Probabilistic Binary Offloading for Wireless Powered Mobile Edge Computing System

Takuya Kobayashi and Koichi Adachi
The Univeristy of Electro-Communications, Tokyo, Japan
E-mail: {t.kobayashi, adachi}@awcc.uec.ac.jp

***Abstract*—In recent years, with the advancement of the Internet-of-Things (IoT), some problems are listed. First, the computation resource is limited because of miniaturization and decreasing costs for IoT devices and sensors. Mobile edge computing (MEC) that can compute alternatively heavy tasks of wireless devices (WDs) has been proposed for solving this problem. Second, the WDs' battery management is troublesome because of the increasing number of IoT devices and sensors. Wireless power transfer (WPT) has been proposed for solving this problem. In WPT, an access point (AP) wirelessly charges batteries of WDs. Wireless powered-mobile edge computing (WP-MEC) system combining WPT and MEC is expected to solve these problems. This paper proposes a probabilistic binary offloading (PBO) system that selects one of two modes for each WD; offloading and local computing. The mode selection is determined by probabilistic control instead of the centralized control by the AP. Moreover, we propose a mode switching method that switches from offloading to local computing when the WD fails to offload tasks. We aim to reduce energy consumption of WDs and task processing delay and improving communication quality.**

I. Introduction

In recent years, the Internet-of-Things (IoT) is popularized with communication standards for wireless sensor networks [1]. In IoT networks, a large number of sensors are deployed within a small area. There are some critical issues; the limitation of computing power due to the miniaturization of IoT devices and their battery charge management. Mobile edge computing (MEC) that computes heavy tasks of wireless devices (WDs) is a solution for the problem of computing [2]. This system deploys a high-performance server at an access point (AP) or base station (BS) to process the task of the wireless devices (WDs) alternatively. In conventional cloud computing [3], the offloading communication delay is too long because of the large physical distance between WDs and a cloud server. Wireless power transfer (WPT) that charges batteries of WDs from the AP is a solution for battery management [4]. Radio frequency (RF) signal is transmitted from an AP to WDs and charge WDs' batteries wirelessly. The sensors can operate without wired charging by applying the harvested energy to the communication and computation circuits. Recently, wireless powered MEC (WP-MEC) systems combining WPT and MEC have been proposed [5]. Some conventional works considered dynamic optimization through centralized control at the AP [6] [7]. This system incurs the overheads related to energy consumption and processing delay because of exchanging control information.

There are two types of offloading strategies in MEC: partial offloading and binary offloading [2]. Partial offloading splits one task of WDs into two parts: the offloading part and the local computing part. Although this scheme can be time-efficient, task splitting may not be suitable for some applications. Binary offloading processes the entire task with either offloading mode or local computing mode. This scheme cannot process on MEC and WDs at the same time but can eliminate the division process and can simplify entire systems.

This paper proposes a decentralized probabilistic binary offloading (PBO) strategy. By PBO, each WD probabilistically selects one of two modes, i.e., offloading and local computing, instead of centralized control by an AP. In addition, we propose a mode switching method that switches from offloading mode to local computing mode when the AP fails to receive offloaded tasks. Numerical evaluation shows that the proposed PBO strategy can reduce the WDs' energy consumption and task processing delay of WP-MEC systems. Furthermore, the PBO strategy leads to improving communication quality.

The rest of the paper is organized as follows. Section 2 introduces the system model of the wireless powered MEC. Section 3 describes the proposed scheme, PBO, and mode switching. Section 4 discusses numerical results. Finally, we conclude the paper and future directions in Section 5.

II. System Model

We consider an AP equipped with a MEC server and K WDs (set $\mathcal{K} = \{1, ..., K\}$) in the system.

A. Task Generation Model

The computation tasks of each WD are randomly generated according to the Poisson process [8]. Task i of WD $k \in \mathcal{K}$ is generated at

$$t_{k,i} = t_{k,i-1} - \frac{\log X}{\lambda}, \tag{1}$$

where λ denotes the expected value of Poisson process [/sec] and X is the random variable following the uniform distribution $(0, 1)$. From now on, the expected value is called as the *task generating rate*.

B. Channel Model

Free space path loss is used for the channel model of the offloading uplink/downlink and the WPT. The effective channel gain between WD k and the AP is given by

$$g_k^{\mathrm{t}} = \left[\frac{\lambda^{\mathrm{t}}}{4\pi(d_k + 1)}\right]^2 G, \tag{2}$$

where G denotes the product of the gain of the transmit antenna and receive antenna. [9] Here, we set $\mathsf{t} = \{\text{offl}, \text{wpt}\}$, which are the parameters related to the offloading carrier and the WPT carrier.

C. Binary Mode for Task Processing

Binary offloading is one of the offloading schemes that chooses one of binary modes per task: if offloading mode 0 ($\mathcal{M}_0$) is selected, WD offloads its task, otherwise it computes the task locally ($\mathcal{M}_1$) [7] . Each WD selects one of two modes for each task. Let $m_{k,i} \in \{\mathcal{M}_0, \mathcal{M}_1\}$ denote the mode of task i of WD k.

1) Task Offloading ($m_{k,i} \in \mathcal{M}_0$): Selecting mode 0, each WD offloads a task to the AP. This paper adopts the ALOHA method [10] as the random access protocol for task offloading. All WDs use the same channel for offloading. If at least one of the following conditions is met, the offloading is assumed to fail:

- More than one tasks are offloaded simulateneously, i.e., packet collision.
- Offloaded task arrives at the AP while the AP is transmitting the computation result to a WD.

The transmission of the computation result is assumed to be always successful. When a WD offloads its task, the WD can determine the success of offloading based on the computation result is returned or not. When WDs offload tasks, the WDs can determine the success or the failure of offloading by whether the result is returned or not.

We assume that the WDs can control offloading transmit power ideally. Power control is performed so that the offloading power doesn't exceed the maximum capacity given by a modulation scheme. From Shannon's channel capacity theorem [11], the offloading power is given by

$$P_k^{\text{offl}} = \begin{cases} \frac{(2^{C^{\max}}-1)\sigma^2}{g_k^{\text{offl}}} & (\text{if } C_k \geq C^{\max}) \\ P_{\text{WD}}^{\max} & (\text{otherwise}), \end{cases} \tag{3}$$

where σ^2 denotes the noise power [W], $C^{\max}$ denotes the maximum capacity [bit/sec/Hz], and $P_{\text{WD}}^{\max}$ denotes the maximum offloading power of the WD [W].

The duration required for offloading task i from WD $k \in \mathcal{K}$ to the AP is given by

$$\tau_{k,i}^{\text{offl}} = \frac{L_{k,i}}{R_k} = \frac{L_{k,i}}{B\log_2\left(1+\frac{g_k^{\text{offl}}P^{\text{offl}}}{\sigma^2}\right)}, \tag{4}$$

where $L_{k,i}$ denotes the task size [bits], R_k denotes the offloading rate [bits/sec] and B denotes the bandwidth [Hz]. The energy required for offloading is given by

$$E_{k,i}^{\text{offl}} = \tau_k^{\text{offl}}P_k^{\text{offl}} = \frac{L_k}{B\log_2\left(1+\frac{g_k^{\text{offl}}P_k^{\text{offl}}}{\sigma^2}\right)}P_k^{\text{offl}}. \tag{5}$$

Upon successful offloading, MEC computation and the return of the computation results are performed. The computation time of task i offloaded from WD k at MEC server is given by

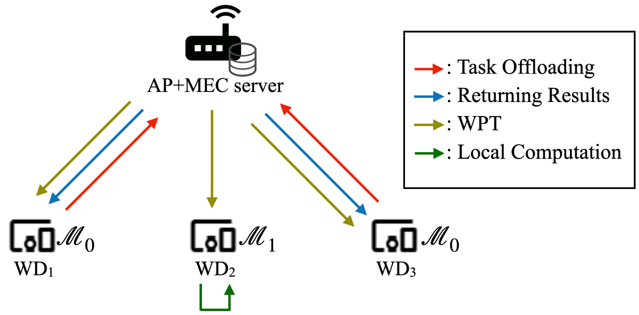


Fig. 1. A schematic diagram of WP-MEC with binary offloading.

$$\tau_{k,i}^{\text{M}} = \frac{A_kL_{k,i}}{f^{\text{M}}}, \tag{6}$$

where f^{M} denotes the CPU frequency of the MEC server [Hz] ([CPU cycles/sec]) and A_k denotes the number of CPU cycles per bit of processing [CPU cycles/bit], which is called as the *task type*.

The return time for the computation result is given by

$$\tau_{k,i}^{\text{re}} = \frac{H_{k,i}}{R^{\text{ap}}} = \frac{H_{k,i}}{B\log_2\left(1+\frac{g_k^{\text{offl}}P^{\text{ap}}}{\sigma^2}\right)}, \tag{7}$$

where $H_{k,i}$ denotes the size of the computation results of task i of WD k [bits], R^{ap} denotes the transmit rate of the AP [bits/sec] and P^{ap} denotes the return power of the calculation results [W]. The transmit power of the AP shall be controlled not to exceed the maximum capacity in the same way as the transmit power of the WD.

2) Local Computing ($m_{k,i} \in \mathcal{M}_1$): Selecting mode 1, each WD performs local computing with its own internal computation circuit. The local computing time for task i of WD k is given by

$$\tau_{k,i}^{\text{loc}} = \frac{A_kL_{k,i}}{f_k}, \tag{8}$$

where f_k denotes the CPU frequency of WD k [Hz] ([CPU cycles/sec]). Using CPU effective capacitance coefficient ζ_k, the local computing power is expressed as $\zeta_k f_k^3$ [12]. The energy required for the local computing is given by

$$E_{k,i}^{\text{loc}} = \zeta_k f_k^3 \tau_{k,i}^{\text{loc}} = \zeta_k f_k^2 A_k L_{k,i}. \tag{9}$$

D. WPT

The WPT is performed using beamforming for each WD [7]. The beam energy harvesting is performed in a time division manner. The amount of harvested energy of WD k by WPT is expressed as

$$E_k^{\text{wpt}} = \tau_k^{\text{wpt}}\eta_k g_k^{\text{wpt}}P^{\text{wpt}}, \tag{10}$$

where τ_k^{wpt} denotes run time of the WPT [sec], η_k denotes the EH efficiency of the WDs and P^{wpt} denotes the WPT power of the AP [W].

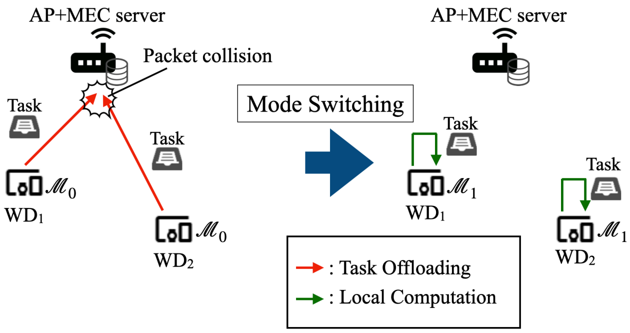


Fig. 2. A schematic diagram of the mode switching.

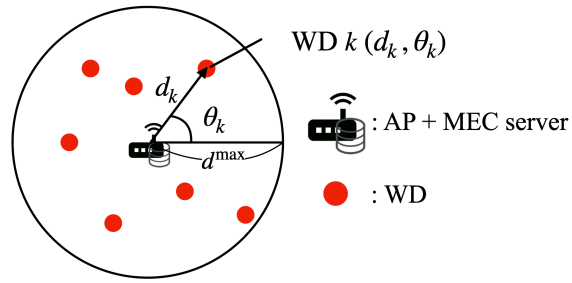


Fig. 3. Simulation area.

E. Total Processing Time and Energy Consumption

The total processing time is defined as the time from the task generation to receiving results or as the local computing time. The total processing time of task i of WD k is

$$T_{k,i} = \begin{cases} \tau_{k,i}^{\text{offl}} + \tau_{k,i}^{\text{M}} + \tau_{k,i}^{\text{re}} & (m_{k,i} \in \mathcal{M}_0) \\ \tau_{k,i}^{\text{loc}} & (m_{k,i} \in \mathcal{M}_1). \end{cases} \tag{11}$$

Moreover, it is assumed that the energy storage of WD depends only on offload transmission, local computing and EH. The energy storage of WD k after the processing of task i is expressed as

$$\varepsilon_{k,i} = \begin{cases} \max\left(\varepsilon_{k,i-1} + E_k^{\text{wpt}} - E_{k,i}^{\text{offl}}, 0\right) & (m_{k,i} \in \mathcal{M}_0) \\ \max\left(\varepsilon_{k,i-1} + E_k^{\text{wpt}} - E_{k,i}^{\text{loc}}, 0\right) & (m_{k,i} \in \mathcal{M}_1), \end{cases} \tag{12}$$

where $\varepsilon_{k,0}$ denotes initial energy storage [J].

III. Proposed Method

This section describes the details of the proposed PBO strategy and mode switching, which further improve the performance of WP-MEC.

A. Probabilistic Binary Offloading (PBO)

The proposed PBO strategy probabilistically selects one of the binary modes in an autonomous manner. An offloading probability, $p_k \in [0, 1]$, is set for each WD. A schematic diagram of the application of WP-MEC with binary offloading is shown in Fig. 1.

Each WD determines its own mode based on the offloading probability, which is decided in a decentralized manner. For example, if $p_k = 1.0$, WD k always selects mode 0 and if $p_k = 0.5$, WD k has 50% chance selecting mode 0 or mode 1. This paper assumes that the offloading probability is fixed for each WD. Since the required energy for offloading becomes larger as the distance between a WD and the AP becomes longer. It is more reasonable to have a larger offloading probability for the WD closer to the AP. Based on this rationale, this paper proposes an offloading probability p_k, which is given by

$$p_k = 1 - \frac{f(d_k)}{f(d_{\max})}, \tag{13}$$

where d_k denotes the distance between WD k and an AP [m] and $d_{\max}$ denotes the maximum communication distance. Function $f(x)$ can be variously considered. Eq. (5) shows that the offloading energy consumption follows $O\left(\left[\log_2\left(1 + \frac{1}{(d_k+1)^2}\right)\right]^{-1}\right)$. This paper uses the following equation:

$$f(x) = \frac{1}{\log_2\left(1 + \frac{1}{(x+1)^2}\right)}. \tag{14}$$

Suppose the offloading energy required for the WDs located at the coverage edge is larger than that for local computing. Letting those WDs compute the tasks locally, their battery life can be extended compared to always offloading. Furthermore, it is possible to reduce the simultaneous transmissions. This results in reducing the number of packet collision.

B. Mode Switching in Case of Offloading Failure

In the general random access method, a WD will receive an ACK (acknowledgment) signal from an AP when its offloaded task is correctly received by the AP. If a WD could not receive the ACK signal, it retransmits the same packet. However, this results in a higher packet collision. Thus we propose a mode switching method. If a WD could not receive a returning result, the WD switches from mode 0 to mode 1 instead of packet retransmission. This method enables us to process all tasks without any failure. Fig. 2 shows a schematic diagram of the mode switching at the time of packet collision.

IV. Numerical Results

A. Simulation Parameters

The simulation parameters are shown in Table 1 [6] [7]. Here, the position of WD $k \in \mathcal{K}$ is determined by the distance $d_k \sim \mathcal{U}(0, 50)$ [m] and angle $\theta_k \sim \mathcal{U}(0, 2\pi)$ [rad]. $\mathcal{U}(a, b)$ denotes the uniform distribution $[a, b]$. The WDs' locations are assumed to be fixed and available at the AP. The model of the simulation area is shown in Fig.3.

The highest modulation scheme of the WDs' offloading packets is quadrature phase shift keying (QPSK), so the maximum channel capacity is $C^{\max} = 2$ [bit/sec/Hz]. The bandwidth is set to $B = 2$ [MHz] and the maximum transmit rate is $R^{\max} = 4$ [Mbits/sec]. The maximum transmit power

TABLE I
SIMULATION PARAMETERS

Prameter	Value
Maximum communication distance $d_{\max}$	50 [m]
The number of WDs K	50
Task type A_k	10^3 [CPU cycles/bit]
Return result standby time $T^{\rm stb}$	0.5 [sec]
Occuring task size L_k	$[10, 100]$ [kbits]
Resulting task size H_k	$[L_k/100, L_k/10]$ [bits]
CPU capacitance coefficient ζ_k	10^{-29}
EH efficiency η_k	0.3
Task generating rate λ	0.1 [/sec]
CPU frequency of WDs f_k	$[10, 50]$ [MHz]
CPU frequency of MEC $f^{\rm M}$	10 [GHz]
Run time of WPT $\tau_k^{\rm wpt}$	1.0 [sec]
Noise power σ^2	10^{-11} [W]
WPT transmit power $P^{\rm wpt}$	3 [W]
WPT carrier frequency $f_c^{\rm wpt}$	915 [MHz]
Offloading carrier frequency $f_c^{\rm offl}$	2.4 [GHz]
Product of antenna gain G	4.11
Initial energy storage $\varepsilon_{k,0}$	10^{-3} [J]

for WD offloading is $P_{\rm WD}^{\max} = 20$ [mW] and the maximum transmit power of the AP is $P_{\rm M}^{\max} = 100$ [mW].

B. Simulation Results

1) Performance Metrics: Slotted offloading system is considered as benchmarks. In this system, the AP decides the offloading strategies in a centralized manner to minimize delay or consumption. The AP has complete knowledge of task generation, channel condition, and battery level of each WD. The AP calculates the processing delay or energy consumption of offloading and local computing for each task. Based on the calculation, the AP selects one of the binary modes for minimizing the processing delay or energy consumption. The processing delay of a task in the slotted offloading system is expressed as

$$T_{k,i}^{\rm s} = T_{k,i} + \tau_{k,i}^{\rm p} - \tau_{k,i}^{\rm g}, \tag{15}$$

where $\tau_{k,i}^{\rm p}$ denotes the time of starting task processing [sec] and $\tau_{k,i}^{\rm g}$ denotes the time of task generating.

The processing delay for this slotted offloading system is shown in Fig. 4. Since multiple offloading cannot be performed simultaneously, the WDs offload one by one in the generated order. When multiple tasks need to be computed locally in the same slot at a WD, the tasks are processed in the generated order. It is assumed that information exchange for scheduling is performed ideally, i.e., with no power consumption and no delay.

2) Packet Delivery Rate (PDR): Fig. 5 shows the PDR performances of the proposed approach (PBO with mode switching) and that of PBO with retransmission as a function of task generating rate λ. As a comparison, the case with $p_k = 1$ is included. Proposed approach can improve PDR as task generating rate λ increases compared to retransmission and fixed $p_k = 1$. Mode switching with PBO can reduce the number of offloading failure by approximately 35% over retransmisssion. and mode switching with fixed $p_k = 1$ can

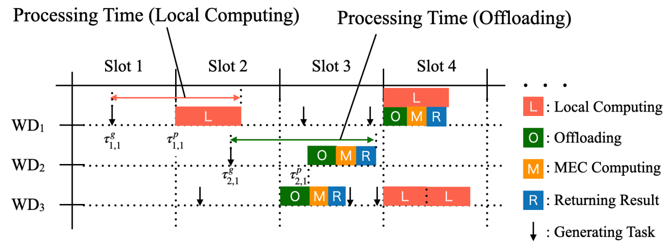


Fig. 4. The processing delay for slotted offloading system.

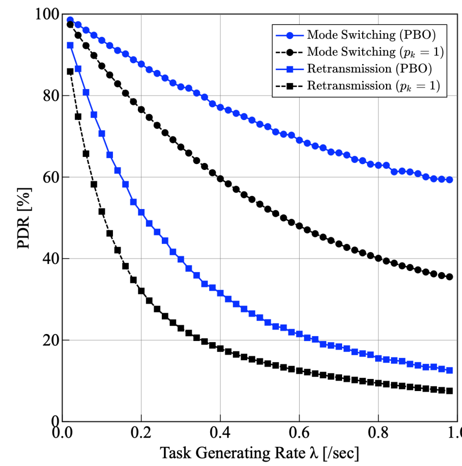
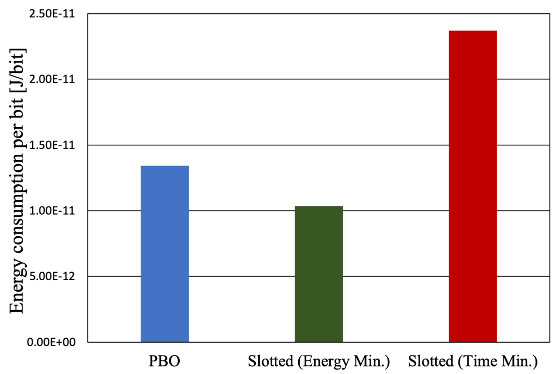


Fig. 5. Comparison of PDR

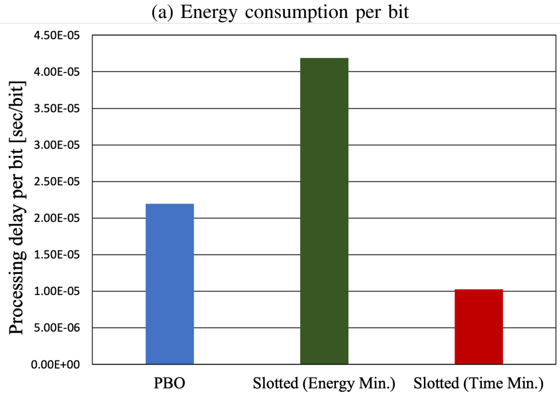
reduce by approximately 50% retransmission. Furthermore, PBO with mode switching can reduce by approximately 50% fixed $p_k = 1$. From these results, mode switching and PBO can reduce the number of packet collision as task generating rate λ increases. This is because the mode switching and PBO reduce the number of task offloading and then improve the communication quality.

3) Comparison with Ideal Slotted Offloading System: Next, the performances of the proposed PBO strategy are compared against those of the slotted system. The slot length is set as $\tau_s = 1.0$ [sec]. Fig. 6 (a) shows that the proposed PBO strategy requires about 1.3 times more energy than the slotted system with energy minimization. However, it requires about 0.57 times less energy than the slotted system with delay minimization. Fig. 6 (b) shows that the PBO strategy requires about 2.2 times longer delay than the slotted system with delay minimization. However, it requires about 0.52 times shorter delay than the slotted system with energy minimization. Although the proposed PBO strategy requires no information exchange among the WDs and the AP, these results show that the PBO strategy can balance the trade-off between energy consumption and processing delay.

4) Battery Lifetime CDF: Fig. 7 shows the cumulative distribution function (CDF) when one of the WDs reaches zero battery first. In the figure, "Random" denotes the system



(a) Energy consumption per bit



(b) Processing time per bit

Fig. 6. Comparison of the proposed method and the ideal slotted method.

with random offloading probability for each WD. With PBO, battery life can approach the ideal case with energy minimizing and can be extended compared to the other two methods. The figure shows that the proposed PBO strategy can lengthen battery life, and it is close to the slotted system with energy minimization. We can also say that the proposed PBO strategy with mode switching can balance the trade-off between energy consumption and processing time.

V. Conclusion

In this paper, we proposed the probabilistic binary offloading (PBO) strategy for a WP-MEC system. In the proposed PBO strategy, each WD probabilistically selects either offloading or local computing. To further enhance the performance, mode switching. If a WD knows that the offloading failed, it switches from offloading mode to local computing mode. The simulation results elucidated that the proposed PBO strategy with mode switching can improve PDR significantly. We showed that the proposed system could balance the trade-off between energy consumption and processing delay compared with the centralized slotted offloading system.

Acknowledgment

This work was supported by the European Commission in the framework of the H2020-EUJ-02-2018 project 5G-

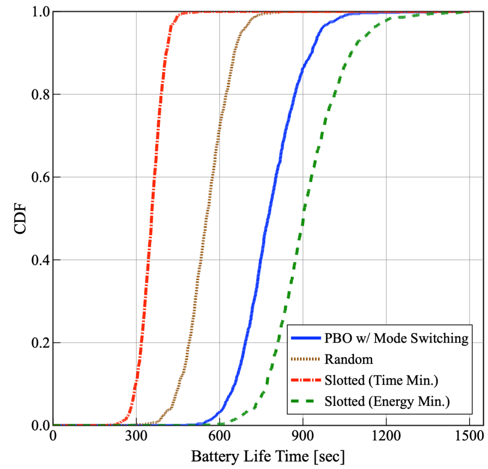


Fig. 7. Comparison of CDF.

Enhance (Grant agreement no.815056) and the Ministry of Internal Affairs and Communications (MIC) of Japan.

References

[1] A. A. Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications,” *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[2] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, “A Survey on Mobile Edge Computing: The Communication Perspective,” *IEEE Commun. Surveys Tuts*, vol. 19, no. 4, pp. 2322-2358, Aug. 2017.

[3] Q. Zhang, L. Cheng, and R. Boutaba, “Cloud Computing: State-of-the-Art and Research Challenges,” *J. Internet Services Appl.*, vol. 1, no. 1, pp. 7–18, May 2010.

[4] Y. Zeng, B. Clerckx, and R. Zhang, “Communications and Signals Design for Wireless Power Transmission,” *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2264-2290, May 2017.

[5] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, “Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4569-4581, Sep. 2013.

[6] F. Wang, J. Xu, X. Wang, and S. Cui, “Joint Offloading and Computing Optimization in Wireless Powered Mobile-Edge Computing Systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1784-1797, Mar. 2018.

[7] S. Bi and Y. J. Zhang, “Computation Rate Maximization for Wireless Powered Mobile-Edge Computing with Binary Computation Offloading,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4177-4190, Jun. 2018.

[8] A. F. Molisch, *Wireless Communications Second Edition*, United States of America: Wiley-IEEE Press, 2011.

[9] H. Kawabata, K. Ishibashi, S. Vuppala, and G. Abreu, “Robust Relay Selection for Large-Scale Energy-Harvesting IoT Networks,” *IEEE Internet of Things journal*, vol. 4, no. 2, pp. 384-392, Apr. 2017.

[10] B. Sklar and F. J. Harris, *Digital Communications: Fundamentals and Applications*, Prentice-hall Englewood Cliffs, NJ, 1988.

[11] C. E. Shannon, “A Mathematical Theory of Communication,” *The Bell System Technical Journal*, vol. 27, pp. 379-423, Jul., 1948.

[12] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, “Mobile-Edge Computing: Partial Computation Offloading Using Dynamic Voltage Scaling,” *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4268–4282, Oct. 2016.