# Self-organized Relay Selection for Cooperative Transmission in Vehicular Ad-hoc Networks

Daxin Tian, *Senior Member, IEEE,* Jianshan Zhou, Zhengguo Sheng, Min Chen, *Senior Member, IEEE,*
Qiang Ni, *Senior Member, IEEE,* and Victor C.M. Leung, *Fellow, IEEE*

*Abstract*—**Cooperation is a promising paradigm to improve spatial diversity in vehicular ad-hoc networks. In this paper, we pose a fundamental question: how the greediness and selfishness of individual nodes impact cooperation dynamics in vehicular ad-hoc networks. We map the self-interest-driven relay selection decision-making problem to an automata game formulation and present a non-cooperative game-theoretic analysis. We show that the relay selection game is an ordinal potential game. A decentralized self-organized relay selection algorithm is proposed based on a stochastic learning approach where each player evolves toward a strategic equilibrium state in the sense of Nash. Furthermore, we study the exact outage behavior of the multi-relay decode-and-forward cooperative communication network. Closed-form solutions are derived for the actual outage probability of this multi-relay system in both independent and identically distributed channels and generalized channels, which need not assume an asymptotic or high signal-to-noise ratio. Two tight approximations with low computational complexity are also developed for the lower bound of the outage probability. With the exact closed-form outage probability, we further develop an optimization model to determine optimal power allocations in the cooperative network, which can be combined with the decentralized learning-based relay selection. The analysis of the exact and approximative outage behaviors and the convergence properties of the proposed algorithm toward a Nash equilibrium state are verified theoretically and numerically. Simulation results are also given to demonstrate that the resulting cooperative network induced by the proposed algorithm achieves high energy efficiency, transmission reliability, and network-wide fairness performance.**

*Index Terms*—**Vehicular ad-hoc networks, cooperative communication, relay selection, outage probability, energy efficiency, noncooperative game.**

## I. Introduction

D. Tian, J. Zhou are with Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems & Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: dtian@buaa.edu.cn, jianshanzhou@foxmail.com)

Z. Sheng is with Department of Engineering and Design, the University of Sussex, Richmond 3A09, UK.(E-mail: z.sheng@sussex.ac.uk)

M. Chen is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China. (E-mail: minchen2012@hust.edu.cn)

Q. Ni is with School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK. (e-mail:q.ni@lancaster.ac.uk)

V. Leung is with Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, B.C., V6T 1Z4 Canada.(E-mail: vleung@ece.ubc.ca)

Recently, vehicular ad-hoc networks (VANETs) have attracted significant research interest because of their potential to leverage many significant telematic applications, including safety-oriented and entertainment applications, for traffic information and intelligent transportation systems [1], [2]. A generalized VANET, also called connected vehicles, comprises wireless communication links among vehicles and road-side access points (APs), and supports both vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications [3], [4]. However, due to vehicular mobility, multi-path propagation, and other time-varying effects, there exists inherent space time frequency variability in vehicular channels [5]–[7]. Recent interest in the development of a novel class of communication architectures to fully exploit *diversity* has increased. *Cooperative communications* have been widely acknowledged in various communities as a powerful technique for performance improvement of wireless relay transmission systems from physical-layer or cross-layer perspectives [8]–[10]. Emerging cooperative communication technologies suggest broader designs and protocols that can, to some extent, relax the fast signal fading problem [11].

Generally, cooperative wireless networks have a set of communication terminals, each having a single antenna and its own information to transmit. These terminals are spatially distributed and expected to cooperatively share antennas based on the typical relay channel model [12], so as to create a virtual antenna array at the physical layer, which can achieve spatial diversity (also called cooperative diversity) and coding gains to mitigate fast channel fading [13]. Cooperative communication systems can provide spatial diversity and performance enhancement similar to that provided by multi-antenna systems, such as multiple-input multiple-output (MIMO) systems [14]–[16]. In addition, such systems are more practical to be deployed with existing distributed hardware and limited resources (e.g., on-board battery and data buffer). Thus, cooperative communications can motivate novel network solutions to boost VANET performance in terms of certain network metrics such as outage probability, network capacity, and energy efficiency [8]–[11].

Despite many successful application reports about VANETs integrating cooperative communications, there still exist some fundamental issues that previous studies have not explicitly considered and that need to be reconsidered and fully explored from a more realistic view point. These basic issues primarily include the following. i) If nodes are assumed to be selfish and unwilling to cooperate with other transmitters without any (noticeable) benefit, can cooperative communications still

be applied? ii) If so, how should this be mapped into the mathematical design of an appropriate relay selection method? iii) How will this impact the selection dynamics of cooperative relays in a network? iv) Is there a desired solution state whereby the resulting cooperative network can achieve high performance in terms of not only network metrics but also fair resource allocation (i.e., benefit equilibrium)? If such a cooperative network exists, how should such a desired solution state be represented mathematically? v) What constitutes an effective algorithm that enables nodes to make cooperation decisions that depend solely on their historical information and local network states, such that nodes can learn dynamically and co-evolve asymptotically to the desired performance state, i.e., in a self-organized and distributed manner?

To this end, in this study, we investigate the fundamental questions raised above, focusing on the decentralized learning-based self-organized relay selection (DLbSo-RS) mechanism to leverage a cooperative VANET. We explicitly consider the selection of multiple relay nodes. Because the decode-and-forward (DF) cooperative communication protocol has good scalability in terms of practical implementation and can provide full second-order diversity [8], [10], [13], we exploit the DF protocol with multiple relays helping some transmitters in a VANET. While a cooperative vehicular network has various performance metrics, we consider terminal energy consumption and transmission reliability. Transmission reliability is characterized by a QoS-related transmission rate and an outage probability. More importantly, we present a powerful formulation, based on the game theory of learning automata [17], that facilitates mapping the cooperation problem with respect to resource utilization and interest equilibrium in vehicular networks into a decentralized learning-based decision making framework.

The remainder of this paper is organized as follows. In Section II, our cooperative vehicular network model is presented. In Section III, we develop the game-theoretic framework and propose a decentralized learning-based adaptation for the relay selection decision. We also derive key theoretical properties of the proposed algorithm under the game-theoretic framework. In Section IV, we derive our exact theoretical expression of the outage probability of a multi-relay DF cooperative scheme. Based on this, we propose an energy-efficiency optimization model and determine the optimal power allocation for the multi-relay DF cooperative transmission scheme, which is integrated with the proposed learning-based relay selection adaptation. In Section V, numerical results are presented to validate our theoretical developments and to evaluate the comprehensive performance of our method. Finally, conclusions are drawn in Section VI.

## II. SYSTEM MODEL

In this study, we consider VANET wherein a certain number of vehicles (mobile nodes) are assumed to be uniformly distributed in a given road region. Each vehicular communication terminal is equipped with a single antenna. We employ repetition-based cooperative diversity, i.e., the DF cooperative communication protocol [13], [18], such that
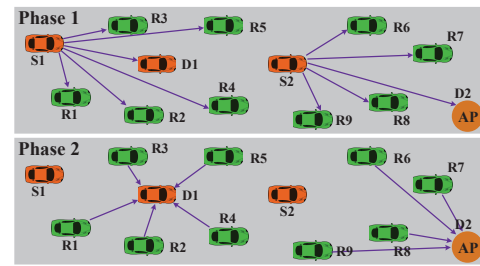


Fig. 1. Two phases of DF cooperative communication with multiple relays in a VANET. In the first phase, the source nodes (e.g., $S1$ and $S2$) broadcast their information to the corresponding relays (e.g., $\{R1, R2, \ldots, R9\}$) as well as to their destinations (e.g., $D1$ and $D2$). Then, these relays can re-transmit the information to the corresponding destinations on orthogonal sub-channels during the second phase.

vehicular transmitters can exploit spatial diversity to mitigate signal fading arising from the inherent variability of vehicular channels across time, frequency, and space. We consider that this network is associated with a set of source nodes, denoted by $\mathcal{I} = \{i | i = 1, 2, \ldots, n\}$, and a set of candidate relay nodes, denoted by $\mathcal{J} = \{j | j = 1, 2, \ldots, m\}$. Each source $i \in \mathcal{I}$ transmits its message to a corresponding destination node $d(i) \notin \mathcal{I}, \mathcal{J}$ with the assistance of some potential nodes in $\mathcal{J}$. The destination may be either a road-side AP or a vehicular node. There are $m_i$ candidate relay nodes in the neighborhood of the source $i$, denoted by the set $\mathcal{N}_i \subset \mathcal{J}$, i.e., $|\mathcal{N}_i| = m_i$. We also assume that any $j \in \mathcal{J}$ can be a neighbor of several source nodes. Let the set of $j$'s neighboring sources be $\mathcal{A}_j \subset \mathcal{I}$ and its cardinal $n_j$, i.e., $|\mathcal{A}_j| = n_j$. Thus, we have $j \in \bigcap_{i \in \mathcal{A}_j} \mathcal{N}_i$, $\bigcup_{i=1}^{n} \mathcal{N}_i = \mathcal{J}$ and $\bigcup_{j=1}^{m} \mathcal{A}_j = \mathcal{I}$. Any $j$ is allowed to independently determine which source to cooperate with, i.e., autonomously selecting a source from $\mathcal{A}_j$, denoted $s_j \in \mathcal{A}_j$, as its service target. Throughout this paper, we focus on multi-relay cooperative communications, and although there are algorithmic differences between single and multi-relay-based cooperations, multi-relay-based algorithms can be easily extended to single-relay-based cooperative communications. Thus, assuming that there are multiple relays (from $\mathcal{N}_i$) serving source $i$, we define the set of $i$'s actual relays by $\mathcal{R}_i$, i.e., $i \in \bigcap_{j \in \mathcal{R}_i} \mathcal{A}_j$ and $\mathcal{R}_i \subset \mathcal{N}_i$. We denote the cardinal of $\mathcal{R}_i$ as $N_i$ for any source $i \in \mathcal{I}$, i.e., $|\mathcal{R}_i| = N_i$. We adopt a time-division-multiple-access protocol with $N_i + 1$ time slots for the medium access control in the multi-relay DF cooperative communications to guarantee orthogonal transmissions [19]. In this way, each cooperative node transmits its decoded information in a fraction $1/(N_i + 1)$ of the total degrees of freedom in the channel, which ensures that nodes conform to the half-duplex constraint and avoid collisions.

In a typical scenario of vehicular cooperative communications such as Fig. 1, multi-relay DF cooperative communications operate during two transmission phases. The first phase consists of the first time slot, in which the source transmits a signal to its corresponding destination and to the cooperative nodes (its relays). The second phase consists of the following $N_i$ time slots, in which the relays repeat the message from their source in a pre-specified order [19]. Let $a_{i,d(i)}$, $a_{i,j}$, and $a_{j,d(i)}$ be the coefficients that capture the effects arising

from path loss, shadowing, and frequency non-selective fading on the $i$-to-$d(j)$, $i$-to-$j$, and $j$-to-$d(i)$ transmission channels, respectively. Generally, these channel coefficients $a_{Tx,Rx}$ can be statistically modeled as zero-mean, independent, circularly symmetric complex Gaussian random variables with variances $1/\lambda_{Tx,Rx}$, such that $|a_{Tx,Rx}|^2$ are exponentially distributed with parameter $\lambda_{Tx,Rx}$ [8], [10], [11]. Following the literature, we also assume that the quantities that capture the effects of noise and other interference at the receiver are modeled as zero-mean mutually independent, circularly symmetric, complex Gaussian random sequences with variance $\sigma_0^2$. We denote $p_{Tx}$ as the discrete-time transmission power in $Joule/2 - D$. Thus, the corresponding SNR can be represented as $SNR_{Tx} = p_{Tx}/\sigma_0^2$, so that the channel model is parameterized by such random SNR variables

$$SNR_{Tx}|a_{Tx,Rx}|^2 = \frac{p_{Tx}}{\sigma_0^2}|a_{Tx,Rx}|^2 \qquad (1)$$

for any $Tx$-to-$Rx$ link pair.

With the above channel model, the relay selection in multi-relay DF cooperative communications is equivalent to constructing a proper relay set $\mathcal{R}_i$ to help the source $i \in \mathcal{I}$ and to optimize transmission power allocation over the source $i$ and relay set $\mathcal{R}_i$ in terms of transmission link reliability or other network metrics. Without loss of generality, we divide continuous time into a series of discrete intervals indexed by $t \in \mathbb{Z}_{\geq 0}$. Each time interval consists of a sequence of time slots. Throughout this paper, the power level adopted by $i$ ($j$) during the time interval $t$ is denoted by $p_{i,t}$ ($p_{j,t}$), the residual energy of a node $i \in \mathcal{I}$ ($j \in \mathcal{J}$) at the end of the time interval $t$ is denoted as $E_{i,t}$ ($E_{j,t}$), and the maximum transmission power level at $i$ ($j$) is $p_{i,\max}$ ($p_{j,\max}$).

## III. Self-organized Relay Selection Framework

### A. Relay Selection Game Formulation

To map the relay selection into a multi-player non-cooperative strategic-form game, we consider that the selection decision is made from the perspective of the candidate relay node. That is, the candidate relay nodes in $\mathcal{J}$ are treated as players (i.e., decision makers or automata) and the residual energy of both sources and relays (after power allocation in each round of the game) is considered the external state. Note that players choose an appropriate source. The normal form of the game can be presented as a four tuple:

$$\mathcal{G} = \langle \mathcal{C}, \mathcal{J}, \mathcal{A}, \mathcal{U} \rangle \qquad (2)$$

where $\mathcal{C}$ denotes the space of the external states, i.e., $\mathcal{C} = \{E_{Tx,t}|Tx \in \mathcal{I} \cup \mathcal{J}, t \in \mathbb{Z}_{\geq 0}\}$. $\mathcal{J}$ is the player set where $j \in \mathcal{J}$ is the index of an independent player in the game $\mathcal{G}$. $\mathcal{A}$ represents the discrete finite-action space, which collects all individual actions, i.e., $\mathcal{A} = \times_{j=1}^{m} \mathcal{A}_j$. Similarly, the action profiles for all players except $j$ can be represented by the set $\mathcal{A}_{-j} = \times_{j' \neq j} \mathcal{A}_{j'}$. In addition, we denote an action profile with respect to $j$ as $\boldsymbol{s}_j = (s_j, \boldsymbol{s}_{-j})$ and $\boldsymbol{s}_j \in \mathcal{A}$, where $s_j$ is player $j$'s action (i.e., $s_j \in \mathcal{A}_j$) and $\boldsymbol{s}_{-j}$ denotes a vector collecting the actions taken by the other $m - 1$ players, i.e., $\boldsymbol{s}_{-j} \in \mathcal{A}_{-j}$. For any player $j \in \mathcal{J}$, we define a utility function

as $u_j : \mathcal{A} \to \mathbb{R}$, which captures $j$'s preferences over the action profiles $\mathcal{A}$. $\mathcal{U}$ denotes a vector that collects all individual utility functions, i.e., $\mathcal{U} = (u_1, u_2, \ldots, u_m) : \mathcal{A} \to \mathbb{R}^m$.

In vehicular interactions, to account for significant changes in the neighborhood structure, we consider that game $\mathcal{G}$ is played repeatedly, such that it can evolve over time $t$ and adapt to the decision-making behaviors. In each round of $\mathcal{G}$, which is also indexed by $t$, we denote the current action taken by player $j \in \mathcal{J}$ as $s_{j,t} \in \mathcal{A}_j$ and the current action profile as $\boldsymbol{s}_j(t) = (s_{j,t}, \boldsymbol{s}_{-j,t}) \in \mathcal{A}$, where $\boldsymbol{s}_{-j,t} \in \mathcal{A}_{-j}$. To facilitate game-theoretic mapping from the selection decision of self-interest-driven nodes in cooperative communication interactions to a proper decentralized learning-based decision-making formulation, we propose a utility function for each player $j \in \mathcal{J}$. In realistic scenarios, due to their selfish and greedy nature, each player tends to maximize its payoff in the game $\mathcal{G}$. The player's payoff specifies the benefit received from the resulting cooperative network and the cost incurred by cooperative transmission. Thus, the general form of each player's utility function can be formulated as the difference between the received benefit and the incurred cost, so that it captures the benefit-cost trade off and maps the player's action profile to a payoff. For any $j \in \mathcal{J}$, the associated utility function $u_j$ is given as follows:

$$u_j(\boldsymbol{s}_j(t)) = f_j(\boldsymbol{s}_j(t)) - g_j(\boldsymbol{s}_j(t)) \qquad (3)$$

where the function $f_j : \mathcal{A} \to \mathbb{R}$ denotes the benefit $j$ can gain when $\boldsymbol{s}_j(t)$ is deployed and $g_j : \mathcal{A} \to \mathbb{R}$ denotes the cost incurred by cooperative communication.

Naturally, the cost incurred by player $j$ in cooperative communication, $g_j$, can be formulated by capturing the power consumption level and the individual residual energy. To be specific, let $q_{j,t} = p_{j,t} + p_{s_{j,t},t}$ be the sum of the transmission power levels used by $j$ and $s_{j,t}$. The cost component $g_j$ is given as follows:

$$g_j(\boldsymbol{s}_j(t)) = \alpha_1 q_{j,t} \frac{Q_{j,0}}{Q_{j,t}} \qquad (4)$$

where $\alpha_1$ is a nonnegative weighting coefficient, and $Q_{j,t}$ is the sum of the residual energy of $j$ and $s_{j,t}$ at time $t$, i.e., $Q_{j,t} = E_{j,t} + E_{s_{j,t},t}$ (particularly, $Q_{j,0} = E_{j,0} + E_{s_{j,t},0}$). From (4), it can be seen that consuming more power, i.e., larger $q_{j,t}$, or having less energy remaining, i.e., a smaller $E_{j,t}$, leads to a higher $g_j$, indicating a higher cost incurred in cooperative communication.

On the other hand, to formulate the component $f_j$, we characterize the benefit perceived by player $j$ with the link reliability of the cooperative communication originating from $s_{j,t}$, the transmission power capacity, and the degree of energy utilization balance. A specific $f_j$ for each $j \in \mathcal{J}$ is given by

$$f_j(\boldsymbol{s}_j(t)) = H_j(\boldsymbol{s}_j(t))\left(\alpha_1 q_{j,\max} \frac{Q_{j,0}}{Q_{j,t}} + \alpha_2 \overline{W_j(s_{j,t})}\right) \quad (5)$$

where $\alpha_2$ is also a nonnegative weighting coefficient, and $q_{j,\max}$ is defined as the sum of the usable maximum power levels at $j$ and $s_{j,t}$, i.e., $q_{j,\max} = p_{j,\max} + p_{s_{j,t},\max}$, which indicate the transmission power capability of $j$ and $s_{j,t}$ (Note that a larger $p_{j,\max}$ ($p_{s_{j,t},\max}$) implies a higher transmission

potential of $j$ ($s_{j,t}$)). $H_j (s_j(t))$ indicates the transmission reliability of the resulting cooperative network when the action profile $s_j(t)$ is adopted, and $\overline{W_j (s_{j,t})}$ is the average of the residual energy of $j$'s local neighboring nodes.

More specifically, we introduce the indicator function

$$h_j (s_{j,t}) = \begin{cases} 1, & q_{j,t} \geq (w_{j,t} + w_{s_{j,t},t}) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

and then formulate $H_j (s_j(t))$ as follows:

$$H_j (s_j(t)) = \prod_{j \in \mathcal{J}} (h_j (s_{j,t})) \quad (7)$$

where $w_{j,t}$ and $w_{s_{j,t},t}$ denote the minimum transmission power level needed at $j$ and $s_{j,t}$, respectively, to establish and maintain reliable cooperative transmission. In this paper, we consider that a QoS requirement (cooperative like quality) is characterized by a tolerable maximum outage probability $\beta_{s_{j,t}}$ and the required transmission data rate (spectral efficiency) $r_{s_{j,t}}$ (in $bits/s/Hz$). Hence, there must exist an optimal allocation of transmission power levels among the transmitters $\{s_{j,t}\} \cup \mathcal{R}_{s_{j,t}}$ given that the transmission reliability requirement is satisfied. That is, $w_{s_{j,t},t}$ and $w_{j,t}$ are solved using the following optimization model:

$$\min_{\boldsymbol{p}(s_{j,t})} \quad p_{s_{j,t},t} + \sum_{j' \in \mathcal{R}_{s_{j,t}}} p_{j',t}$$

$$s.t. \quad \begin{cases} Pr \left( I_{s_{j,t},d(s_{j,t})} < r_{s_{j,t}} \right) \leq \beta_{s_{j,t}} \\ p_{s_{j,t},t} \leq p_{s_{j,t},\max} \text{ and } \forall p_{j',t} \leq p_{j',\max} \\ \boldsymbol{p} (s_{j,t}) = \{ p_{s_{j,t},t}, p_{j',t} | \forall j' \in \mathcal{R}_{s_{j,t}} \} \end{cases} \quad (8)$$

where $I_{s_{j,t},d(s_{j,t})}$ represents the maximum average mutual information between source $s_{j,t}$ and its destination $d (s_{j,t})$ in the multi-relay DF cooperative communication, and $Pr \left( I_{s_{j,t},d(s_{j,t})} < r_{s_{j,t}} \right)$ denotes the outage probability that this mutual information $I_{s_{j,t},d(s_{j,t})}$ falls below the QoS-oriented spectral efficiency $r_{s_{j,t}}$. Note that the channel average mutual information is a function of several factors including the coding protocol, the relay selection scheme for constructing the decoding set $\mathcal{R}_{s_{j,t}}$, and the fading coefficients of the channel [13], [18], [20]. $I_{s_{j,t},d(s_{j,t})}$ is also a random variable.

In addition, the average residual energy dynamics of the local neighboring nodes in the network are captured by the term $\overline{W_j (s_{j,t})}$, which is formulated as follows

$$\overline{W_j (s_{j,t})} = \frac{1}{|\mathcal{A}_j| + |\mathcal{N}_{s_{j,t}}|} \left( \sum_{i' \in \mathcal{A}_j} \frac{E_{i',t}}{E_{i',0}} + \sum_{j' \in \mathcal{N}_{s_{j,t}}} \frac{E_{j',t}}{E_{j',0}} \right) \quad (9)$$

By including $\overline{W_j (s_{j,t})}$ into the benefit component $f_j$ in (5), player $j$ is rewarded for improving the average energy utilization level. This means that this player is motivated to contribute to balanced energy utilization when establishing a cooperative network while maximizing its own payoff in the normal game induced by the selfish and greedy nature:

$$\mathcal{G} : \quad \max_{s_{j,t} \in \mathcal{A}_j} \{ u_j (s_{j,t}, \boldsymbol{s}_{-j,t}) \} \quad \forall j \in \mathcal{J} \quad (10)$$

In addition, from (4-9), we derive the following results to present the properties of mapping the benefit cost tradeoff to

an instantaneous payoff any player $j$ receives by playing game $\mathcal{G}$ with the individual utility function $u_j$ proposed in (3).

**Result 1:** The indicator function of the transmission reliability of the overall established network, $H_j (s_j(t))$, is a nondecreasing function with respect to the transmission power level $q_{j,t}$. That is, if $q_{j,t} \geq q'_{j,t}$ then $H_j (s_{j,t}, \boldsymbol{s}_{-j,t}) \geq H_j (s'_{j,t}, \boldsymbol{s}_{-j,t})$ for any $j \in \mathcal{J}$ where $s_{j,t}$ and $s'_{j,t}$ are two different actions, and $s_{j,t}, s'_{j,t} \in \mathcal{A}_j$.

*Proof:* It is easy to understand that a larger transmission power, which enhances the transmission signal, usually leads to better coverage and reduces the probability that a communication link is impaired or interrupted by signal interferences. In other words, a higher transmission power level $q_{j,t}$ can result in a lower outage probability of the transmission link. From the formulation of the indicator function (6), $h_j (s_{j,t}) = 1$ if and only if $q_{j,t}$ is greater than the minimum power level required to guarantee the transmission reliability; otherwise, $h_j (s_{j,t}) = 0$. If $q_{j,t} \geq q'_{j,t}$, $h_j (s_{j,t}) \geq h_j (s'_{j,t})$ holds, then

$$\begin{aligned} & H_j (s_{j,t}, \boldsymbol{s}_{-j,t}) \\ & = h_j (s_{j,t}) \prod_{j' \in \mathcal{J} - \{j\}} (h_{j'} (s_{j',t})) \\ & \geq h_j (s'_{j,t}) \prod_{j' \in \mathcal{J} - \{j\}} (h_{j'} (s_{j',t})) = H_j (s'_{j,t}, \boldsymbol{s}_{-j,t}) \end{aligned} \quad (11)$$

■

**Result 2:** For any player $j \in \mathcal{J}$, the overall reward it receives from the cooperative network can be positive, i.e., $u_j (s_j(t)) \geq 0$, if and only if this resulting network can maintain the QoS-constrained transmission reliability; i.e., $H_j (s_j(t)) = 1$ holds.

*Proof:* From (5) and (7), it can be easily found that if there exists at least one player $j$ adopting a certain action $s_{j,t}$ such that $q_{j,t}$ falls below the minimum power level required to guarantee the transmission reliability, i.e., $q_{j,t} < w_{j,t} + w_{s_{j,t},t}$, then $h_j (s_{j,t}) = 0$, thereby leading to the overall term $H_j (s_j(t)) = 0$. The overall payoff of $j$ is reduced to $u_j (s_j(t)) = -g_j (s_j(t)) < 0$. This means that a player cannot gain any positive benefit from an unreliable network, even though it pays the cost $g_j (s_j(t))$ in the game $\mathcal{G}$.

In contrast, when $H_j (s_j(t)) = 1$ is satisfied, the overall payoff of $j$ can be re-expressed as follows:

$$u_j (s_j(t)) = \alpha_1 \frac{Q_{j,0}}{Q_{j,t}} (q_{j,\max} - q_{j,t}) + \alpha_2 \overline{W_j (s_{j,t})} \quad (12)$$

Given $q_{j,\max} \geq q_{j,t}$, it is obvious that $u_j (s_j(t)) \geq 0$. Thus, in this situation, the player is expected to receive a positive payoff from the resulting reliable network. ■

**Result 3:** With $u_j (s_j(t))$, an action profile $s'_j(t) = (s'_{j,t}, \boldsymbol{s}'_{-j,t})$ is a NE of the game $\mathcal{G}$ if $u_j (s'_{j,t}, \boldsymbol{s}'_{-j,t}) \geq u_j (s_{j,t}, \boldsymbol{s}'_{-j,t})$ holds for $\forall j \in \mathcal{J}$ and $\forall s_{j,t} \in \mathcal{A}_j$.

*Proof:* This result is in accordance with the mathematical definition of a NE [21]. When no unilateral deviation in the action adopted by $j \in \mathcal{J}$ can be profitable for itself, the associated profile $s'_j(t) = (s'_{j,t}, \boldsymbol{s}'_{-j,t})$ is considered a NE. ■

## B. Game-Theoretic Analysis

Based on the game-theoretic mapping presented in the previous subsection, we provide a comprehensive analysis of the relay selection game $\mathcal{G}$. First, to study the optimal solution of $\mathcal{G}$ in the sense of Nash, we introduce two lemmas.

*Lemma 1:* If there exists a real-value function $U : \mathcal{A} \to \mathbb{R}$ such that for $\forall j \in \mathcal{J}$, $\forall \boldsymbol{s}_{-j,t} \in \mathcal{A}_{-j}$, and $s'_{j,t}, s''_{j,t} \in \mathcal{A}_j$, $U\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) - U\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) > 0$ if and only if $u_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) - u_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) > 0$ holds. $\mathcal{G}$ is OPG, and $U$ is the ordinal potential function (OPF) of $\mathcal{G}$ [22].

*Lemma 2:* If the action space of an OPG is compact then this OPG is guaranteed to possess at least a NE in pure strategies [22].

These lemmas indicate that an OPG requires variations in the payoffs of all players and in the potential function having the same increasing or decreasing direction. Next, we show that the constructed game $\mathcal{G}$ with (3) is an OPG.

*Theorem 1:* The game $\mathcal{G}$ proposed in (10), where each player's instantaneous-payoff mapping is specified by the individual utility function given in (3), is an OPG. An associated OPF can be formulated as follows:

$$U\left(\boldsymbol{s}_j(t)\right) = \sum_{j \in \mathcal{J}} H_j\left(\boldsymbol{s}_j(t)\right)\left(\alpha_1 q_{j,t}\frac{Q_{j,0}}{Q_{j,t}} + \alpha_2 \overline{W_j\left(s_{j,t}\right)}\right) \\ - \sum_{j \in \mathcal{J}} \alpha_1 q_{j,t}\frac{Q_{j,0}}{Q_{j,t}} \tag{13}$$

*Proof:* Suppose that for any player $j \in \mathcal{J}$, $s'_{j,t} \in \mathcal{A}_j$ and $s''_{j,t} \in \mathcal{A}_j$ are two different actions and their corresponding transmission power levels are $q'_{j,t}$ and $q''_{j,t}$, respectively. We can evaluate deviation in the individual utility function $u_j$ when this player unilaterally changes its action from $s'_{j,t}$ to $s''_{j,t}$ as follows:

$$\Delta u_j = \alpha_1 q_{j,\max}\frac{Q_{j,0}}{Q_{j,t}}\left(H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) - H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right)\right) \\ - \alpha_1 \frac{Q_{j,0}}{Q_{j,t}}\left(q'_{j,t} - q''_{j,t}\right) \\ + \alpha_2\left(\begin{array}{c} H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right)\overline{W_j\left(s'_{j,t}\right)} \\ - H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right)\overline{W_j\left(s''_{j,t}\right)} \end{array}\right) \tag{14}$$

Similarly, we also derive the deviation in the asserted function (13) as follows:

$$\Delta U = U\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) - U\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) \\ = \Delta u_j + \sum_{j' \in \mathcal{J}, j' \neq j}\left[H_{j'}\left(s'_{j,t}, \boldsymbol{s}_{j,t}\right) - H_{j'}\left(s''_{j,t}, \boldsymbol{s}_{j,t}\right)\right] \\ \times\left(\alpha_1 q_{j',\max}\frac{Q_{j',0}}{Q_{j',t}} + \alpha_2 \overline{W_{j'}\left(s_{j',t}\right)}\right) \tag{15}$$

Next, we show the properties of $\Delta u_j$ and $\Delta U$ under two possible situations: i) $q'_{j,t} \geq q''_{j,t}$ and ii) $q'_{j,t} < q''_{j,t}$. According to *Result 1*, the transmission reliability indicator function $H_j\left(s_{j,t}, \boldsymbol{s}_{-j,t}\right)$ is a non-decreasing function with respect to the transmission power level $q_{j,t}$. That is, when $q'_{j,t} \geq q''_{j,t}$,

we have $H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) \geq H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right)$. According to the proof of *Result 2*,

$$f_j\left(s_{j,t}, \boldsymbol{s}_{-j,t}\right) = \alpha_1 q_{j,\max}\frac{Q_{j,0}}{Q_{j,t}} + \alpha_2 \overline{W_j\left(s_{j,t}\right)} \\ \geq \alpha_1 q_{j,t}\frac{Q_{j,0}}{Q_{j,t}} = g_j\left(s_{j,t}, \boldsymbol{s}_{-j,t}\right) \tag{16}$$

Note that $H_j \in \{0, 1\}$. Therefore, it follows from (14) that

$$\Delta u_j \begin{cases} \geq 0, & \text{if } q'_{j,t} \geq q''_{j,t}, H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) > H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) \\ \geq 0, & \text{if } q'_{j,t} \geq q''_{j,t}, H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) = H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) \\ < 0, & \text{if } q'_{j,t} \geq q''_{j,t}, H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) = H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) \end{cases} \tag{17}$$

When $q'_{j,t} < q''_{j,t}$, i.e., $H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) \leq H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right)$, it follows from (14) that

$$\Delta u_j \begin{cases} \leq 0, & \text{if } q'_{j,t} < q''_{j,t}, H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) < H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) \\ \leq 0, & \text{if } q'_{j,t} < q''_{j,t}, H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) = H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) \\ > 0, & \text{if } q'_{j,t} < q''_{j,t}, H_j\left(s'_{j,t}, \boldsymbol{s}_{-j,t}\right) = H_j\left(s''_{j,t}, \boldsymbol{s}_{-j,t}\right) \end{cases} \tag{18}$$

Accordingly, the sign of the second term on the right side of (15) is the same as that of $\Delta u_j$ for the first case in both situations, i.e., (17) and (18). For the last two cases, (either situation (17) or (18)), because the transmission reliability profile $H_j$ remains unchanged, the second term on the right side of (15) is equal to zero, which further indicates that $\Delta U = \Delta u_j$. To summarize, for all cases of both (17) and (18), the variations in $u_j$ and $U$ due to player $j$'s unilateral deviation ($\Delta u_j$ and $\Delta U$) have the same sign, i.e., $\text{sgn}(\Delta u_j) = \text{sgn}(\Delta U)$. This implies that $U$ is an OPF and $\mathcal{G}$ is an OPG according to *Lemma 1*. ∎

*Theorem 2:* The game $\mathcal{G}$ in (10), where each player's instantaneous-payoff mapping is specified by the individual utility function given in (3), has at least a pure-strategy NE that coincides with a local maximizer of the OPF $U$.

*Proof:* This conclusion naturally follows *Theorem 1* and *Lemma 2*. ∎

## C. Learning-Based Relay Selection Adaptation

As shown in [22], the potential maximizers of an OPF form a subset of the NE of the corresponding OPG. This implies that to identify the NE of the OPG, we can solve for the maximizers of the associated OPF, i.e., $\max_{s_{j,t} \in \mathcal{A}_j, \forall j \in \mathcal{J}}\{U\left(s_{1,t}, s_{2,t}, \ldots, s_{m,t}\right)\}$. However, it is quite difficult or even impossible to solve the OPF over a network because the OPF is a global function that requires complete information about the global networking individuals. In real scenarios, the dynamic nature of the vehicular network leads to the time-varying topological structure of each player and the decision is made independently by each player in each play. The absence or impracticality of centralized control and infrastructure would render some existing algorithms that require complete network information (e.g., conventional numerical global optimization algorithms, and better response algorithms [22]) inapplicable. Thus, we propose a decentralized learning-based self-organized algorithm based on a learning automata approach (the linear reward-inaction approach [17]) for each player to learn its own optimal strategy with

incomplete information. This algorithm can adapt the decision-making behavior of each player and be applied for networking players to learn their NE strategies from their individual action-reward history and the residual energy states of local neighboring nodes in a distributed manner.

To present the development of the decentralized learning-based self-organized algorithm, we let a mixed strategy of any player $j \in \mathcal{J}$ $\boldsymbol{x}_j(t) = \left(x_{j,1}(t), x_{j,2}(t), \ldots, x_{j,|\mathcal{A}_j|}(t)\right)^{\mathrm{T}}$ be the selection probability vector over the action set $\mathcal{A}_j$, where $x_{j,s_j}(t)$ denotes the probability of player $j$ selecting an action $s_j \in \mathcal{A}_j$ at time $t$. Then, we update the selection probability as

$$x_{j,s_j}(t+1) =$$
$$\begin{cases} x_{j,s_j}(t) + \delta \widetilde{u}_j\left(s_{j,t}\right)\left(1 - x_{j,s_j}(t)\right), & \text{if } s_j = s_{j,t} \\ x_{j,s_j}(t) - \delta \widetilde{u}_j\left(s_{j,t}\right) x_{j,s_j}(t), & \text{otherwise} \end{cases} \quad (19)$$

for $\forall s_j \in \mathcal{A}_j$, where $\delta \in (0,1)$ is a learning rate that should be re-specified sufficiently small, and $\widetilde{u}_j\left(s_{j,t}\right)$ is an instantaneous reward that player $j$ receives at the previous time $t$ when it takes the action $s_{j,t}$, which is normalized in the interval $(0,1)$ and evaluated based on its utility function (3):

$$\widetilde{u}_j\left(s_{j,t}\right) = \frac{u_j\left(s_{j,t}, \boldsymbol{s}_{-j,t}\right) - u_{j,t}^{\mathrm{lower}}}{u_{j,t}^{\mathrm{upper}} - u_{j,t}^{\mathrm{lower}}} \quad (20)$$

where $u_{j,t}^{\mathrm{upper}}$ and $u_{j,t}^{\mathrm{lower}}$ are the upper and lower records of player $j$'s utility function up to time $t$, respectively:

$$\begin{cases} u_{j,t}^{\mathrm{upper}} = \max_{0 \le \tau \le t}\left\{u_j\left(s_{j,\tau}, \boldsymbol{s}_{-j,\tau}\right)\right\} \\ u_{j,t}^{\mathrm{lower}} = \min_{0 \le \tau \le t}\left\{u_j\left(s_{j,\tau}, \boldsymbol{s}_{-j,\tau}\right)\right\} \end{cases} \quad (21)$$

The proposed DLbSo-RS algorithm is given in **Algorithm 1**.

***Result 4:*** In the DLbSo-RS algorithm, the updates given by (22) and (23) do not change the non-negativity and the normalization of a probability vector $\boldsymbol{x}_j(t)$ $(\forall j \in \mathcal{J})$.

*Proof:* Let $x'_{j,s_j}(t+1)$ denote the update in the right term of (22) for $\forall s_j \in \mathcal{A} - \mathcal{A}_j^{out}$ and $x'_{j,s'_j}(t+1)$ denote the update in the right term of (23) for $\forall s'_j \in \mathcal{A}_j^{in}$. Clearly, $x'_{j,s_j}(t+1) \ge 0$ and $x'_{j,s'_j}(t+1) \ge 0$ because $x_{j,s_j}(t) \ge 0$ for $\forall s_j \in \mathcal{A}_j$.

Furthermore, to simplify the expression, let $\mathcal{A}'_j = \mathcal{A}_j - \mathcal{A}_j^{out}$ and $\mathcal{A}''_j = \mathcal{A}_j - \mathcal{A}_j^{out} + \mathcal{A}_j^{in}$. It is evident that $|\mathcal{A}''_j| = |\mathcal{A}'_j| + |\mathcal{A}_j^{in}|$. For $\mathcal{A}'_j$ and $\mathcal{A}''_j$, the equation

$$\sum_{s_j \in \mathcal{A}''_j} x_{j,s_j}(t+1) = \sum_{s_j \in \mathcal{A}'_j} x'_{j,s_j}(t+1) + \sum_{s'_j \in \mathcal{A}_j^{in}} x'_{j,s'_j}(t+1)$$
$$= \frac{|\mathcal{A}'_j|}{|\mathcal{A}''_j|} \frac{\sum_{s_j \in \mathcal{A}'_j} x_{j,s_j}(t+1)}{\sum_{s \in \mathcal{A}'_j} x_{j,s}(t+1)} + \frac{\sum_{s'_j \in \mathcal{A}_j^{in}} 1}{|\mathcal{A}''_j|}$$
$$= \frac{|\mathcal{A}'_j| + |\mathcal{A}_j^{in}|}{|\mathcal{A}''_j|} = 1 = \sum_{s_j \in \mathcal{A}_j} x_{j,s_j}(t+1)$$
$$(24)$$

holds, which yields Result 4. ∎

From the linear reward-inaction-based learning approach in ***Algorithm 1***, it can be seen that a larger reward can result in a higher selection probability in the next strategic update. Note that when the selection game is played repeatedly, the

---

**Algorithm 1** DLbSo-RS

1: **Initialization**: Set the time period index as $t = 0$, and let each transmitter $Tx \in \mathcal{I} \cup \mathcal{J}$ discover its neighborhood, i.e., setting $\mathcal{N}_{Tx}$ for $\forall Tx \in \mathcal{I}$ and $\mathcal{A}_{Tx}$ for $\forall Tx \in \mathcal{J}$. Then, set the selection probability vector as $x_{j,s_j} = \frac{1}{|\mathcal{A}_j|}$, for $\forall j \in \mathcal{J}$ and $s_j \in \mathcal{A}_j$.

2: **Adaptation**: At every $t$,
   1) each player $j \in \mathcal{J}$ selects an action $s_{j,t}$ based on the selection probability profile $\boldsymbol{x}_j(t)$;
   2) according to the selections made by the players, every source $s_{j,t} \in \mathcal{A}_j$ can form a relay set $\mathcal{R}_{s_{j,t}}$ consisting of the players selecting the same action $s_{j,t}$;
   3) based on the optimization model (8), every source $s_{j,t}$ can solve for the optimal transmission power allocation $\left\{w_{s_{j,t},t}, w_{j',t} | j' \in \mathcal{R}_{s_{j,t}}\right\}$ for $\{s_{j,t}\} \cup \mathcal{R}_{s_{j,t}}$;
   4) each source $s_{j,t}$ and its relays $j' \in \mathcal{R}_{s_{j,t}}$ use the optimal transmission power levels $q_{s_{j,t},t} = w_{s_{j,t},t}$ and $q_{j',t} = w_{j',t}$, respectively, to perform the multi-relay DF cooperative communication;
   5) each player $j$ receives the instantaneous reward $u_j\left(s_{j,t}, \boldsymbol{s}_{-j,t}\right)$ specified by (3) and then evaluates the reward normalization $\widetilde{u}_j\left(s_{j,t}\right)$ based on (20);
   6) each player $j$ updates its selection probability vector by (19), i.e., deriving a new $\boldsymbol{x}_j(t+1)$.

3: **Update**: Each transmitter updates its own residual energy $E_{Tx,t}$ as well as the set of its local neighboring nodes, i.e., updating $\mathcal{N}_i$ for $\forall i \in \mathcal{I}$ and $\mathcal{A}_j$ for $\forall j \in \mathcal{J}$. In addition, due to the dynamic topology of the network, the selection probability vector of each player $j$, i.e., $\boldsymbol{x}_j(t+1)$, should be further updated as follows (let $\mathcal{A}_j^{out}$ be the set of nodes that move out of player $j$'s neighborhood set $\mathcal{A}_j$, and let $\mathcal{A}_j^{in}$ be the set of the nodes moving into $\mathcal{A}_j$).
   1) for $\forall s_j \in \mathcal{A}_j - \mathcal{A}_j^{out}$, player $j$ updates

$$x_{j,s_j}(t+1) \leftarrow \frac{|\mathcal{A}_j - \mathcal{A}_j^{out}|}{|\mathcal{A}_j - \mathcal{A}_j^{out} + \mathcal{A}_j^{in}|}$$
$$\times \left(\frac{x_{j,s_j}(t+1)}{\sum_{s \in \mathcal{A}_j - \mathcal{A}_j^{out}} x_{j,s}(t+1)}\right) \quad (22)$$

   2) for $\forall s_j \in \mathcal{A}_j^{in}$, the player $j$ updates:

$$x_{j,s_j}(t+1) \leftarrow \frac{1}{|\mathcal{A}_j - \mathcal{A}_j^{out} + \mathcal{A}_j^{in}|} \quad (23)$$

Then, set $t \leftarrow t + 1$, $\mathcal{A}_j \leftarrow \mathcal{A}_j - \mathcal{A}_j^{out} + \mathcal{A}_j^{in}$, and go proceed to **Adaptation** to repeat the next round of the game.

---

players are treated as multiple automata, and the instantaneous reward is treated as a reinforcement signal to incentivize each automata to adapt its decision-making behavior independently and in a distributed manner.

***Theorem 3:*** The DLbSo-RS algorithm converges to a NE of the game $\mathcal{G}$ when the learning rate $\delta$ is sufficiently small.

*Proof:* Denote the mixed strategy profile over $\mathcal{J}$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2715328, IEEE Transactions on Vehicular Technology

7

by the selection probability matrix $\boldsymbol{X}(t)$, i.e., $\boldsymbol{X}(t) = (\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), \dots, \boldsymbol{x}_m(t)) \in [0,1]^{\sum_{j \in \mathcal{J}} |\mathcal{A}_j|}$, and denote a unit probability vector of the same dimension as that of $\boldsymbol{x}_j(t)$ by $\boldsymbol{e}_{j,s_j}(t)$ for each $j \in \mathcal{J}$, where the $s_j$-th equals 1 and other elements equal zero. The mixed strategy profile $\boldsymbol{X}(t)$ can also be expressed as $\boldsymbol{X}(t) = (\boldsymbol{x}_j(t), \boldsymbol{X}_{-j}(t))$, where $\boldsymbol{X}_{-j}(t)$ is the mixed strategy profile over $\mathcal{J} - \{j\}$, i.e., $\boldsymbol{X}_{-j}(t) \in [0,1]^{\sum_{j' \in \mathcal{J}, j' \neq j} |\mathcal{A}_{j'}|}$. In addition, let $\phi_j(\boldsymbol{X}(t))$ be the expectation of the instantaneous reward received by the player $j$, i.e., $\phi_j(\boldsymbol{X}(t)) = \mathbf{E}_{\boldsymbol{X}(t)}[u_j]$, and let $\Phi(\boldsymbol{X}(t)) = \mathbf{E}_{\boldsymbol{X}(t)}[U]$ be the expected OPF of the game $\mathcal{G}$. When the learning rate $\delta$ is sufficiently small, it follows previous analysis [17] by which the linear reward-action learning approach in (19) can be represented by an ordinary differential equation (ODE) system (note that the normalization of the probability vector always holds, i.e., $\sum_{s_j' \in \mathcal{A}_j} x_{j,s_j'}(t) = 1$ for $t \in \mathbb{Z}_{\geq 0}$). In other words, according to [17], the selection probability $x_{j,s_j}(t)$ converges to the solution of the following ODE:

$$
\begin{aligned}
&\frac{\mathrm{d}x_{j,s_j}(t)}{\mathrm{d}t} \\
&= x_{j,s_j}(t) \sum_{s_j' \in \mathcal{A}_j} x_{j,s_j'}(t) \begin{pmatrix} \phi_j\left(\boldsymbol{e}_{j,s_j}(t), \boldsymbol{X}_{-j}(t)\right) \\ - \phi_j\left(\boldsymbol{e}_{j,s_j'}(t), \boldsymbol{X}_{-j}(t)\right) \end{pmatrix}
\end{aligned} \tag{25}
$$

for $\forall j \in \mathcal{A}$ and $\forall s_j \in \mathcal{A}_j$. Note that $x_{j,s_j}(t)$ is treated as a continuous-time extension of the discrete-time probability $x_{j,s_j}(t)$ given in (19).

We have $\Phi(\boldsymbol{X}(t)) = \sum_{s_j \in \mathcal{A}_j} x_{j,s_j}(t) \Phi\left(\boldsymbol{e}_{j,s_j}(t), \boldsymbol{X}_{-j}(t)\right)$, whose partial differential with respect to $\boldsymbol{x}_{j,s_j}(t)$ is $\partial \Phi(\boldsymbol{X}(t))/\partial x_{j,s_j}(t) = \Phi\left(\boldsymbol{e}_{j,s_j}(t), \boldsymbol{X}_{-j}(t)\right)$. Here, to simplify the expressions, let $\Phi_{j,s_j} = \Phi\left(\boldsymbol{e}_{j,s_j}(t), \boldsymbol{X}_{-j}(t)\right)$ and $\Delta\phi_{j,s_j} = \phi_j\left(\boldsymbol{e}_{j,s_j}(t), \boldsymbol{X}_{-j}(t)\right) - \phi_j\left(\boldsymbol{e}_{j,s_j'}(t), \boldsymbol{X}_{-j}(t)\right)$. Thus, we obtain the derivative of $\Phi(\boldsymbol{X}(t))$ with respect to $t$ as follows:

$$
\begin{aligned}
\frac{\mathrm{d}\Phi(\boldsymbol{X}(t))}{\mathrm{d}t} &= \sum_{j \in \mathcal{J}} \sum_{s_j \in \mathcal{A}_j} \frac{\partial \Phi(\boldsymbol{X}(t))}{\partial x_{j,s_j}(t)} \frac{\mathrm{d}x_{j,s_j}(t)}{\mathrm{d}t} \\
&= \sum_{j \in \mathcal{J}} \sum_{s_j \in \mathcal{A}_j} \sum_{s_j' \in \mathcal{A}_j} x_{j,s_j}(t) x_{j,s_j'}(t) \Phi_{j,s_j} \Delta\phi_{j,s_j}
\end{aligned} \tag{26}
$$

Similarly, let $\Delta\Phi_{j,s_j} = \Phi_{j,s_j} - \Phi_{j,s_j'}$ and $\Delta\phi_{j,s_j'} = \phi_j\left(\boldsymbol{e}_{j,s_j'}(t), \boldsymbol{X}_{-j}(t)\right) - \phi_j\left(\boldsymbol{e}_{j,s_j}(t), \boldsymbol{X}_{-j}(t)\right) = -\Delta\phi_{j,s_j}$. The equation (26) above can be rearranged as

$$
\begin{aligned}
\frac{\mathrm{d}\Phi(\boldsymbol{X}(t))}{\mathrm{d}t} &= \sum_{j \in \mathcal{J}} \sum_{s_j \in \mathcal{A}_j} \sum_{s_j' \in \mathcal{A}_j} x_{j,s_j}(t) x_{j,s_j'}(t) \Delta\Phi_{j,s_j} \Delta\phi_{j,s_j} \\
&\quad - \sum_{j \in \mathcal{J}} \sum_{s_j \in \mathcal{A}_j} \sum_{s_j' \in \mathcal{A}_j} x_{j,s_j}(t) x_{j,s_j'}(t) \Phi_{j,s_j'} \Delta\phi_{j,s_j'}
\end{aligned} \tag{27}
$$

Combining (26) and (27) can further yield

$$
\begin{aligned}
&\frac{\mathrm{d}\Phi(\boldsymbol{X}(t))}{\mathrm{d}t} \\
&= \frac{1}{2} \sum_{j \in \mathcal{J}} \sum_{s_j \in \mathcal{A}_j} \sum_{s_j' \in \mathcal{A}_j} x_{j,s_j}(t) x_{j,s_j'}(t) \Delta\Phi_{j,s_j} \Delta\phi_{j,s_j} \geq 0
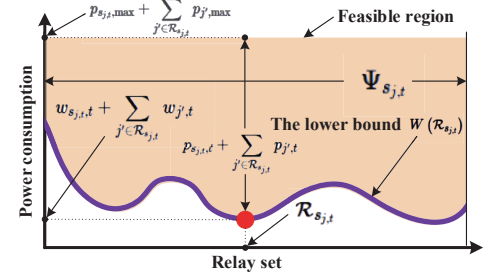\end{aligned} \tag{28}
$$



Fig. 2. A sketch of the power consumption bound curve.

where the nonnegativity always holds because, from the properties of the OPG given in **Theorem 1** and **Theorem 2**, the signs of $\Delta\Phi_{j,s_j}$ and $\Delta\phi_{j,s_j}$ are always the same, and the existence of a NE is guaranteed. $\Phi(\boldsymbol{X}(t))$ is shown to be nondecreasing along the phase trajectories $\boldsymbol{X}(t)$ of the ODE system. In addition, since $\boldsymbol{X}(t)$ is bounded, $\Phi(\boldsymbol{X}(t))$ is also bounded. Therefore, according to [17], the convergence of the DLbSo-RS algorithm to a NE of the game $\mathcal{G}$ is guaranteed. ∎

*Theorem 4:* The NE cooperative network resulting from the DLbSo-RS algorithm preserves transmission reliability and its overall power consumption is maintained at the lower bound of the feasible power solution region.

*Proof:* As in the DLbSo-RS algorithm, the transmission power levels used by the transmitters are set as $p_{s_{j,t},t} = w_{s_{j,t},t}$ and $p_{j,',t} = w_{j',t}$ for $\forall j' \in \mathcal{R}_{s_{j,t}}$, which can always satisfy the constraint on the outage probability $Pr\left(I_{s_{j,t},d(s_{j,t})} < r_{s_{j,t}}\right) \leq \beta_{s_{j,t}}$. This naturally follows the definition of the transmission reliability indicator function given in (6) in that $h_j(s_{j,t}) = 1$ always holds at each round of the game $\mathcal{G}$, such that the instantaneous reward each player receives is nonnegative, i.e., $u_j \geq 0$ as shown in **Result 2**. Thus, the resulting network in the NE state guarantees transmission reliability.

In the relay selection problem, the relay set $\mathcal{R}_{s_{j,t}}$ constructed at each round of the game $\mathcal{G}$ can be treated as a combination of the candidate relays of $s_{j,t}$. Let $\Psi_{s_{j,t}}$ be the set consisting of all possible combinations of the candidate relays in $s_{j,t}$'s neighborhood $\mathcal{N}_{s_{j,t}}$, such that $\mathcal{R}_{s_{j,t}} \in \Psi_{s_{j,t}}$. At this point, $\Psi_{s_{j,t}}$ indeed contains all feasible relay selection solutions and is finite. Recall that $\{w_{s_{j,t},t}\} \cup \{w_{j',t}, j' \in \mathcal{R}_{s_{j,t}}\}$ are solved from model (8), which are an optimal power allocation for $\{s_{s,t}\} \cup \mathcal{R}_{s_{j,t}}$. In other words, $w_{s_{j,t},t} + \sum_{j' \in \mathcal{R}_{s_{j,t}}} w_{j',t} \leq p_{s_{j,t},t} + \sum_{j' \in \mathcal{R}_{s_{j,t}}} p_{j',t}$ holds for a specific $\mathcal{R}_{s_{j,t}}$. Note that different relay sets correspond to different optimal power allocation solutions. Treating $\mathcal{R}_{s_{j,t}}$ as a variable ranging within $\Psi_{s_{j,t}}$, we can formulate a function of $\mathcal{R}_{s_{j,t}}$ that indicates the lower-bound of the overall power consumption in a multi-relay DF cooperative communication system where the source $s_{j,t}$ transmits to its destination $d(s_{j,t})$ with the help of the relays in $\mathcal{R}_{s_{j,t}}$ as $W(\mathcal{R}_{s_{j,t}}) = w_{s_{j,t},t} + \sum_{j' \in \mathcal{R}_{s_{j,t}}} w_{j',t}$. The power consumption bound curve is given in Fig.2. ∎

## IV. ENERGY-EFFICIENT OPTIMIZATION

As presented in the previous section, the transmission power allocation optimization plays a significant role in the game-theoretic relay selection method. In this section, we propose

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2715328, IEEE Transactions on Vehicular Technology

8

an outage probability analysis for the multi-relay DF cooperative communication system. A closed-form solution of the outage probability is first derived, and then the optimal power allocation model is developed.

### A. Mutual Information and Outage Behavior Analysis

Rather than introducing additional complications, we use the notation $i \in \mathcal{I}$ (Section II) to substitute the notation $s_{j,t}$ (Section III); the subscript $t$ has been omitted for brevity in the following analysis. Based on the proposed relay selection method, each source node $i$ can construct a relay set $\mathcal{R}_i$. Furthermore, if the SNR of the channel between $i$ and a relay $j \in \mathcal{R}_i$ is sufficiently large for this relay to successfully decode $i$'s information, then $j$ becomes an active cooperating relay that can further forward $i$'s information to the corresponding destination $d(i)$. Let $\mathcal{D}_i$ be the set of active cooperating relays from $\mathcal{R}_i$ (this $\mathcal{D}_i$ can also be called $i$'s decoding set [19]), i.e., $\mathcal{D}_i \subset \mathcal{R}_i$. Based on (1), the mutual information between $i$ and $j$ is expressed as [13], [19]

$$I_{i,j} = \frac{1}{1+N_i} \log_2 \left( 1 + \frac{p_i}{\sigma_0^2} |a_{i,j}|^2 \right) \quad (29)$$

and the mutual information of the multi-relay DF cooperative transmission is

$$I_{i,d(i)} = \frac{1}{1+N_i} \log_2 \left( 1 + \frac{p_i}{\sigma_0^2} |a_{i,d(i)}|^2 + \sum_{j\in\mathcal{D}_i} \frac{p_j}{\sigma_0^2} |a_{j,d(i)}|^2 \right) \quad (30)$$

Using the law of total probability, the outage probability in model (8) can be further presented as

$$Pr\left(I_{i,d(i)} < r_i\right) = \sum_{\mathcal{D}_i \subset \mathcal{R}_i} Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i\right) Pr\left(\mathcal{D}_i\right) \quad (31)$$

where $Pr\left(\mathcal{D}_i\right)$ is formulated as

$$Pr\left(\mathcal{D}_i\right) = \prod_{j\in\mathcal{D}_i} \left(1 - Pr\left(I_{i,j} < r_i\right)\right) \prod_{j'\in\mathcal{R}_i-\mathcal{D}_i} Pr\left(I_{i,j'} < r_i\right) \quad (32)$$

Recalling that $|a_{i,j}|^2 \sim \exp(\lambda_{i,j})$ for any $j \in \mathcal{R}_i$, we further derive

$$Pr\left(I_{i,j'} < r_i\right) = 1 - \exp\left\{ -\lambda_{i,j} \frac{\left(2^{r_i(1+N_i)} - 1\right)\sigma_0^2}{p_i} \right\} \quad (33)$$

which results in

$$Pr\left(\mathcal{D}_i\right) = \prod_{j\in\mathcal{D}_i} \exp\left\{ -c\frac{\lambda_{i,j}}{p_i} \right\} \times \prod_{j'\in\mathcal{R}_i-\mathcal{D}_i} \left( 1 - \exp\left\{ -c\frac{\lambda_{i,j'}}{p_i} \right\} \right) \quad (34)$$

where $c = \left(2^{r_i(1+N_i)} - 1\right)\sigma_0^2$.

To derive the formulation of $Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i\right)$, we introduce the following lemmas.

**Lemma 3:** If a random variable $X \sim \exp(\lambda)$, then $kX \sim \exp(\lambda/k)$.

**Lemma 4:** Given that $f_X(x)$ and $f_Y(y)$ are two probability density functions (pdfs) with respect to the two independent random variables $X$ and $Y$, respectively, the pdf with respect

to $Z = X + Y$ is the convolution of their separate density functions

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) \mathrm{d}x \quad (35)$$

**Lemma 5:** Given that $\{\lambda_k, k = 1, 2, \ldots, n\}$ are $n$ real numbers that satisfy $\lambda_k \neq \lambda_{k'}$ for $k \neq k'$, equation

$$\sum_{k=1}^{n} A(k)B(n,k) = 0 \quad (36)$$

always holds for $n \in \mathbb{Z}_{\geq 2}$ where $A(k)$ is defined as

$$A(k) = \begin{cases} 0, & k < 1 \text{ or } k > n \\ 1, & k = 1 \\ \frac{(-1)^{k-1}}{\prod_{j=1}^{k-1}(\lambda_k - \lambda_j)}, & 2 \leq k \leq n \end{cases} \quad (37)$$

and $B(n,k)$ is defined as

$$B(n,k) = \begin{cases} 0, & k < 1 \text{ or } k > n \\ 1, & k = n \\ \frac{1}{\prod_{i=k+1}^{n}(\lambda_i - \lambda_k)}, & 1 \leq k \leq n-1 \end{cases} \quad (38)$$

**Lemma 3** is indeed a basic property of the exponential distribution, and **Lemma 4** is a direct result of the pdf definition. More detailed proofs of these two lemmas can be referred to in most current monographs on probability theory and statistics. Here, we only detail the proof of **Lemma 5**.

*Proof:* Using mathematical induction, we prove **Lemma 5** via the following three steps.

1) when $n = 2$, $A(1) = 1$ and $B(2,2) = 1$ according to the definitions given in (37) and (38). Thus, the left side of equation (36) is reduced to

$$B(2,1) + A(2) = \frac{1}{\lambda_2 - \lambda_1} - \frac{1}{\lambda_2 - \lambda_1} = 0 \quad (39)$$

which always holds.

2) Suppose that equation (36) holds for $3 \leq n \leq m$ and $m \in \mathbb{Z}_{\geq 3}$, i.e., $\sum_{k=1}^{m} A(k)B(m,k) = 0$.

3) Following the above supposition 2), we show that (36) holds for $m+1$ as follows. To simplify the expressions, we first introduce $C(n,k)$ as follows:

$$C(n,k) = \begin{cases} 0, & k \geq n \text{ or } k < 1 \\ \frac{1}{\lambda_n - \lambda_k}, & 1 \leq k \leq n-1 \end{cases} \quad (40)$$

From (40),

$$B(m+1,k) = C(m+1,m) \left[ \begin{array}{l} B(m,k) \\ - C(m+1,k)B(m-1,k) \end{array} \right] \quad (41)$$

holds for $1 \leq k \leq m$. Based on (41), and noting that $B(m+1, m+1) = 1$, $B(m-1, m) = 0$ as given in (38), and $\sum_{k=1}^{m} A(k)B(m,k) = 0$ as given in supposition 2),

we obtain the following:

$$\sum_{k=1}^{m+1} A(k)B(m+1,k)$$

$$= A(m+1) + \sum_{k=1}^{m} A(k)B(m+1,k)$$

$$= A(m+1)$$

$$- C(m+1,m)\sum_{k=1}^{m-1} A(k)B(m-1,k)C(m+1,k)$$

$$(42)$$

Note that, from (38) and (40),

$$B(m-1,k)C(m+1,k)$$

$$= C(m+1,m-1)\begin{bmatrix} B(m-1,k) \\ -B(m-2,k)C(m+1,k) \end{bmatrix}$$

$$(43)$$

holds for $1 \le k \le m-1$.

Substituting (43) into (42) and then using the supposition $\sum_{k=1}^{m-1} A(k)B(m-1,k) = 0$ obtains:

$$\sum_{k=1}^{m+1} A(k)B(m+1,k)$$

$$= A(m+1)$$

$$+ (-1)^2 C(m+1,m)C(m+1,m-1)$$

$$\times \sum_{k=1}^{m-2} A(k)B(m-2,k)C(m+1,k)$$

$$(44)$$

Similar to (43) and (44), we perform the algebraic transformation $m-1$ times and derive the following:

$$\sum_{k=1}^{m+1} A(k)B(m+1,k)$$

$$= A(m+1)$$

$$+ (-1)^{m-1} \prod_{l=1}^{m-1} C(m+1,m-l+1)$$

$$\times C(m+1,1)$$

$$= A(m+1) + (-1)^{m-1} \prod_{l=1}^{m} C(m+1,l) = 0$$

$$(45)$$

where the last equation always holds because $A(m+1) = (-1)^m \prod_{l=1}^{m} C(m+1,l)$. At this point, (36) always holds for $n = m+1$. Thus, the lemma is proven. ∎

**Theorem 5:** Suppose that $\{X_k, k = 1, 2, \ldots, n; n \ge 2\}$ are some mutually independent random variables that follow different exponential distributions, i.e., $X_k \sim \exp(\lambda_k)$ where $\lambda_k \ne \lambda_{k'}$ for $k \ne k'$. The pdf of the sum of these random variables, i.e., $Y_n = \sum_{k=1}^{n} X_k$, is as follows:

$$f_{Y_n}(y) = \begin{cases} C_n \sum_{k=1}^{n} A(k)B(n,k)\exp(-\lambda_k y), & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$(46)$$

where $C_n$ is defined as $C_n = (\prod_{k=1}^{n} \lambda_k)$.

*Proof:* We employ the law of mathematical induction to prove **Theorem 5**.

1) When $n = 2$, it follows via **Lemma 3** and **Lemma 4** that equation (46)

$$f_{Y_2}(y) = \int_0^y \lambda_1 \lambda_2 \exp\{-\lambda_2 y - (\lambda_1 - \lambda_2)x\}\,dx$$

$$= C_2 \begin{Bmatrix} A(1)B(2,1)\exp(-\lambda_1 y) \\ + A(2)B(2,2)\exp(-\lambda_2 y) \end{Bmatrix}$$

$$(47)$$

holds.

2) Suppose that (46) always holds for $3 \le n \le m$ and $m \in \mathbb{Z}_{\ge 3}$.

3) Let $Y_{m+1} = Y_m + X_{m+1}$ where $Y_m$ and the $(m+1)$-th random variable $X_{m+1}$ are mutually independent, and let $X_{m+1} \sim \exp(\lambda_{m+1})$ where $\lambda_{m+1} \ne \lambda_k$ for $k = 1, 2, \ldots, m$. Based on **Lemma 4** and the above supposition 2), we obtain

$$f_{Y_{m+1}}(y)$$

$$= \int_0^y f_{X_{m+1}}(x)f_{Y_m}(y-x)\,dx$$

$$= \int_0^y \begin{Bmatrix} \lambda_{m+1}\exp(-\lambda_{m+1}x) \\ \times C_m \sum_{k=1}^{m} A(k)B(m,k)\exp(-\lambda_k(y-x)) \end{Bmatrix}\,dx$$

$$(48)$$

Note that

$$C_{m+1} = \lambda_{m+1}C_m \qquad (49)$$

and

$$\int_0^y \exp(-\lambda_{m+1}x)\exp(-\lambda_k(y-x))\,dx$$

$$= \frac{\exp(-\lambda_k y) - \exp(-\lambda_{m+1}y)}{\lambda_{m+1} - \lambda_k}$$

$$= C(m+1,k)\left[\exp(-\lambda_k y) - \exp(-\lambda_{m+1}y)\right]$$

$$(50)$$

Substituting (49) and (50) into (48) gets

$$f_{Y_{m+1}}(y)$$

$$= C_{m+1} \sum_{k=1}^{m} A(k)B(m,k)C(m+1,k)\exp(-\lambda_k y)$$

$$- C_{m+1}\exp(-\lambda_{m+1}y) \sum_{k=1}^{m} A(k)B(m,k)C(m+1,k)$$

$$(51)$$

According to $B(m,k)C(m+1,k) = B(m+1,k)$ and using **Lemma 5**, we obtain $\sum_{k=1}^{m+1} A(k)B(m+1,k) = 0$, which is equivalent to $A(m+1)B(m+1,m+1) = -\sum_{k=1}^{m} A(k)B(m+1,k)$. Substituting this result into (51) further derives

$$f_{Y_{m+1}}(y)$$

$$= C_{m+1} \sum_{k=1}^{m+1} A(k)B(m+1,k)\exp(-\lambda_k y).$$

$$(52)$$

Therefore, (46) also holds for $n = m+1$, which proves **Theorem 5** ∎

Next, we derive the exact closed-form expression of the conditional outage probability in (31), $Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i\right)$, with the following theorem.

***Theorem 6:*** The outage probability conditional on the active cooperating set $\mathcal{D}_i$, $Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i\right)$, is given as

$$
\begin{aligned}
&Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i\right) \\
&= C_{|\mathcal{D}_i|+1} \sum_{k=1}^{|\mathcal{D}_i|+1} \frac{A(k)B(|\mathcal{D}_i|+1,k)\left(1 - \exp(-\lambda_k c)\right)}{\lambda_k}
\end{aligned}
\tag{53}
$$

where $c = 2^{(1+N_i)r_i} - 1$, and let

$$
\lambda_k = \begin{cases} \frac{\lambda_{i,d(i)}\sigma_0^2}{p_i}, & k = 1 \\ \frac{\lambda_{j,d(i)}\sigma_0^2}{p_j}, & k = j+1; j = 1,2,\ldots,|\mathcal{D}_i| \end{cases}
\tag{54}
$$

and $\lambda_k \neq \lambda_k'$ for $k \neq k'$.

***Proof:*** According to ***Lemma 3***, $(p_{Tx}|a_{Tx,d(i)}|^2)/\sigma_0^2 \sim \exp\left\{(\lambda_{Tx,d(i)}\sigma_0^2)/p_{Tx}\right\}$, given $|a_{Tx,d(i)}|^2 \sim \exp(\lambda_{Tx,d(i)})$ for $Tx = i$ and $Tx = j \in \mathcal{D}_i$. For simplification, let

$$
X_k = \begin{cases} \frac{p_i|a_{i,d(i)}|^2}{\sigma_0^2}, & k = 1 \\ \frac{p_j|a_{j,d(i)}|^2}{\sigma_0^2}, & k = j+1; j = 1,2,\ldots,|\mathcal{D}_i| \end{cases}
\tag{55}
$$

Using the notations in (54), the random variables $X_k \sim \exp(\lambda_k)$ are mutually independent and non-identically distributed for $k = 1,2,\ldots,|\mathcal{D}_i|+1$. Let $Y_{|\mathcal{D}_i|+1} = \sum_{k=1}^{|\mathcal{D}_i|+1} X_k$. Hence, the conditional outage probability $Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i\right)$ is equivalent to $Pr\left(Y_{|\mathcal{D}_i|+1} < c | \mathcal{D}_i\right)$, which can be derived based on ***Theorem 5***:

$$
Pr\left(Y_{|\mathcal{D}_i|+1} < c | \mathcal{D}_i\right) = \int_0^c f_{Y_{|\mathcal{D}_i|+1}}(y)\mathrm{d}y
\tag{56}
$$

The integral equation (56) directly leads to (53). At this point, the proof of ***Theorem 6*** is completed. ∎

Combining (34) and (53) into (31) can give the outage probability $Pr\left(I_{i,d(i)} < r_i\right)$. Note that the derivation of the outage probability proposed here does not depend on the assumption, as usually adopted in the current literature, that the cooperative transmissions are performed in high $SNR_{Tx}$ regimes (i.e., with consideration of the asymptotical behavior of $SNR_{Tx}$: $SNR_{Tx} \to +\infty$). Moreover, we do not assume the identical transmission power levels. The transmission power of the transmitters, $p_{Tx}$, can differ. Hence, the exact and general closed-form formulation of the outage probability (31) generalizes the evaluation of the multi-relay DF cooperative communication in regimes where $SNR_{Tx}$ is usually non-asymptotic or finite.

On the other hand, the computation of the outage probability in (31) may be very complex because the summation in (31) is carried out over all possible $\mathcal{D}_i$. Recall $\mathcal{D}_i \subset \mathcal{R}_i$ and $|\mathcal{R}_i| = N_i$. In total, there are $2^{N_i}$ different subsets of $\mathcal{R}_i$, which implies that the summation in (31) must evaluate $2^{N_i}$ items. Therefore, we consider some specific conditions that are practical in actual applications and further specify the corresponding closed-form outage probability or derive the tight approximation, which makes the computation more tractable.

***Lemma 6:*** Given that $\{X_k, k = 1, 2, \ldots, n\}$ are some random variables independently following an identical exponential distribution with parameter $\lambda$, i.e., $X_k \sim \exp(\lambda)$ for $\forall k$, the sum of these random variables, $Z_n = \sum_{k=1}^n X_k$, follows an Erlang distribution with the shape parameter $n$ and the rate $\lambda$, i.e., $Z_n \sim \mathrm{Erlang}(n, \lambda)$, whose pdf is given as

$$
f_{Z_n}(z) = \begin{cases} \frac{\lambda^n z^{n-1}\exp(-\lambda z)}{(n-1)!}, & z > 0 \\ 0, & \text{otherwise} \end{cases}
\tag{57}
$$

***Proof:*** Under the given condition, we can perform the convolution calculation in (35) $(n-1)$ times directly, which results in ***Lemma 6***. ∎

***Theorem 7:*** Given that the channel fading parameters $a_{i,j}$ from source $i$ to any cooperative relay $j \in \mathcal{R}_i$ are mutually independent and identically distributed, i.e., $\lambda_{i,j} = \lambda_{i,j'}$ for $j \neq j'$; $j, j' \in \mathcal{R}_i$, and the transmission power level used by $i$, $p_i$, and by $j$, $p_j$, can make the fading parameters from $i$ and $j$ to the destination $d(i)$ satisfy $\left(\lambda_{i,d(i)}\sigma_0^2\right)/p_i = \left(\lambda_{j,d(i)}\sigma_0^2\right)/p_j$ for $\forall j \in \mathcal{R}_i$, the outage probability (31) is given as

$$
\begin{aligned}
&Pr\left(I_{i,d(i)} < r_i\right) \\
&= \sum_{k=0}^{N_i} \left\{ \begin{aligned} &\left[ 1 - \sum_{l=0}^k \left( \frac{(c\lambda)^l \exp(-\lambda c)}{l!} \right) \right] \\ &\times \binom{N_i}{k}\left(\exp(-\lambda' c)\right)^k \left(1 - \exp(-\lambda' c)\right)^{N_i-k} \end{aligned} \right\}
\end{aligned}
\tag{58}
$$

where $c = 2^{(1+N_i)r_i} - 1$, $\lambda' = \left(\lambda_{i,j}\sigma_0^2\right)/p_i$ for $\forall j \in \mathcal{R}_i$, and $\lambda = \left(\lambda_{i,d(i)}\sigma_0^2\right)/p_i = \left(\lambda_{j,d(i)}\sigma_0^2\right)/p_j$ for $\forall j \in \mathcal{R}_i$.

***Proof:*** Under the given condition that $\left(\lambda_{i,j}\sigma_0^2\right)/p_i$ are identical and $\left(\lambda_{i,d(i)}\sigma_0^2\right)/p_i = \left(\lambda_{j,d(i)}\sigma_0^2\right)/p_j$ for $\forall j \in \mathcal{R}_i$, the outage probability $Pr\left(I_{i,d(i)} < r_i\right)$ given in (31) can be rewritten as follows:

$$
\begin{aligned}
&Pr\left(I_{i,d(i)} < r_i\right) \\
&= \sum_{k=0}^{N_i} Pr\left(I_{i,d(i)} < r_i | |\mathcal{D}_i| = k\right) Pr\left(|\mathcal{D}_i| = k\right)
\end{aligned}
\tag{59}
$$

According to (32), we derive

$$
Pr(|\mathcal{D}_i| = k) = \binom{N_i}{k}\left(\exp(-\lambda' c)\right)^k \left(1 - \exp(-\lambda' c)\right)^{N_i-k}
\tag{60}
$$

where $c = 2^{(1+N_i)r_i} - 1$. In addition, referring to the notations in (55) and using ***Lemma 6***, we calculate the following:

$$
\begin{aligned}
&Pr(I_{i,d(i)} < r_i | |\mathcal{D}_i| = k) = Pr(\sum_{l=1}^{k+1} X_l < c | |\mathcal{D}_i| = k) \\
&= \int_0^c f_{Y_{k+1}}(y)\mathrm{d}y = 1 - \sum_{l=0}^k \frac{(c\lambda)^l \exp(-\lambda c)}{l!}
\end{aligned}
\tag{61}
$$

Combining (60) and (61) into (59) results in (58). ∎

To reduce the potential computational complexity, an alternative method is to adopt appropriate tight approximation on the computation of the outage probability.

***Lemma 7:*** Given that $\{X_k, k = 1, 2, \ldots, n\}$ are some mutually independent random variables, the probability of

$X_{\max} = \{X_k, k = 1, 2, \ldots, n\}$, $Pr\left(X_{\max} < x\right)$, is given as

$$Pr\left(X_{\max} < x\right) = \prod_{k=1}^{n} Pr\left(X_k < x\right) \qquad (62)$$

*Proof:* The probability $Pr\left(X_{\max} < x\right)$ is equivalent to $Pr\left(X_1 < x, X_2 < x, \ldots, X_n < x\right)$. As these random variables are assumed to be mutually independent, this leads to (62). ∎

Using the same notations given in **Theorem 6**, we derive two different lower bounds of the outage probability of the multi-relay DF cooperative communication in general situations where the channel fading parameters are mutually independent but distributed non-identically, i.e., $\lambda_k \neq \lambda'_k$ for $\forall k$.

*Theorem 8:* One lower bound of the outage probability $Pr\left(I_{i,d(i)} < r_i\right)$, i.e., $Pr_1^{lower}\left(I_{i,d(i)} < r_i\right)$, can be given as

$$
\begin{aligned}
&Pr_1^{lower}\left(I_{i,d(i)} < r_i\right) \\
&= \left[\prod_{j \in \mathcal{R}_i} \left(1 - \exp\left(-c\frac{\lambda_{i,j}\sigma_0^2}{p_i}\right)\right)\right] (1 - \exp(-c\lambda_1)) \\
&\quad + \left[1 - \prod_{j \in \mathcal{R}_i} \left(1 - \exp\left(-c\frac{\lambda_{i,j}\sigma_0^2}{p_i}\right)\right)\right] \\
&\quad \times \left(C_{N_i+1} \sum_{k=1}^{N_i+1} \frac{A(k)B(N_i+1, k)\left(1 - \exp(-c\lambda_k)\right)}{\lambda_k}\right)
\end{aligned}
$$

$$(63)$$

Alternatively, another lower bound $Pr_2^{lower}\left(I_{i,d(i)} < r_i\right)$ can be formulated as

$$
\begin{aligned}
&Pr_2^{lower}\left(I_{i,d(i)} < r_i\right) \\
&= \left[\prod_{j \in \mathcal{R}_i} \left(1 - \exp\left(-c\frac{\lambda_{i,j}\sigma_0^2}{p_i}\right)\right)\right] (1 - \exp(-c\lambda_1)) \\
&\quad + \left[1 - \prod_{j \in \mathcal{R}_i} \left(1 - \exp\left(-c\frac{\lambda_{i,j}\sigma_0^2}{p_i}\right)\right)\right] \\
&\quad \times \left[\prod_{k=1}^{N_i+1} \left(1 - \exp\left(-\frac{c\lambda_k}{N_i+1}\right)\right)\right]
\end{aligned}
$$

$$(64)$$

where $c = 2^{(1+N_i)r_i} - 1$. In addition, $Pr_1^{lower}\left(I_{i,d(i)} < r_i\right) \geq Pr_2^{lower}\left(I_{i,d(i)} < r_i\right)$ always holds.

*Proof:* According to (31), it holds that

$$
\begin{aligned}
&Pr\left(I_{i,d(i)} < r_i\right) \\
&\geq Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i = \emptyset\right) Pr\left(\mathcal{D}_i = \emptyset\right) \\
&\quad + Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i = \mathcal{R}_i\right) Pr\left(\mathcal{D}_i \neq \emptyset\right)
\end{aligned}
$$

$$(65)$$

Applying the results in **Theorem 6** and (34) to (65) can immediately obtain (63).

Besides, since

$$
\begin{aligned}
&Pr\left(I_{i,d(i)} < r_i | \mathcal{D}_i = \mathcal{R}_i\right) = Pr\left(\sum_{k=1}^{N_i+1} X_k < c \,\middle|\, \mathcal{D}_i = \mathcal{R}_i\right) \\
&\geq Pr\left((N_i + 1)X_{\max} < c | \mathcal{D}_i = \mathcal{R}_i\right)
\end{aligned}
$$

$$(66)$$

where the random variables $X_k$ are defined as given in **Theorem 6**, $X_{\max} = \max\{X_k, k = 1, \ldots, N_i + 1\}$.

For each $k$, $Pr\left(X_k < \frac{c}{N_i+1}\right) = 1 - \exp\left(-\frac{c\lambda_k}{N_i+1}\right)$. Thus, using **Lemma 7** yields the following:

$$
\begin{aligned}
&Pr\left((N_i + 1)X_{\max} < c | \mathcal{D}_i = \mathcal{R}_i\right) \\
&= \prod_{k=1}^{N_i+1} \left(1 - \exp\left(-\frac{c\lambda_k}{N_i+1}\right)\right)
\end{aligned}
$$

$$(67)$$

Finally, substituting the result of (66) and (67) into (65), and then, applying (34), we can obtain another lower bound as given in (64). According to (66), it is obvious that the first lower bound in (63) is tighter than the second in (64). ∎

### B. Optimum Power Allocation for Multi-relay DF Cooperative Communication

Here, we detail the optimum transmission power allocation strategy for the multi-relay DF cooperative communication based on model (8) and the outage probability formulations presented in the previous subsection. Considering the practical computation of the outage probability, we focus on the alternative modeling approaches presented in **Theorem 7** and **Theorem 8**.

Following the conditions specified in **Theorem 7**, we denote $\lambda' = \lambda\left(\lambda_{i,j}/\lambda_{i,d(i)}\right)$ and $p_i = \left(\lambda_{i,d(i)}\sigma_0^2\right)/\lambda$, $p_j = \left(\lambda_{j,d(i)}\sigma_0^2\right)/\lambda$ for $\forall j \in \mathcal{R}_i$, and then substitute these results into (58). We find that the outage probability given in (58) can be treated as a function with respect to the single parameter $\lambda$, i.e., denoted $Pr\left(\lambda\right) = Pr\left(I_{i,d(i)} < r_i\right)$. Model (8) is reduced to the following formation with the single decision variable $\lambda$:

$$
\begin{aligned}
&\min_{\lambda} \quad \frac{\sigma_0^2\left(\lambda_{i,d(i)} + \sum_{j \in \mathcal{R}_i} \lambda_{j,d(i)}\right)}{\lambda} \\
&s.t. \quad \begin{cases} Pr\left(\lambda\right) \leq \beta_i \\ \max\left\{\frac{\sigma_0^2\lambda_{i,d(i)}}{p_{i,\max}}, \frac{\sigma_0^2\lambda_{j,d(i)}}{p_{j,\max}}, j \in \mathcal{R}_i\right\} \leq \lambda \end{cases}
\end{aligned}
$$

$$(68)$$

When $\lambda$ increases, $p_i$ and $p_j$ decrease for $\forall j \in \mathcal{R}_i$, which can reduce the level of mutual information $I_{i,d(i)}$ according to its mathematical definition. This improves $Pr\left(I_{i,d(i)} < r_i\right)$, which implies that $Pr\left(I_{i,d(i)} < r_i\right)$ is a monotonically increasing function of $\lambda$. At this point, $\lambda$ can attain its single maximum value $\lambda_{\max}$ if and only if $Pr\left(\lambda\right) = \beta_i$. That is, the optimal decision variable $\lambda_{\max}$ maximizing model (68) can be obtained by solving the equation $Pr\left(\lambda\right) - \beta_i = 0$, which can be easily achieved using many existing highly effective root-finding numerical methods such as the well-known Newton's method, Sidi's generalized secant method, and Brent's method [23].

An alternative way to relax the assumption that the channel fading is independent and identically distributed while avoiding complexity when calculating the exact outage probability, we can adopt the lower bound (63) or (64) given in **Theorem 8** as an approximation of the outage probability. According to (63) and (64), the lower bound of the outage probability can also be treated as a function of the multiple parameters $\{\lambda_k, k = 1, 2, \ldots, N_i + 1\}$. Thus, let $Pr\left(\lambda_1, \lambda_2, \ldots, \lambda_{N_i+1}\right) = Pr_1^{lower}\left(I_{i,d(i)} < r_i\right)$ or

$$Pr\left(\lambda_1, \lambda_2, \ldots, \lambda_{N_i+1}\right) = Pr_2^{lower}\left(I_{i,d(i)} < r_i\right).$$ A variation in model (8) can be given as

$$\min_{\lambda_1,\ldots,\lambda_{N_i+1}} \quad \frac{\sigma_0^2 \lambda_{i,d(i)}}{\lambda_1} + \sum_{j=1}^{N_i} \frac{\sigma_0^2 \lambda_{j,d(i)}}{\lambda_{j+1}}$$

$$s.t. \quad \begin{cases} Pr\left(\lambda_1, \lambda_2, \ldots, \lambda_{N_i+1}\right) \leq (1-\eta)\beta_i \\ \frac{\sigma_0^2 \lambda_{i,d(i)}}{p_{i,\max}} \leq \lambda_1 \\ \frac{\sigma_0^2 \lambda_{j,d(i)}}{p_{j,\max}} \leq \lambda_{j+1}, j = 1, \ldots, N_i \end{cases}$$

$$(69)$$

where the scalar factor $\eta \in (0,1)$ is introduced to make the QoS-oriented constraint much tighter because the lower bound of the outage probability $Pr\left(I_{i,d(i)} < r_i\right)$ is used to make the computation more tractable when considering general channel-fading situations (i.e., mutually independent and non-identical distributed fading).

Note that a feasible power allocation solution satisfying $Pr\left(\lambda_1, \ldots, \lambda_{N_i+1}\right) \leq \beta_i$ may not be feasible for the constraint on the exact outage probability $Pr\left(I_{i,d(i)} < r_i\right) \leq \beta_i$ because $Pr\left(I_{i,d(i)} < r_i\right) \geq Pr\left(\lambda_1, \ldots, \lambda_{N_i+1}\right)$. The factor $\eta$ should be properly pre-specified to compensate the gap between $Pr\left(I_{i,d(i)} < r_i\right)$ and $Pr\left(\lambda_1, \ldots, \lambda_{N_i+1}\right)$. It seems difficult to choose an exact $\eta$. However, from the perspective of actual engineering implementations, the restriction on the outage probability, $\beta_i$, is usually pre-defined according to the QoS requirements of communication applications, which need not be very exact. A tighter $\beta_i$ can also be adopted to avoid the violation of QoS-constrained transmission reliability. The approximation optimization of the transmission power allocation as presented in (69), makes more sense than the exact model (8) because it facilitates the practical power allocation optimization in general channel-fading situations.

Furthermore, similar to $Pr\left(\lambda\right)$ given in model (68), $Pr\left(\lambda_1, \ldots, \lambda_{N_i+1}\right)$ is also a non-decreasing function with respect to each parameter $\lambda_k$ $(k = 1, \ldots, N_i + 1)$. This implies that the optimum of model (69) can be attained only when these $\lambda_k$ maximize $Pr\left(\lambda_1, \ldots, \lambda_{N_i+1}\right)$ simultaneously. That is, the optimal solution $\lambda_k^*$ can be solved using the equation $Pr\left(\lambda_1, \ldots, \lambda_{N_i+1}\right) = (1-\eta)\beta_i$ based on existing numerical methods. The optimal transmission power levels are then determined immediately as $w_i = \left(\sigma_0^2 \lambda_{i,d(i)}\right)/\lambda_1^*, w_j = \left(\sigma_0^2 \lambda_{j,d(i)}\right)/\lambda_{j+1}^*, j = 1, \ldots, N_i$.

## V. NUMERICAL RESULTS

### A. Outage Behavior of Multi-relay DF Cooperative Communication

In the following simulations we set the QoS-oriented spectral efficiency $r_i = 1$bit/s/Hz for all $i$. To show the outage probability of the multi-relay DF cooperative communication with different numbers of cooperative nodes under the independent and identically distributed (i.i.d) channel condition stated in **Theorem 7**, we set the relay number $N_i \in \{1, 3, 5, 7, 9\}$ and $\lambda_{i,d(i)} = \lambda_{j,d(i)} = 1$ and $\lambda_{i,j} = 2$ for $\forall j \in \mathcal{R}_i$. We then calculate the theoretical outage probability with each $N_i$ based on **Theorem 7**. The obtained results are shown in Fig.3a, which can help us gain a better understanding of how the number of cooperative nodes and the normalized transmission power level impact the outage
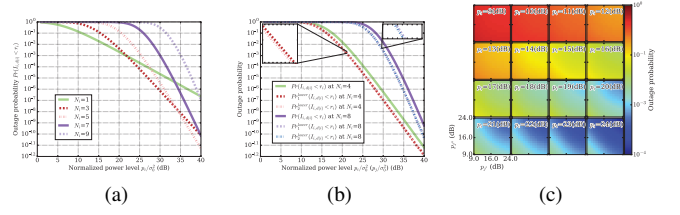


Fig. 3. (a)Outage probability of multi-relay DF cooperative communication for the i.i.d channel case; (b)Outage probability and two tight lower bounds of multi-relay DF cooperative communication for the non-i.i.d channel case where the power levels of different transmitters are assumed to be identical; (c)Outage probability of multi-relay DF cooperative communication for the non-i.i.d channel case where the power level of any cooperative node differs.

behavior of a multi-relay DF cooperative system. The figure shows that when given a certain transmission power level and a specified finite space for the cooperative number (e.g., $\{1, 3, 5, 7, 9\}$), there exists an optimal number of cooperative nodes, which can differ at different power levels. For example, when the normalized power level of source $i$, $p_i/\sigma_0^2$, is less than approximately 22.5dB (recall that the normalized power level of other cooperative nodes can be determined by $\left(\lambda_{j,d(i)}/\lambda_{i,d(i)}\right) \times p_i/\sigma_0^2$ according to **Theorem 7**), the optimal number of relays is $N_i = 1$. When $p_i/\sigma_0^2$ increases to 25 or 35dB, the optimal $N_i$ is 3 or 5, respectively.

We also analyze the actual outage behavior of the multi-relay DF cooperative communication system and the two tight lower bounds under a general channel condition. For the non-i.i.d. channel case stated in **Theorem 8**, we set $\lambda_{i,d(i)} = 1$ while $\lambda_{i,j}$ and $\lambda_{j,d(i)}$ are generated for each $j$ by following a uniform random distribution in the interval $[1, 2]$. Assuming that the power levels of the source and its relays are identical, we compare these two lower bounds of the outage probability obtained based on **Theorem 8** with the actual one obtained from the exact closed-form formulation (31) based on **Theorems 5** and **6** as demonstrated in Fig.3b. As can be seen, even though these two lower bounds are nearly identical, the first bound obtained by approximation (63) is slightly tighter than the second obtained by (64), suggesting that it can provide better approximation of the outage behavior.

Furthermore, fixing $N_i$ at $N_i = 2$, we generalize the power levels of source $i$ and relays $j$ to study the outage behavior under different transmission power conditions as shown in Fig.3c. As can be seen, the outage probability of the multi-relay DF cooperative communication is a decreasing function of the power levels of the transmitters. In other words, increasing the transmission power of any transmitter can reduce the outage probability. This fact, as confirmed by the numerical results shown in Figs.3a, 3b, and 3c, helps solve an optimal power allocation as formulated in (68) or (69).

### B. Evolution of Learning-based Game-theoretic Strategies and Nash Equilibrium

It is noted that the generalized closed-form expression, i.e., (31), can be used to evaluate the actual outage probability of the cooperative transmission system in a general channel situation (i.e., non-i.i.d. channel cases); however, this

TABLE I
SIMULATION PARAMETERS

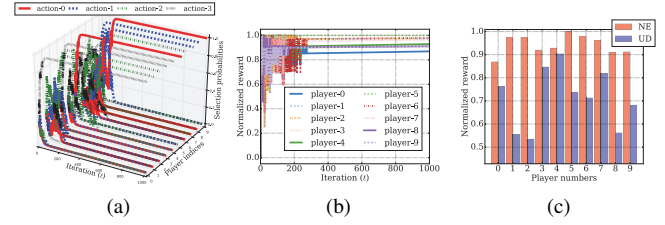| IDM parameters | Values |
|---|---|
| Simulation iterations $D$ | $10^3$ |
| Period of each iteration $\mathrm{d}t$ | 0.01s |
| Desired speed $v_0$ | 120.0/3.6m/s |
| Safe time headway $T$ | 1.6s |
| Acceleration $A$ | 0.73m/s$^2$ |
| Deceleration $B$ | 1.67m/s$^2$ |
| Minimum spacing $Space_0$ | 2.0m |
| Vehicle length $l_a$ | 5.0m |
| Exponent factor $b$ | 4.0 |
| Mobility parameters | Values |
| Initial vehicle speed $v_a(0)$ | $N(12.86, 1.5)$m/s |
| Initial inter-spacing $Space_{a,a-1}(0)$ | $LN(0.685, 0.618)$m |
| DLbSo-RS parameters | Values |
| Required spectral efficiency $r_i$ | 1.0bit/s/Hz |
| Outage probability threshold $\beta_i$ | 0.01 |
| Learning rate $\delta$ | 0.1 |
| Weights $\alpha_1, \alpha_2$ | 1.0, 5.0 |
| Channel parameter $\lambda_{i,d(i)}$ | 1.0 |
| Channel parameter $\lambda_{i,j}, \lambda_{j,d(i)}$ | $U[0, 2]$ |
| Maximum power $p_{i,\max}, p_{j,\max}$ | 40.0dB |
| Initial energy $E_{i,0}, E_{j,0}$ | $Ddtp_{i,\max}$J |
| Maximum transmission range | 300.0m |



Fig. 4. (a)Evolution of the mixed strategies of all the 10 players; (b)Evolution of the normalized reward received by each of all 10 players; (c)Comparison between the normalized reward resulting from the final steady strategy profile and that obtained by unilateral deviation from the resulting strategy profile of each of all 10 players.

$Space^*(v_a(t), \Delta v_{a,a-1}(t))$ is defined as follows:

$$Space^*(v_a(t), \Delta v_{a,a-1}(t)) = Space_0 + v_a(t)T$$
$$+ \frac{v_a(t)\Delta v_{a,a-1}(t)}{2\sqrt{AB}}. \quad (71)$$

Here, $v_0$, $Space_0$, $T$, $A$, and $B$ are model parameters that characterize the desired speed, minimum spacing, desired time headway between any two neighboring vehicles, maximum acceleration, and comfortable braking deceleration, respectively.

In addition, according to Yin et al.'s [25] statistics-based study, the inter-vehicle spacing can be well fitted by the log-normal distribution for free traffic flow during non-peak hours and the vehicle speed can be modeled by following the normal distribution. Thus, in consideration of a realistic non-peak-hour traffic situation, we use the log-normal distribution and normal distribution to initialize the inter-vehicle spacing and the vehicle speed at the beginning of our VANET simulations, respectively. The specific vehicular mobility parameters and other algorithm-related parameters are summarized in Table I. Note that the parameters relevant to the IDM and the vehicular mobility used in our VANET simulations are typical and adapted from the literature [24], [25].

To verify the proposed game-theoretic relay selection model, we first generate a set of moving vehicles distributed on the two lanes, each with seven vehicles. In each lane, two vehicles are randomly selected as the transmission source nodes, while the others are treated as relay candidates. Thus, the overall cooperative VANET consists of four sources ($|\mathcal{I}| = 4$ actions) and 10 relays ($|\mathcal{J}| = 10$ players). Fig. 4a shows the evolution of the selection probabilities of all players' actions (i.e., mixed strategy) for determining which source to cooperate with using the proposed DLbSo-RS algorithm. With equal probabilities at initialization, we can observe that all selection probabilities can converge to pure strategies before the 400-th iteration. The dynamics of the normalized reward of each player are shown in Fig. 4b, which indicates that the comprehensive reward received by each player increases with a lower initial level and finally converges to a high level. In Fig. 4c, we show the unilateral deviation in the decision-making behavior of each of the 10 players. The results UD obtained by the unilateral deviation of each player are compared to those NE obtained from the resulting cooperative network induced by the proposed DLbSo-RS algorithm. As can be seen, the unilateral deviation leads to lower rewards for all players, suggesting that a NE state is reached by the proposed algorithm.
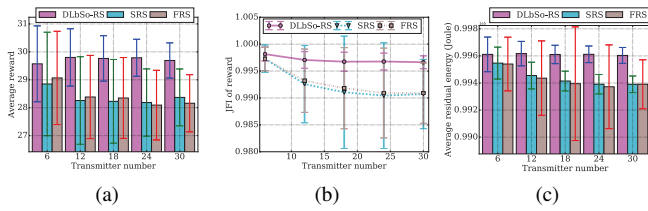
may involve significant computational burden. Therefore, for demonstration, we assume the i.i.d. channel condition in the following simulations, such that we can calculate the exact outage probability using the closed-form formulation (58) given in **Theorem 7** with low computational complexity.

Throughout the following simulations, we consider the transmission reliability constraint by setting the minimum spectral efficiency required as $r_i = 1$bit/s/Hz and the outage probability threshold as $\beta_i = 0.01$ for all $i$. We also set $\lambda_{i,d(i)} = 1.0$ for all $i$ and randomly generate all $\lambda_{i,j}$ and $\lambda_{j,d(i)}$ for $\forall j \in \mathcal{R}_i$ by following a uniform distribution in the interval $[0, 2]$. The power allocation optimization model (68) depending on (58) is adopted to determine the optimal transmission power levels of all $i$ and $j$. In addition, to evaluate the game-theoretic framework under VANETs, we consider a realistic vehicular mobility scenario in urban areas where vehicles are assumed to move on a two-lane road. Specifically, we adopt a well-known car-following model to simulate vehicular mobility, i.e., the intelligent driver model (IDM), which is widely used for the simulation of freeway and urban traffic [24]. The IDM is presented by a differential equation system that describes the dynamics of the positions and speeds of multiple vehicles moving on a road. We denote the position, speed, and length of a vehicle $a$ at time $t$ as $x_a(t)$, $v_a(t)$, and $l_a$, respectively. The net distance between $a$ and the vehicle directly in front of $a$, $a - 1$, is $Space_{a,a-1}(t) = x_{a-1}(t) - x_a(t) - l_{a-1}$, and the approaching rate is $\Delta v_{a,a-1}(t) = v_a(t) - v_{a-1}(t)$. The dynamics of vehicle $a$ can be modeled by the IDM as follows:

$$\begin{cases} \frac{\mathrm{d}x_a(t)}{\mathrm{d}t} = v_a(t) \\ \frac{\mathrm{d}v_a(t)}{\mathrm{d}t} = A\left(1 - \left(\frac{v_a(t)}{v_0}\right)^b - \left(\frac{Space^*(v_a(t), \Delta v_{a,a-1}(t))}{Space_{a,a-1}(t)}\right)^2\right) \end{cases}$$
$$(70)$$

where $b$ is an exponent factor usually set to 4, and

Fig. 5. (a)Average reward of the resulting cooperative VANET versus the total relay (players) number for the proposed relay selection method and the conventional schemes, the error bar denotes the standard deviation of the average reward; (b)JFI of rewards received by all players from the resulting cooperative VANET versus the total relay (players) number, the error bar denotes the standard deviation of JFI; (c)Average residual energy of all transmitters in the resulting cooperative VANET versus the total transmitter number, the error bar denotes the standard deviation of the residual energy.

## C. Energy Efficiency of Decentralized Learning-based Self-organized Relay Selection

Finally, to demonstrate the performance of the proposed method, we conduct a series of Monte Carlo simulations. Specifically, the vehicles (transmitters) are randomly generated and distributed in the two lanes by following the log-normal distribution. The number of source vehicles moving in each lane is set from 1 to 5, i.e., the total source number $|\mathcal{I}|$ is 2 to 10, and the number of relaying vehicles in each lane is set to be double of the source number in the same lane, i.e., the total relay number $|\mathcal{J}|$ ranges from 4 to 20. Thus, the total number of transmitters involved in the network ranges from 6 to 30. In addition, the Monte-Carlo simulations in our experiments are performed with 100 replications per simulation point (transmitter number), and the numerical results are shown with the corresponding standard deviations. For performance comparison, we compare the proposed method DLbSo-RS with two representative schemes: stochastic relay selection SRS and fixed relay selection FRS.

Fig. 5a shows the average energy benefit of players in the resulting cooperative network induced by the different relay selection algorithms, which is evaluated by the average reward of the cooperative network versus different transmitter numbers. In Fig. 5b, the fairness performance of the different algorithms is compared under different numbers of transmitters. The fairness of the energy benefit received among multiple players is measured by the well-known Jain's fairness index (JFI) [26]. From these figures, it can be seen that the proposed DLbSo-RS algorithm outperforms the existing schemes under different numbers of nodes. Furthermore, as shown in Fig. 5c, we provide a performance comparison by evaluating the average residual energy of all transmitters in the resulting cooperative network. As can be seen, a higher level of residual energy on average can be achieved by the proposed algorithm.

## VI. CONCLUSION

In this study, we formulated the multi-relay selection problem as a non-cooperative game and adopted game-theoretic analysis to address the self-interest-driven decision-making issue in VANETs. We proved that it is an OPG and proposed a decentralized learning-based relay selection algorithm to construct and adapt the cooperative network in a distributed

and self-organized manner. Numerical results were provided to prove our theoretical development. Through a series of Monte-Carlo simulations, the proposed method has been demonstrated to outperform two typical conventional schemes in terms of several metrics that indicate the energy efficiency from different perspectives, i.e., energy consumption, energy benefit, and fairness among nodes. Moreover, all the figures illustrate that smaller standard deviations, indicated by the shorter error bars at relevant simulation points, have been achieved by our method. This implies that the randomness in simulation scenarios has slighter influence on the performance of our method, when compared to those of the other two algorithms. Thus, it is confirmed that our algorithm can guarantee better reliability in the vehicular mobility scenario.

## REFERENCES

[1] E. Hossain, G. Chow, V. C. M. Leung, R. D. McLeod, J. Mii, V. W. S. Wong, and O. Yang, "Vehicular telematics over heterogeneous wireless networks: A survey," *Computer Communications*, vol. 33, no. 7, pp. 775–793, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366410000022

[2] S. Chen, J. Hu, Y. Shi, and L. Zhao, "LTE-V: A TD-LTE-based V2X solution for future vehicular network," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 997–1005, Dec 2016.

[3] D. Tian, J. Zhou, Y. Wang, G. Zhang, and H. Xia, "An adaptive vehicular epidemic routing method based on attractor selection model," *Ad Hoc Networks*, vol. 36, Part 2, pp. 465 – 481, 2016, vehicular Networking for Mobile Crowd Sensing. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870515001237

[4] R. Yu, J. Ding, X. Huang, M. T. Zhou, S. Gjessing, and Y. Zhang, "Optimal resource sharing in 5G-enabled vehicular networks: A matrix game approach," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7844–7856, Oct 2016.

[5] C. Campolo, A. Molinaro, A. Vinel, and Y. Zhang, "Modeling prioritized broadcasting in multichannel vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 687–701, Feb 2012.

[6] Q. Wang, S. Leng, H. Fu, and Y. Zhang, "An IEEE 802.11p-based multichannel MAC scheme with channel coordination for vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 449–458, June 2012.

[7] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.

[8] A. Scaglione, D. L. Goeckel, and J. N. Laneman, "Cooperative communications in mobile ad hoc networks," *IEEE Signal Processing Magazine*, vol. 23, no. 5, pp. 18–29, 2006.

[9] A. Chakrabarti, A. Sabharwal, and B. Aazhang, *Cooperative Communications*. Dordrecht: Springer Netherlands, 2006, pp. 29–68. [Online]. Available: http://dx.doi.org/10.1007/1-4020-4711-8_2

[10] Z. Sheng, K. K. Leung, and Z. Ding, "Cooperative wireless networks: from radio to network protocol designs," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 64–69, 2011.

[11] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74–80, 2004.

[12] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[13] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[14] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[15] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[16] D. Tian, J. Zhou, Z. Sheng, and V. C. M. Leung, "Robust energy-efficient mimo transmission for cognitive vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3845–3859, June 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2715328, IEEE Transactions on Vehicular Technology

15

[17] P. S. Sastry, V. V. Phansalkar, and M. A. L. Thathachar, "Decentralized learning of nash equilibria in multi-person stochastic games with incomplete information," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 5, pp. 769–777, 1994.

[18] J. N. Laneman, "Limiting analysis of outage probabilities for diversity schemes in fading channels," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 3, 2003, Conference Proceedings, pp. 1242–1246 vol.3.

[19] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, 2003.

[20] K. G. Seddik, A. K. Sadek, W. Su, and K. J. R. Liu, "Outage analysis and optimal power allocation for multinode relay networks," *IEEE Signal Processing Letters*, vol. 14, no. 6, pp. 377–380, 2007.

[21] D. Fudenberg and J. Tirole, *Game Theory*. New York: MIT Press, 1991.

[22] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0899825696900445

[23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing (3rd ed.)*. New York: Cambridge University Press, 2007.

[24] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, 2000, pRE. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.62.1805

[25] S. Yin, Z. Li, Y. Zhang, D. Yao, Y. Su, and L. Li, "Headway distribution modeling with regard to traffic status," in *Intelligent Vehicles Symposium, 2009 IEEE*, 2009, Conference Proceedings, pp. 1057–1062.

[26] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *Computer Science*, vol. cs.ni/9809099, 1998.

**Min Chen** is a professor in the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China. His research interests include Cyber Physical Systems, IoT Sensing, 5G Networks, Mobile Cloud Computing, SDN, Healthcare Big Data, Medica Cloud Privacy and Security, Body Area Networks, Emotion Communications and Robotics, etc.



**Qiang Ni** is a professor in the School of Computing and Communications, Lancaster University. His current research interests are Wireless Networking and Communications, including Energy-Efficient Green Communications, Cognitive Radio Networks, Broadband Wireless, Intelligent Communication Techniques, Context-User-Aware SDN Networks, Fog-Cloud computing, Smart Grids Communication, Vehicular Networks, Quantum Communication and Mobile Positioning.



**Daxin Tian** [M'13, SM'16] is an associate professor in the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include mobile computing, intelligent transportation systems, vehicular ad hoc networks, and swarm intelligent.
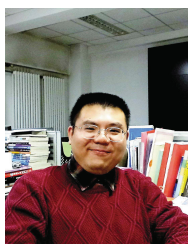


**Jianshan Zhou** received the B.Sc. and M.Sc. degrees in traffic information engineering and control in 2013 and 2016, respectively. He is currently working towards the Ph.D. degree with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests are focused on wireless communication, artificial intelligent system, and intelligent transportation systems.



**Zhengguo Sheng** is currently a Lecturer with the Department of Engineering and Design, University of Sussex, U.K. He has authored over 50 international conference and journal papers. His current research interests cover IoT/M2M, vehicular communications, and edge/cloud computing.



**Victor C. M. Leung** [S'75, M'89, SM'97, F'03] received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (UBC) in 1977, and was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at UBC on a Canadian Natural Sciences and Engineering Research Council Postgraduate Scholarship and received the Ph.D. degree in electrical engineering in 1982.

From 1981 to 1987, Dr. Leung was a Senior Member of Technical Staff and satellite system specialist at MPR Teltech Ltd., Canada. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He returned to UBC as a faculty member in 1989, and currently holds the positions of Professor and TELUS Mobility Research Chair in Advanced Telecommunications Engineering in the Department of Electrical and Computer Engineering. Dr. Leung has co-authored more than 1000 journal/conference papers, 37 book chapters, and co-edited 12 book titles. Several of his papers had been selected for best paper awards. His research interests are in the broad areas of wireless networks and mobile systems.

Dr. Leung is a registered Professional Engineer in the Province of British Columbia, Canada. He is a Fellow of IEEE, the Royal Society of Canada, the Engineering Institute of Canada, and the Canadian Academy of Engineering. He was a Distinguished Lecturer of the IEEE Communications Society. He is serving on the editorial boards of the IEEE Wireless Communications Letters, IEEE Transactions on Green Communications and Networking, IEEE Access, Computer Communications, and several other journals, and has previously served on the editorial boards of the IEEE Journal on Selected Areas in Communications C Wireless Communications Series and Series on Green Communications and Networking, IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, IEEE Transactions on Computers, and Journal of Communications and Networks. He has guest-edited many journal special issues, and provided leadership to the organizing committees and technical program committees of numerous conferences and workshops. He received the IEEE Vancouver Section Centennial Award and 2011 UBC Killam Research Prize. He is the recipient of the 2017 Canadian Award for Telecommunications Research. He is a co-author of the paper that has won the 2017 IEEE ComSoc Fred W. Ellersick Prize.