Mobile Netw Appl
DOI 10.1007/s11036-016-0767-8

Moderate Incentive Design for Delay-Constrained Device-to-Device Relaying

Xuying Zhou[1] · Wei Wang[1] · Yitu Wang[1] · Lin Chen[2] · Zhaoyang Zhang[1]

Abstract Device-to-device (D2D) relaying in a store-carry-forward manner can efficiently expand the transmission range of D2D traffic offloading in cellular systems, which needs external incentives to promote the cooperation of relay nodes who are tend to be selfish. Different to the existing incentive mechanisms which usually adopt large enough incentives, we propose a moderate incentive-compatible data forwarding mechanism based on the Markov decision process (MDP) framework with the principal-agent model. The main idea of this mechanism is to dynamically adjust the payment to incentivize the relay nodes to forward the data with an appropriate radius such that the system utility is maximized. Due to the curse of dimensionality in solving MDP, we propose a greedy algorithm which considers the past information only and further prove its optimality. For discussing the implementation of the proposed solution, we propose an infrastructure-assisted D2D relaying protocol for cellular systems. Simulation results show that our proposed moderate incentive mechanism can achieve a better performance on system utility compared to existing incentive mechanisms.

Keywords D2D communications · Game theory · Incentive compatible · Principal-agent model

✉ Wei Wang
wangw@zju.edu.cn

1 College of Information Science and Electronic Engineering, Zhejiang Key Laboratory of Information Network Technology, Zhejiang University, Hangzhou 310027, PR, China

2 Laboratory de Recherche en Informatique (LRI), University of Paris-Sud 11, Orsay 91405, France

1 Introduction

Device-to-device (D2D) communication as an underlay coexistence with cellular networks [1] allows mobile devices in close proximity to communicate directly, which offloads the cellular traffic. Relaying with D2D communications can further enhance the system performance [2, 3]. With "store-carry-forward" relaying by mobile nodes, the range of D2D traffic offloading can be expanded to get rid of the proximity requirement of source-destination distance.

D2D relaying needs cooperation of relay nodes, but the limitation of the communication resources could easily lead to selfish behaviors. Due to the rationality of relay nodes, they do not always cooperate but need appropriate incentive mechanisms to promote the cooperation. Therefore, several incentive mechanisms are proposed to address the selfish issue and stimulate cooperation for D2D relaying. However, the existing works do not consider the cost of providing the incentives. With the consideration of the incentive cost, it is not necessary to incentivize all the relay nodes to try their best to forward the data, especially for the mobile nodes with limited resources.

In this paper, to balance the data forwarding performance and the incentive cost, we propose the idea of providing *moderate incentives* for delay-constrained D2D relaying. The incentives should be designed considering the incentive compatibility of all relay nodes that the source node meets, so we embrace the principal-agent model to guarantee the incentive compatibility and Markov decision process (MDP) to handle the stochastic optimization problem involving the random node meeting time in cellular networks. We construct a *principal-agent MDP framework* and design moderate incentives for the cooperation of rational

Published online: 29 September 2016

relay nodes. Our contributions in this paper are listed as follows:

– **Principal-Agent MDP Framework**: We construct an MDP framework that dynamically offers relay nodes incentives according to the principal-agent model [4, 5]. Relay nodes can choose one of multiple modes (e.g., transmit power) to forward packet copies. The principal-agent MDP framework provides an efficient approach for designing the appropriate incentives. By doing so, the relay nodes choose the mode that the system expects, which achieves incentive compatibility between the system and relay nodes.
– **Optimal Greedy Implementation**: To overcome the curse of dimensionality of MDP so as to take the causality into consideration, we propose a greedy algorithm to design the incentives greedily based on the past information only. We prove the optimality of this greedy algorithm by theoretic derivation and show its performance improvement compared to existing incentive mechanisms by simulation.
– **Infrastructure-Assisted D2D Relaying Protocol**: Based on the structure of the proposed algorithm, we further discuss its implementation issue which provides a new perspective for such incentive mechanisms on D2D relaying assisted by the infrastructure of cellular systems.

The rest of this paper is organized as follows. Section 2 discusses related incentive schemes and existing literature on data forwarding. Section 3 describes the system model. Section 4 presents the principal-agent MDP model. In Section 5, we propose an optimal greedy incentive-compatible D2D relaying scheme based on the principal-agent MDP model. Following this, Section 6 discusses the implementation issues for the proposed scheme. Section 7 evaluates the performance by simulation. Finally, the paper is concluded in Section 8.

2 Related works

2.1 Data offloading via D2D relaying

Current cellular networks do not have sufficient capacity to accommodate the exponential growth of mobile data requirements. Relaying by mobile nodes has been considered an offloading technique for alleviating the traffic load and energy efficiency. Data offloading has been investigated in many papers [6–10] for various scenarios.

We focus on D2D relaying which has attracted a lot of attentions [11, 12]. In [11], Monte Carlo simulation results showed that the cell edge percent throughput and the percent throughput coverage were improved by more than 150 % by using idle user equipments (UEs) as mobile relays. [12] exploited both multichannel diversity and multicast gain to improve the throughput performance by performing one-to-many D2D, which could achieve up to 40 % gain. The above works assumed that the relay nodes forward the data packets for other nodes without considering any costs, such as battery energy, time, and so on, which is not always true in practice. With considering the costs, it is reasonable for the relay nodes to choose a selfish behavior considering both outcomes and costs to maximize their own benefits. Therefore, it is of high importance to design incentive mechanisms to stimulate nodes' cooperation for D2D relaying.

2.2 Incentive taxonomies for D2D relaying

The incentives for relaying include the direct and indirect reciprocity based incentive. One type is the direct reciprocity based incentive, such as the relaying services from other nodes [13, 14], e.g., two encountering nodes exchange the same amount of messages when they encounter. The nodes violating the cooperation will not obtain the relaying services from other nodes. This kind of incentives does not have any cost if the malicious users can be detected accurately, but has non-ignorable incentive costs on system performance without accurate reputation information [15, 16]. The other type of incentive is the indirect reciprocity based incentive, e.g., the credits with the help of virtual bank or credit clearance service [17–19]. A large enough credit-based payment is adopted to incentivize the relay nodes to make their "best efforts" on data forwarding, which is not always necessary if the incentive cost is considered.

For the cooperation of D2D relaying, there are also two types of incentives as mentioned before. When applying incentive mechanisms for D2D communications, the base station can be regarded as a centralized controller which can be used for signaling exchange.

In indirect reciprocity based schemes, the users' profit from cooperation is reflected by virtual tags [20], and other users will choose to cooperate with one user or not according to its virtual tag. The direct reciprocity based schemes can solve the unfair issue which rewards user's cooperation directly. In direct reciprocity based scheme, relay users are rewarded with time, power resources [21–23]. Especially, a contract-theoretic approach is proposed to solve the problem of providing incentives (storage, energy) for D2D communication in cellular networks [21]. In [22], each time a device receives help from a relay, it pays the relay with a token, which the relay can use to get relay service in the future. [23] proposed a sub-optimal joint relay and power allocation to maximize the proportional fairness. When users are chosen to cooperate with each other, more time resource and less power are allocated to the users as cooperative rewards to improve their energy efficiencies, which provide an effective incentive to stimulate the cooperation among users. However, the existing works do not consider the cost of providing the incentives. With the consideration of the incentive cost, it is not necessary to incentivize all the relay nodes to try their best to forward the data, which motivate us to design a moderate incentive for delay-constrained D2D relaying.

3 System model

We consider a cellular network consisting of N mobile nodes. These nodes move independently with a speed v in an $L \times L$ area and their mobility patterns are independent and identically distributed [24]. Any two nodes can establish a connection and transmit packets once moving into each other's transmission range via D2D communications. Let R denote the maximum transmission radius of the source node. A "store-carry-forward" relaying example is illustrated in Fig. 1, where user 1 as the source node generates a message needed to reach user 3 as the destination node within time T. By D2D relaying, user 1 forwards the packet copy to user 2 as a relay node via link 1, and determines the payment of successful relaying for user 2. When the relay user 2 meets user 3 during moving, the message copy is forwarded via link 2 if user 3 does not receive the message. User 2 gets paid if relaying finishes before time T. Otherwise, user 3 will request the message from base station through link 3.

The wireless channel model we adopt is the large-scale path loss model. As the signal-to-interference ratio (SIR) should be higher than the threshold at the receiver for successful transmission, we consider the energy consumption as a cost of transmission. The average large-scale path loss (PL) for transmitter-receiver separation is expressed as a function of the distance: $\overline{PL}(d) \propto (\frac{d}{d_0})^n$, where n is the path loss exponent, d_0 is the close-in reference distance which is determined by measurements close to the transmitter, and d is the distance between the transmitter and the receiver. For a given received power, the transmission power, denoted as P_t, can be expressed according to [25] as

$$P_t = d^n \cdot c, \tag{1}$$

where c is a constant which varies among different scenarios. It is assumed that the energy consumption of receiving packets can be ignored.

Time is divided into slots. The source node probes if any relay node moves into its maximum transmission radius at the beginning of each slot until T. Let the contact time $t_k, k = 0, 1, 2, \cdots, K$ denote the time when the source node meets a relay node without a packet copy for the k-th time and we number this relay node by k. If the source node finds more than one relay nodes inside its transmission range, the source node should forward the packet copy to the nearest one.

Packet forwarding between two nodes only occurs at meeting times and are assumed to be instantaneous [25]. The inter-meeting time (IMT) is assumed to be exponentially distributed with parameter λ. The validity of this assumption for synthetic mobility models (e.g., random walk, random direction, random waypoint, etc.) has been discussed in [26]. As for random waypoint model, The meeting rate λ of any two nodes is calculated as

$$\lambda \approx \frac{2\omega r v}{L^2}, \tag{2}$$

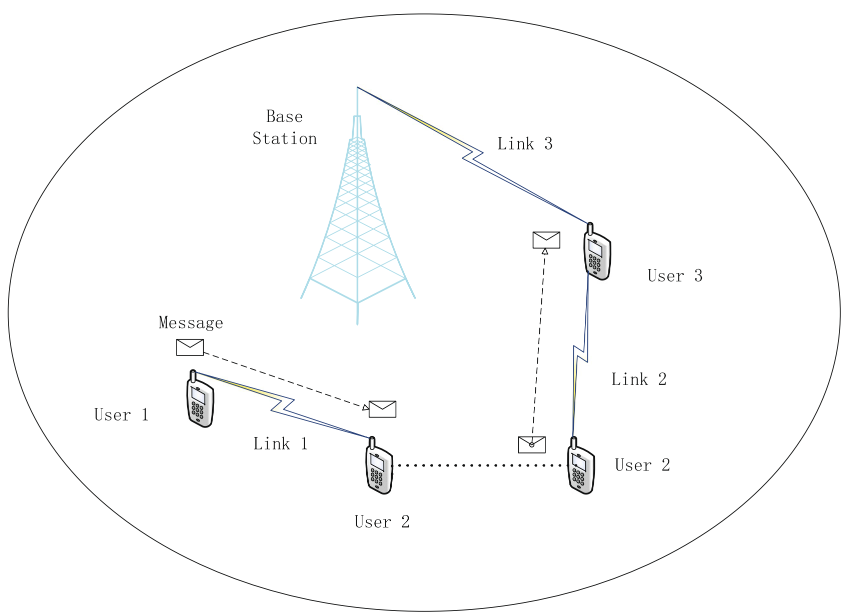


Fig. 1 Illustration of "store-carry-forward" D2D relaying

where ω is a constant set by [26], r is the transmission distance, and v is the average relative speed between two nodes.

Let D_k denote the delivery predictability, which is the ratio that the packet is delivered within time T predicted by the source node at time t_k. If the source node forwards the packet copy to a relay node at time t_k, according to [25], the delivery predictability of the current relay is

$$D_k = 1 - e^{-\lambda_{k,d}(T-t_k)}, \tag{3}$$

where $\lambda_{k,d}$ denotes the meeting rate between the relay node k and the destination node. At the initial time, only the source node carries the packet, so the initial delivery predictability D_0 is

$$D_0 = 1 - e^{-\lambda_{0,d}T}, \tag{4}$$

where $\lambda_{0,d}$ denotes the meeting rate between the source node and the destination node.

Without loss of generality, we assume the source node encounters a relay node at time t_k. Before forwarding the packet to this relay node, the corresponding delivery predictability is D_k. After the source node forwards the packet copy to the relay node, the delivery predictability changes to D_{k+1}, which can be derived as

$$D_{k+1} = 1 - (1 - D_k)e^{-\lambda_{k,d}(T-t_k)}. \tag{5}$$

Considering both the successful probability of D2D relaying and the reward paying to relay nodes, we define the system utility U_k at time t_k as

$$U_k = (D_{k+1} - D_k)(1 - \alpha_k). \tag{6}$$

where α_k is the incentive factor at time t_k. Denote α as the set of the incentive factors for the encountered relay nodes over time, i.e., $\alpha = (\alpha_1, \cdots, \alpha_K)$.

Relay nodes have multiple modes to forward packet copies. According to Eq. 1, the transmit power is used to distinguish different modes, which implies that relay nodes can forward packet copies with different radius.

Suppose a relay node k meets the destination node when it has not received the packet yet and forwards the packet copy successfully with probability P_k. The destination node will pay the relay node according to the incentive factor α_k as reward. In this case, other relay nodes meeting the destination node later cannot receive the reward. The cost of the relay node is its transmit power in Eq. 1. The utility of relay node k at time t_k, denoted as u_k, is its received payment minus the cost, i.e.,

$$u_k = (\alpha_k - d_k^n c)P_k. \tag{7}$$

When the source node encounters relay nodes at time t_k, the source node determines the value of α_k, which the relay nodes are aware of. Relay nodes then choose the mode which can achieve the maximum utility under such a α_k, which affects the system's utility. Denote S_K as the utility of the system at the last contacting time t_K, which can be calculated as the sum of the rewards for all contact time. The system objective can be formulated as

$$\max_{\alpha} S_K = \sum_{k=1}^{K} U_k. \tag{8}$$

4 Principal-agent MDP framework for D2D relaying

In this section, we model the D2D relaying problem as a principal-agent MDP model, where the principal-agent model guarantees the incentive compatibility and the MDP model handles the stochastic optimization problem involving the random node meeting time. In this model, the "principal" is the system that wants to accomplish packets forwarding, and the "agents" are the relay nodes for forwarding packet copies. Under the principal-agent model, Fig. 2 presents a graphical description of the D2D relaying by demonstrating the problem as a two-stage Stackelberg game [27]. At every contact time t_k, the principal moves first and offers a payment scheme (contract) to the agent. The payment scheme will reward the relay if it is the first one to forward the data to the destination node. The relay node decides whether to accept this contract, and determines the forwarding strategy to maximize its own utility u_k. The relay nodes may forward the data in a lazy way, e.g., using a low power, which reduces the system utility. However, the system does not exactly know the actions of the relays, which needs to addressed by the principal-agent model. Thus, we construct the principal-agent MDP model to take both the current and future rewards into account and motivate the agents to act on behalf of the principal.

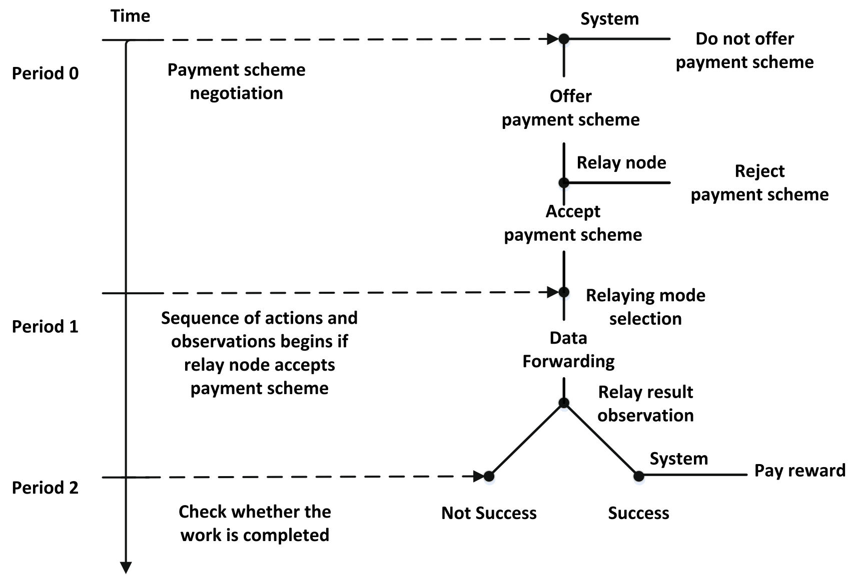
To construct the principal-agent MDP framework, we discuss the key components in a MDP framework. Note that the actions in this MDP does not affect the transition probability and reward directly, but has an impact via the behavior of relay nodes, which is handled by the incentive design according to the principal-agent model.

State The system state at the k-th meeting between the source node and the relay nodes includes the delivery predictability D_k and the contact time t_k, i.e.,

$$S_k = (D_k, t_k). \tag{9}$$

Action The source node's action at time t_k is the incentive factor α_k which is set by the source node for the k-th encountering relay node k. According to the principal-agent model, the source node's action maps to the mode that relay node k chooses. Denote d_k is the forwarding radius that

Fig. 2 Graphical representation of D2D relaying



relay node k chooses. As relay nodes are assumed to be rational, the mode of the relay node is selected as

$$d_k = \arg\max_{d} u_k(d, \alpha_k). \tag{10}$$

Transition Probability The state transition of t_k depends on the node mobility and is independent of the action α_k. The transition probability of t_k is

$$P(t_{k+1}|t_k) = \begin{cases} 1 - e^{-(N-k)\bar{\lambda}(t_{k+1}-t_k)} & \text{if } t_k < t_{k+1} \le T \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

where $\bar{\lambda}$ is the average meeting rate in the network.

The delivery predictability state D_{k+1} depends on both the current D_k and the action of relay node. The action of relay node is incentivized by the incentive factor α_k according to the principal-agent model. The transition probability of D_k can be formulated as

$$P(D_{k+1}|D_k, \alpha_k) = \begin{cases} 1 & \text{if } D_{k+1} \text{ satisfies} \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Eq. 5.

Combining Eqs. 11 and 12, we have the transition probability of the system state S_k as

$$P(S_{k+1}|S_k) = P(t_{k+1}|t_k) \cdot P(D_{k+1}|D_k, \alpha_k) \tag{13}$$

Reward The immediate principal's reward $R(S_k, \alpha_k)$ is the increment of the delivery predictability in state S_k under action α_k. According to Eqs. 5 and 6, by removing the incentive paying to the relay, $R(S_k, \alpha_k)$ can be calculated as

$$R(S_k, \alpha_k) = (1 - D_k)(1 - e^{-\lambda_{k,d}(T-t_k)})(1 - \alpha_k). \tag{14}$$

Value Function Define π as the policy factor which is a mapping from the source state S_k to the action a_k for all time t_k. The value function $V_\pi(S_k)$ is defined as the cumulative reward for starting in state S_k and acting according to π thereafter. Based on the Bellman equation, the value function is given as follows:

$$V_\pi(S_k) = R(S_k, \alpha_k) + \sum_{S_{k+1}} P(S_{k+1}|S_k, \alpha_k) V_\pi(S_{k+1}). \tag{15}$$

The optimal value function V^* is the unique solution of Bellman equation:

$$V_\pi^*(S_k) = \max_{\alpha_k} \left\{ R(S_k, \alpha_k) + \sum_{S_{k+1}} P(S_{k+1}|S_k, \alpha_k) V_\pi(S_{k+1}) \right\}. \tag{16}$$

The corresponding optimal actions in each contact time can be calculated by backward induction, and then stored in a table. By searching the table in each source state for the corresponding optimal action, the optimal opportunistic forwarding scheme for maximizing the delivery predictability is obtained.

5 Optimal greedy solution for principal-agent MDP

In this section, we propose a greedy algorithm which provides an optimal solution for the above principal-agent MDP problem. A two-mode example is first explored and is generalized to a general form for incentive design. Finally, we prove the optimality of the proposed greedy algorithm.

5.1 Greedy algorithm

As analyzed previously, the system utility during the data forwarding process can be calculated according to Eq. 16. However, the conventional algorithm needs to obtain the value function $V_\pi(S_k)$ by calculating over all possible state transitions, which leads to the curse of dimensionality. Since the practical application of the optimal scheme is severely limited due to its exponential computation complexity, it is necessary to provide a low-complexity scheme to achieve the balance between performance and computation complexity.

In this subsection, we propose a greedy solution with the system reward in Eq. 14 where only the past information is used. We will prove later that this greedy algorithm achieves optimal performance.

Simplify Eq. 5,

$$\frac{1-D_{k+1}}{1-D_k} = e^{-\lambda_{k,d}(T-t_k)}. \tag{17}$$

By summing up $\frac{1-D_{k+1}}{1-D_k}$ over all $j < k$, we obtain D_k as

$$D_k = (1-D_0)\prod_{j=1}^{k-1} e^{-\lambda_{j,d}(T-t_j)}. \tag{18}$$

Substituting Eq. 18 into Eq. 6, we obtain the system utility for each source-relay meeting as

$$U_k = (1-D_0)\prod_{j=1}^{k-1} e^{-\lambda_{j,d}(T-t_j)}[1-e^{-\lambda_{k,d}(T-t_k)}](1-\alpha_k). \tag{19}$$

Instead of maximizing the total system utility $\sum_{k=1}^{K} U_k$, we maximize the current instantaneous system utility U_k to achieve a greedy solution for Eq. 8. The optimization problem for maximizing U_k is formulated as

$$\max_{\alpha_k}(1-D_0)\prod_{j=1}^{k-1} e^{-\lambda_{j,d}(T-t_j)}[1-e^{-\lambda_{k,d}(T-t_k)}](1-\alpha_k). \tag{20}$$

5.2 How to achieve incentive compatibility?

To solve the optimization problem of Eq. 20 under the principal-agent model, the incentive factor α_k should satisfy the following two essential constraints.

- *Participation constraint*: The principal provides a non-negative expended utility to the agents, i.e., $u_k \geq 0, \forall k$.
- *Incentive compatibility constraint*: The agent achieves the highest expected utility which also obeys the principal's preference.

We first derive the expression of agents' utility u_k in Eq. 7 where the probability P_k needs to be determined. Consider the relay node k encounters the source node at time t_k. The source node only has the past information before t_k due to the causality. In the case that relay node k is the first node that forwards the packet copy to the destination node, the $k-1$ relay nodes that the source node meets before t_k have not reached the destination yet. Thus, P_k can be calculated as

$$\begin{aligned} P_k &= \int_{t_k}^{T} \lambda_{k,d} e^{-\lambda_{k,d}(t-t_k)} \prod_{j=1}^{k-1} e^{-\lambda_{j,d}(t-t_j)} dt \\ &= \lambda_{k,d} e^{\lambda_{k,d} t_k} \prod_{j=1}^{k-1} e^{\lambda_{j,d} t_j} \int_{t_k}^{T} e^{-\lambda_{k,d} t} \prod_{j=1}^{k-1} e^{-\lambda_{j,d} t} dt \\ &= \lambda_{k,d} e^{\lambda_{k,d} t_k} \frac{e^{t_k \sum_{j=1}^{k}(-\lambda_{j,d})} - e^{T \sum_{j=1}^{k}(-\lambda_{j,d})}}{\sum_{j=1}^{k}(\lambda_{j,d})}, \end{aligned} \tag{21}$$

Based on the above expression of P_k, we obtain the agents' utility.

Next, we investigate the key parameter in incentive design, i.e., the incentive factor α. The existence of incentive factor α_k establishes the relevance of U_k and u_k. Figure 3 analyzes the effects of α on the utilities of both the system and the relay node, when the source node encounters the k-th relay node and determines the incentive factor α_k. As

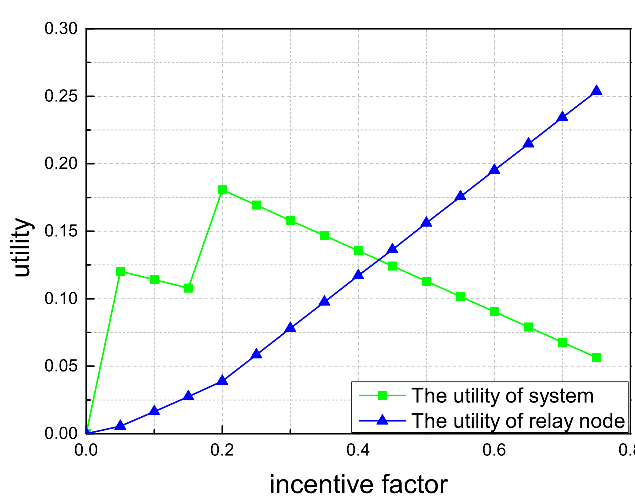


Fig. 3 Effect of the incentive factor α

α_k increases, the utility of the relay node increases, demonstrating that the appropriate design of α_k can strengthen the relay node's utility. There are two jumps in the system utility curve because the relay node changes its actions with the increasing of α_k. Therefore, we set the value of α_k within the jumping points which are the minimum values to maximize the system utility.

5.2.1 A two-mode example

Now, we consider the cases that the source node has different preference on the mode of encountering relay nodes. In order to simplify the problem, we first consider two modes for relay nodes as an example. In Mode 1 and Mode 2, the relay node forwards the packet copy in radius d_e and d_l, respectively. Here, $d_e > d_l \geq 0$.

1) The source node prefers Mode 1

In this case, $U_k(d_e) > U_k(d_l)$ and $U_k(d_e) > 0$ which makes sure that the system can achieve a positive utility by choosing d_e. According to these two essential constraints in the principal-agent model, α_k should satisfy the following two inequalities:

$$\begin{cases} u_k(d_e) \geq 0, \\ u_k(d_e) \geq u_k(d_l). \end{cases} \tag{22}$$

By solving the above inequations, α_k should satisfy

$$\begin{cases} \alpha_k \geq \frac{p_k(d_e)(d_e)^n c - p_k(d_l)(d_l)^n c}{p_k(d_e) - p_k(d_l)}, \\ \\ \alpha_k \geq (d_e)^n c. \end{cases} \tag{23}$$

We find that it always holds that

$$\frac{p_k(d_e)(d_e)^n c - p_k(d_l)(d_l)^n c}{p_k(d_e) - p_k(d_l)} > (d_e)^n c. \tag{24}$$

Combined with Eq. 20, the value of α is set as

$$\alpha_k = \frac{p_k(d_e)(d_e)^n c - p_k(d_l)(d_l)^n c}{p_k(d_e) - p_k(d_l)}. \tag{25}$$

2) The source node prefers Mode 2

In this case, $U_k(d_l) \geq U_k(d_e)$ and $U_k(d_l) > 0$ which makes sure that the system can achieve a positive utility by choosing d_l. Using the approach similar to that for Mode 1, we obtain

$$\alpha_k = (d_l)^n c. \tag{26}$$

5.2.2 Incentive design

We extend the above two modes into the general multi-mode case. Denote M as the set of modes and $d_m, \forall m \in M$, as the transmission radius of the relay node in Mode m. When the source node prefers mode m, α_k should satisfy the following inequalities:

$$\begin{cases} u_k(d_m) \geq 0, \\ u_k(d_m) \geq u_k(d_q),\ \forall q \in M, q \neq m. \end{cases} \tag{27}$$

By solving the above inequations, α_k is obtained as

$$\begin{cases} \alpha_k \geq \frac{p_k(d_m)(d_m)^n c - p_k(d_q)(d_q)^n c}{p_k(d_m) - p_k(d_q)}, \text{if} d_m > d_q, \forall q \in M, \\ \alpha_k \leq \frac{p_k(d_q)(d_q)^n c - p_k(d_m)(d_m)^n c}{p_k(d_q) - p_k(d_m)}, \text{if} d_m < d_q, \forall q \in M, \\ \alpha_k \geq (d_m)^n c. \end{cases} \tag{28}$$

Referring to the above analysis, we design the incentive factor α_k as follows: When d_m is the smallest radius that relay nodes can choose, we set

$$\alpha_k = (d_m)^n c. \tag{29}$$

Otherwise, the value of α_k is set as

$$\alpha_k = \max_{q \in M, d_q < d_m} \frac{p_k(d_m)(d_m)^n c - p_k(d_q)(d_q)^n c}{p_k(d_m) - p_k(d_q)}. \tag{30}$$

Due to these two constraints of the principal-agent model, each different α_k can guide relay nodes to choose the mode that the system expects.

Here, we prove that the incentive factors α_k according to Eqs. 29 and 30 exist for all modes.

Theorem 1 (Existence of Moderate Incentives α) *Given the condition $(d_m)^n c \leq 1, \forall m \in M$, the incentive factors $\alpha_k \in [0, 1]$ obtained by Eqs. 29 and 30 exist for all modes.*

Proof As the utilities of the system and the relay nodes should be non-negative, according to Eqs. 6 and 7, we can find the following inequality:

$$(d_m)^n c \leq 1, \forall m \in M \tag{31}$$

which means the forwarding radius for all modes should satisfy Eq. 31 and the incentive factor according to Eq. 29 exists.

Since $P_k(d)$ in Eq. 30 is a increasing function of d, by combining it with Eq. 31, we obtain that it always holds that

$$\frac{p_k(d_m)(d_m)^n c - p_k(d_q)(d_q)^n c}{p_k(d_m) - p_k(d_q)} \leq 1, \tag{32}$$

Considering both cases, it can be concluded that the system utility U_k is non-negative. Therefore, the existence of the proposed moderate incentives is proved. □

By appropriate incentive design, we can find an one-to-one mapping relationship between the actions of source node and relay nodes. The details of the proposed algorithm are described using pseudo code as Algorithm 1.

Algorithm 1 Greedy Incentive Design Algorithm

for each contact time **do**
if the relay node that the source node meets does not have a packet copy **then**
  the source forwards the packet copy to the relay k
  the source calculates utility $U_k(d_m)$ for every $m \in M$
  if $U_k(d_m) > U_k(d_q)$ for every $q \in M$ and $q \neq m$ and $U_k(d_m) > 0$ **then**
    the system prefers to Mode i and sets α_k according to Eqs. 29 and 30
  end if
end if
if the relay node k with a packet copy moves meets the destination node that has not received this packet yet **then**
  the relay node forwards the packet copy to the destination node and receives the payment according to the incentive α_k.
end if

5.3 Algorithm optimality

Greedy algorithm is a common method to reduce the computational complexity which can not ensure optimality. In [28], the authors proposed a greedy online algorithm that does not need future arrival information. The greedy algorithm makes decisions based on the information available in the current slot, which achieves at least 50 % of the optimal performance. In this subsection, we prove that our proposed greedy algorithm achieves the optimal performance.

Theorem 2 (Optimality of Greedy Algorithm) *The incentive factor α satisfying*

$$\alpha_k^* = \arg\max_{\alpha_k}(1-D_0)\prod_{j=1}^{k-1} e^{-\lambda_{j,d}(T-t_j)}[1-e^{-\lambda_{k,d}(T-t_k)}](1-\alpha_k) \tag{33}$$

maximizes the total system utility in Eq. 8.

Proof For the case that the source encounters a relay node k at time t_k and makes an action

$$\alpha_k = \frac{p_k(d_m)(d_m)^n c - p_k(d_q)(d_q)^n c}{p_k(d_m) - p_k(d_q)} \tag{34}$$

because of $U_k(d_m) > U_k(d_q)$ for every $q \in M$ and $q \neq m$ according to the information before t_k.

If we change the decision-making time of action α_k to when the source node encounters the last relay node within time T, we denote the action as α_k'. In this case, the source node has the information over all time slots. The increment of delivery predictability after making strategy α_k' is

$$\Delta D' = (1-D_0)\prod_{j=1, j\neq k}^{K} e^{-\lambda_{j,d}(T-t_j)}(1-e^{-\lambda_{k,d}(T-t_k)}). \tag{35}$$

The parameter P_k' required for calculating incentive factor α_k' should be changed:

$$\begin{aligned} P_k{}' &= \int_{t_k}^{T} \lambda_{k,d} e^{-\lambda_{k,d}(t-t_k)} \prod_{j=1, j\neq k}^{K} e^{-\lambda_{j,d}(t-t_j)} dt \\ &= \lambda_{k,d} e^{\lambda_{k,d} t_k} \prod_{j=1, j\neq k}^{K} e^{\lambda_{j,d} t_j} \int_{t_k}^{T} e^{-\lambda_{k,d} t} \prod_{j=1, j\neq k}^{K} e^{-\lambda_{j,d} t} dt \\ &= \lambda_{k,d} e^{\lambda_{k,d} t_k} \frac{e^{-t_k \sum_{j=1}^{K} \lambda_{j,d}} - e^{-T \sum_{j=1}^{K} \lambda_{j,d}}}{\sum_{j=1}^{K} \lambda_{j,d}} \end{aligned} \tag{36}$$

Calculating the system's utility when relay node k chooses every mode with $\Delta D'$ and $P_k{}'$, we still find that $U_k'(d_m) > U_k'(d_q)$ for every $q \in M$ and $q \neq m$. Therefore, it is proved that whenever the source node makes a decision does not affect the performance, and thus, the optimality of the proposed greedy algorithm is proved. □

6 Implementation consideration for D2D relaying

In this section, we discuss the implementation issues of the proposed incentive compatible delay-constrained D2D relaying algorithm. Specifically, the D2D destination node requests data from the D2D source node, and receives data from either the base station or the D2D relay nodes. The D2D relay nodes are incentivized to participate in the data forwarding by the incentive payment α_k, which could enhance the efficiency of D2D relaying and alleviate the traffic load of the base station.

D2D communications adopt the base station as a centralized controller with signaling exchange, which provides more convenience to the payment process of the incentive determined by the source node and executed when the data are transmitted to the destination node successfully.

An incentive compatible D2D relaying procedure is described in Fig. 4, which includes the following stages:

– The D2D destination first requests the data from the base station via the physical uplink control channel [29]. As soon as the base station receives the service request message, the time counter is initialized to zero and is started.
– When a D2D relay node moves into the transmission range of the D2D source node, this D2D relay node

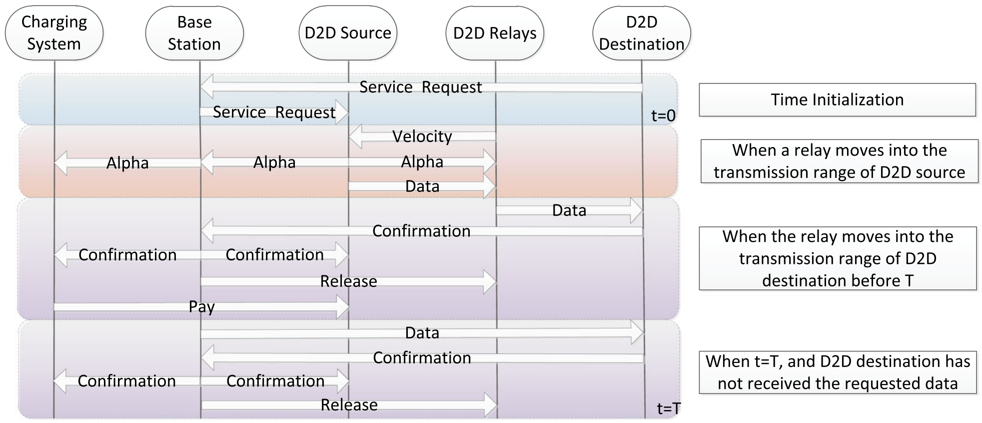


Fig. 4 Signaling for D2D relaying

reports its velocity to the D2D source node. Then the D2D source node determines a payment α according to the proposed algorithm, which is send to base station, charging system and the D2D relay node afterwards. Also, the requested data is sent to the D2D relay node.

– When the D2D relay node moves into the transmission range of the D2D destination node before T, it forwards the data to the destination, and then receives the payment from charging system after the data is confirmed. When the base station receives the confirmation message, it broadcasts release message to all D2D relay nodes to release the data from their buffer.
– When $t = T$ but the D2D destination node has not received the requested data, the base station sends the data to the D2D destination node directly to ensure the delay constraints. Similarly, the data cached in D2D relays are released after data confirmation.

7 Simulation

In this section, we evaluate the performance of the proposed incentive design algorithm for D2D relaying by simulation using Matlab. The simulation includes two scenarios: the first part under random waypoint model and the other one under the real VANET scenario. Without loss of generality, we consider two modes in this simulation.

For performance comparison, four algorithms are adopted as baselines.

– *Baseline 1* [17]: The system sets a threshold to select optimal relay nodes.
– *Baseline 2*: The system sets the value of the incentive factor $\alpha_k = (d_l)^n c$ for every k such that relay nodes choose Mode 2 only.
– *Baseline 3*: The system adopts a constant incentive factor.
– *Baseline 4* [19]: The system sets the value of the incentive factor $\alpha_k = \frac{p_k(d_e)(d_e)^n c - p_k(d_l)(d_l)^n c}{p_k(d_e) - p_k(d_l)}$ such that relay nodes can choose Mode 1 only.

Note that Baselines 1 and 4 are the conventional threshold-based scheme and best effort scheme respectively, and Baselines 2 and 3 adopt our proposed framework but with simple incentive design.

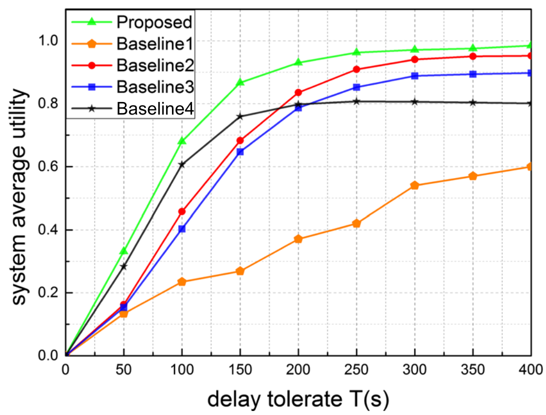
7.1 Performance under random waypoint model

In the simulation, 40 mobile nodes are deployed in the network, in which one source-to-destination pair is investigated for collecting simulation results. Each node's movement is independent following random waypoint mobility model. The transmission radius of the source node is 20m and those of the relay nodes in Mode 1 and Mode 2 are 20m and 10m, respectively. We run each algorithm for 10000 times and consider the average utility of system as the performance metric. In the simulation results in Fig. 5, we consider the effect of three factors including tolerate delay T, the moving speed of nodes v and the size of square length L.

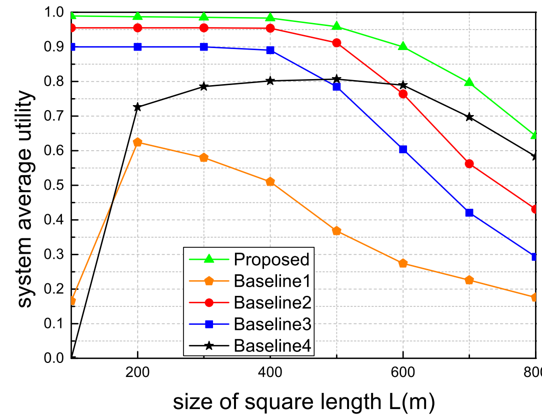
Figure 5a shows the simulation results with different tolerate delay T under $L = 500$m and $v = 10$m/s. The values of tolerate delay T are [0,400] and the interval is 50s. Our proposed incentive design algorithm always achieves a better average system utility than all baseline algorithms.

Figure 5b indicates the effect of nodes' moving speed to the system average utility under $L = 500$m and $T = 250$s. The moving speed values from 0 to 40m/s and the interval is set to 5s. All the five curves increase significantly when v increases from 0m/s to 15m/s then all curves remain stable, because the increase of moving speed can increase the encountering chance between nodes while the effect is limited. It is worth noting that our proposed algorithm also has a higher average system utility than all baseline algorithms.

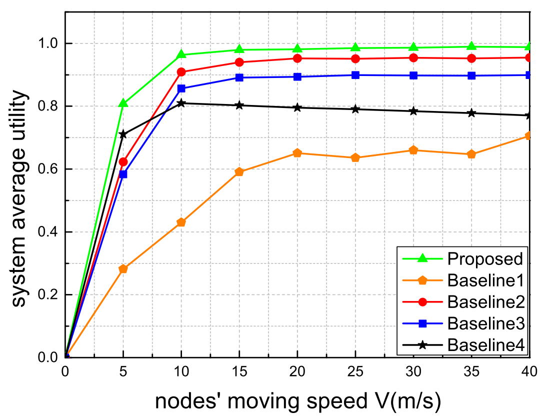
Figure 5c shows the influence of size of square length L under $v = 10$m/s and $T = 200$s. L is discrete from 100m



(a) Effect of tolerate delay T



(b) Effect of nodes' moving speed v



(c) Effect of the size of square length L

Fig. 5 Performance comparison under random waypoint model

to 800m and the gap is 100. We notice that three curves decrease when L increases while the curves of Baselines 1 and 4 increase for small L and decreases for large L. As we know that the larger the nodes' moving range is, the smaller probability the node meeting occurs with. Especially, we would like to discuss the first point of Baselines 1 and 4. When $L = 100$m, the system should pay a large incentive to let relay nodes to forward packet copies. Thus, the system expects relay nodes not forward the packet copies and it gains no utility. Again, the results show that our proposed algorithm achieves a higher average system utility than baseline algorithms as above.

7.2 Performance under real VANET scenario

The vehicle moving trajectory data set of city Koln [30] is adopted for providing more practical results. The mobile trajectory of 3400 vehicles covered about 400km^2 area of the city's center and suburbs. The forwarding radiuses of the relay nodes in Mode 1 and Mode 2 are 200m and 100m, respectively. Figure 6 shows the performance comparison by simulation using this vehicle moving trajectory data set. We can obvious find that the performance gain of

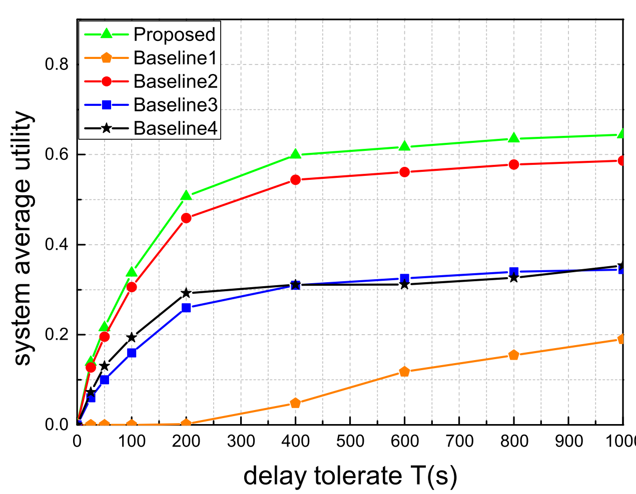


Fig. 6 Performance comparison under real VANET scenario

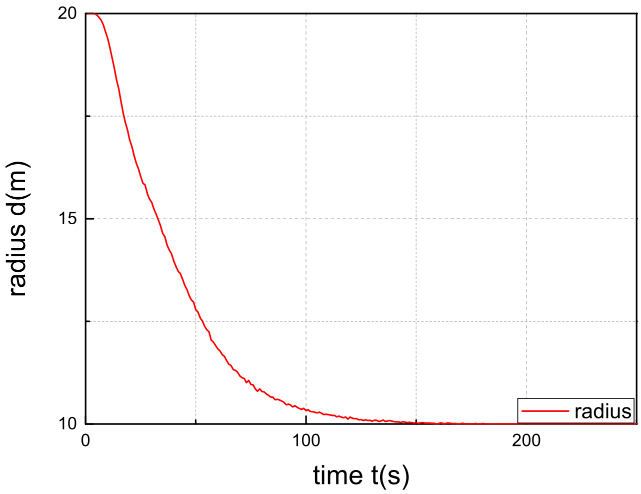


Fig. 7 The system's preference on forwarding radius

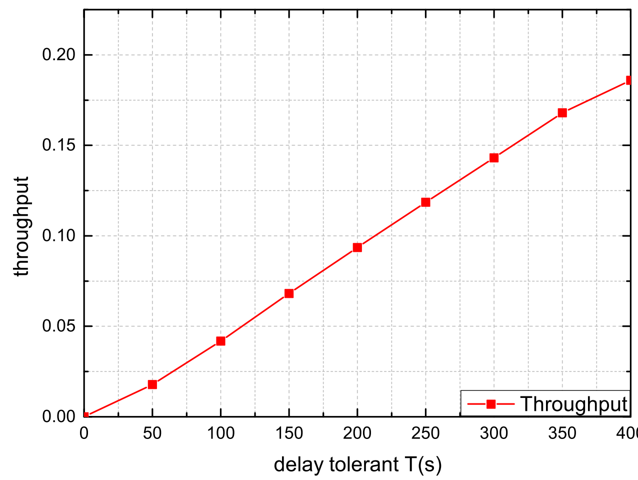


Fig. 9 System throughput

our proposed algorithm increases as the delay tolerate time T increases.

7.3 Forwarding radius preference

Figure 7 shows the system's preference on the forwarding radiuses that the relay nodes choose. Here, we set L = 500m, T = 250s and v = 10m/s. The average radius is obtained by repeating the transmission for 30000 times. It can be observed that the system prefers to the larger radius at the beginning. With the increasing of time, the proportion of large radius reduces. The system turns to choose the smaller radius at all when the time approaches T. It is mainly because that the incentive payment of relaying with a small radius is smaller than that with a large radius. It is also verified that it's not necessary to let the relay nodes transmitting with the largest radius over all time.

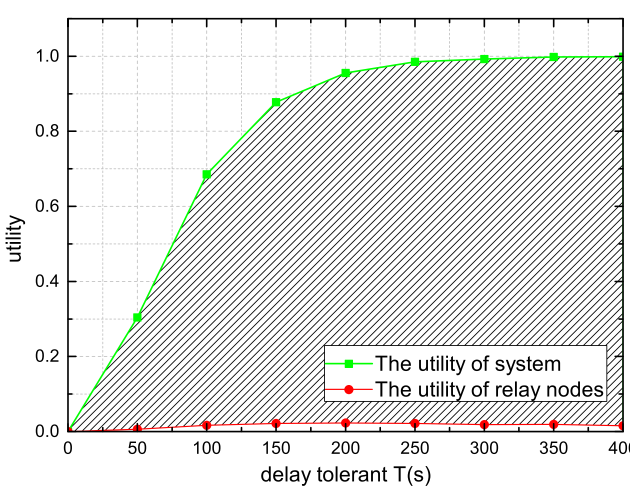


Fig. 8 Successful probability of D2D relaying and the payment

7.4 Successful probability and system throughput

In Fig. 8, both the successful probability and the payment increase first and then tend to a constant value as the tolerate delay T increases. The reason is that when T increases, the source node meets more relay nodes for data forwarding. We notice that the probability to meet the delay constraint is close to 1 when T exceeds 250s. The shaded part between two curves is the system utility.

Figure 9 shows the system throughput, which is defined as the times of successful D2D relaying during a unit of time. Assume that the throughput is 0 while no packet be forwarded within T. It is obviously that the system throughput increases as T increases because of the growth of successful probability.

8 Conclusion

In this paper, we propose a moderate incentive design for delay-constrained D2D relaying. This design embraces both the principal-agent model and the MDP. The system achieves maximum utility by dynamically adjusting the incentives to relay nodes. The opportunistic forwarding problem is formulated as a principal-agent MDP framework. With implementation consideration, we propose a greedy algorithm which needs the past information only. Furthermore, we prove the optimality of this greedy algorithm and propose an infrastructure-assisted D2D relaying protocol for cellular systems. Simulation results confirm that, compared with the existing inventive mechanisms, our proposed algorithm achieves a higher average system utility. Especially, we find that the harder working of relay nodes (e.g., choosing a larger forwarding radius) does not always lead to a higher utility to the system.

Acknowledgments This work is supported in part by National Natural Science Foundation of China (No. 61571396), Zhejiang Provincial Natural Science Foundation of China (No. R17F010006), and National Hi-Tech R&D Program (No. 2014AA01A702).

References

1. Wu D, Cai YM, Hu RQY, Qian Y (2015) Dynamic distributed resource sharing for mobile D2D communications. IEEE Trans Wirel Commun 14(10):5417–5429
2. Liu P, Tao ZF, Lin ZN, Erkip E, Panwar S (2006) Cooperative wireless communications : A cross-layer approach. IEEE Wirel Commun Mag 13(4):84–92
3. Hakola S, Chen T, Lehtomaki J, Koskela T (2010) Device-to-device (D2D) communication in cellular network-performance analysis of optimum and practical communication mode selection
4. Laffont J, Martimort D (2001) The Theory of Incentives I: The Principal-agent Model. Princeton University Press, Princeton NJ, U.S.
5. Bolton P, Dewatripont M (2004) Contract Theory. MIT Press
6. Bo H, Pan H, Anil Kumar VS, Marathe MV, Shao JH, Aravind S (2011) Mobile Data Offloading through Opportunistic Communications and Social Participation. IEEE Trans Mob Comput 11(5):821–834
7. Apostolaras A, Iosifidis G, Chounos K, Korakis T (2015) C2M: Mobile data offloading to mesh networks . Proc IEEE GLOBECOM:4877–4883
8. Mehmeti F, Spyropoulos T Performance Analysis of Mobile Data Offloading in Heterogeneous Networks. IEEE Trans Mob Comput. to appear
9. Si P, He Y, Yao H, Yang R (2016) DaVe: Offloading Delay-Tolerant Data Traffic to Connected Vehicle Networks. IEEE Trans Veh Technol 65(6):3941–3953
10. Wang L, Peng T, Yang Y, Wang W (2012) Interference constrained relay selection of D2D communication for relay purpose underlaying cellular networks. Proc IEEE WiCOM
11. Vanganuru K, Ferrante S, Sternberg G (2012) System capacity and coverage of a cellular network with D2D mobile relays. Proc of IEEE MILCOM
12. Zhou B, Hu H, Huang S, Chen H (2013) Intracluster device-to-device relay algorithm with optimal resource utilization. IEEE Trans Veh Technol 62(5):2315–2326
13. Buttyn L, Dra L, Flegyhzi M, Vajda I (2010) Barter trade improves message delivery in opportunistic networks. Ad Hoc Netw 8(1):1–14
14. Li N, Das SK (2010) RADON: reputation-assisted data forwarding in opportunistic networks. Proc ACM International Workshop on Mobile Opportunistic Networking:8–14
15. Wang W, Chen L, Shin KG, Duan L (2014) Secure cooperative spectrum sensing and access against intelligent malicious behaviors. Proc IEEE INFOCOM
16. Wang W, Chen L, Shin KG, Duan L (2015) Thwarting intelligent malicious behaviors in cooperative spectrum sensing. IEEE Trans Mob Comput 14(11):2392–2405
17. Wen D, Cai Y, Li Z (2013) An incentive compatible two-hop multi-copy routing protocol in DTNs. Proc IEEE MSN:140–146
18. Shevade U, Song H, Qiu L, Zhang Y (2008) Incentive-aware routing in DTNs. Proc ICNP:238–247
19. Brun O, El-Azouzi R, Prabhu B, Seregina T (2014) Modeling rewards and incentive mechanisms for Delay Tolerant Networks. International Symposium on Modeling and Optimization in Mobile Ad Hoc, and Wireless Networks:233–240
20. Li P, Guo S (2015) Incentive mechanisms for device-to-device communications. IEEE Netw 29(4):75–79
21. Zhang Y, Song L, Saad W, Dawy Z, Han Z (2015) Contract-Based Incentive Mechanisms for Device-to-Device Communications in Cellular Networks. IEEE J Sel Areas Commun 33(10):1–1
22. Mastronarde N, Patel V, Xu J, Van DSM (2013) Learning relaying strategies in cellular D2D networks with token-based incentives. IEEE Globecom Workshops:163–169
23. Sun Q, Tian L, et al. (2015) Energy efficient incentive resource allocation in D2D cooperative communications. Proc IEEE ICC:2632–2637
24. Camp T, Boleng J, Davies V (2002) A survey of mobility models for ad hoc network research. Wirel Commun Mob Comput 2(5):483–502
25. Lu Y, Wang W, Chen L, Zhang Z, Huang A (2014) Distance-based energy-efficient opportunistic forwarding in mobile delay tolerant networks. Proc IEEE ICC:3564–3569
26. Groenevelt R, Nain P, Koole G (2005) The message delay in mobile ad hoc networks. Perform Eval 62(1-4):210–228
27. Qin YH (2012) A stackelberg-game model in a two-stage supply chain. Syst Eng Procedia 3:268–274
28. Li S, Zheng ZZ, Ekici E, Shroff NB (2013) Maximizing social welfare in operator-based Cognitive Radio Networks under spectrum uncertainty and sensing inaccuracy. Proc of IEEE INFOCOM
29. Wang W, Lau VKN (2014) Delay-aware cross-layer design for device-to-device communications in future cellular systems. IEEE Commun Mag 52(6):133–139
30. TAPASCologne project. http://sourceforge.net/apps/mediawiki\/sumo/index.php? title=TAPASCologn