. This simulation confirms that the social utility at the nominal NE is very sensitive to variations in system parameters. In contrast, the social utility at the RNE is stable, which means that at the RNE, users' utilities are protected under any condition of error. Moreover, when variations in system parameters are greater than 10%, the mean of social utility at the RNE is greater than that at the nominal NE.

2) Effect of Size of Uncertainty Region on the RNE: To demonstrate the results of Theorem 2 and Section VII on social utilities at the nominal NE and at the RNE, in Fig. 4, we compare the effect of uncertainty when Theorem 2 holds with that of the case when it does not hold, in terms of the ratio of social utilities at the RNE and at the nominal NE for different amounts of uncertainty. Simulations are performed for Rayleigh fading channels and bounded and uniformly generated errors for each cross-subchannel gain. To satisfy the nominal NE's uniqueness condition (i.e., $\boldsymbol{\Upsilon}$ in (6) being a P-matrix), channel gains are such that $h_{nm}^k < 0.01 h_{nn}^k$, and for multiple nominal NEs, we have $h_{nm}^k > 0.5 h_{nn}^k$. The ratio of the social utility in Fig. 4 is obtained by averaging over 100 channel realizations. When $\boldsymbol{\Upsilon}$ in (6) is a P-matrix (i.e., when the NE and the RNE are unique), the social utility of the robust game at its RNE is decremented when the uncertainty region is expanded, as expected from Theorem 2. However, for the case of multiple nominal NEs, no uniformity in the social utility is observed, and the social utility at the RNE may exceed that of the corresponding nominal NE, as expected from the discussion in Section VII. For example, when $\varepsilon = 10\%$, the social utility at the RNE is higher than that at the nominal NE, and when $\varepsilon = 20\%$, the ratio substantially falls. The trend is not monotonic for different values of uncertainty, e.g., the social utility at the RNE is higher for $\varepsilon = 50\%$ as compared with those for $\varepsilon = 40\%$ and $\varepsilon = 60\%$. This simulation is similar to the numerical validation of Theorem 2 in Section IV-C, where it was shown that when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix (i.e., when the NE and the RNE are unique), the social utility of the robust game is reduced when the uncertainty region is expanded, which also supports Proposition 2 by showing that the social utility at the RNE can

TABLE III
STATISTICS OF SOCIAL UTILITY AT NOMINAL NE AND RNE FOR DIFFERENT BOUNDS ON VARIATIONS IN SYSTEM PARAMETERS

Bound on System Parameters' Variations	Up to 10%		Up to 30%		Up to 50%		Up to 70%	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Social utility at the nominal NE	2.83	0.031	2.65	0.037	2.35	0.042	2.19	0.06
Social utility at RNE	2.76	3.9×10^{-6}	2.76	3.6×10^{-5}	2.74	2.1×10^{-4}	2.73	3.8×10^{-4}

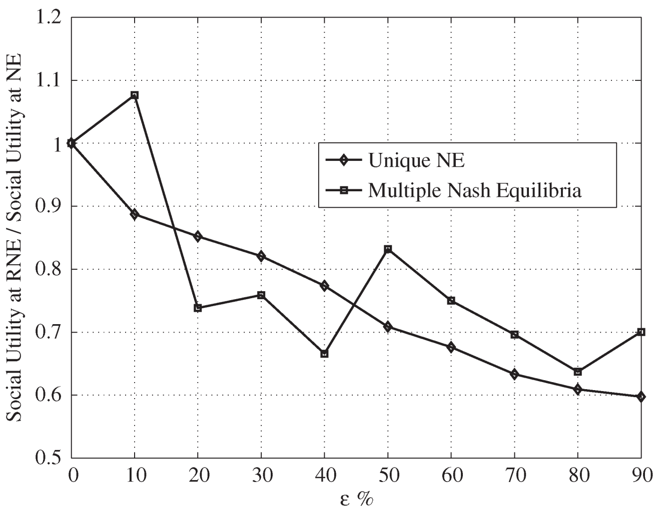


Fig. 4. Ratio of social utilities at the RNE and at the nominal NE versus ε for a unique nominal NE and for multiple nominal NEs.

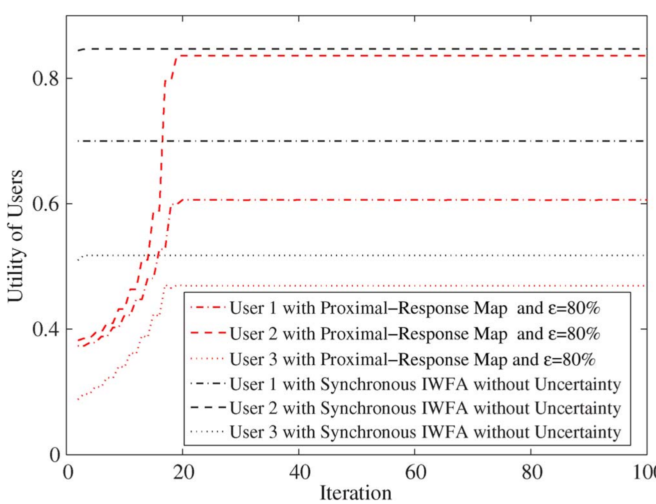


Fig. 5. Convergence times of the proximal point method and the IWFA.

be higher than that at its corresponding nominal NE for the case of multiple nominal NEs.

In Fig. 5, the users' utilities versus iteration numbers are shown for the proximal-point method when $\varepsilon = 80\%$. Note that the convergence time of the proximal-point method is longer than that of the IWFA with no uncertainty for users. In Table IV, for 10 000 channel realizations and $\zeta = 0.01$, the average number of iterations for convergence of the power allocation game is shown. The number of iterations is increased when the uncertainty region is expanded but remains below that of the nominal game and [10]. For small values of ε, the convergence time of the RACG and that of the IWFA ($\varepsilon = 0$ in Table IV) are close to each other.

3) Performance of the Opportunistic Algorithm: In Fig. 6, we compare the efficiency of the opportunistic approach with that of the nominal game, defined by $\eta = (\sum_{n\in\mathcal{N}} u_n^{\mathrm{OP}} - v^*)/v^*$, where u_n^{OP} is the achieved utility of user n at the end of the opportunistic algorithm. The value of η is obtained by averaging over 1200 channel realizations for $N = [4, 6, 8]$, and $K = [32, 64]$ for different values of h_{nm}^k. In this simulation, we assume $L_1 = 100$ and $\delta = 0.01$ and only consider the results when the nominal game converges to a local nominal NE. Note that for very high interference levels, i.e., when $h_{nm}^k \gg h_{nn}^k$, the efficiency of the proposed opportunistic approach is considerably higher than that of moderate interference levels, i.e., when $h_{nm}^k \geq h_{nn}^k$. This is in line with Proposition 2. As shown in Fig. 6, the efficiency of the opportunistic algorithm is significantly better in high interference levels and when the number of subchannels is high. The reason being that for a higher number of channels, the probability of convergence to orthogonal power allocation at the RNE is increased, resulting in less interference between users and in a higher social utility, as also shown in [19].

From Theorem 2 and as shown in Fig. 4, when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix, the social utility at the RNE is less than that at the nominal NE. In addition, the distributed algorithm for reaching the RNE needs more message passing compared with the distributed algorithm for reaching the nominal NE. Moreover, in Fig. 5, the convergence time for the distributed algorithm at the RNE is longer than that at the nominal NE. All of these constitute the costs of robustness. In contrast, at the RNE, the utility of each user is stable for different values of uncertainty in the system, resulting in reliable communications for all users, as shown in Fig. 3.

B. Other Examples of ACG

Finally, we introduce two other examples of the ACG, namely, downlink transmit power allocation in DSLAMs and routing delay minimization in Jackson networks. For the first example, the basic system model is similar to the power control problem in interference channels except that here, intercarrier interference (ICI) exists among different frequency bins. For this example, we have $f_n^k(\mathbf{a}_{-n}, \mathbf{s}_n) = \sum_{m\neq n}\sum_{j=1}^{K}(\gamma(k-j)h_{mn}^k a_m^k) + \sigma_n^k$, where

$$\gamma(j) = \begin{cases} 1, & \text{if } j = 0 \\ \frac{2}{K^2 \sin^2\left(\frac{\pi}{K} j\right)}, & \text{if } -\frac{K}{2} \leq j \leq \frac{K}{2},\ j \neq 0 \end{cases}$$

is an ICI coefficient that is symmetric and circular, i.e., $\gamma(j) = \gamma(-j) = \gamma(K-j)$. The utility of user n is the same as (2), and its strategy is the allocated power in different frequency bins (carriers). Since the results for this example are similar to the results for the power control problem in interference channels that have already been presented, for brevity, we refrain from further discussions.

TABLE IV
MEAN NUMBER OF ITERATIONS FOR CONVERGENCE OF THE POWER CONTROL GAME VERSUS ε

ε	0	10%	20%	30%	40%	50%	60%	70%	80%	90%
Mean number of iterations	13.5	16.6	19.6	22.8	25.2	26.3	27.6	28.33	28.64	28.9

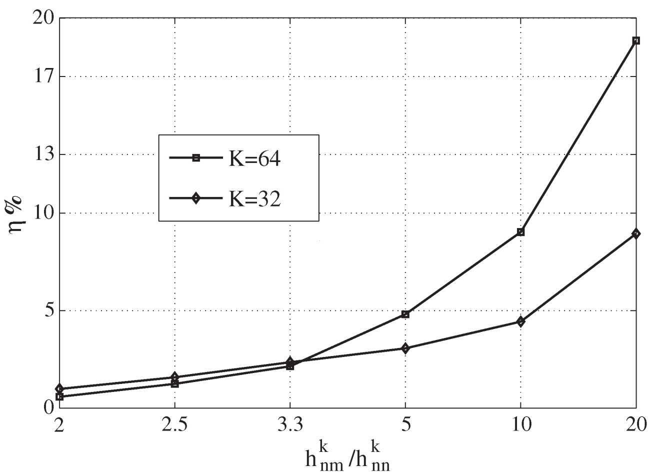


Fig. 6. Efficiency of the opportunistic algorithm in the power control game versus interference levels.

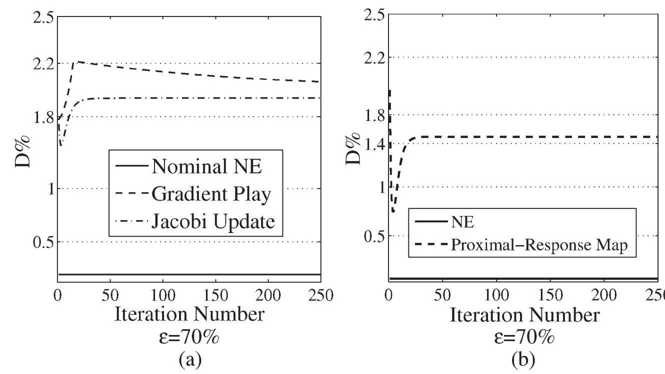


Fig. 7. Total delay in Jackson network when $\boldsymbol{\varepsilon} = 70\%$. (a) At nominal NE via gradient play and Jacobi update [6]. (b) At RNE via proximal point method.

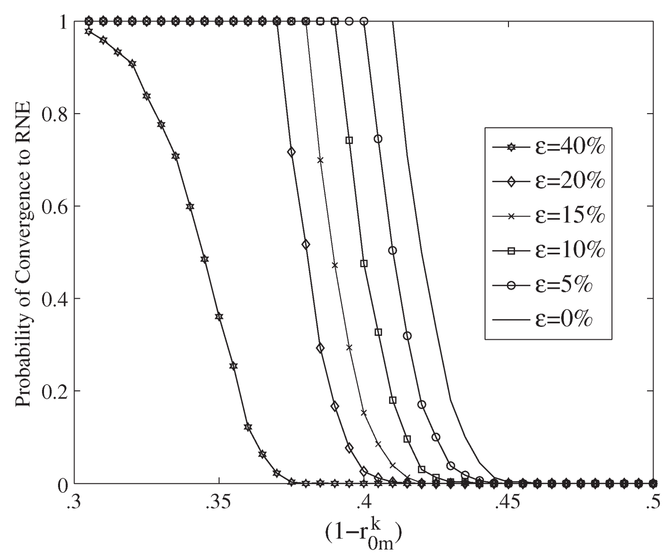


Fig. 8. Probability of convergence to the RNE when $\boldsymbol{\Upsilon}$ in (6) is a P-matrix for different amounts of uncertainty versus $(1 - r_{0m}^k)$ for $N = 5$ and $K = 3$.

Delay minimization in routing of packets in Jackson networks is another example of ACG in communication networks [6], [21]. In this network, incoming packets to node n are split into $\mathcal{K} = \{1, \ldots, K\}$ traffic classes. For class k, the input rate and the service rate are ψ_n^k and ρ_n^k, respectively. Each node is a player, and ψ_n^k is the action of player n for class k of input traffic. The total rate is subject to the minimum rate constraint $\sum_{k=1}^{K} \psi_n^k \geq \psi_n^{\min}$. A class k packet destined to node m is routed to node n with probability r_{nm}^k, or exit the network with probability $r_{0m}^k = 1 - \sum_{n=1}^{N} r_{nm}^k$. We have $[R^k]_{nm} = r_{nm}^k$, $\mathbf{\Theta}^k = (1 - \mathbf{R}^k)^{-1}$, and $\nu_{nm}^k = [\Theta^k]_{nm}$. It can be easily shown that the user's utility for minimizing $M/M/1$ queueing delay can be expressed by $d_n(\mathbf{\Psi}) = \sum_{k=1}^{K} (1/(\rho_n^k - \sum_{m=1}^{N} \nu_{nm}^k \psi_m^k))$, where $\mathbf{\Psi} = [\Psi_1, \ldots, \Psi_N]$, and $\Psi_n = [\psi_n^1, \ldots, \psi_n^K]\ \forall n \in \mathcal{N}$. The optimization problem can be rewritten by maximizing $v_n(\mathbf{a}_n, \mathbf{f}_n)$ subject to the minimum data rate constraint for each user, where $v_n(\mathbf{a}_n, \mathbf{f}_n) = \sum_{k=1}^{K} (\rho_n^k - \sum_{m=1}^{N} \nu_{nm}^k \psi_m^k)$. In this network, $\sum_{m=1}^{N} \nu_{nm}^k \psi_m^k$ is the arrival rate to node n from other nodes (i.e., f_n^k), which can be uncertain. The condition of Theorem 4 holds for this utility function, and the proximal-point method can be used to solve the robust game of the Jackson network in a distributed manner.

To show the effect of uncertainty on the delay in Jackson networks, consider a network with $N = 5$ nodes, $K = 3$ traffic classes, and $\varepsilon = 70\%$. Fig. 7(a) and (b) shows the effect of uncertainty in ψ_n^k on the convergence of Jacobi update and gradient play for reaching the NE and on the convergence of the proximal-point method for reaching the RNE, respectively. In these figures, we show $D = (\breve{d} - d^*)/(d^*\%)$, where $\breve{d}$ is the total delay under perturbation, and d^* is the delay at the NE. Note that when robustness is not applied, neither the Jacobi update nor the gradient play converge to the nominal NE; however, the RNE converges to the vicinity of the nominal NE. In this case, the total delay with Jacobi update and gradient play increases up to 2% at the nominal NE, as compared with a total delay of about 1.4% at the RNE. This shows that the delay at the RNE is about the same as that at the nominal NE. In addition, the proximal-point method converges very fast to the RNE.

Fig. 8 shows the probability of RNE's convergence versus the total routing probability (i.e., $(1 - r_{0m}^k)$ for all m in $\mathcal{N}$) for different uncertainty regions. Note that by increasing uncertainty, the system converges to the RNE for a smaller value of $(1 - r_{0m}^k)$ as compared with that of the nominal NE (i.e., $\varepsilon = 0$). For example, if $\varepsilon = 40\%$, only for $(1 - r_{0m}^k) < 0.3$, the system converges to its equilibrium, whereas for $\varepsilon = 10\%$, the value of $(1 - r_{0m}^k)$ can be up to 0.5 for convergence to the RNE. Fig. 8 shows that the effect of uncertainty is more profound in a network with a high value of $(1 - r_{0m}^k)$, causing poor performance, i.e., large delays in the network. Therefore, when the system encounters a high degree of uncertainty, a lower value of $(1 - r_{0m}^k)$ should be considered to maintain the performance of the system.

IX. Conclusion

We have studied the RNE for a wide range of problems in communication networks when each user's utility depends on its actions and is additively coupled to other users' actions. In this game, the impact of other users on each user is uncertain, and each user optimizes its own utility using the worst case robust optimization. We showed that the theory of finite-dimensional variational inequalities can be used to obtain the sufficient conditions for the existence and uniqueness of the RNE. We have also proposed a distributed algorithm for reaching the RNE. In the case of multiple nominal NEs, simulations showed that at the RNE, the social utility may be higher than that at the corresponding NE, and we derived the sufficient conditions under which this phenomenon can be observed. Simulations confirmed our analysis for two examples, namely, power control in interference channels and delay minimization in Jackson networks.

Appendix A
Proof of Lemma 1

1) Function $\mathbf{f}_n$ is a linear function of other users' strategies and system parameters. Moreover, the norm function is a convex function bounded to ε_n (see [34, Sec. 2.2.2]). Hence, as per [34, Sec. 3.2], $\widetilde{\mathcal{A}}_n(\mathbf{a}_{-n})$ is a convex, bounded, and closed set.
2) To prove the concavity of (13) with respect to $\mathbf{a}_n$, we follow [34, Sec. 3.1]. Consider $\mathbf{a}_n = \mu\mathbf{a}_n^1 + (1-\mu)\mathbf{a}_n^2$ for any positive value $\mu \in [0,1]$. We have

$$\begin{aligned}\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n}) &= \min_{\widetilde{\mathbf{f}}_n\in\Re_n(\mathbf{a}_{-n})} u_n\left(\mu\mathbf{a}_n^1+(1-\mu)\mathbf{a}_n^2,\widetilde{\mathbf{f}}_n\right)\\ &\geq \min_{\widetilde{\mathbf{f}}_n\in\Re_n(\mathbf{a}_{-n})} \mu u_n\left(\mathbf{a}_n^1,\widetilde{\mathbf{f}}_n\right)+(1-\mu)u_n\left(\mathbf{a}_n^2,\widetilde{\mathbf{f}}_n\right)\\ &= \mu\Psi_n\left(\mathbf{a}_n^1,\mathbf{a}_{-n}^1\right)+(1-\mu)\Psi_n\left(\mathbf{a}_n^2,\mathbf{a}_{-n}^2\right).\end{aligned} \tag{A.1}$$

Since (A.1) is based on the convexity of $u_n(\mathbf{a}_n,\widetilde{\mathbf{f}}_n)$ with respect to $\mathbf{a}_n$, the function $\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})$ is concave in $\mathbf{a}_n$. The same is true for the convexity of $\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})$ with respect to $\mathbf{f}_n$. Since $\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})$ is concave, the Lagrange dual function for (12) in the uncertainty region is

$$\begin{aligned}&L(\mathbf{a}_n,\widetilde{\mathbf{f}}_n,\lambda_n)\\ &= \sum_{k=1}^{K} u_n^k\left(a_n^k,\widetilde{f}_n^k\right)-\lambda_n\left(\varepsilon_n^2-\sum_{k=1}^{K}\left(\widetilde{f}_n^k-f_n^k\right)^2\right)\end{aligned} \tag{A.2}$$

where λ_n is the nonnegative Lagrange multiplier that satisfies

$$\lambda_n \times \left(\varepsilon_n^2-\sum_{k=1}^{K}\left(\widetilde{f}_n^k-f_n^k\right)^2\right)=0. \tag{A.3}$$

The solution to (A.2) for $\widetilde{f}_n^k$ is obtained from $((\partial L(\mathbf{a}_n,\widetilde{\boldsymbol{f}}_n,\lambda_n))/(\partial\widetilde{f}_n^k))=0$ [34], [35], which yields $((\partial u_n^k(a_n^k,\widetilde{f}_n^k))/(\partial\widetilde{f}_n^k))=-2\lambda_n\times(\widetilde{f}_n^k-f_n^k)$ for all $k\in\mathcal{K}$. The latter can be rewritten as $(\widetilde{f}_n^k-f_n^k)=(1/(-2\lambda_n))\times((\partial u_n^k(a_n^k,\widetilde{f}_n^k))/(\partial\widetilde{f}_n^k))$. When this value is used in (A.3), the value of λ_n can be obtained, which is $\lambda_n=(1/(2\varepsilon_n))\times\sqrt{\sum_{n=1}^N((\partial u_n^k(a_n^k,\widetilde{f}_n^k))/(\partial\widetilde{f}_n^k))^2}$. As such, the uncertain parameter is $\widetilde{\mathbf{f}}_n^*=\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n$, where $\widetilde{\mathbf{f}}_n^*=[\widetilde{f}_n^*,\ldots,\widetilde{f}_n^{K*}]$, $\boldsymbol{\vartheta}_n=[\vartheta_n^1,\ldots,\vartheta_n^K]$, and ϑ_n^k is (13). Using ϑ_n^k in the utility function u_n, we have $\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})=u_n(\mathbf{a}_n,\mathbf{f}_n)|_{\widetilde{\mathbf{f}}_n^*=\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n}$. Comparing the value $\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})$ with $v_n(\mathbf{a}_n,\mathbf{f}_n)$ indicates that the difference between Ψ_n and the utility function of the nominal game is the extra term $\varepsilon_n\boldsymbol{\vartheta}_n$. From Assumption **A2** in Section II, $\varepsilon_n\boldsymbol{\vartheta}_n$ is continuous. Therefore, $\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})$ is continuous with respect to $\mathbf{a}_n$. The derivative of Ψ_n with respect to $\mathbf{a}_n$ is

$$\begin{aligned}&\nabla_{\mathbf{a}_n}\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})\\ &\quad=\nabla_{\mathbf{a}_n}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)+\nabla_{\widetilde{\mathbf{f}}_n}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)\\ &\qquad\times\mathbf{1}_K\times\nabla_{\mathbf{a}_n}\widetilde{\mathbf{f}}_n\times\mathbf{1}_K^{\mathrm{T}}\\ &\quad=\nabla_{\mathbf{a}_n}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)-\varepsilon_n\nabla_{\widetilde{\mathbf{f}}_n}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)\\ &\qquad\times\mathbf{1}_K\times\nabla_{\mathbf{a}_n}\boldsymbol{\vartheta}_n\times\mathbf{1}_K^{\mathrm{T}}\end{aligned} \tag{A.4}$$

where $\mathbf{1}_K$ is a $1\times K$ vector whose elements are equal to one. The last term in (A.4) contains $((\partial^2 u_n^k)/(\partial a_n^k\partial f_n^k))$. From Assumption **A3** in Section II, $\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})$ is differentiable with respect to $\mathbf{a}_n$. Now, the optimization problem for each user is rewritten as $\widetilde{u}_n=\max_{\mathbf{a}_n\in\mathcal{A}_n}\Psi_n(\mathbf{a}_n,\mathbf{a}_{-n})$, and the game is reformulated as $\{\mathcal{N},(\Psi_n)_{n\in\mathcal{N}},\mathcal{A}\}$.

Appendix B
Proof of Lemma 2

For the RACG, we have $GVI(\mathcal{A},\widetilde{\mathcal{F}})$, and $\widetilde{\mathcal{F}}(\mathbf{a})=(\widetilde{\mathcal{F}}_n(\mathbf{a}))_{n=1}^N$, where $\widetilde{\mathcal{F}}_n(\mathbf{a})$ is obtained by (A.4) for user n. Variations in system parameters and in other users' strategies cause variations in $\widetilde{\mathbf{f}}_n$ for user n. When both of these two variations are zero, we have $\|\widehat{\mathbf{e}}_n\|=\varepsilon_n=0$. The Taylor series of $\widetilde{\mathcal{F}}_n(\mathbf{a})$ around ε_n is $\widetilde{\mathcal{F}}_n(\mathbf{a})=[\widetilde{\mathcal{F}}_n(\mathbf{a})]_{\varepsilon_n}+[\sum_{i=1}^{\infty}(1/(i!))(\varepsilon_n)^i(\nabla_{\widetilde{\mathbf{f}}_n}^i\widetilde{\mathcal{F}}_n)]_{\varepsilon_n}$. From (A.4), this Taylor series is

$$\begin{aligned}&\widetilde{\mathcal{F}}_n(\mathbf{a})\\ &\quad=-\left[\nabla_{\mathbf{a}_n}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)\right]_{(\varepsilon_n)}\\ &\qquad-\varepsilon_n\Big[\nabla^2_{\mathbf{a}_n\widetilde{\mathbf{f}}_n}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)\\ &\qquad\qquad\times\left(\mathbf{1}_K^{\mathrm{T}}-\varepsilon_n\nabla_{\widetilde{\mathbf{f}}_n}\boldsymbol{\vartheta}_n\times\mathbf{1}_K^{\mathrm{T}}\right)\Big]_{(\varepsilon_n)}\\ &\qquad-\frac{\varepsilon_n^2}{2}\Big[\nabla^3_{\mathbf{a}_n\widetilde{\mathbf{f}}_n^2}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)\\ &\qquad\qquad\times\left(\mathbf{1}_K^{\mathrm{T}}-\varepsilon_n\nabla_{\widetilde{\mathbf{f}}_n}\boldsymbol{\vartheta}_n\times\mathbf{1}_K^{\mathrm{T}}\right)\end{aligned}$$

$$\times\left(\mathbf{1}_K^{\mathrm{T}}-\varepsilon_n\nabla_{\widetilde{\mathbf{f}}_n}\boldsymbol{\vartheta}_n\times\mathbf{1}_K^{\mathrm{T}}\right)$$
$$+\nabla^2_{\mathbf{a}_n\widetilde{\mathbf{f}}_n}u_n(\mathbf{a}_n,\mathbf{f}_n-\varepsilon_n\boldsymbol{\vartheta}_n)$$
$$\times\left(\varepsilon_n\nabla^2_{\widetilde{\mathbf{f}}_n\widetilde{\mathbf{f}}_n}\boldsymbol{\vartheta}_n\times\mathbf{1}_K^{\mathrm{T}}\right)\Big]_{(\varepsilon_n)}+o(\cdot).\quad\text{(B.1)}$$

From (3), and the fact that $\widetilde{\mathbf{f}}_n=\mathbf{f}_n$ for $\varepsilon_n=0$, (B.1) can be rewritten as

$$\widetilde{\mathcal{F}}_n(\mathbf{a})=\mathcal{F}_n(\mathbf{a})-\varepsilon_n\left[\nabla^2_{\mathbf{a}_n\mathbf{f}_n}v_n(\mathbf{a}_n,\mathbf{f}_n)\times\mathbf{1}_K^{\mathrm{T}}\right]$$
$$-\frac{\varepsilon_n^2}{2}\left[\nabla^3_{\mathbf{a}_n\mathbf{f}_n^2}v_n(\mathbf{a}_n,\mathbf{f}_n)\times\mathbf{1}_K^{\mathrm{T}}\right]+o(\cdot).\quad\text{(B.2)}$$

From Assumption **A1** in Section II, all derivatives of $v_n(\mathbf{a}_n,\mathbf{f}_n)$ are bounded. Therefore, the last three terms on the RHS of (B.1) are bounded, and $\widetilde{\mathcal{F}}_n(\mathbf{a})$ is the bounded perturbed version of $\mathcal{F}_n(\mathbf{a})$.

APPENDIX C
PROOF OF THEOREM 2

1) From Lemma 2, recall that in $GVI(\mathcal{A},\widetilde{\mathcal{F}})$, $\widetilde{\mathcal{F}}(\mathbf{a})$ is a set-valued mapping, and $\widetilde{\mathcal{F}}(\mathbf{a})$ is a perturbed version of $\mathcal{F}(\mathbf{a})$. The perturbed region of $\widetilde{\mathcal{F}}(\mathbf{a})$ is $\mathcal{Q}=\|\mathcal{F}(\mathbf{a})-\widetilde{\mathcal{F}}(\mathbf{a})\|_2\quad\forall\,\mathbf{a}\in\mathcal{A}$. Since the strategy space of all users in each dimension is bounded as in (1), and the uncertainty region is bounded and convex, this region is also bounded, i.e., $q^{\max}\triangleq\max_{\mathbf{a}\in\mathcal{A}}\|\mathcal{F}(\mathbf{a})-\widetilde{\mathcal{F}}(\mathbf{a})\|_2\leq\infty$. From Assumption **A2** in Section II, we have

$$\widetilde{\mathcal{F}}(\mathbf{a})\geq\mathcal{F}(\mathbf{a})+q^{\max}\left(\mathbf{1}_K^{\mathrm{T}}\right)_1^N.\quad\text{(C.1)}$$

Lemma 3: When the mapping $\mathcal{F}(\mathbf{a})$ is strongly monotone and $\widetilde{\mathcal{F}}(\mathbf{a})$ is a bounded perturbed version of $\mathcal{F}(\mathbf{a})$, $\widetilde{\mathcal{F}}(\mathbf{a})$ is strongly monotone.

Proof: To establish strong montoncity of $\widetilde{\mathcal{F}}(\mathbf{a})$, we need to show that there exists $c>0$ such that $(\mathbf{a}_1-\mathbf{a}_2)(\widetilde{\mathcal{F}}(\mathbf{a}_1)-\widetilde{\mathcal{F}}(\mathbf{a}_2))\geq c\|\mathbf{a}_1-\mathbf{a}_2\|_2$, $\forall\,\mathbf{a}_1,\,\mathbf{a}_2\in\mathcal{A}$. To do so, when $\widetilde{\mathcal{F}}(\mathbf{a})$ is a bounded perturbed version of $\mathcal{F}(\mathbf{a})$, (C.1) holds, and via simple mathematical manipulations, we have $(\mathbf{a}_1-\mathbf{a}_2)(\widetilde{\mathcal{F}}(\mathbf{a}_1)-\widetilde{\mathcal{F}}(\mathbf{a}_2))\geq(\mathbf{a}_1-\mathbf{a}_2)(\mathcal{F}(\mathbf{a}_1)-\mathcal{F}(\mathbf{a}_2))$. Now, when $\mathcal{F}(\mathbf{a})$ is strongly monotone, there exists $c_1>0$ such that $(\mathbf{a}_1-\mathbf{a}_2)(\mathcal{F}(\mathbf{a}_1)-\mathcal{F}(\mathbf{a}_2))\geq c_1\|\mathbf{a}_1-\mathbf{a}_2\|_2$. From the given discussion, we write $(\mathbf{a}_1-\mathbf{a}_2)(\widetilde{\mathcal{F}}(\mathbf{a}_1)-\widetilde{\mathcal{F}}(\mathbf{a}_2)\geq c_1\|\mathbf{a}_1-\mathbf{a}_2\|_2$, which establishes strong monotonicity of $\widetilde{\mathcal{F}}(\mathbf{a})$. ■

When (6) is a P-matrix, $v_n(\mathbf{a}_n,\mathbf{a}_{-n})$ is uniformly strong convex (see [4, Sec. 12.4.1]), and its gradient column vector $\mathcal{F}(\mathbf{a})$ defined in (3) is strongly monotone [3]. Hence, from Lemma 3, $\widetilde{\mathcal{F}}(\mathbf{a})$ is strongly monotone. Moreover, we showed in Lemma 1 that $\widetilde{\mathcal{F}}(\mathbf{a})$ is continuous. Thus, the assumptions of [13, Th. 4.3] hold, and $GVI(\mathcal{A},\widetilde{\mathcal{F}})$ has a unique solution.

2) As was shown in Lemma 2, the RNE is a perturbed solution to $VI(\mathcal{A},\mathcal{F})$, which can be rewritten as $VI(\mathcal{A},\mathcal{F}+\mathbf{q})$, where $\mathbf{q}=(\mathbf{q}_n)_{n=1}^N$, $\mathbf{q}_n=[q_n^1,\ldots,q_n^K]^{\mathrm{T}}$, and each q_n^k is bounded such that $\|\mathbf{q}\|_2\in\mathcal{Q}$. Recall that when (6) is a P-matrix, $\mathcal{F}(\mathbf{a})$ is strongly monotone, and the utility is strongly convex. Since $\mathcal{A}$ is convex in $\mathbb{R}^K$, and $\mathcal{F}(\mathbf{a}):K\to\mathbb{R}^K$ is a continuous mapping on $\mathcal{A}$, the solution to $VI(\mathcal{A},\mathcal{F}+\mathbf{q})$ is always a compact and convex set (see [3, Corollary 2.6.4]). This solution set contains $\mathbf{a}^*$ and $\widetilde{\mathbf{a}}^*$, and we have

$$(\widetilde{\mathbf{a}}^*-\mathbf{a}^*)\mathcal{F}(\mathbf{a}^*)>0\quad\text{(C.2)}$$
$$(\mathbf{a}^*-\widetilde{\mathbf{a}}^*)\left(\mathcal{F}(\widetilde{\mathbf{a}}^*)+\mathbf{q}\right)>0.\quad\text{(C.3)}$$

If we subtract (C.2) from (C.3), we have

$$(\widetilde{\mathbf{a}}^*-\mathbf{a}^*)\left(\mathcal{F}(\mathbf{a}^*)-(\mathcal{F}(\widetilde{\mathbf{a}}^*)+\mathbf{q})\right)>0.\quad\text{(C.4)}$$

Since from (C.1), the term $(\mathcal{F}(\mathbf{a}^*)-(\mathcal{F}(\widetilde{\mathbf{a}}^*)+\mathbf{q}))$ in (C.4) is negative, we should have $\widetilde{\mathbf{a}}^*<\mathbf{a}^*$. Now, when the utility of user n at the RNE is greater than that at the nominal NE (i.e., when $\Psi_n(\widetilde{\mathbf{a}}_n^*,\widetilde{\mathbf{a}}_{-n}^*)>v_n(\mathbf{a}_n^*,\mathbf{f}_n(\mathbf{a}_{-n}^*,\mathbf{s}_n)))$, we have $\widetilde{\mathcal{F}}_n(\widetilde{\mathbf{a}}^*)<\mathcal{F}_n(\mathbf{a}^*)$. Therefore, $\widetilde{\mathcal{F}}(\mathbf{a})$ is not strongly monotone, which contradicts Lemma 3. This contradiction implies that each user's utility at the RNE is less than that at the nominal NE. Consequently, the social utility at the RNE is less than that at the nominal NE.

3) Since $\widetilde{\mathcal{F}}(\mathbf{a})$ is strongly monotone, there is a unique solution [3], denoted by $\widetilde{\mathbf{a}}^*=\Phi^*(\mathbf{q})$, which can be considered as the worst case robust solution to $\widetilde{\mathcal{G}}$ for $\|\mathbf{q}\|_2\leq\|\boldsymbol{\Delta}\|_2$. Now, both $\mathbf{a}_n^*$ and $\widetilde{\mathbf{a}}_n^*$ must satisfy

$$0\leq\left(\Phi^*(\mathbf{q})-\Phi^*(\mathbf{0})\right)\left(\mathcal{F}\left(\Phi^*(\mathbf{0})\right)\right)\quad\text{(C.5)}$$
$$0\leq\left(\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\right)\left(\mathcal{F}\left(\Phi^*(\mathbf{q})\right)+\mathbf{q}\right)\quad\text{(C.6)}$$

where $\mathbf{0}=(\mathbf{0}_K)_1^N$, and $\mathbf{0}_K$ is the $K\times1$ all-zero vector. By subtracting (C.5) from (C.6), we get

$$\left(\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\right)\left(\mathcal{F}\left(\Phi^*(\mathbf{0})\right)-\mathcal{F}\left(\Phi^*(\mathbf{q})\right)\right)$$
$$\leq\left(\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\right)\mathbf{q}.\quad\text{(C.7)}$$

Using Schwartz inequality for the RHS, we have $\|(\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q}))\mathbf{q}\|_2\leq\|\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\|_2\|\mathbf{q}\|_2$. Since $\Phi^*(\mathbf{q})$ is the cocoercive function of $\mathbf{q}$ (see [3, Proposition 2.3.11]), the left-hand side of (C.7) is always greater than $c_{\mathrm{sm}}\|\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\|_2^2$. Therefore, (C.7) can be rewritten as $c_{\mathrm{sm}}\|\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\|_2^2\leq\|\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\|_2\|\mathbf{q}\|_2$, which is simplified to $c_{\mathrm{sm}}\|\Phi^*(\mathbf{0})-\Phi^*(\mathbf{q})\|_2\leq\|\mathbf{q}\|_2$. Recall that $\Phi^*(\mathbf{0})$ and $\Phi^*(\mathbf{q})$ correspond to $\mathbf{a}_n^*$ and $\widetilde{\mathbf{a}}_n^*$, respectively, and the upper bound on $\mathbf{q}$ is $\boldsymbol{\Delta}$. Hence, we have $c_{\mathrm{sm}}\|\mathbf{a}_n^*-\widetilde{\mathbf{a}}_n^*\|_2\leq\|\boldsymbol{\Delta}\|_2$, which is the same as (14).

APPENDIX D
PROOF OF (16)

Since $\widetilde{\mathcal{F}}(\mathbf{a})$ is a bounded perturbed version of $\mathcal{F}(\mathbf{a})$, the difference between the utilities of user n at the RNE and at the nominal NE can be approximated by the first term of the Taylor series of $u_n^k(\mathbf{a}_n,\mathbf{f}_n)$ with respect to all variations in the strategies of user n and other users, i.e., $u_n^k(a_n^k,f_n^k)-v_n^k(a_n^k,f_n^k)\approx\varepsilon_n(((\partial v_n^k(a_n^k,f_n^k))/(\partial a_n^k))\times((\partial a_n^k)/(\partial\varepsilon_n))\ +$

$((\partial v_n^k(a_n^k, f_n^k))/(\partial f_n^k)) \times ((\partial f_n^k)/(\partial \varepsilon_n))) \quad \forall n \in \mathcal{N},\ k \in \mathcal{K}$, which is

$$u_n^k\left(a_n^k, f_n^k\right) - v_n^k\left(a_n^k, f_n^k\right) \approx \varepsilon_n \left(\partial \frac{v_n^k\left(a_n^k, f_n^k\right)}{\partial a_n^k} \times \frac{\partial a_n^k}{\partial \varepsilon_n} + \frac{\partial v_n^k\left(a_n^k, f_n^k\right)}{\partial f_n^k} \times \sum_{m \neq n} x_{nm}^k \frac{\partial a_m^k}{\partial \varepsilon_n}\right). \tag{D.1}$$

When ε_n is sufficiently small, $((\partial a_n^k)/(\partial \varepsilon_n)) = \lim_{\varepsilon_n \to 0}((\tilde{a}_n^{*k} - a_n^{*k})/(\varepsilon_n))$. By considering (D.1) for all users, we have

$$\|v(\mathbf{a}^*) - u(\tilde{\mathbf{a}}^*)\|_2 \approx \|\mathbf{W}(\mathbf{a}^*)\|_2 \times \|\mathbf{a}^* - \widetilde{\mathbf{a}}^*\|. \tag{D.2}$$

By replacing (14) into (D.2), approximation (16) is obtained. In the given formula, we use the first term in Taylor polynomial for v_n to approximate $\|v(\mathbf{a}^*) - u(\tilde{\mathbf{a}}^*)\|_2$. The remainder of this difference for user n is always less than or equal to $((\varepsilon_n^2)/(2!))((\partial^2 v_n)/(\partial^2 \mathbf{a}_n))$ [36]. For all users, this remainder is upper bounded to $((\|J(\mathcal{F})\|_2 \|\boldsymbol{\Delta}\|_2^2)/2)$.

Appendix E
Proof of Theorem 3

1) From Lemma 1, the mapping for the RACG is the perturbed mapping of the NACG. Since the mapping for the NACG is linear for utility function (18), the perturbed mapping is

$$\widetilde{\mathcal{M}}_n^k(\mathbf{a}) = \varpi_n^k + \sum_{m=1}^{N} \frac{\widetilde{x}_{nm}^k}{x_{nn}^k} a_m^k \quad \forall \widetilde{x}_{nm}^k \in \Re_n^k \tag{E.1}$$

where $\widetilde{\mathcal{M}}_n^k(\mathbf{a})$ and ϖ_n^k are the kth elements of $\widetilde{\mathcal{M}}_n(\mathbf{a})$ and $\boldsymbol{\varpi}_n$, respectively. Now, (E.1) can be rewritten as $\widetilde{\mathcal{M}}_n^k(\mathbf{a}) = \varpi_n^k + \sum_{m=1}^N ((x_{nm}^k/x_{nn}^k) + ((\widetilde{x}_{nm}^k - x_{nm}^k)/x_{nn}^k))a_m^k \leq \varpi_n^k + \sum_{m=1}^N (x_{nm}^k/x_{nn}^k)a_m^k + \epsilon_n^k\|\mathbf{a}_{-n}^k\|$. Therefore, the mapping at the RNE is (22).

2) Since a_m^k is bounded in $[a_{mk}^{\min}, a_{mk}^{\max}]$, and the uncertainty region is bounded, the value of $\boldsymbol{\epsilon}_n^k\|\mathbf{a}_{-n}^k\|_2$ is bounded. Hence, for any bounded uncertainty region in the RACG, its AVI is $AVI = (\mathcal{A}, \mathcal{M} + \mathbf{m})$, where $\mathbf{m} = (\mathbf{m}_n)_{n=1}^N = (\boldsymbol{\varpi}_n + \widehat{\mathcal{M}}_n)_{n=1}^N$, $\|\mathbf{m}\|_2 < \infty$, and the RNE is the perturbed solution to $AVI = (\mathcal{A}, \mathcal{M})$. From [3, Th. 4.3.2], when $\mathbf{M}$ is semicopositive (matrix $\mathbf{M}$ is semicopositive if for any positive vector $\boldsymbol{\varrho}$, we have $\varrho_i(\mathbf{M}\boldsymbol{\varrho})_i > 0$, where ϱ_i is the ith element of $\boldsymbol{\varrho}$), AVI has a unique solution for any $\mathbf{m}$. Thus, the RACG has a unique solution for any bound on the uncertainty region.

3) Part 3 can be proved the same as Part 2 of Theorem 1. Recall that $\mathcal{M}$ is strongly monotone on $\mathcal{A}$ when there exists c_{sm}, such that for all $\mathbf{a}^1 = (\mathbf{a}_n^1)_{n\in\mathcal{N}}$ and $\mathbf{a}^2 = (\mathbf{a}_n^2)_{n\in\mathcal{N}}$, we have $(\mathbf{a}^1 - \mathbf{a}^2)(\mathcal{M}(\mathbf{a}^1) - \mathcal{M}(\mathbf{a}^2)) \geq c_{\text{sm}}\|\mathbf{a}^1 - \mathbf{a}^2\|$. When $d_n^k = (a_n^{k1} - a_n^{k2})$, we have

$$\begin{aligned}
&\left(\mathbf{a}_n^1 - \mathbf{a}_n^2\right)\left(\mathcal{M}_n(\mathbf{a}^1) - \mathcal{M}_n(\mathbf{a}^1)\right)\\
&= (\mathbf{a}_n^1 - \mathbf{a}_n^2)\left(\sum_{m=1}^{N}\left(\mathbf{M}_{nm}(\mathbf{a}_m^1)^{\mathrm{T}} - \mathbf{M}_{nm}(\mathbf{a}_m^2)^{\mathrm{T}}\right)\right)\\
&= \sum_{k=1}^{K}\left(a_n^{k1} - a_n^{k2}\right)\left(\sum_{m=1}^{N} M_{nm}^{kk}\left(a_n^{k1} - a_n^{k2}\right)\right)\\
&\geq \sum_{k=1}^{K}\left(d_n^k\right)^2 - \sum_{m=1, m\neq n}^{N}\left|\sum_{k=1}^{K} e_m^k \frac{x_{nm}^k}{x_{mm}^k} d_n^k\right|\\
&\geq \sum_{k=1}^{K}\left(d_n^k\right)^2 - \sum_{m=1,\, m\neq n}^{N}\left(\sum_{k=1}^{K} d_n^k\right)^2 \max_{k\in K}\frac{x_{nm}^k}{x_{mm}^k} \times \left(\sum_{k=1}^{K}\left(d_m^k\right)^2\right)^2\\
&\geq \|\mathbf{d}_n\|_2 \sum_{m=1}^{N}[\mathbf{M}^{\max}]_{nm}\|\mathbf{d}_m\|_2
\end{aligned} \tag{E.2}$$

where $\mathbf{d}_n = [d_n^1 \cdots d_n^k]^{\mathrm{T}}$. Thus, $(\mathbf{a}^1 - \mathbf{a}^2)(\mathcal{M}(\mathbf{a}^1) - \mathcal{M}(\mathbf{a}^2)) \geq \mathbf{d}^{\mathrm{T}}\mathbf{M}^{\max}\mathbf{d}$ for all $n \in \mathcal{N}$ and $\mathbf{d}^{\mathrm{T}} = [\mathbf{d}_1^{\mathrm{T}}, \ldots, \mathbf{d}_N^{\mathrm{T}}]$. Since $\mathbf{d}^{\mathrm{T}}\mathbf{M}^{\max}\mathbf{d} \geq \lambda_{\min}(\mathbf{M}^{\max})\|\mathbf{d}\|_2$, matrix $\mathbf{M}^{\max}$ is positive semidefinite. Hence, $c_{\text{sm}}(\mathcal{F}) = \lambda_{\min}(\mathbf{M}^{\max})$, and from (14), Part 3 of Theorem 3 is proved.

Appendix F
Proof of Theorem 4

When the solution to (24) is obtained by contraction mapping, the distributed algorithm for the proximal-point method converges. For any vector $\mathbf{z} \in \mathcal{A}$ in (24), we have

$$(\mathbf{z} - \widehat{\mathbf{a}}(\mathbf{b}_1))\left[\widetilde{\mathcal{F}}\left(\widehat{\mathbf{a}}(\mathbf{b}_1), \mathbf{b}_1\right) + \left(\widehat{\mathbf{a}}(\mathbf{b}_1) - \mathbf{b}_1\right)^{\mathrm{T}}\right] \geq 0 \tag{F.1}$$

$$(\mathbf{z} - \widehat{\mathbf{a}}(\mathbf{b}_2))\left[\widetilde{\mathcal{F}}\left(\widehat{\mathbf{a}}(\mathbf{b}_2), \mathbf{b}_2\right) + \left(\widehat{\mathbf{a}}(\mathbf{b}_2) - \mathbf{b}_2\right)^{\mathrm{T}}\right] \geq 0. \tag{F.2}$$

Considering $\mathbf{z} = \widehat{\mathbf{a}}(\mathbf{b}_2)$ in (F.1) and $\widehat{\mathbf{z}} = \widehat{\mathbf{a}}(\mathbf{b}_1)$ in (F.2), from the given two inequalities, we get

$$\begin{aligned}
0 \leq &\left(\widehat{\mathbf{a}}(\mathbf{b}_2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\right)\left[\widetilde{\mathcal{F}}\left(\widehat{\mathbf{a}}(\mathbf{b}^1), \mathbf{b}^1\right) + \left(\widehat{\mathbf{a}}(\mathbf{b}^1) - \mathbf{b}^1\right)^{\mathrm{T}}\right]\\
&+ \left(\widehat{\mathbf{a}}(\mathbf{b}^1) - \widehat{\mathbf{a}}(\mathbf{b}^2)\right)\left[\widetilde{\mathcal{F}}\left(\widehat{\mathbf{a}}(\mathbf{b}^2), \mathbf{b}^2\right) + \left(\widehat{\mathbf{a}}(\mathbf{b}^2) - \mathbf{b}^2\right)^{\mathrm{T}}\right]\\
= &\left(\widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\right)\left[\widetilde{\mathcal{F}}\left(\widehat{\mathbf{a}}(\mathbf{b}^1), \mathbf{b}^1\right) - \widetilde{\mathcal{F}}\left(\widehat{\mathbf{a}}(\mathbf{b}^2), \mathbf{b}^2\right)\right]\\
&- \|\widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\| + \left(\widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\right)(\mathbf{b}^1 - \mathbf{b}^2)^{\mathrm{T}}.
\end{aligned} \tag{F.3}$$

Recall that $\widetilde{\mathcal{F}}_n = -\nabla_{\mathbf{a}_n} u_n(\mathbf{a}_n, \mathbf{f}_n + \varepsilon_n \boldsymbol{\vartheta}_n) - \varepsilon_n \nabla_{\widetilde{\mathbf{f}}_n} u_n(\mathbf{a}_n, \mathbf{f}_n + \varepsilon_n \boldsymbol{\vartheta}_n) \times \mathbf{1}_K \times \nabla_{\mathbf{a}_n} \boldsymbol{\vartheta}_n \times \mathbf{1}_k^{\mathrm{T}}$, $\nabla_{\mathbf{a}_n} \widetilde{\mathcal{F}}_n = -\nabla_{\mathbf{a}_n, \mathbf{a}_n}^2 u_n + \varepsilon_n \times \nabla_{\mathbf{a}_n \mathbf{a}_n \mathbf{f}_n}^3 u_n$, and $\nabla_{\mathbf{a}_m} \widetilde{\mathcal{F}}_n = -\nabla_{\mathbf{a}_n \mathbf{a}_m}^2 u_n + \varepsilon_n \times \nabla_{\partial \mathbf{a}_n \partial^2 \mathbf{f}_n}^3 u_n \times \mathbf{1}_K^{\mathrm{T}} \times \mathbf{x}_{nm}$, when $((\partial^3 v_n)/(\partial \mathbf{a}_n \partial^2 \mathbf{f}_n)) = ((\partial^3 v_n)/(\partial^2 \mathbf{a}_n \partial \mathbf{f}_n)) = 0$. We rewrite (F.3) as

$$\begin{aligned}
&\left(\widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\right)\left[\sum_{n\in\mathcal{N}} -\nabla_{\mathbf{a}_n \mathbf{a}_n}^2 u_n\right]\left(\widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\right)^{\mathrm{T}}\\
&+ \left(\widehat{\mathbf{a}}(\mathbf{b}^1) - \widehat{\mathbf{a}}(\mathbf{b}^2)\right)\left[\sum_{m\in\mathcal{N}, m\neq n} -\nabla_{\mathbf{a}_n \mathbf{a}_m}^2 u_n\right](\mathbf{b}^1 - \mathbf{b}^2)^{\mathrm{T}}\\
&- \|\widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\| + \left(\mathbf{a}(\mathbf{b}^2) - \mathbf{a}(\mathbf{b}^1)\right)\left(\mathbf{b}^1 - \mathbf{b}^2\right)^{\mathrm{T}} \geq 0.
\end{aligned} \tag{F.4}$$

Consider $\widetilde{\alpha}_n(\mathbf{a}) \triangleq$ smallest eigenvalue of $-\nabla^2_{\mathbf{a}_n} u_n(\mathbf{a}_n, \mathbf{f}_n)$, $\widetilde{\beta}_{nm}(\mathbf{a}) \triangleq \| -\nabla_{\mathbf{a}_n \mathbf{a}_m} u_n(\mathbf{a}_n, \mathbf{f}_n)\|$, $\forall\, n \neq m$, and $\mathbf{z} \triangleq \tau(\mathbf{a}^1(\mathbf{b}^1), \mathbf{b}^1) + (1-\tau)(\mathbf{a}^2(\mathbf{b}^2), \mathbf{b}^2)$.

From (F.4), we get

$$(1+\widetilde{\alpha}_n(\mathbf{z}))\, \|\widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\left\| \leq \sum_{n=1}^{N} \widetilde{\beta}_{nm}(\mathbf{z}_n) \right\| \mathbf{b}^1_{-n} - \mathbf{b}^2_{-n}\|. \tag{F.5}$$

On the other hand, $-\nabla^2_{\mathbf{a}_n \mathbf{a}_n} u_n = -\nabla_{\mathbf{a}_n \mathbf{a}_n} u_n(\mathbf{a}_n, \mathbf{f}_n - \varepsilon_n \boldsymbol{\vartheta}_n) + \varepsilon_n \nabla_{\widetilde{\mathbf{f}}_n} u_n(\mathbf{a}_n, \mathbf{f}_n - \varepsilon_n \boldsymbol{\vartheta}_n) \times \mathbf{1}_K \times \nabla_{\mathbf{a}_n} \boldsymbol{\vartheta}_n$. Since the utility is convex with respect to $\mathbf{a}_n$, we have $-\nabla_{\mathbf{a}_n \mathbf{a}_n} u_n(\mathbf{a}_n, \mathbf{f}_n - \varepsilon_n \boldsymbol{\vartheta}_n) > 0$ and $\|\nabla^2_{\mathbf{a}_n \mathbf{a}_n} u_n\| \geq \|\nabla^2_{\mathbf{a}_n \mathbf{a}_n} v_n\|$. Moreover, $\nabla^2_{\mathbf{a}_n \mathbf{a}_m} u_n = \nabla_{\mathbf{a}_n \mathbf{a}_m} u_n(\mathbf{a}_n, \mathbf{f}_n + \varepsilon_n \boldsymbol{\vartheta}_n) - \varepsilon_n \nabla_{\widetilde{\mathbf{f}}_n} u_n(\mathbf{a}_n, \mathbf{f}_n - \varepsilon_n \boldsymbol{\vartheta}_n) \times \mathbf{1}_K \times \nabla_{\mathbf{a}_n} \boldsymbol{\vartheta}_n$, which leads to $\|\nabla^2_{\mathbf{a}_n \mathbf{a}_m} u_n\| \leq \|\nabla^2_{\mathbf{a}_n \mathbf{a}_m} v_n\|$. From these two inequalities, (F.5) can be rewritten as $(1 + \alpha_n(\mathbf{z}))\, \|\, \widehat{\mathbf{a}}(\mathbf{b}^2) - \widehat{\mathbf{a}}(\mathbf{b}^1)\, \| \leq \sum_{n=1}^{N} \beta_{nm}(\mathbf{z}^n)\, \|\, \mathbf{b}^1_{-n} - \mathbf{b}^2_{-n}\|$, which shows that when (6) is a P-matrix, (25) is a contraction mapping (see [4, Proposition 12.17 in Sec. 12]), and converges to a unique RNE.

Appendix G
Negative Definiteness of Matrix $\mathbf{M}$ in Affine VI

Consider $AVI(\mathcal{A}, \mathcal{M} + \mathbf{m})$, where $\mathcal{A}$ and $\mathcal{M}(\mathbf{a})$ were defined in Proposition 1, $\mathbf{a} \in \mathcal{A}$, and $\mathbf{m}$ is a vector with bounded positive values. When $\mathcal{M}(\mathbf{a})$ is strongly monotone, the solution to $AVI(\mathcal{A}, \mathcal{M} + \mathbf{m})$, denoted by the row vector $\Phi(\mathbf{m})$, is monotone (see [3, Exercise 2.9.17]). From [3, Definition 1.1.1], we have

$$(\Phi(\mathbf{m}) - \Phi(\mathbf{0}))\, \mathcal{M}\,(\Phi(\mathbf{0})) > 0 \tag{G.1}$$

$$(\Phi(\mathbf{0}) - \Phi(\mathbf{m}))\, (\mathcal{M}(\Phi(\mathbf{m})) + \mathbf{m}) > 0. \tag{G.2}$$

Subtracting (G.1) from (G.2), we get $(\Phi(\mathbf{0}) - \Phi(\mathbf{m}))$ $(\mathcal{M}(\Phi(\mathbf{m})) - \mathcal{M}(\Phi(\mathbf{0}))) + (\Phi(\mathbf{0}) - \Phi(\mathbf{m})) > 0$. Since the solution of affine VI is assumed to be a monotone and a decreasing function (Remark 5), we have $\Phi(\mathbf{0}) \geq \Phi(\mathbf{m})$. Hence, $\mathcal{M}(\Phi(\mathbf{m})) > \mathcal{M}(\Phi(\mathbf{0}))$, which together from Part **1** in Proposition 1, yields $\mathbf{M}\Phi^{\mathrm{T}}(\mathbf{m}) > \mathbf{M}\Phi^{\mathrm{T}}(\mathbf{0})$, where $\mathbf{M} = (\mathbf{M}_n)_{n=1}^{N}$ and $\mathbf{M}_n = \sum_{m=1}^{N} \mathbf{M}_{nm}$. Now, since $\mathbf{g} = \Phi(\mathbf{m}) - \Phi(\mathbf{0})$ is a negative vector, we get $\mathbf{g}\mathbf{M}\mathbf{g}^{\mathrm{T}} < 0$, which means that $\mathbf{M}$ is a negative definite matrix.

References

[1] S. Parsaeefard, A. R. Sharafat, and M. van der Schaar, "Robust equilibria in additively coupled games in communication networks," in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1–6.

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

[3] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. New York, NY, USA: Springer-Verlag, 2003.

[4] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[5] M. Felegyhazi and J. P. Hubaux, "Game theory in wireless networks: A tutorial," École Polytech. Féd. de Lausanne, Laussane, Switzerland, EPFL Tech. Rep. LCA-REPORT-2006-002, Feb. 2006.

[6] Y. Su and M. van der Schaar, "Structural solutions for additively coupled sum constrained games," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3779–3796, Dec. 2012.

[7] F. Meshkati, A. J. Goldsmith, H. V. Poor, and S. C. Schwartz, "A game-theoretic approach to energy-efficient modulation in CDMA networks with delay QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1069–1078, Aug. 2007.

[8] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[9] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory—Part I: Nash equilibria," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1230–1249, Mar. 2008.

[10] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory—Part II: Algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1250–1267, Mar. 2008.

[11] M. Aghassi and D. Bertsimas, "Robust game theory," *Math. Programm.*, vol. 107, no. 1, pp. 231–273, Jun. 2006.

[12] S. Hayashi, N. Yamashita, and M. Fukushima, "Robust Nash equilibria and second-order cone complementarity problems," *J. Nonlin. Convex Anal.*, vol. 6, no. 2, pp. 283–296, 2005.

[13] R. Nishimura, S. Hayashi, and M. Fukushima, "Robust Nash equilibria in N-person non-cooperative games: Uniqueness and reformulation," *Pac. J. Optim.*, vol. 5, no. 2, pp. 237–259, May 2009.

[14] A. Ben-Tal and A. Nemirovski, "Selected topics in robust convex optimization," *Math. Programm.*, vol. 112, no. 1, pp. 125–158, Jul. 2007.

[15] J. Wang, G. Scutari, and D. P. Palomar, "Robust MIMO cognitive radio via game theory," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1183–1201, Mar. 2011.

[16] P. Setoodeh and S. Haykin, "Robust transmit power control for cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 915–939, May 2009.

[17] A. B. Gershman and N. D. Sidiropoulos, *Space–Time Processing for MIMO Communications*. Hoboken, NJ, USA: Wiley, 2005.

[18] R. H. Gohary and T. J. Willink, "Robust IWFA for open-spectrum communications," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4964–4970, Dec. 2009.

[19] A. J. G. Anandkumar, A. Anandkumar, S. Lambotharan, and J. A. Chambers, "Robust rate-maximization game under bounded channel uncertainty," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4471–4486, Nov. 2011.

[20] H. Tembine, "Dynamic robust games in MIMO systems," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 990–1002, Aug. 2011.

[21] H. Chen and D. Yao, *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*. New York, NY, USA: Springer-Verlag, 2001.

[22] S. Parsaeefard and A. R. Sharafat, "Robust worst-case interference control in underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 8, pp. 3731–3745, Oct. 2012.

[23] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Found. Trends Netw.*, vol. 2, no. 4, pp. 381–533, Jul. 2008.

[24] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.

[25] E. A. Gharavol, Y.-C. Liang, and K. Mouthaan, "Robust downlink beamforming in multiuser MISO cognitive radio networks with imperfect channel-state information," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2852–2860, Jul. 2010.

[26] J.-S. Pang, G. Scutari, D. Palomar, and F. Facchinei, "Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3251–3271, Feb. 2010.

[27] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[28] Z.-Q. Luo and J.-S. Pang, "Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, p. 024012, Mar. 2006.

[29] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory in multiuser communication systems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 35–49, May 2010.

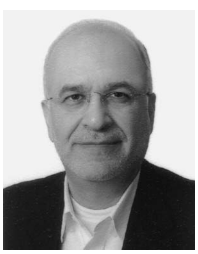
[30] F. Giannessi, A. Maugeri, and P. M. Pardalos, Eds., *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models (Nonconvex Optimization and Its Applications)*. Berlin, Germany: Springer-Verlag, 2002.

[31] M. Fukushima and G.-H. Lin, "Smoothing methods for mathematical programs with equilibrium constraints," in *Proc. 12th Int. Conf. Inf. Res. Develop. Knowl. Soc. Infrastruct.*, Jul. 2004, pp. 206–213.

[32] C. A. Floudas and P. M. Pardalos, Eds., *Encyclopedia of Optimization*. Norwell, MA, USA: Kluwer, 2001.
[33] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.
[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[35] W. A. Brock and S. N. Durlauf, "Local robustness analysis: Theory and application," *J. Econ. Dyn. Control*, vol. 29, no. 11, pp. 2067–2092, Nov. 2005.
[36] P. D. Miller, *Applied Asymptotic Analysis*. Providence, RI, USA: AMS, 2006.

Saeedeh Parsaeefard (S'09) received the B.Sc. and M.Sc. degrees from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2003 and 2006, respectively, and the Ph.D. degree in electrical and computer engineering from Tarbiat Modares University, Tehran, in 2012.

From November 2010 to October 2011, she was a Visiting Ph.D. Student with the Department of Electrical Engineering, University of California, Los Angeles, CA, USA. Her current research interests include robust resource allocation and applications of robust game theory in wireless networks.



Ahmad R. Sharafat (S'75–M'81–SM'94) received the B.Sc. degree from Sharif University of Technology, Tehran, Iran, in 1975 and the M.Sc. and Ph.D. degrees from Stanford University, Stanford, CA, USA, in 1976 and 1981, respectively, all in electrical engineering.

He is a Professor of electrical and computer engineering with Tarbiat Modares University, Tehran. His research interests include advanced signal processing techniques and communication systems and networks.

Dr. Sharafat is a Senior Member of Sigma Xi.

Mihaela van der Schaar (SM'04–F'10) received the M.Sc. and Ph.D. degrees from Eindhoven University of Technology, Eindhoven, The Netherlands.

She is a Chancellor's Professor of electrical engineering with the University of California, Los Angeles (UCLA), Los Angeles, CA, USA. She is also the Founding Director of the UCLA Center for Engineering Economics, Learning, and Networks (see netecon.ee.ucla.edu). She is the holder of 33 U.S. patents. Her research interests include engineering economics and game theory, strategic design, expert and social networks, online reputation and social media, dynamic multiuser networks and system designs, wireless networks, online learning, real-time stream mining, meta-learning, and multimedia networking and systems. (For more information about her research visit: http://medianetlab.ee.ucla.edu/.)

Dr. van der Schaar is an IEEE Fellow and was a Distinguished Lecturer of the Communications Society during 2011–2012, the Editor-in-Chief of the IEEE TRANSACTIONS ON MULTIMEDIA, and a member of the Editorial Board of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. She received a National Science Foundation Career Award (2004), the Best Paper Award from the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2005), the Okawa Foundation Award (2006), the IBM Faculty Award (2005, 2007, and 2008), the Most Cited Paper Award from EURASIP: *Image Communications Journal* (2006), the Gamenets Conference Best Paper Award (2011), and the 2011 IEEE Circuits and Systems Society Darlington Best Paper Award. She received three ISO awards for her contributions to the MPEG video compression and streaming international standardization activities.