Energy Efficient Resource Allocation and Admission Control for D2D-aided Collaborative Mobile Clouds

Foad Hajiaghajani†, Ramtin Davoudi§, and Mehdi Rasti†
†Department of Computer Engineering and IT, Amirkabir University of Technology, Tehran, Iran
§Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran
Email: {f.hajiaghajani, r_davoudi, rasti}@aut.ac.ir

***Abstract*—In this paper, we investigate the joint admission control and resource allocation problem for collaborative computation under fading channels. We develop an Internet-of-Things (IoT) framework in which establishing Device-to-Device (D2D) communications, resource-poor wearable Source Mobile Terminals (SMTs) may offload their computations to resource-rich Processing Mobile Terminals (PMTs), or execute them locally, so as to save energy. Considering the offloading scenario, first, a probabilistic admission control algorithm is proposed for Mobile Terminals (MTs) taking both the deadline and energy harvesting constraints into account. Then, the joint CPU clock frequency/transmit power allocation and collaborative pair selection problem for MTs is addressed mathematically. For local execution scenario, optimal CPU clock frequencies are obtained for SMTs. Finally, based on energy consumption and outage imposed by each scenario, SMTs decide whether to offload their computations or execute them locally. Simulation results demonstrate that the proposed D2D-aided Collaborative Mobile Cloud (DCMC) approach attains a near-optimal energy expenditure in a semi-feasible system while effectively mitigating outage ratio of MTs.**

***Index Terms*—collaborative mobile cloud, D2D communication, resource allocation, wireless networks.**

I. Introduction

The realization of Internet-of-Things (IoT) [1] will connect tens of billions of mobile terminals (MTs), such as wearable gadgets and smart phones, to Internet via 5G cellular networks. In particular, wearable MTs are defined as technology to be worn on the body able to perform various tasks, e.g., image processing, diet tracking, mental stress detection or in general, monitoring physiological functions and providing biofeedback [2]. Such IoT applications are delay sensitive and CPU-intensive wherein energy is a very precious resource for battery-operated wearable devices. Consequently, prolonging battery lives of MTs along with improving their computing performance pose significant challenges for designing next-generation wireless networks.

Energy harvesting by Wireless Power Transfer (WPT) [3] is one of the promising technologies which realizes the capability of self-maintenance in wireless networks. Specifically, WPT enables MTs to harvest energy through capturing microwave radiations powered by Base Station (BS) and convert them into a direct current voltage. By recent developments, it has been shown that a wireless signal may carry both energy and data at the same time introducing simultaneous wireless information and power transfer (SWIPT) research area. Executing tasks on local CPU, however, may violate the delay requirement of CPU-intensive applications, specifically in real-time health monitoring systems where data is time sensitive. A similar example is face recognition which takes as long as 38 minutes if executed on smart glasses [4]. Collaborative Mobile Cloud (CMC) [5] is another approach aiming to reduce both the computation delay and energy consumption through offloading computations to nearby powerful MTs. Despite traditional Mobile Cloud Computing (MCC) which necessitates the involvement of the BS to exchange tasks between MT and cloud, in CMCs, MTs cooperatively form a local cloud and contribute to execution of tasks of each other. In conventional CMCs, MTs actively use two wireless interfaces: one to communicate with the BS (LTE), and the other to cooperate with other MTs (WiFi) [7]. With the advent of Device-to-Device (D2D) communications [6],[16], MTs reach higher data rate and communication range (compared to WiFi), while D2D links between MTs can be established over the same interface as the one used for cellular communications. In this paper, these technologies are effectively integrated so as to develop a novel framework for IoT.

Mobile cloud computing has been an attractive research area in computer science where the focus of most works in the literature is on designing energy efficient computation offloading and resource allocation approaches. Previous studies mainly addressed the resource allocation problem for MTs based on Lyapunov optimization [7],[8], Lagrange relaxation under delay constraint [9]–[11], game theory [12],[14], and heuristic approaches [13],[15]. In [7], the authors considered different types of workloads and proposed a joint scheduling and resource allocation scheme for tasks and CPU speed, respectively. In [8], a delay-sensitive offloading strategy is proposed considering a network in which the achievable data rates of MTs are subject to change over time. In recent work, however, communications are taken place either over 3G or through WiFi networks. Considering WPT, two frameworks for CPU clock frequency and radio resource allocation are presented in [9] and [11], respectively. The resource allocation scheme proposed in [9], optimizes CPU clock frequency and data transmission scheduling by solving corresponding convex optimization problems in both local execution and cloud computation scenarios. Nevertheless, [8],[9],[11] did not consider allocation of both network and CPU resources, simultaneously. In [12], adopting the Slotted Aloha medium access control protocol, the authors studied the time and energy minimization problems for computation offloading using game theory. Likewise, in [14], a coalitional game is developed to minimize sum-energy cost of MTs, albeit execution delay constraint is ignored. In [13] and [15], the authors developed heuristic approaches based on user-provided

978-3-901882-90-6/17 ©2017 IFIP

resources platform and greedy algorithm, respectively, to extend lifetime of MTs.

In summary, most of previous works on computation offloading are restrictive in some ways: (i) As in conventional MCC, MTs are allowed to offload their computations only to cloud servers, [7],[9],[10],[12]. However, not only does such an approach rely on network compatibility, but the large delay and energy waste due to long-range BS-MT communications may also not be worth the offloading effort. (ii) Perfect channel state information (CSI) of all links are required to be available at BS [8],[11],[13]–[15], which incur excessive signalling overhead. (iii) Unrealistic system models are considered, e.g., there is only one MT in the system, [7],[9],[10], or channel gains/receiving power of all MTs are assumed the same at different locations, [8],[13],[15]. Meanwhile, major works considered absolute feasible systems which are not the case for reality.

In this paper, we develop a D2D-aided collaborative mobile cloud (DCMC) by integrating overlaying D2D communications, MCC and WPT technologies with a modern delay sensitive IoT application. In contrast to existing works, our framework eliminates the need for cloud servers and long-range communications between the BS and MTs. We consider a multi-user network in which precise CSI between MTs are unavailable at BS and therefore, resources are allocated to MTs using statistical CSI to avoid large signalling overheads in the network. Moreover, every MT compensates for CPU energy dissipation by harvesting energy transferred from the BS. We study the problem of CPU and network resources allocation of MTs in a network where differently from most previous studies, local execution of tasks may not be feasible for every MT. Thus, a wearable cellular MT has to decide whether to offload its computation to a nearby resource-rich MT, or execute it locally, so as to minimize the energy consumption and outage ratio. We first formally state the joint CPU clock frequency/transmission power allocation and pair selection problem for computation offloading scenario under fading channels. Considering both the deadline and energy harvesting constraints, a probabilistic distance-based admission control algorithm is proposed to maximize the number of admitted MTs. Then, optimal CPU clock frequencies and transmission power of MTs are obtained mathematically to minimize the sum-energy consumption. Second, for local execution scenario, we derive the optimal CPU clock frequencies of MTs to minimize the computation energy consumption under deadline and energy harvesting constraints. Finally, taking the tradeoff between the energy dissipation and outage ratio of MTs into account, MTs decide whether to offload their computations or execute them locally. Simulation results demonstrate the effectiveness of the proposed DCMC scheme in terms of sum-energy consumption and outage ratio of MTs. Fig. 1, illustrates the overall procedure of our framework.

The rest of the paper is organized as follows. In Section II, we describe the system model. In Section III, the proposed joint probabilistic admission control, resource allocation and pair selection scheme is presented under computation offloading and local execution scenarios. We present our simulation results in Section IV. Finally, conclusions are drawn in Section V.

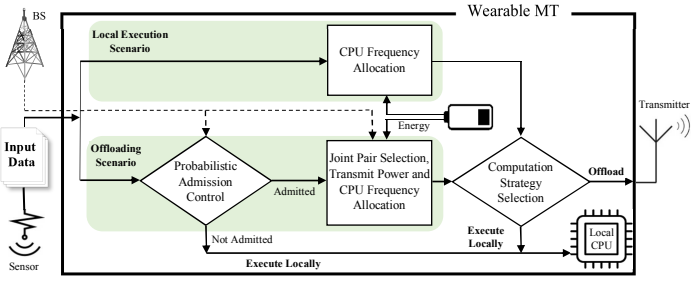


Fig. 1: Overall procedure of the proposed DCMC framework.

II. System Model

A. Network Model

Consider an OFDMA-based circular cellular cell with radius R including a central Base Station (BS) and N active Mobile Terminals (MTs) which together form the set $\mathcal{N}$. As shown in Fig. 2, MTs are divided into two sub-sets: i) a set of resource-poor wearable Source Mobile Terminals (SMTs) denoted by $\mathcal{N}_s \subseteq \mathcal{N}$ where $\mid \mathcal{N}_s \mid = N_s$, and ii) a set of resource-rich Processing Mobile Terminals (PMTs) denoted by $\mathcal{N}_p \subseteq \mathcal{N}$ where $\mid \mathcal{N}_p \mid = N_p$. Apparently, $\mathcal{N} = \mathcal{N}_s \bigcup \mathcal{N}_p$ and $\mathcal{N}_s \bigcap \mathcal{N}_p = \emptyset$. In this paper, it is assumed that $N_p \geq N_s$. Every SMT $i \in \mathcal{N}_s$ has an executable application which is abstracted by an application profile $A_i(L_i, \widetilde{L}_i, T_i)$ comprising of the following parameters:

- Input data size L_i: Number of bits in input data for application assigned to SMT i (size of code is negligible).
- Output data size $\widetilde{L}_i$: Number of bits in output data after execution of the application belonging to SMT i. Depending on the application, the size of output data may be larger or smaller than the input data. We assume that $\widetilde{L}_i \triangleq \lceil \theta L_i \rceil$ for some $\theta \in (0, 1]$ where $\lceil x \rceil$ denotes the ceiling of x.
- Completion Deadline T_i: The delay threshold for SMT i before which its application should be completed.

A mobile application belonging to SMT i can be either executed locally or offloaded to a potential PMT $j \in \mathcal{N}_p$ by establishing a half-duplex Device-to-Device (D2D) communication. PMTs are assumed to be powerful idle devices (e.g., smart phones) which, due to possessing rich resources, can be selected by wearable SMTs for collaborative computation to maximize energy saving. Let $\boldsymbol{\rho} \triangleq \{\rho_{i,j}\}_{N_s \times N_p}$ be the computation assignment matrix where $\rho_{i,j} = 1$, if the computation belonging to SMT i is offloaded to PMT j, and $\rho_{i,j} = 0$, otherwise.

B. Channel Model

We adopt independent block fading channels for D2D links, i.e., channel states are assumed to remain constant during a single system snapshot while are independent of previous states. The channel gain between SMT i and PMT j is modeled as

$$h_{i,j} = \zeta_{i,j}^{-1} \overline{h}_{i,j} = K \zeta_{i,j}^{-1} d_{i,j}^{-\alpha}, \tag{1}$$

where K and α are pathloss constant and exponent, respectively, $\zeta_{i,j}$ is the channel fading coefficient, and $d_{i,j}$ is the distance between SMT i and PMT j. Throughout this paper, we use $\overline{h}_{i,j} = K d_{i,j}^{-\alpha}$ to indicate the channel gain without fading coefficient. D2D communications between SMTs and PMTs are

assumed to be approximately symmetric, i.e., $h_{i,j} \simeq h_{j,i}$. In computation offloading, the BS is in charge of synchronization and resource allocation for MTs for which certain CSI is needed. In practice, $h_{B,i}$ and $h_{B,j}$, the channel gains from the BS to SMT i and PMT j, respectively, can be obtained at BS using classical channel estimation methods, since these links are directly connected to the BS. Acquiring instantaneous $h_{i,j}$, however, incurs high signaling overhead for frequent CSI updates which therefore is not practical. For this reason, we utilize the statistical characteristic of SMT-PMT links to avoid instantaneous CSI feedbacks while nullifying fading effect. We assume that the fading coefficient $\zeta_{i,j}$ is ergodic and stationary with *probability density function* (pdf), $f(\zeta_{i,j})$, and *cumulative density function* (cdf), $F(\zeta_{i,j})$. We denote by σ_N^2, the power of additive white Gaussian noise (AWGN) on every channel. In this paper, we also made the following assumptions.

- SMT i and PMT j contributed to a collaborative computation are allocated the same transmission power, $P_{i,j}$.
- The BS has sufficient spectrum resources and D2D communications are established on dedicated resources.

C. QoS and Energy Consumption Models

Adopting the common approach [9], the number of CPU cycles for an application assigned to SMT i can be expressed as $W_i = L_i X_i$, where X_i is the number of CPU cycles required for computing 1-bit data. As in [10], X_i can be modeled as a random variable with Gamma distribution. The D2D-aided collaborative mobile cloud framework in this paper follows the following energy models:

1) *CPU Computation Energy*: the energy consumed in a single CPU cycle for computing 1-bit input data by a MT is given by γf^2 where γ is a constant determined by the switched capacitance and f is the CPU clock frequency.
2) *D2D Communication Energy*: the energy consumed by MT i for transmitting 1-bit input data to MT j is expressed as $P_{i,j}[log_2(1+\frac{P_{i,j}h_{i,j}}{\sigma_N^2})]^{-1}$.
3) *Idle Energy*: the energy consumption of MT i during the time at which MT j is computing 1-bit data offloaded by MT i is given by $P_i^{\text{idle}}(f_{j,i}^{-1})$, where P_i^{idle} is the CPU idle power for MT i.
4) *Harvested Energy*: the energy harvested by MT i in one time unit is given by $\delta P_{B,i} h_{B,i}$ where the constant $0 < \delta \leq 1$ represents the energy converting efficiency and $P_{B,i}$ is BS's transmission power for MT i.

We assume that CPU clock frequency remains static in every cycle during the computation. We denote by f_i^{s} and $f_{j,i}^{\text{p}}$, the CPU frequency of SMT i (local execution scenario) and PMT j collaborating with SMT i (offloading scenario), respectively.

The resource allocation problem has to satisfy delay constraint for SMT i, i.e., the completion time $t_i^{\text{loc}} \triangleq W_i f_i^{s\,-1} \leq T_i$, if the application is executed locally. In case of offloading, however, since the channel fading coefficient, $\zeta_{i,j}$ is unavailable at BS, precise channel capacity cannot be determined. Hence, we define the following delay violation probability to guarantee a minimum Quality-of-Service (QoS) for SMT i when its computation is offloaded to PMT j,

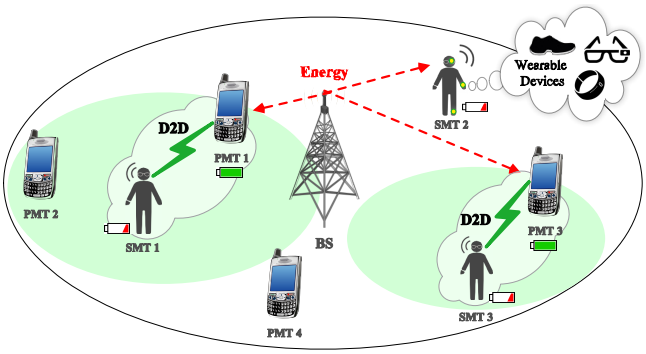


Fig. 2: System model of D2D-aided collaborative mobile cloud.

$$Pr\left\{ t_{i,j}^{\text{off}} \triangleq \frac{\lceil L_i(1+\theta)\rceil}{log_2(1+\frac{P_{i,j}h_{i,j}}{\sigma_N^2})} + W_i f_{j,i}^{p\,-1} > T_i \right\} \leq \varphi, \quad (2)$$

the first and second terms of $t_{i,j}^{\text{off}}$ account for communication (sending L_i and receiving $\widetilde{L}_i$) and computation times, respectively. φ is the maximum acceptable delay violation probability. In absence of perfect CSI of D2D link, delay violation occurs whenever the SMT does not meet the deadline, i.e., $t_{i,j}^{\text{off}} > T_i$.

As mentioned earlier, every MT executing an application compensates for CPU energy dissipation by harvesting energy transmitted from the BS. Specifically, the energy consumed by CPU should not exceed the energy harvested from BS during the same time period as computation takes, i.e.,

$$W_i \gamma f_i^{s\,2} \leq \delta P_{B,i} h_{B,i} (\frac{W_i}{f_i^s}), \ \forall i \in \mathcal{N}_s, \quad (3)$$

$$W_i \gamma f_{j,i}^{p\,\,2} \leq \delta P_{B,j} h_{B,j} (\frac{W_i}{f_{j,i}^p}), \ \forall i \in \mathcal{N}_s, j \in \mathcal{N}_p. \quad (4)$$

The energy can be harvested during the regular downlink transmissions of BS and can be consumed concurrently for computation based on MT's energy saving plan [9].

III. Energy Efficient Resource Allocation for MTs

A SMT should choose a computation strategy which enables it to save more energy (or expend less energy) while meeting a minimum QoS. To this end, in this section, we address the energy minimization problem for SMTs under two independent scenarios: A) computation offloading, and B) local execution.

A. Computation Offloading Scenario

Under this scenario, SMTs have to offload their computations to potential PMTs in order to minimize their energy consumption. The expected energy consumption for SMT i collaborating with PMT j, is expressed as

$$\mathbb{E}\left\{ [\xi_{i,j}^{\text{off}} \triangleq \frac{P_{i,j}(L_i + \widetilde{L}_i h_{i,j})}{log_2(1+\frac{P_{i,j}h_{i,j}}{\sigma_N^2})} + P_i^{\text{idle}}(W_i f_{j,i}^{\text{p}\,-1})] \mid t_{i,j}^{\text{off}} \leq T_i \right\}, \quad (5)$$

note that the receiving power at SMT i is $P_{i,j}h_{i,j}$. Thus, the first term of $\xi_{i,j}^{\text{off}}$ includes the energy consumption of both transmitting and receiving L_i and $\widetilde{L}_i$ bits, respectively. Now, we formally state the problem of joint pair selection and resource allocation for MTs in offloading scenario as follows,

$$\min_{\rho_{i,j},f^{\mathrm{p}}_{j,i},P_{i,j}} \sum_{i\in\mathcal{N}_s}\sum_{j\in\mathcal{N}_p} \rho_{i,j}\mathbb{E}\left\{[\xi^{\mathrm{off}}_{i,j}] \mid t^{\mathrm{off}}_{i,j}\le T_i\right\}, \quad (6)$$

$$\text{s.t.}\quad Pr\left\{\frac{\lceil L_i(1+\theta)\rceil}{log_2(1+\frac{P_{i,j}h_{i,j}}{\sigma_N^2})}+\frac{W_i}{f^{\mathrm{p}}_{j,i}}>T_i\right\}\le\varphi,\forall i\in\mathcal{N}_s, \quad (C1)$$

$$W_i\gamma {f^{p}_{j,i}}^2\le\delta P_{B,j}h_{B,j}(\frac{W_i}{f^{p}_{j,i}}),\forall j\in\mathcal{N}_p, \quad (C2)$$

$$\sum_{i\in\mathcal{N}_s}\rho_{i,j}\le 1,\forall j\in\mathcal{N}_p,\quad \sum_{j\in\mathcal{N}_p}\rho_{i,j}\le 1,\forall i\in\mathcal{N}_s, \quad (C3)$$

$$P_{j,i}\in(0,\overline{P}^{\max}], f^{\mathrm{p}}_{j,i}\in(0,\overline{f}^{\max}_j],\forall i\in\mathcal{N}_s,j\in\mathcal{N}_p, \quad (C4)$$

$$\rho_{i,j}\in\{0,1\},\forall i\in\mathcal{N}_s,j\in\mathcal{N}_p, \quad (C5)$$

where $\overline{P}^{\max}$ and $\overline{f}^{\max}_j$ represent the maximum transmission power and CPU frequency for PMT j, respectively. As the combinatorial optimization problem above is with nonlinear constraints including both discrete and continuous variables, it is difficult to obtain its solution directly (see [18], Chap. 4). Especially, its complexity dramatically grows with the number of MTs. To address (6), we divide it into two sub-problems and solve them one by one: 1) admission control, and 2) resource allocation and pair selection for collaborative computation.

1) Probabilistic Admission Control for MTs: From the SMT's viewpoint, choosing unsuitable PMTs strongly influences on transmission energy and execution velocity. Hence, a SMT first needs to determine whether a PMT is admissible or not. A PMT j can be admitted to collaboration with SMT i, if constraints (C1), (C2), and (C4) in (6) hold concurrently. At this stage, the major goal is to determine a policy for SMT-PMT pairs upon which initiating a collaboration is conditional. Meanwhile, the policy should allow as many reliable PMTs as possible to be admitted to collaboration with SMTs.

Theorem 1. The number of admitted PMTs for collaboration is maximized when $P_{i,j}$ and $f^{\mathrm{p}}_{j,i}$ reach their maximum level.

Proof. Maximizing the number of admitted PMTs is equivalent to minimizing the ratio of unsatisfied MTs contributing to collaborative computation (i.e., outage ratio). Let's assume that (C2) in (6) holds for given $f^{\mathrm{p}}_{j,i}$. Therefore, to minimize the outage ratio, the probability in (C1) should be minimized, i.e.,

$$\min_{P_{i,j},f^{\mathrm{p}}_{j,i}} Pr\left\{\frac{\lceil L_i(1+\theta)\rceil}{log_2(1+\frac{P_{i,j}\overline{h}_{i,j}}{\sigma_N^2\zeta_{i,j}})}+\frac{W_i}{f^{\mathrm{p}}_{j,i}}>T_i\right\}$$
$$=\min_{P_{i,j},f^{\mathrm{p}}_{j,i}}(1-\int_0^u[\zeta_{i,j}]d\zeta_{i,j})\equiv\max_{P_{i,j},f^{\mathrm{p}}_{j,i}}\int_0^u[\zeta_{i,j}]d\zeta_{i,j}, \quad (7)$$

where $.\equiv.$ denotes the equivalence relation and u is given by,

$$u=\frac{1}{\sigma_N^2}\left[\frac{P_{i,j}\overline{h}_{i,j}}{e^{\ln(2)(\frac{\lceil L_i(1+\theta)\rceil}{T_i-W_i{f^{\mathrm{p}}_{j,i}}^{-1}})}-1}\right], \quad (8)$$

obviously, (7) is maximized with the pair $(\overline{P}^{\max},\overline{f}^{\max}_j)$. ■

Without considering the channel fading coefficient ($\zeta_{i,j}=1$), inspiring by the idea in [16], a criteria can be derived to identify the maximum admissible communication distance between pair of SMT-PMT. Therefore, according to whether (C1) and (C2) in (6) can be met at $\overline{P}^{\max}$ and $\overline{f}^{\max}_j$, or not, the maximum distance criteria, $\overline{d}^{\max}_{i,j}$ between SMT i and PMT j is obtained by substituting (1) into (C1) and (C2) in (6), that is,

$$\overline{d}^{\max}_{i,j}\triangleq\left[\frac{K\overline{P}^{\max}}{\sigma_N^2(e^{\ln(2)\omega_{i,j}}-1)}\right]^{\frac{1}{\alpha}}, \quad (9)$$

where

$$\omega_{i,j}=\begin{cases}\frac{\lceil L_i(1+\theta)\rceil}{T_i-W_i\overline{f}^{\max-1}_j}, & \text{if } d_{B,j}\le\left[\frac{P_{B,j}K\delta}{\gamma\overline{f}^{\max 3}_j}\right]^{\frac{1}{\alpha}},\\ \frac{\lceil L_i(1+\theta)\rceil}{T_i-W_i\left[\frac{P_{B,j}h_{B,j}\delta}{\gamma}\right]^{-\frac{1}{3}}}, & \text{if } d_{B,j}>\left[\frac{P_{B,j}K\delta}{\gamma\overline{f}^{\max 3}_j}\right]^{\frac{1}{\alpha}},\end{cases} \quad (10)$$

in which $d_{B,j}$ represents the distance between BS and PMT j. Based on (9), $\overline{d}^{\max}_{i,j}$ is the distance at which both (C1) and (C2) inequalities in (6) turn into equality forms. Hence, PMT j is not admissible to collaborate with SMT i, if $d_{i,j}>\overline{d}^{\max}_{i,j}$, due to violation of (C1) and (C2). Obviously, once the channel fading coefficient $\zeta_{i,j}$ is considered, (9) may not hold. Thus, the maximum distance criteria, $\overline{d}^{\max}_{i,j}$ should be reduced by a fraction ψ to counteract the fading effect and therefore to satisfy (C1) in (6), that is, $\widetilde{d}^{\max}_{i,j}=\overline{d}^{\max}_{i,j}/\psi$. Let $\widetilde{h}_{i,j}\triangleq K\zeta^{-1}_{i,j}(\widetilde{d}^{\max}_{i,j})^{-\alpha}$, then, the delay violation probability constraint can be re-written as

$$Pr\left\{t^{\mathrm{off}}_{i,j}>T_i\right\}=Pr\left\{\frac{\lceil L_i(1+\theta)\rceil}{log_2(1+\frac{P_{i,j}\widetilde{h}_{i,j}}{\sigma_N^2})}+\frac{W_i}{f^{\mathrm{p}}_{j,i}}>T_i\right\}$$
$$=Pr\left\{\frac{\lceil L_i(1+\theta)\rceil}{log_2(1+\frac{P_{i,j}\widetilde{h}_{i,j}}{\sigma_N^2})}>\frac{\lceil L_i(1+\theta)\rceil}{log_2(1+\frac{P_{i,j}\overline{h}_{i,j}}{\sigma_N^2})}\right\}$$
$$=Pr\left\{\widetilde{h}_{i,j}<\overline{h}_{i,j}\right\}=Pr\left\{K\zeta^{-1}_{i,j}(\frac{d_{i,j}}{\psi})^{-\alpha}<\overline{h}_{i,j}\right\}$$
$$=Pr\left\{\psi^{\alpha}\zeta^{-1}_{i,j}\overline{h}_{i,j}<\overline{h}_{i,j}\right\}=Pr\left\{\zeta_{i,j}>\psi^{\alpha}\right\}\le\varphi. \quad (11)$$

Rayleigh fading channels are considered in this paper wherein $\zeta_{i,j}$ is distributed exponentially with mean κ. The cdf of Rayleigh distribution is given by

$$F(\zeta_{i,j})=1-e^{-\kappa\zeta_{i,j}},\zeta_{i,j}\ge 0, \quad (12)$$

by substituting (12) into (11), we obtain the maximum distance criteria under Rayleigh fading channel as

$$\widetilde{d}^{\max}_{i,j}=\left[\frac{-\ln(\varphi)}{\kappa}\right]^{-\frac{1}{\alpha}}\overline{d}^{\max}_{i,j}. \quad (13)$$

Hence, $d_{i,j}\le\widetilde{d}^{\max}_{i,j}$ should be satisfied so as to PMT j be admitted to collaborative computation with SMT i. We denote by $\mathcal{A}_i$, the set of admitted PMTs for every SMT i, that is,

$$\mathcal{A}_i=\{j\mid j\in\mathcal{N}_p,d_{i,j}\le\widetilde{d}^{\max}_{i,j}\},\forall i\in\mathcal{N}_s. \quad (14)$$

2) Resource Allocation and Pair selection for MTs: So far, SMTs are provided with the set of nearby PMTs potential for collaborative computation. We now focus on optimizing transmission power and CPU frequencies of MTs to further discover the optimal PMTs for collaboration. The expected energy consumption in (6) is calculated as $\mathbb{E}\left\{[\xi^{\mathrm{off}}_{i,j}]\mid t^{\mathrm{off}}_{i,j}\le T_i\right\}$

$$= \mathbb{E}\left\{ [\xi_{i,j}^{\text{off}}] \mid \frac{\lceil L_i(1+\theta)\rceil}{log_2(1+\frac{P_{i,j}\overline{h}_{i,j}}{\sigma_N^2\zeta i,j})} + \frac{W_i}{f_{j,i}^{\text{p}}} \leq T_i \right\}$$

$$= \mathbb{E}\left\{ [\frac{P_{i,j}(L_i+\widetilde{L}_i\overline{h}_{i,j}\zeta i,j^{-1})}{log_2(1+\frac{P_{i,j}\overline{h}_{i,j}}{\sigma_N^2\zeta i,j})}] \mid \zeta_{i,j} \leq u \right\} + P_i^{\text{idle}}(\frac{W_i}{f_{j,i}^{\text{p}}})$$

$$= \int_0^u [\frac{P_{i,j}(L_i+\widetilde{L}_i\overline{h}_{i,j}\zeta i,j^{-1})}{log_2(1+\frac{P_{i,j}\overline{h}_{i,j}}{\sigma_N^2\zeta_{i,j}})}]\frac{f(\zeta_{i,j})}{F(u)}d\zeta_{i,j} + P_i^{\text{idle}}(\frac{W_i}{f_{j,i}^{\text{p}}}), \tag{15}$$

where u is the upper bound of $\zeta_{i,j}$ defined in (8), implying that fading coefficients larger than u cause delay violation for SMT i. Without considering the bounds $(\overline{P}^{\max}, \overline{f}_j^{\max})$, the energy consumption of SMTs in offloading scenario is strongly vulnerable to transmission power of MTs contributing to a D2D communication, given that

$$\begin{cases} \lim\limits_{P_{i,j}\to\infty} \frac{P_{i,j}(L_i+\widetilde{L}_i h_{i,j})}{log_2(1+\frac{P_{i,j}h_{i,j}}{\sigma_N^2})} = +\infty, \\ \lim\limits_{f_{j,i}^{\text{p}}\to\infty} P_i^{\text{idle}}(\frac{W_i}{f_{j,i}^{\text{p}}}) = \epsilon, \end{cases} \tag{16}$$

where $0 < \epsilon \ll 1$. Taking the actual upper bounds into account, however, the behaviour of the objective function in (6) should be traced, independently, as $\xi_{i,j}^{\text{off}}$ is not monotonically increasing with $P_{i,j}$. Since $\frac{\partial \xi_{i,j}^{\text{off}}}{\partial f_{j,i}^{\text{p}}} < 0$, $\xi_{i,j}^{\text{off}}$ is strictly decreasing function on $f_{j,i}^{\text{p}}$ and reaches minimum when CPU frequency is at peak level. From (C2) and (C4) in (6), a feasible upper bound for $f_{j,i}^{\text{p}}$, denoted by $\widetilde{f}_{j,i}^{\max}$, is obtained as

$$\widetilde{f}_{j,i}^{\max} \triangleq min\left\{ \left[\frac{P_{B,j}h_{B,j}\delta}{\gamma}\right]^{\frac{1}{3}}, \overline{f}_j^{\max} \right\}. \tag{17}$$

As shown in Fig. 3a, by increasing $P_{i,j}$ in $\xi_{i,j}^{\text{off}}$, the pace at which the linear function $\phi_1(P) \triangleq P_{i,j}(L_i + \lceil\theta L_i\rceil h_{i,j})$ grows is initially lower than that of the non-linear function $\phi_2(P) \triangleq log_2(1+\frac{P_{i,j}h_{i,j}}{\sigma_N^2})$ (striped area). But, this rate further rebounds and increases severely which therefore leads to increase in energy expenditure (dotted area). In order to obtain the optimal $P_{i,j}$, first, by substituting (17) into (C1) in (6), a lower bound for $P_{i,j}$ can be defined as follows,

$$\widetilde{P}_{i,j}^{\min} \triangleq \frac{\sigma_N^2(e^{\ln(2)\widetilde{\omega}}-1)F^{-1}(1-\varphi)}{\overline{h}_{i,j}}, \tag{18}$$

where $\widetilde{\omega} = \frac{\lceil L_i(1+\theta)\rceil}{T_i - \frac{W_i}{\widetilde{f}_{j,i}^{\max}}}$. Consequently, to minimize the energy consumption due to communication, we need to find the point at which the distance between the line $\phi_1(P)$ and the curve $\phi_2(P)$ in striped area of Fig. 3a is maximized, that is,

$$\arg\max_{P_{i,j}} S_{i,j}(P) \triangleq log_2(1+\frac{P_{i,j}h_{i,j}}{\sigma_N^2}) - P_{i,j}(L_i+\widetilde{L}_i h_{i,j}). \tag{19}$$

Let I be the intersection point of $\phi_1(P)$ and $\phi_2(P)$ functions. Define the positive constant $\lambda \triangleq \ln(2)(L_i+\widetilde{L}_i h_{i,j})\frac{\sigma_N^2}{h_{i,j}}$, the corresponding transmission power P_I is then obtained as

$$P_I = \frac{-\sigma_N^2}{h_{i,j}}\left[\frac{\mathbb{W}(-\lambda e^{-\lambda})}{\lambda} + 1\right], \tag{20}$$

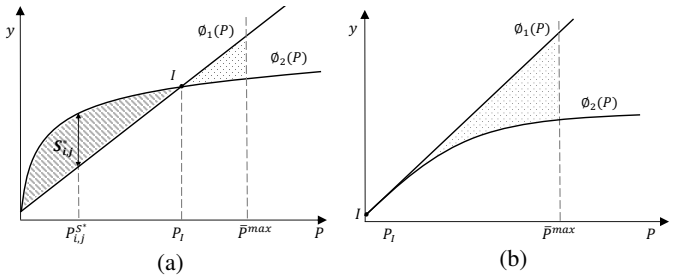


Fig. 3: Feasible area of $P_{i,j}$. Striped area is where $\phi_2(P) > \phi_1(P)$, dotted area is where $\phi_2(P) < \phi_1(P)$, and $\phi_2(P) = \phi_1(P)$ at I.

where $\mathbb{W}(y)$ is the Lambert function defined as the solution for $\mathbb{W}(y)e^{\mathbb{W}(y)} = y$. Let $S_{i,j}^*$ be the maximum vertical distance $S_{i,j}$. Then, feasibility of the corresponding transmission power, $P_{i,j}^{S^*}$, is contingent to P_I as shown in the following theorem.

Theorem 2. The problem in (19) has a feasible solution $P_{i,j}^{S^*} > 0$, if $P_I \neq (\frac{h_{i,j}}{\sigma_N^2})^2[\frac{L_i+\widetilde{L}_i h_{i,j} - \ln(2)\frac{h_{i,j}}{\sigma_N^2}}{\ln(2)}]$ holds.

Proof. The problem has no solution in striped area of Fig. 3a when $\int_\epsilon^{P_I}[S_{i,j}(P)]dP = 0$ for some negligible ϵ, implying that $\phi_1(P)$ and $\phi_2(P)$ are tangent to each other at the point I (Fig. 3b). In that case, P_I would not be sufficiently large to counteract the fading effect or equivalently to satisfy (C1) in (6). The first-order derivative of $\phi_2(P)$ gives us $\frac{\partial\phi_2(P)}{\partial P_I} = \frac{h_{i,j}}{\ln(2)\sigma_N^2(1+\frac{P_I h_{i,j}}{\sigma_N^2})}$. Thus, the tangent line equation at point P_I is obtained as

$$\overline{\phi}_1(P) = \frac{h_{i,j}(P_{i,j}-P_I)}{\ln(2)\sigma_N^2(1+\frac{P_I h_{i,j}}{\sigma_N^2})} + \log_2(1+\frac{P_I h_{i,j}}{\sigma_N^2}), \tag{21}$$

which should not share the same slope with $\phi_1(P)$ so as to guarantee the feasibility of (19). That is,

$$P_I \neq (\frac{h_{i,j}}{\sigma_N^2})^2\left[\frac{L_i+\widetilde{L}_i h_{i,j} - \ln(2)\frac{h_{i,j}}{\sigma_N^2}}{\ln(2)}\right]. \tag{22}$$

■

Corollary 1. According to analysis above, if (22) violates, $P_{i,j}^{S^*} = P_I = \epsilon$, and therefore, (C1) in (6) will not be satisfied. It follows from (17), (18), and (C4) in (6) that the optimal CPU frequency of admitted PMT j collaborating with SMT i is given by ${f_{j,i}^{\text{p}}}^* = \widetilde{f}_{j,i}^{\max}$ and the optimal transmission power of SMT i and admitted PMT j is expressed as

$${P_{i,j}}^* = \min\left\{\overline{P}^{\max}, \max\left\{P_{i,j}^{S^*}, \widetilde{P}_{i,j}^{\min}\right\}\right\}, \tag{23}$$

where $P_{i,j}^{S^*}$, the optimal solution to (19), is obtained as,

$$P_{i,j}^{S^*} = \left[\frac{1}{\ln(2)(L_i + \frac{\lceil\theta L_i\rceil\overline{h}_{i,j}}{F^{-1}(1-\varphi)})} - \frac{\sigma_N^2 F^{-1}(1-\varphi)}{\overline{h}_{i,j}}\right]^+. \tag{24}$$

Once optimal resources are obtained, the following integer programming sub-problem should be solved to determine the optimal pair (PMT) for collaboration with every SMT i.

$$\min_{\rho_{i,j}} \sum_{i\in\mathcal{N}_s}\sum_{j\in\mathcal{A}_i} \rho_{i,j}\mathbb{E}\left\{[\xi_{i,j}^{\text{off}}]\right\}, \text{ s.t. (C3) and (C5) in (6).} \tag{25}$$

When there is only one PMT admitted to collaboration with SMT i, the computation is offloaded to the admitted PMT.

Algorithm 1 DCMC Algorithm

Initialize: Zero set $\mathcal{A}_i$ and matrix $\boldsymbol{\rho}$=$\{\rho_{i,j}\}_{N_s \times N_p}$, $\forall i \in \mathcal{N}_\text{s}, j \in \mathcal{N}_\text{p}$.

Admission control and resource allocation under two scenarios:

for each SMT i **do**
  Obtain optimal $f_i^{\text{s}*}$ by (27);
  for each PMT j **do**
    Obtain probability-based distance criteria $\widetilde{d}_{i,j}^{\max}$ using (13);
    if $d_{i,j} \leq \widetilde{d}_{i,j}^{\max}$ **then**
      Update $\mathcal{A}_i \leftarrow \mathcal{A}_i \cup \{j\}$; ▷ *admit PMT j*
      Obtain $f_{j,i}^{\text{p}}{}^*$ and $P_{i,j}^*$ by (17) and (23), respectively;
    end if
  end for
end for
Get optimal $\boldsymbol{\rho}^*$ matrix applying Kuhn-Munkres algorithm [17];

Computation strategy selection:

for each SMT i **do**
  Get ξ_i^{off} and ξ_i^{loc}, pick a computation strategy based on (28);
end for

However, when $\mid \mathcal{A}_i \mid > 1, \forall i \in \mathcal{N}_s$, the optimal solution to the problem in (25) can be obtained applying Kuhn-Munkres algorithm [17]. As a result, there would be an admitted PMT j' selected for collaboration with SMT i such that $\xi_i^{\text{off}} = \{\xi_{i,j'}^{\text{off}} \mid j' \in \mathcal{A}_i, \rho_{i,j'} = 1\}$. Note that a communication between D2D transmitter and receiver is initiated based on D2D pair discovery and session setup protocols [6] which are beyond this paper.

B. Local Execution Scenario

In local execution scenario, we aim to minimize the energy consumed by local CPUs with guaranteeing delay and energy harvesting constraints for SMTs. The resource allocation problem for SMT i in local execution is formally stated as

$$\min_{f_i^{\text{s}}} \ \xi_i^{\text{loc}} \triangleq W_i \gamma (f_i^{\text{s}})^2 \tag{26}$$

$$\text{s.t.} \quad W_i f_i^{\text{s}-1} \leq T_i, \forall i \in \mathcal{N}_s, \tag{C1}$$

$$W_i \gamma f_i^{\text{s}2} \leq \delta P_{B,i} h_{B,i} (\frac{W_i}{f_i^{\text{s}}}), \forall i \in \mathcal{N}_s, \tag{C2}$$

$$f_i^{\text{s}} \in (0, \overline{f}_i^{\max}], \forall i \in \mathcal{N}_s. \tag{C3}$$

Regardless of whether local execution is feasible or not, we address the above problem in the following lemma.

lemma 1. The problem in (26) is convex and its solution occurs at a feasible boundary within the interval of f_i^{s}.

Proof. It can be clearly observed that the objective function in (26) is convex. Multiplying both sides of (C1) and (C2) in (26) by f_i^{s} would then result in a convex problem. From the first-order derivative of the objective function, it can be realized that the only critical point exists at $f_i^{\text{s}} = 0$ which obviously is not feasible. Forasmuch as SMTs are considered as resource-poor MTs, the feasibility of (C2) in (26) strongly depends on BS's transmission power. Hence, define $\overline{P}_{B,i} \triangleq \frac{\gamma (W_i T_i^{-1})^3}{\delta h_{B,i}}$ as the threshold power of BS for SMT i, the optimal CPU frequency of SMT i exists at a boundary within the interval of f_i^{s}, that is,

$$f_i^{\text{s}*} = \begin{cases} \min\{\frac{W_i}{T_i}, \overline{f}_i^{\max}\}, & \text{if } P_{B,i} \geq \overline{P}_{B,i}, \\ \min\{\left[\frac{\delta P_{B,i} h_{B,i}}{\gamma}\right]^{\frac{1}{3}}, \overline{f}_i^{\max}\}, & \text{if } P_{B,i} < \overline{P}_{B,i}. \end{cases} \tag{27}$$

■

Table I: Simulation Parameters

Parameter	Value
System	Single cell, R=250 m, BW=1 MHz
Number of users	SMTs: $1 \sim 25$, PMTs: $1 \sim 120$
Maximum transmit power	BS: 46 dBm, MT: 23 dBm
Maximum CPU frequency	SMTs: 200 MHz, PMTs: 400 MHz
CPU energy constant (γ)	SMTs: 10^{-20}, PMTs: 10^{-28}
CPU idle power	10 dBm
Completion Deadline (T)	$0.165 \sim 0.215$ s
Channel model	$\sigma_N^2 = -180$ dBm, exponent $\alpha = 3$
Data parameters	$L = 10$ bits, $\theta = 0.5$
Number of CPU cycles/bit	Gamma distribution: $\alpha = 4, \beta = 200$

While $f_i^{\text{s}*}$ minimizes the objective function in (26), it would be a feasible solution if and only if $P_{B,i} \geq \overline{P}_{B,i}$ and $\frac{W_i}{T_i} \leq \overline{f}_i^{\max}$.

C. Computation Strategy Selection

Based on optimal resources allocated to MTs, SMTs have to choose whether to offload their computation, or to execute them locally. Define the maximum energy saving for SMT i as $\xi_i^{\text{save}} \triangleq \xi_i^{\text{off}} - \xi_i^{\text{loc}}$, local execution is dominant if $\xi_i^{\text{save}} > 0$ while offloading is preferred, otherwise. Selecting computation strategy only based on energy consumption, as considered in most previous works, is effective only when both local execution and computation offloading scenarios are feasible. In practice, however, potential PMTs may not be available close enough to every SMT for collaboration. On the other hand, as described in previous subsection (III-B), the QoS of wearable SMTs in local execution scenario is vulnerable to transmission power of BS; i.e., SMT i experiences a delay violation unless $P_{B,i} \geq \overline{P}_{B,i}$ and $\frac{W_i}{T_i} \leq \overline{f}_i^{\max}$. Especially, cell-edge SMTs are more severely exposed to delay violation as receiving low power from BS. For these SMTs, however, consuming even more energy, the delay violation can be eliminated by offloading computation to potential PMTs in vicinity. Therefore, taking both the energy and outage factors into account, the collaborative computation strategy, s_i^{col}, for SMD i is selected as $s_i^{\text{col}} =$

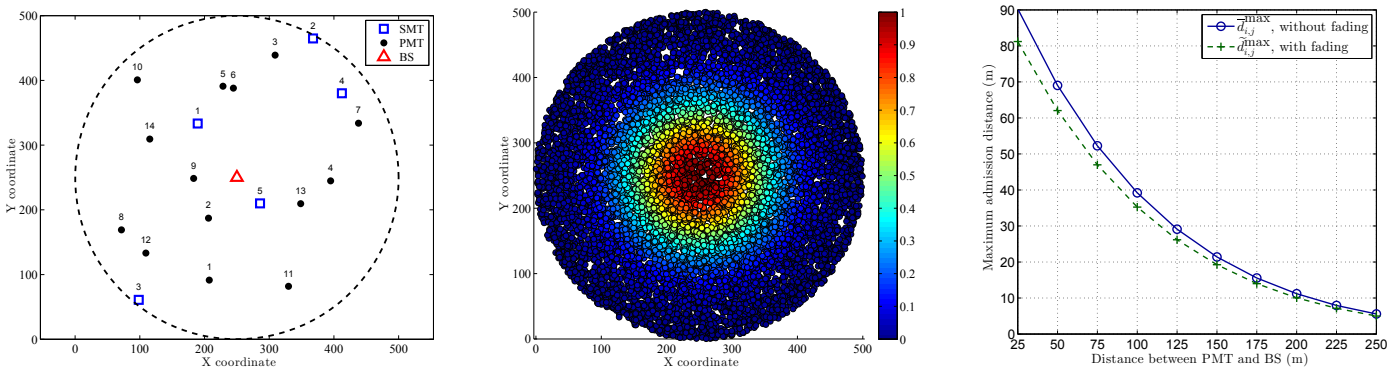
$$\begin{cases} \text{offload}, & \text{if } \mid \mathcal{A}_i \mid \geq 1, \{P_{B,i} < \overline{P}_{B,i} \ \mid \ \frac{W_i}{T_i} > \overline{f}_i^{\max}\} \\ \text{compute locally}, & \text{if } \mid \mathcal{A}_i \mid < 1, \{P_{B,i} \geq \overline{P}_{B,i} \ \& \ \frac{W_i}{T_i} \leq \overline{f}_i^{\max}\} \\ \text{choose based on } \xi_i^{\text{save}}, & \text{otherwise.} \end{cases} \tag{28}$$

The proposed DCMC scheme is summarized in Algorithm 1.

IV. Simulation Results

In this section, the performance of the proposed DCMC algorithm is evaluated by simulation. To achieve average results, we ran every simulation 500 times. We also included additional approaches in our simulations referred to as: (i) *optimal solution* which can be obtained when fading coefficients are known, (ii) *shortest-path offloading* in which SMTs offload their computations to the nearest PMTs, and (iii) *random offloading* wherein a PMT is selected randomly by every SMT for collaboration. We refer to either proposed forced strategy as *offload-only* or *local-only*. Simulation parameters are summarized in Table I.

Fig. 4a shows a sample system model where we uniformly scattered MTs in a circular cell with $R = 250m$. Fig. 4b depicts the 500 times random distribution of 40 MTs where different



(a) (b) (c)

Fig. 4: (a) Sample snapshot of network. (b) Density of energy harvested from BS. (c) Maximum admissible SMT-PMT distance for collaboration

colors on MTs specify the density of energy harvested from BS. Clearly, the shorter the distance to BS, the higher the BS's radiation and thus, the higher the chance to be satisfied with constraints in (3) and (4). In Fig. 4c, we randomly distributed a single SMT in the vicinity of BS (50m), then moved a PMT from center to cell-edge and measured the proposed distance criteria. The PMT is most likely ignored for admission when is far from BS. This is due to the presence of high communication delay and low energy harvested from BS in large distances.

Fig. 5a, shows the cumulative outage ratio with respect to delay violation probability when $N_s = 25$ and $N_p = 30$. Cumulative outage ratio is defined as the fraction of number of unsatisfied SMTs and PMTs to total number of MTs. It can be observed that there is an optimal point of threshold for delay violation probability at which the highest possible number of MTs are satisfied with their energy harvesting and delay requirements. φ directly affects the number of PMTs admitted to collaboration since it influences on $\widetilde{d}_{i,j}^{\max}$. Furthermore, our DCMC scheme outperforms forced local execution or offloading scenarios. This is because SMTs in the proposed DCMC algorithm take both the energy and outage factors into consideration for choosing a strategy. Fig. 5b illustrates the average minimum transmission power of BS required for satisfying QoS of every MT (i.e., to maintain zero outage) versus deadline threshold when $N_s = 20$. The results reveal that when the number of PMTs increases ($N_p > N_s$), the BS needs lower transmission power to maintain a feasible network under strict deadlines. This is due to the increase in number of potential PMTs admitted to collaboration for every SMT. In addition, the DCMC scheme may considerably reduce the BS's minimum transmission power in comparison to forced local execution. Since in DCMC scheme, those SMTs harvesting low energy levels from BS may offload their computations to potential PMTs which are most likely closer to BS.

Fig. 6 shows the cumulative outage ratio and energy consumption versus number of PMTs. We set $\varphi = 0.2$ and $N_s = 10$. The proposed DCMC scheme achieves a performance close to the optimal solution when the number of PMTs increases and offloading is less competitive. Applying the probabilistic admission control method, in our proposed scheme, inefficient PMTs are ignored by SMTs. In addition, SMTs choose between local execution and offloading strategies not only to save energy, but also to mitigate the outage. Shortest-path and random offloading schemes do not check whether a PMT is feasible or not. That is why the cumulative outage ratio of the proposed scheme is lower than that of shortest-path and random offloading schemes even when N_s equals N_p. Based on our observations, as the number of PMTs increases, SMTs are provided with more potential PMTs admitted to collaboration. As a result, aiming to meet QoS requirements, unsatisfied SMTs located in cell-edge may choose offloading strategy as they barely harvest energy, while SMTs close to BS may choose local execution strategy to ignore communication energy loss. At lower number of PMTs, however, shortest-path scheme slightly surpasses DCMC scheme in energy consumption. This is obviously because in competitive situation, our proposed scheme compensates for outage of MTs as shown in Fig. 5a.

Fig. 7, illustrates a more detailed comparison of energy consumption and cumulative outage ratio of the proposed DCMC scheme, optimal solution and conventional approaches with respect to deadline T. In this experiment, we set the number of PMTs and SMTs to 25 and 30, respectively, and vary the deadline within a large scale from infeasible to feasible values. We can draw several observations by comparing the results in Fig. 7a and 7b. First, when infeasible deadlines are applied, our proposed DCMC scheme tends to maintain low energy consumption through local execution as no PMTs are admitted for collaboration. However, shortest-path and random offloading schemes waste energy through offloading. Once the system partially becomes feasible, a little increase can be observed in energy consumption of DCMC scheme. This is reasonable since unsatisfied SMTs find the opportunity to meet their deadline and energy requirements by switching to offloading strategy. Moreover, when T exceeds 200 ms, no further decrease in energy consumption is observed in either offloading approaches, since no further change occurs in pair selection step. When the deadline further increases, the proposed DCMC scheme allows potential SMTs to switch to local execution strategy to consume lower energy. The figure demonstrates that DCMC scheme improves outage ratio up to 47% and 66% comparing to shortest-path offloading and local execution schemes, respectively, while achieving a near optimal energy consumption.

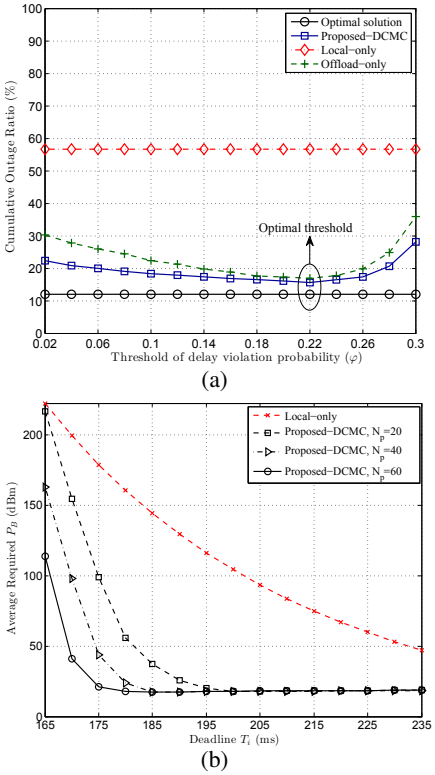


Fig. 5: (a) Cumulative outage ratio versus maximum delay violation probability. (b) Average required transmission power of BS

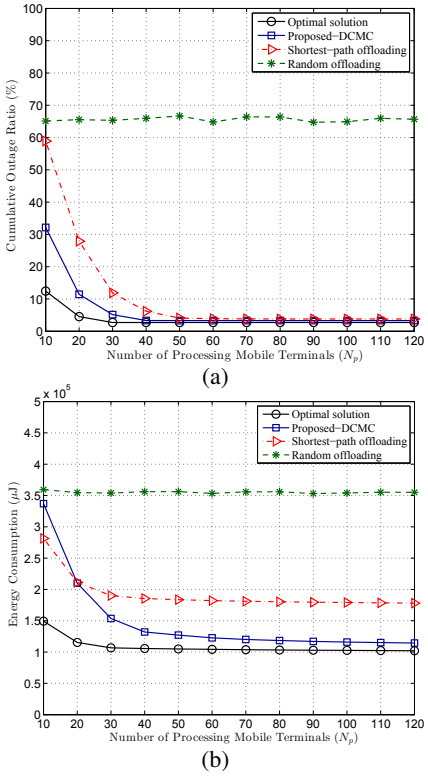


Fig. 6: Cumulative outage ratio and sum-energy consumption of SMTs with varying number of PMTs, $N_s = 10$ and $\varphi = 0.2$.

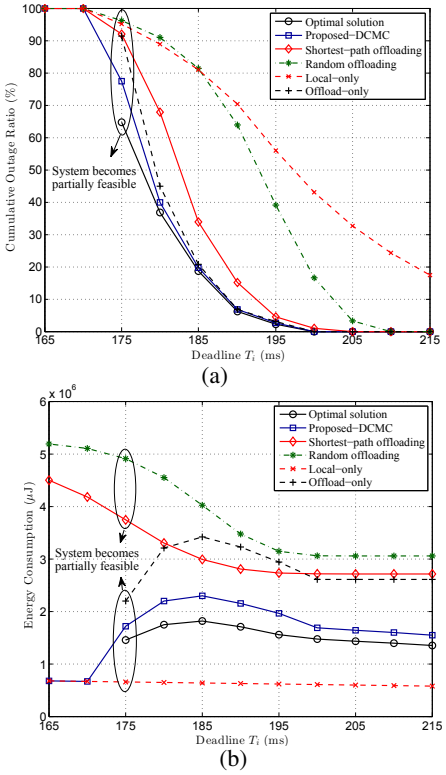


Fig. 7: Cumulative outage ratio and sum-energy consumption of SMTs with varying deadline T_i, $N_s = 25$, $N_p = 30$, $\varphi = 0.2$.

V. Conclusion

The joint admission control and resource allocation problem for D2D-aided collaborative mobile cloud (DCMC) has been studied in an IoT-enabled network under channel uncertainty. Considering both deadline and energy harvesting constraints, we addressed the energy minimization problem under computation offloading and local execution scenarios. For offloading scenario, we derived a probabilistic admission control criteria and addressed the problem of joint CPU frequency/transmission power allocation and pair selection for MTs. Then, we obtained optimal CPU frequencies of source MTs under local execution scenario. Finally, a computation strategy selection criteria has been proposed based on energy and outage of MTs. Simulation results demonstrated the effectiveness of the proposed scheme.

References

[1] B. Ahlgren, M. Hidell and E. C. H. Ngai, "Internet of Things for Smart Cities: Interoperability and Open Data," in *IEEE Internet Computing*, vol. 20, no. 6, pp. 52-56, Nov.-Dec. 2016.

[2] K. Tehrani and M. Andrew. (2014, Mar.) Wearable technology and wearable devices: Everything you need to know. [Online].

[3] W. C. Brown, The history of power transmission by radio waves, *IEEE Trans. on Microwave Theory and Techniques*, vol. 32, no. 9, pp. 12301242, Sep. 1984.

[4] R. LiKamWa, Z. Wang, A. Carroll, F. X. Lin, and L. Zhong, Draining our glass: An energy and heat characterization of Google Glass, *CoRR*, vol. abs/1404.1320, 2014.

[5] M. Y. Pederson and F. H. P. Fitzek, "Mobile Clouds: The New Content Sharing Platform," *Proceeding of IEEE*, Vol. 100, pp. 1400-1403, 2012.

[6] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Commun. Mag.*, vol. 47, pp. 42-49, 2009.

[7] J. Kwak, Y. Kim, J. Lee and S. Chong, "DREAM: Dynamic Resource and Task Allocation for Energy Minimization in Mobile Cloud Systems," in *IEEE Journal on Selected Areas in Commun.*, vol. 33, no. 12, pp. 2510-2523, Dec. 2015.

[8] D. Huang, P. Wang and D. Niyato, "A Dynamic Offloading Algorithm for Mobile Computing," in *IEEE Trans. on Wireless Commun.*, vol. 11, no. 6, pp. 1991-1995, Jun. 2012.

[9] C. You, K. Huang and H. Chae, "Energy Efficient Mobile Cloud Computing Powered by Wireless Energy Transfer," in *IEEE Journal on Selected Areas in Commun.*, vol. 34, no. 5, pp. 1757-1771, May 2016.

[10] 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. on Wireless Commun.*, vol.12, no.9, pp. 4569-4581, Sep. 2013.

[11] Z. Chang, J. Gong, T. Ristaniemi and Z. Niu, "Energy-Efficient Resource Allocation and User Scheduling for Collaborative Mobile Clouds With Hybrid Receivers," in *IEEE Trans. on Vehicular Technology*, vol. 65, no. 12, pp. 9834-9846, Dec. 2016.

[12] L. Tang and Q. Li, "Energy and time optimization for wireless computation offloading," *International Conference on Wireless Commun. and Signal Processing (WCSP)*, Nanjing, 2015, pp. 1-5.

[13] X. Liu, C. Yuan, Z. Yang and Z. Hu, "An Energy Saving Algorithm Based on User-Provided Resources in Mobile Cloud Computing," *IEEE Vehicular Technology Conference (VTC Fall)*, Las Vegas, NV, 2013, pp. 1-5.

[14] L. Xiang, B. Li and B. Li, "Coalition Formation Towards Energy-Efficient Collaborative Mobile Computing," *International Conference on Computer Commun. and Networks (ICCCN)*, Las Vegas, NV, 2015, pp. 1-8.

[15] S. Cao, X. Tao, Y. Hou and Q. Cui, "An energy-optimal offloading algorithm of mobile computing based on HetNets," *International Conference on Connected Vehicles and Expo (ICCVE)*, Shenzhen, 2015, pp. 254-258.

[16] F. Hajiaghajani, R. Davoudi, M. Rasti, "A Joint Channel and Power Allocation Scheme for Device-to-Device Communications Underlaying Uplink Cellular Networks," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2016, San Francisco, CA, pp. 473-478.

[17] D. West *et al., Introduction to Graph Theory*. Upper Saddle River, NJ: Prentice Hall, 2001.

[18] A. Schrijver, Combinatorial Optimization: Polyhedra and Efficiency, *Springer*, 2003.