Full-Duplex Machine-to-Machine Communication for Wireless-Powered Internet-of-Things

Yong Xiao*, Zixiang Xiong†, Dusit Niyato‡, Zhu Han* and Luiz A. DaSilva§
* Department of Electrical and Computer Engineering, University of Houston, TX
† Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX
‡ School of Computer Engineering, Nanyang Technological University, Singapore
§CONNECT, Trinity College Dublin, Ireland

***Abstract*—This paper considers machine-to-machine (M2M) communication for wireless-powered Internet-of-Things (IoT) based networking systems. Motivated by the observation that transmitting signals generally requires more energy than receiving signals for most IoT-based systems, we study a special wireless-powered M2M communication system in which the receiver can send its surplus energy to the transmitter. We propose a framework of wireless powered full-duplex M2M communication (WP-FD-M2M) in which the energy transfer from the receiver to the transmitter and the data transmission from the transmitter to the receiver take place at the same time over the same frequency. We establish a stochastic game-based model, referred to as the M2M game, to characterize the interaction between autonomous M2M transmitter and receiver. We prove that, if the transmitter and receiver can sequentially optimize their data transmission and energy transfer based on the Markov strategy, it is possible to achieve the maximum long-term performance for M2M communication without a centralized controller or coordination between the transmitter and receiver. Numerical results show that our proposed approach can significantly improve the performance for M2M communication under various situations.**

***Index Terms*—Energy harvesting, wireless energy transfer, machine-to-machine communication, full-duplex, game theory, stochastic game.**

I. Introduction

It is commonly believed that the next generation wireless networks will be based on the concept of Internet-of-Things (IoT) which is a new paradigm promised to bridge the gap between the human and physical world by ubiquitously connecting billions of "things" throughout the Internet. One of the key enablers for IoT-based networks is machine-to-machine (M2M) communication, which allows mobile devices and machines to autonomously establish wireless communication links with each other [1]. The IoT's vision of ubiquitous connectivity needs to be supported by ubiquitous energy supply. Energy harvesting enables mobile devices to power their services with energy harvested from the surrounding environment, which provides a unique opportunity to solve the energy problem for M2M communication in IoT-based network systems.

Most existing works on wireless-powered communication networks focus on the cases in which the wireless devices can either harvest energy from the natural environment such as the sunlight, wind, radio wave, and vibration, or receive energy transferred from dedicated energy sources such as power beacons [2], network access points, and cellular base stations (BSs) [3], [4]. For example, it was recently demonstrated that the energy harvested from ambient radio frequency (RF) signals can support at least 1 kbps wireless communication between two battery-free devices over a distance of 2.5 feet [5]. One of the main challenges for energy harvesting-based communication systems is that the energy supply is generally unreliable due to the uncertainty and unpredictability of the natural environment. Recent development in wireless power transfer technology triggers interest in wireless-powered communication systems supported by dedicated energy sources. More specifically, a hybrid communication system consisting of both cellular mobile networks and dedicated wireless RF power transfer infrastructure that can wirelessly charge nearby mobile devices was studied in [2]–[4], [6], [7].

RF energy transfer has attracted significant interest due to its ability to combine both data signal and wireless energy signal together to achieve simultaneous wireless information and power transfer (SWIPT) [8], [9]. It was shown that a wireless powered full-duplex communication system has the potential to further improve both the reliability of the energy supply and spectrum utilization efficiency for many existing network systems. Specifically, a wireless-powered relaying system was studied in [10] where a full-duplex relay node can simultaneously forward signals and receive RF energy sent by the source. A full-duplex information and energy transmission network was considered in [3] in which a full-duplex cellular BS can simultaneously send an RF energy signal and receive data packets to and from half-duplex mobile devices, respectively. This result was further extended into the cases that mobile devices can also operate in full-duplex mode in [4].

In this paper, we focus on M2M communication for an IoT system supported by energy harvesting. Different from most existing works, which focus on centralized resource allocation controlled by network infrastructure (e.g., cellular BSs) and/or consist of dedicated energy sources deployed by network operators or utility companies, we consider an M2M communication link in which both transmitter and receiver can harvest energy from external energy sources to support autonomous wireless communication. Motivated by the observation that, in many IoT systems, data transmission causes higher energy consumption than that of data

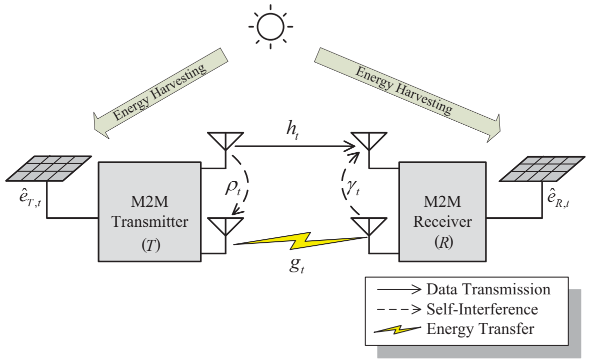

Fig. 1. System model for a wireless-powered full duplex M2M communication link.

reception, we propose a novel framework referred to as wireless-powered full-duplex M2M communication (WP-FD-M2M) for IoT systems. In WP-FD-M2M, if an M2M receiver harvests more energy than it can consume, it can transfer the surplus harvested energy to the transmitter to further improve the reliability of the energy supply for M2M communication. The energy transfer from the receiver to the transmitter, as well as the data transmission from the transmitter to receiver, take place at the same time over the same frequency. We focus on the distributed optimization for WP-FD-M2M in which the transmitter can autonomously schedule its energy use for data transmission and the receiver can also decide by itself whether to transfer its surplus energy to the transmitter in each time slot. Distributed optimization problems for multiple agents with energy-constraints in a time-varying environment are notoriously difficult to solve. In this paper, we study the WP-FD-M2M from a game theoretic perspective. We establish a stochastic game-based model, referred to as the M2M game, to characterize the interaction between the transmitter and the receiver in a time-varying environment. We prove that, if both transmitter and receiver can sequentially optimize their data transmission and energy transfer according to a Markov strategy, it is possible to maximize the long-term performance of M2M communication even when there is no centralized controller to coordinate actions and energy/data transmission between the transmitter and receiver. Numerical results show that our proposed approach can significantly improve the performance of M2M communication.

The remainder of this paper is organized as follows. In Section II, we introduce the system model and problem formulation. The distributed optimization approach is proposed in Section III. We present numerical results to compare our proposed approach with other existing results in Section IV. We conclude the paper in Section V.

II. System Model and Problem Formulation

A. System Model

We consider an M2M communication link consisting of an M2M transmitter T and an M2M receiver R. Both T and R are quipped with energy harvesters which can convert external energy into electric power. We focus on a wireless-powered system in which the M2M communication from T to R is solely supported by harvested energy. We assume that the M2M communication process is slotted and the length of each time slot has been normalized into unity. We can hence use the terms energy and power interchangeably throughout this paper. Let $\hat{e}_{T,t}$ and $\hat{e}_{R,t}$ be the amounts of energy that can be harvested during the tth time slot by T and R, respectively. Motivated by the observation that, in most wireless communication systems, transmitters consume more energy than receivers, we assume that if R harvests more energy than that it can consume, it can wirelessly transfer its surplus energy to T as an RF signal. We propose the *WP-FD-M2M* as illustrated in Figure 1, in which the data transmission from T to R and the energy transfer from R to T take place at the same time over the same frequency. To simplify our description, we assume that both T and R are equipped with two antennas: one for data transmission and receiving and the other for energy transfer and receiving. Our results can be directly extended into more general cases with more antennas installed at T and R, as further discussed later in this paper. Let h_t be the channel gain between the data transmission and receiving antennas of T and R. Let g_t be the wireless energy transfer efficiency between the energy transmit antenna of R and the energy receiving antenna installed at T in time slot t.

The full-duplex transmission of energy and data signals in WP-FD-M2M results in the following two types of self-interference:

SI1) *Self-interference at R*: RF energy signal sent from R to T will cause *self-interference* to the data reception at R. To reduce this self-interference, various interference cancellation techniques have been proposed [11]. Unfortunately, there is still no practical solution that can perfectly cancel all the interference at the receiver. In this paper, we assume that if R transfers $w_{R,t}$ amount of energy to T, the residual self-interference power received by R will be given by $\gamma_t w_{R,t}$ where γ_t is the self-interference cancellation factor in time slot t for $0 \leq \gamma_t < 1$.

SI2) *Self-energy recycling at T*: an M2M data signal sent from T to R will also be received by the RF energy receiving antenna at T. We refer to this signal received by T as the *self-energy recycling signal*. Different from the self-interference observed at R, this self-energy recycling signal is beneficial and can be obtained by T as a part of its received energy. We assume that if the transmit power of T is w_T, the energy that can be received by T from the self-energy recycling signal will be given by $\rho_t w_T$ where ρ_t is the self-energy recycling factor in time slot t for $0 \leq \rho_t < 1$.

We assume that the energy required by R to process and decode the data signals sent by T can be regarded as a constant τ_R. We can then write the surplus energy of R as $e_{R,t} = (\hat{e}_{R,t} - \tau_R)^+$ where we denote $(\cdot)^+ = \max\{0, \cdot\}$. We assume that R cannot store its harvested energy but will either discard its surplus energy or transfer all the surplus energy to T during each time slot. Let $\delta_{R,t}$ be the decision made by R about whether to send its energy to T, i.e., we use $\delta_{R,t} = 1$ (or $\delta_{R,t} = 0$) to mean R will (or will not) send its surplus energy to T in time slot t. We can write the amount of energy transferred from R at the end of time slot t as $\delta_{R,t} e_{R,t}$. The total amount of energy that can be received by T during time slot t is given by

$$\bar{w}_{T,t} = \hat{e}_{T,t} + u_{R,t} + u_{T,t}, \tag{1}$$

where $u_{R,t} = \delta_{R,t} g_t e_{R,t}$ and $u_{T,t} = \rho_t w_{T,t}$. To simplify our description, in this paper, we follow the same line as [10] and ignore the energy converted from the additive noise received by T.

We assume that there is a minimum unit of energy that can be harvested, received and used by T to send each data packet. Specifically, $\hat{e}_{T,t}$, $w_{T,t}$, $u_{R,t}$ and $u_{T,t}$ can be regarded as values that are taken from finite sets $\mathcal{E}_T$, $\mathcal{W}_T$, $\mathcal{U}_R$ and $\mathcal{U}_T$, respectively. We also refer to $\boldsymbol{v}_t = \langle h_t, \gamma_t, g_t \rangle$ as the environmental state of WP-FD-M2M during time slot t. We assume that both T and R can observe the environmental state of the current time slot. Let $\boldsymbol{e}_t = \langle \hat{e}_{T,t}, e_{R,t} \rangle$ be the energy harvesting states in time slot t. Let $\mathcal{E}$ and Υ be the sets of possible values for $\boldsymbol{e}_t$ and υ_t, respectively. T has a battery that can store up to $\bar{e}_T$ units of energy. We denote the set of battery levels for T as $\mathcal{B}$. We focus on WP-FD-M2M with causal constraint and assume that the energy obtained by T during the current time slot can only be used in the data transmission in the following time slots. More specifically, we can write the battery level of T at the beginning of time slot t as

$$b_{T,t} = \min\{\bar{e}_T, \bar{w}_{T,t-1} + b_{T,t-1} - w_{T,t-1}\}, \tag{2}$$

where we ignore the temporal energy storage loss and assume that the energy stored in the battery of T will not decrease with time. We assume that both T and R know the battery level at the beginning of each time slot. This can be achieved by allowing T to include its battery level information in the M2M data packets sent to R. We will discuss how to relax this assumption in Section V. The transmit power of T in time slot t needs to satisfy the following constraint:

$$0 \leq w_{T,t} \leq b_{T,t}. \tag{3}$$

B. Problem Formulation

At the beginning of each time slot, T can schedule its transmit power, and R can also decide whether to send its surplus energy to T during the rest of the time slot. We assume that there is always data for T to transmit to R and the main objective for both T and R is to maximize the long-term discounted payoff, determined by the transmission rate defined as

$$\mathbb{E}\left(\lim_{t\rightarrow\infty} \sum_{l=0}^{t} \alpha^l \varpi_l \left(w_{T,l}, \delta_{R,l}\right)\right), \tag{4}$$

where $\varpi_t \left(w_{T,t}, \delta_{R,t}\right)$ is the transmission rate of the M2M communication in time slot t given by

$$\varpi_t \left(w_{T,t}, \delta_{R,t}\right) = \log\left(1 + \frac{h_t w_{T,t}}{\delta_{R,t} \gamma_t w_{R,t} + \sigma_{R,t}}\right), \tag{5}$$

and $\sigma_{R,t}$ is the additive noise level at R. $\mathbb{E}(\cdot)$ denotes the expectation and α is the discount coefficient satisfying $0 \leq \alpha < 1$.

III. Distributed Optimization for Energy Usage and Transfer Scheduling for WP-FD-M2M

To maximize the long-term discounted payoff in (4), T and R not only need to estimate the current battery level, harvested energy, power transfer efficiency and M2M communication channel gain, but should also take into consideration the future evolution of these parameters in the physical environment. However, the future change of the physical environment can be affected by various factors, most of which are unpredictable and uncontrollable by T or R. Fortunately, it has been observed in that if the duration of each time slot is short enough, it is reasonable to assume that the evolution of the physical environment satisfies the *Markov property*. That is, the environment in the current time slot depends only on that of the previous time slot. In this paper, we assume that the time-varying characteristics of $\boldsymbol{v}_t$ and $\boldsymbol{e}_t$ possess the Markov property and can be characterized by transition functions $\Pr\left(\boldsymbol{v}'|\boldsymbol{v}\right)$ and $\Pr\left(\boldsymbol{e}'|\boldsymbol{e}\right)$, respectively, where $\Pr\left(\boldsymbol{v}'|\boldsymbol{v}\right)$ is the probability distribution of the current environmental state $\boldsymbol{v}'$ when the previous environmental state is given by $\boldsymbol{v}$ for $\boldsymbol{v}, \boldsymbol{v}' \in \Upsilon$ and $\Pr\left(\boldsymbol{e}'|\boldsymbol{e}\right)$ is the probability distribution of current harvested and received energy $\boldsymbol{e}'$ given that the harvested and received energy in the previous time slot is given by $\boldsymbol{e}$. We assume that $\Pr\left(\boldsymbol{v}'|\boldsymbol{v}\right)$ and $\Pr\left(\boldsymbol{e}'|\boldsymbol{e}\right)$ are stationary and can be known by both T and R. One way to achieve this is to allow T and R to learn these probability distributions from their past observations using reinforcement learning approaches proposed in [12], [13].

The long-term payoff of the M2M communication link also depends on the interaction between R and T. For example, R should only send energy to T when it believes that the battery level of T is or will soon be insufficient to support the required M2M data transmission. However, how T's battery level will change also depends on the future transmit power decided by T, which is unknown to R. On the other hand, T can increase the transmit power in the current time slot if it believes that R will send its surplus energy in the next few time slots. Similarly, it is generally difficult for T to perfectly know the future actions of R.

To solve these problems, we formulate a stochastic game-based model, which we refer to as the *M2M game*, to investigate the interaction between T and R in a time-varying environment. We have the following definition.

Definition 1: An *M2M game* is defined as a tuple $\mathcal{G} = \langle \mathcal{P}, \boldsymbol{\mathcal{A}}, \mathcal{S}, \mathcal{T} \rangle$ where $\mathcal{P}$ is the set of players T and R, $\boldsymbol{\mathcal{A}}$ is the action space of T and R, $\mathcal{S}$ is the state space, and $\mathcal{T}$ is the state transition function characterizing the probability distribution of the transition between different states under each possible action.

We give a detailed description of each of the above elements for the M2M game as follows:

- *State Space* $\mathcal{S} = \mathcal{B} \times \Upsilon$: is a finite set of all the possible battery levels and environmental states in each time slot. We write the state of the transmitter in time slot t as $s_t \in \mathcal{S}$ for all $t \geq 0$.
- *Action Space* $\mathcal{A} = \mathcal{W}_T \times \Delta_R$: is a finite set of all the possible combinations of actions for T and R. More specifically, in WP-FD-M2M, the action of T is its decision about the transmit power and the action of R corresponds to its decision on whether to send its surplus energy to T in each time slot. We write the action decided by T and R in time slot t as $a_t = \langle w_{T,t}, \delta_{R,t} \rangle \in \mathcal{A}$ for all t.
- *State transition function* $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$: specifies the probability distribution that, starting at state s and action a in the current time slot, the state ends in s' in the next time slot. T and R can estimate the state transition function using their observed battery level as well as the probability distribution of the future changes of environment state. Specifically, suppose in the current time slot, the environmental state and battery level are given by υ and b, respectively, and the actions of T and R are given by $a = \langle w_T, \delta_R \rangle$. We can calculate the probability of having battery level b' at the beginning of next time slot as follows: If $b' < \bar{b}$, the energy stored and newly obtained by T in the next time slot will not exceed the maximum capacity of its battery. We can hence write the probability of having state s' at the beginning of next time slot as

$$\begin{aligned}
&\Pr\left(s'|s, a\right) \\
&\quad = \quad \Pr\left(\langle b', \upsilon' \rangle | \langle b, \upsilon \rangle, \langle w_T, \delta_R \rangle\right) \\
&\quad = \quad \Pr\left(\langle b' = b + \bar{w}_T - w_T, \upsilon' \rangle | \langle b, \upsilon \rangle, \langle w_T, \delta_R \rangle\right) \\
&\quad = \quad \sum_{\langle \upsilon', \boldsymbol{e}' \rangle \in \Phi} \Pr\left(\boldsymbol{e}' | \boldsymbol{e}\right) \Pr\left(\upsilon' | \upsilon\right), \qquad (6)
\end{aligned}$$

where $\Phi = \{\langle \upsilon', \boldsymbol{e}' \rangle : b' = b + \bar{w}_T - w_T, \ \forall \upsilon' \in \Upsilon, \boldsymbol{e}' \in \mathcal{E}\}$. If $b' = \bar{b}$, the energy that will be stored and obtained by T in the time slot will exceed the maximum capacity of T's battery. Following the same line as the previous case, we can write the probability of the state transition for this case as follows:

$$\Pr\left(s'|s, a\right) = \sum_{\langle \upsilon', \boldsymbol{e}' \rangle \in \Phi'} \Pr\left(\boldsymbol{e}' | \boldsymbol{e}\right) \Pr\left(\upsilon' | \upsilon\right), \qquad (7)$$

where $\Phi' = \{\langle \upsilon', \boldsymbol{e}' \rangle : b + \bar{w}_T - w_T \geq \bar{e}_T, \ \forall \upsilon' \in \Upsilon, \boldsymbol{e}' \in \mathcal{E}\}$.

The state transition probability can be fully specified by combining equations (6) and (7).

It can be observed that the M2M game is a special stochastic game, also called collaborative stochastic game [14], in which all players try to optimize the same payoff function. The main solution for the M2M game is the Nash equilibrium (NE), which is an action profile for all the players such that no player can further improve its expected discounted payoff by unilaterally changing its action [14].

To maximize the long-term average payoff, T needs to evaluate both the current and future payoffs that can be obtained by each of its possible actions under each possible action of R. We define the value function $V_T\left(s_t, w_{T,t}|\delta_{R,t}\right)$ of T as the sum of the current and future expected payoffs when the current states and actions of T and R are given by s_t and $a_t = \langle w_{T,t}, \delta_{R,t} \rangle$, respectively. Suppose the current state is given by s_t. We can write the current payoff $\bar{\varpi}_{T,t}$ when T chooses action $w_{T,t}$ in the current time slot as follows:

$$\bar{\varpi}_{T,t} = \sum_{s_t \in \mathcal{S}} \Pr\left(s_t | s_{t-1}, a_{t-1}\right) \varpi_t\left(w_{T,t}, \delta_{R,t}\right), \qquad (8)$$

where $\varpi_t\left(w_{T,t}, \delta_{R,t}\right)$ is defined in (5).

T should also be able to estimate the future expected payoff using the state transition function. We can hence write $V_T\left(s_t, w_{T,t}|\delta_{R,t}\right)$ as follows:

$$\begin{aligned}
&V_T(s_t, w_{T,t}|\delta_{R,t}) = \bar{\varpi}_t \\
&\quad + \quad \alpha \sum_{s_{t+1} \in \mathcal{S}} \Pr\left(s_{t+1}|s_t, a_t\right) V^*(s_{t+1}, w_{T,t+1}|\delta_{R,t+1}). (9)
\end{aligned}$$

And we can write the optimal value function for T under state s_t given action $\delta_{R,t}$ of R as follows:

$$V_T^*(s_t|\delta_{R,t}) = \max_{a_t \in \mathcal{A}} V_T(s_t, w_{T,t}|\delta_{R,t}). \qquad (10)$$

Note that the value function of T also depends on the action of R. If T always chooses the action that maximizes the above value function for each given action of R, the resulting strategy can also be referred to as the best response for T. Since, in the M2M game, both T and R can observe the same state and also try to maximize the same payoff function, T can estimate the optimal action that will be chosen by R. In particular, T can estimate the best response of R for each of given action of T. More specifically, T can choose its optimal action $w^*_{T,t}$ by

$$w^*_{T,t} \quad = \quad \arg \max_{w_{T,t} \in \mathcal{W}} V(s_t, w_{T,t}|\hat{\delta}^*_{R,t}), \qquad (11)$$

where $\hat{\delta}^*_{R,t}$ is the optimal action of R estimated by T under its action $w_{T,t}$ which can be calculated by T using $\hat{\delta}^*_{R,t} = \arg \max_{\delta_{R,t} \in \Delta} V_T(s_t, w_{T,t}|\delta_{R,t})$.

It can be observed that the optimal action $w^*_{T,t}$ that T would choose during time slot t depends only on current state s_t. This strategy has been commonly referred to as the *Markov strategy* [15].

Similarly, R can also estimate the optimal action of T and decide its Markov strategy by

$$\delta^*_{R,t} \quad = \quad \arg \max_{\delta_{R,t} \in \Delta} V_R(s_t, \delta_{R,t}|\hat{w}^*_{T,t}), \qquad (12)$$

where $\hat{w}^*_{R,t}$ is the optimal action of T estimated by R under its action $\delta_{R,t}$, which can be calculated by R using $\hat{w}^*_{T,t} = \arg \max_{w_{T,t} \in \mathcal{W}} V_R(s_t, \delta_{R,t}|w_{T,t})$.

From the above analysis, we can prove the following result.

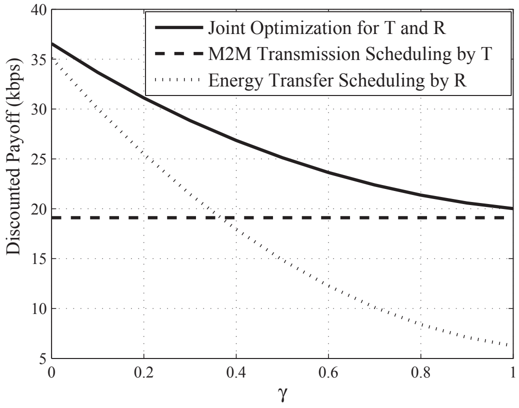

Fig. 2. Comparison of discounted payoff achieved by three optimization methods under different self-interference cancellation efficiencies.

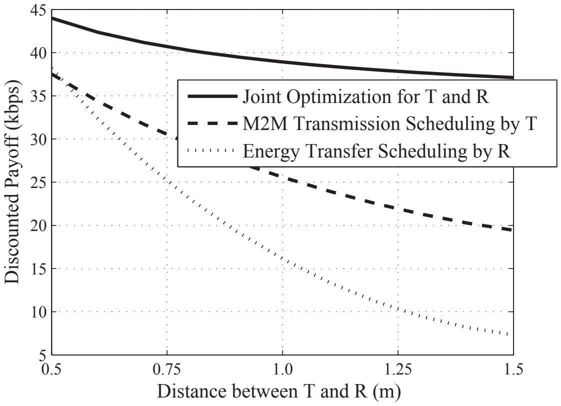

Fig. 3. Comparison of discounted payoff achieved by three optimization methods under different distances between T and R.

Theorem 1: If T and R always choose their Markov strategies using (11) and (12) at the beginning of each time slot, the resulting action profile in each time slot is an NE. In addition, the resulting NE is unique and optimal for the M2M game.

Proof: It can be observed that both T and R have finitely many actions with a limited number of possible states. We can hence claim that an NE always exits in our M2M game. Since the payoff observed by T or R as well as their Markov strategies are fully determined by the state of the system, we can combine players T and R together and create a new player with the combined action space $\mathcal{A}$. In this way, the two-player M2M game has been converted into a single-player stochastic game and (11) and (12) are equivalent to the Bellman equation. We can therefore follow the same line as [16] to prove that (11) and (12) achieve the optimal policy for both T and R. This concludes the proof. ■

IV. Numerical Results

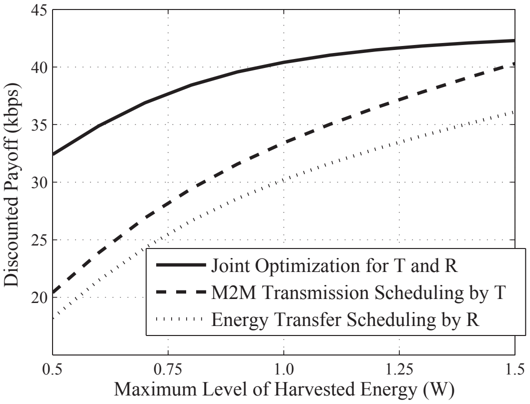

Fig. 4. Comparison of discounted payoff achieved by three optimization methods under different maximum levels of harvested energy.

In this section, we present numerical results to access the performance of our proposed optimization approach for WP-FD-M2M. We assume that the number of energy units that can be harvested by T or R is a discrete uniformly distributed random variable between 0 and 1 W. We ignore the energy consumed by R to receive and decode the M2M data packets sent from T. We assume that the RF energy transfer from R to T follows the Friis equation $g_t = \frac{G_T G_R \nu^2}{(4\pi d)^2}$ where G_T and G_R are the antenna gains of T and R, respectively, d is the distance between devices in the M2M communication link, and ν is the wavelength [17]. The battery installed at T can store up to 1 W of energy. We also assume that the minimum energy that can be used and received by T is 0.5 mW. The distributed optimization method proposed in Section III allows T and R to sequentially optimize their transmit power and energy transfer decisions. Our proposed optimization method can be regarded as a generalization of existing methods that only focus on the optimization of either T or R. In this section, we refer to our proposed method as the joint optimization method. We compare our proposed method with two existing optimization methods: M2M transmission scheduling and energy transfer scheduling. In the first method, R cannot transfer its surplus energy to T. However, T can schedule the use of its harvested energy to further improve the expected discounted payoff. This approach is also referred to as the transmit power/energy scheduling studied in [6], [18]. In the second method, T cannot schedule its energy usage but will always use the energy stored and obtained during previous time slots to send its data packets. R can however optimize its decision on whether to send its surplus energy to T at the beginning of each time slot.

As mentioned previously, the efficiency of the self-interference cancellation technology implemented by T and R plays a vital role on the performance of the full-duplex communication system. We hence compare the payoff of three above mentioned optimization approaches under different self-interference cancellation efficiencies at R in Figure 2. It can be observed that if the self-interference can be perfectly cancelled, the energy transfer from R to T will not cause any performance degradation to the M2M data transmission and R should always transfer its surplus energy to T. In this case, both joint optimization and energy transfer scheduling methods achieve significant payoff gains compared to the M2M transmission scheduling method.

However, if the self-interference cannot be perfectly cancelled, R should limit its energy transfer to avoid intolerable interference for the M2M data transmission. With the increase of the self-interference received by R, R should not send its surplus energy to T for most of the time, and hence the payoff achieved by the joint optimization method will approach that achieved by M2M transmission scheduling method.

It is known that both the transmission rate of M2M communication as well as the wireless energy transfer efficiencies suffer with the increase of transmission distance. Therefore, in Figure 3, we investigate the effect of the distance between T and R on the discounted payoff of WP-FD-M2M. It can be observed that, compared to the M2M transmission scheduling, the performance of energy transfer scheduling decreases at a much faster rate with the transmission distance. This is because, when wireless energy transfer efficiency becomes low, RF energy sent by R can only provide limited contribution to the reliability of energy supply for M2M communication. In addition, the increase of transmission distance also results in lower channel gain for the M2M data transmission as well as relatively higher performance degradation caused by the self-interference of the energy transfer from R to T. If T has to mostly rely on its energy harvested from the natural environment to support M2M data transmission, optimizing the scheduling of energy use according to the future energy availability becomes more important to improve the long-term performance of M2M communication, especially compared to the energy transfer scheduling method.

The energy availability of the surrounding environment of T and R also directly affects the performance of M2M communication. We compare the discounted payoffs of different optimization methods under different maximum levels of harvested energy in Figure 4. We observe that, with the increase of the harvested energy, the performance of M2M transmission scheduling increases faster than that of energy transfer scheduling by R. This is because although the increase of harvested energy by R can result in a larger amount of energy transferred from R to T, the larger amount of energy transferred by R also increases the self-interference to the M2M data receiving at R. We can also observe that by jointly optimizing the data transmission and energy transfer at T and R, the long-term discounted payoff can be significantly improved for WP-FD-M2M.

V. Conclusion and Future Work

In this paper, we have studied M2M communication for an IoT system powered by energy harvesting and wireless energy transfer. We have proposed a novel framework, referred to as the WP-FD-M2M, in which the transmitter and receiver can decide the energy scheduling for data transmission and the energy transfer for improvement of the wireless energy supply reliability. We have developed a stochastic game-based model, referred to as the M2M game, to investigate the interaction between an M2M transmitter and an M2M receiver. We have proved that, if both M2M transmitter and receiver optimize their decisions based on the Markov strategy, the maximum long-term performance for the M2M communication link can be achieved even without centralized control or coordination between the transmitter and receiver. At the moment, both M2M communication and wireless-powered IoT systems are still in the early stage of developments. This paper can serve as a step for future research on the next generation of wireless-powered IoT-based communication networking systems.

References

[1] K.-C. Chen and S.-Y. Lien, “Machine-to-machine communications: Technologies and challenges,” *Ad Hoc Networks*, vol. 18, pp. 3 – 23, Jul. 2014.

[2] K. Huang and V. Lau, “Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 902–912, Feb. 2014.

[3] H. Ju and R. Zhang, “Throughput maximization in wireless powered communication networks,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 418–428, Jan. 2014.

[4] X. Kang, C. Ho, and S. Sun, “Full-duplex wireless-powered communication network with energy causality,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5539–5551, Oct 2015.

[5] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, “Ambient backscatter: Wireless communication out of thin air,” in *ACM SIGCOMM*, Hong Kong, China, Aug. 2013.

[6] Y. Xiao, D. Niyato, Z. Han, and L. A. DaSilva, “Joint optimization for power scheduling and transfer in energy harvesting communication systems,” in *IEEE Global Communications Conference (GLOBECOM)*, San Diego, CA, Dec. 2015.

[7] Y. Xiao, Z. Han, and L. A. DaSilva, “Opportunistic relay selection for cooperative energy harvesting communication networks,” in *IEEE Global Communications Conference (GLOBECOM)*, Austin, TX, Dec. 2014.

[8] L. R. Varshney, “Transporting information and energy simultaneously,” in *IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, Jul. 2008.

[9] X. Zhou, R. Zhang, and C. K. Ho, “Wireless information and power transfer: architecture design and rate-energy tradeoff,” *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Oct. 2013.

[10] Y. Zeng and R. Zhang, “Full-duplex wireless-powered relay with self-energy recycling,” *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 201–204, Apr. 2015.

[11] D. Bharadia, E. McMilin, and S. Katti, “Full duplex radios,” in *ACM SIGCOMM*, Hong Kong, China, Aug. 2013, pp. 375–386.

[12] Y. Xiao, Z. Han, D. Niyato, and C. Yuen, “Bayesian reinforcement learning for energy harvesting communication systems with uncertainty,” in *IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015.

[13] Y. Xiao, D. Niyato, Z. Han, and L. DaSilva, “Dynamic energy trading for energy harvesting communication networks: A stochastic energy trading game,” *IEEE J. Sel. Area Commun. special issue on Green Communications and Networking*, vol. 33, no. 12, pp. 2718–2734, Dec. 2015.

[14] M. Bowling and M. Veloso, “An analysis of stochastic game theory for multiagent reinforcement learning,” Computer Science Department, Carnegie Mellon University, Technical report CMU-CS-00-165, Oct. 2000.

[15] L. S. Shapley, “Stochastic games,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 39, no. 10, pp. 1095–1100, Feb. 1953.

[16] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, ser. Wiley Series in Probability and Statistics. Wiley, 2005.

[17] J. A. Shaw, “Radiometry and the friis transmission equation,” *American Journal of Physics*, vol. 81, no. 1, pp. 33–37, 2013.

[18] J. Yang and S. Ulukus, “Optimal packet scheduling in an energy harvesting communication system,” *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.