188

Performance Optimization and Analysis on P2P Mobile Communication Systems Accelerated by MEC Servers

Xuesong Liang[1], Yongpeng Wu[2], Yujin Huang[1], Derrick Wing Kwan Ng[3], Pei Li[1], and Yingbiao Yao[1]

[1]The School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China
[e-mail: {liangxs, huangyj,lipei,yaoyb}@hdu.edu.cn]
[2]The Department of Electronic Engineering, Shanghai Jiao Tong University, Minhang, Shanghai, China
[e-mail: yongpeng.wu@sjtu.edu.cn]
[3]The School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, N.S.W., Australia
[e-mail: w.k.ng@unsw.edu.au]
*Corresponding author: Xuesong Liang

Abstract

As a promising technique to support tremendous numbers of Internet of Things devices and a variety of applications efficiently, mobile edge computing (MEC) has attracted extensive studies recently. In this paper, we consider a MEC-assisted peer-to-peer (P2P) mobile communication system where MEC servers are deployed at access points to accelerate the communication process between mobile terminals. To capture the tradeoff between the time delay and the energy consumption of the system, a cost function is introduced to facilitate the optimization of the computation and communication resources. The formulated optimization problem is non-convex and is tackled by an iterative block coordinate descent algorithm that decouples the original optimization problem into two subproblems and alternately optimizes the computation and communication resources. Moreover, the MEC-assisted P2P communication system is compared with the conventional P2P communication system, then a condition is provided in closed-form expression when the MEC-assisted P2P communication system performs better. Simulation results show that the advantage of this system is enhanced when the computing capability of the receiver increases whereas it is reduced when the computing capability of the transmitter increases. In addition, the performance of this system is significantly improved when the signal-to-noise ratio of hop-1 exceeds that of hop-2.

.

1. Introduction

With the emerging industry of the Internet of Things (IoT), the scope of mobile communication services has been extended from inter-personal communication to a web of billions of smart devices, e.g., smartphones, sensors, wearable computing devices, laptops, and tablets. Owing to the fast growth of interconnected mobile devices industry, mobile data traffic explosively increases in the past few years and this trend will continue in the future. According to Cisco, monthly Internet traffic is predicted to reach 44 GB per capita by 2022, from 13 GB per capita in 2017 [1]. Moreover, new mobile applications such as face recognition, natural language processing, augmented reality, etc. have become even more popular, which demand not only in high data rate but also in high computational capability [2].

To cope with these challenges, mobile cloud computing (MCC) [3] has been widely used by sharing the cloud computing capability with mobile devices to break through the bottleneck of limited computing resources and reduce the energy consumption. However, the performance of MCC is substantially impaired by the long latency and backhaul bandwidth on the condition that the Internet cloud is usually far away from mobile devices. To address this problem, mobile edge computing (MEC) has attracted great interest recently [4] by offloading intensive mobile computation to proximate MEC servers at the edges of cellular networks. Compared with MCC, MEC has the advantages of low latency, high bandwidth, real-time radio network information, and location awareness. Thus it is widely agreed to be a key technology to realize various visions for the next-generation Internet, such as IoT [5], Internet of Vehicles [6], and the Tactile Internet [4].

The concept of MEC is first proposed by Cisco [7] (which is called fog computing) and ESIT [8], which has enabled the emergence of several new applications, such as virtual reality [9], intelligent video acceleration [10], remote medical diagnosis [11] and IoT gateway [12]. Both academic and industrial communities have conducted extensive researches on several key technologies of MEC, such as computation offloading, resource allocation of MEC servers, cache- enabled MEC, mobility management, and security/privacy issues in the past few years. As one of the key technologies of MEC, computation offloading can enable mobile devices to run new sophisticated applications while reducing their energy consumption by allocating partial or all of their computation-intensive tasks to the proximate MEC servers. In past few years, numerous studies have been made on computation offloading of MEC, including offloading strategy design and optimization of computation and communication resources.

Firstly, the policy of full offloading decision has been widely investigated. Y. Mao in [13] considered single user equipment (UE) adopting the techniques of dynamic voltage and frequency scaling and energy harvesting to minimize its energy consumption and proposed a low-complexity dynamic algorithm based on Lyapunov optimization to reduce the execution delay. Another idea was introduced by M. Kamoun in [14]. The author aims to optimize the computation offloading decision to minimize the energy consumption of the UE while satisfying the execution delay of the applications. The optimal problem was modeled as a constrained Markov decision process in [14] and two offloading strategies, i.e., the online learning strategy and the offline pre-calculated strategy were proposed. The offline offloading strategy shows the superiority over the online strategy in low and medium loads, and it was extended to multi-UEs scenarios in [15]. For the same goal in the multi-cell scenario, S. Sardellitti in [16] jointly optimized the resources of communication and computation and proposed a distributed iterative algorithm to attain local optimal solution using successive convex approximation. In addition, the tradeoff of energy consumption and delay was investigated by X. Chen [17] in the multi-user and multi-channel scenario, the game theory

was used to solve the optimization problem, and a distributed offloading algorithm was proposed to achieve Nash equilibrium.

For another, the case of partial computation offloading, which means there are some parts of data in particular applications that cannot be offloaded to MEC servers, was also investigated in plenty of works. S. Cao in [18] constructed the partial offloading model for a single UE and proposed an optimal adaptive algorithm to reduce the energy consumption of the UE and meet the delay requirement as well. Then Y. Zhao in [19] considered the partial offloading decision problem for the multi-UEs model and proposed a low-complexity algorithm by linear programming to solve the problem. Furthermore, joint optimization of offloading decisions and resource allocation were addressed in the literatures. O. Muñoz in [20] analyzed the tradeoff between the energy consumption and the execution delay in the single UE scenario and indicated that the energy consumption at the UE is inversely correlated with the total execution time. The work in [20] was extended to multi-UEs scenarios in [21], in which the impact of the data rate and the load on the offloading decision was discussed. The optimal resource allocation problem for multi-UEs offloading was studied in two sets of systems, orthogonal frequency-division multiple access (OFDMA) and time-division multiple access (TDMA) in [22]. Simulation results show that the technology of OFDMA can reduce the energy consumption of the UEs for MEC more significantly than TDMA does. The optimal resource allocation problem for multi-UEs offloading was also discussed in [23] with upon stability of the buffer of the MEC server. The authors proposed an online algorithm based on Lyapunov optimization to reduce the power consumption and the execution delay of MEC according to UEs priority. Recently, the computation offloading problem was also studied in combination with the computation resource allocation in non-orthogonal multiple access Communications [24], mobility management of UEs [25], cache-enabled MEC [26], and machine learning [27]. To sum up, offloading strategy and optimal resource allocation are still the key challenges of MEC to run delay-sensitive and computation-intensive applications efficiently.

Although several excellent works have been proposed on computation offloading in MEC networks, all the aforementioned literature focuses on enhancing the computation of mobile devices through offloading strategy designation and optimal resource allocation, i.e., reduce the cost of computing for mobile devices via the radio channels to MEC servers. What is rarely addressed, on the other hand, is using computation resources to accelerate wireless communication between MEC servers and mobile devices. Due to the fact that wireless communication usually consumes an amount of energy for data processing tremendously larger than mobile computing does, i.e., that the energy required for transmission of a single bit is over 1000 times greater than for a single 32-bit computation [28], we can see a considerable boost in energy efficiency with the help of the computing resource of MEC systems to maintain the quality of wireless communication. Recently, a notable model of MEC relaying networks was proposed by M. Qin in [29], which utilized the computation capability of MEC servers to accelerate the peer-to- peer (P2P) transmission process. The authors proposed a MEC-assisted computing and relaying scheme to enhance the throughput of uncompressed data for mobile P2P communications. The tradeoff between the latency and energy consumption during the whole processing was defined by a cost function, and an optimal strategy of transmission and compression was derived in closed-form. It is shown that the proposed [29] model can significantly reduce the system cost in comparison to the conventional P2P communications without MEC resources. The work in [29] gave a notable example of using computing to boost communication in MEC, and the optimal solution was further discussed in several practical scenarios. This work laid a solid foundation with

providing a MEC-assisted computing and relaying model. However, there are still two main limitations in this model:

1. The wireless access channels in this model is too ideal a condition for most of the cases, given that the channels are assumed in static flat fading and perfectly known for receivers. In addition, while the impact of the computing factors is mainly under discussion, i.e., the compression ratios for the transmission data, this model gives little notice to the impact of communication factors, e.g., the data rates and transmit powers. Therefore, the conclusions in [29] of this model may be limited to few practical scenarios.

2. Theoretically, the MEC-assisted relaying model is capable of showing an overall advantage, its performance comparing with that of the conventional relaying model. However, the comparison of two models is incomplete in the light of the cost of data compression and decompression of MEC servers being neglected in the model. Moreover, the condition of when the cost of the MEC-assisted relaying model is less than that of the conventional relaying model is not provided in analytical results, neither in the general case nor in special cases.

Motivated by these observations, we adopt a modified approach to simultaneously optimizes the main factors of computation and communication phases. To go beyond the original model, the influence of both the outage of wireless channels and the computation capability of MEC servers are further investigated. The non-convex optimization problem of our model is efficiently tackled by an iterative algorithm based on the BCD method. Furthermore, the advantage of our model is elaborately discussed with a benchmark of the conventional relaying system The main contributions of the paper are summarized as follows.

1) The MEC-assisted P2P communications between two mobile devices are modeled as a modified computing and relaying system with consideration of the impact of the outage of wireless channels and data processing of MEC servers on the total cost of the system. The factors of both the transmitting phase, (i.e., the data rates and transmit power) and the computing phase (i.e., the compression/decompression ratio) are analyzed. It is revealed that the optima of the factors of the transmitting phase only depend on the condition of communication resources, e.g. channel state and amount of data, while those of the computing phase are decided by the condition of both computation and communication resources.

2) The optimal strategy of our model is challenging, which is shown as a non-convex problem. To tackle this problem, firstly we transform it into a block multi-convex problem. Then we propose a BCD-based algorithm to optimize the factors of the transmitting and computing phases iteratively and provide a local optimum of the original problem. The algorithm proves effective and efficient by guaranteeing global convergence to a local optimum with linear complexity.

3) On the basis of the optimal strategy of transmission and compression, we elaborately analyze the performance of our model compared with the conventional relaying model which is optimized with a similar algorithm. The condition when the MEC-assisted relaying system is superior to the conventional relaying system is derived in closed-form, which provides the criterion for whether the MEC servers should be adopted to accelerate the communication process. Then, we illustrate some special cases to discuss the impact of asymmetric channels and asymmetric computing capabilities of UEs on the advantage of the MEC-assisted relaying system, which can offer valuable insights into the presetting of the model parameters in practical scenarios.

This paper is structured as follows. Section 2 describes the models of the MEC-assisted/non-MEC access P2P relaying systems, based on that a cost function is defined to jointly optimize the time delay and energy consumption of these systems. Section 3 constructs the optimal problems of these two systems to minimize the cost functions, and two iterative algorithms are

proposed to tackle these non-convex problems efficiently. Section 4 investigates the cost gain achieved by the MEC-assisted relaying system comparing to the non-MEC relaying system. The condition of when the cost performance of the MEC-assisted relaying system is better than the non-MEC relaying system is discussed in detail and some special cases related to practical scenarios are also analyzed. Section 5 provides the simulation results and further discussion. Finally, the conclusions are given in Section 6.

2. System Model and Performance Metric

2.1. System Model

(a) The MEC-assisted wireless communication between two mobile users.
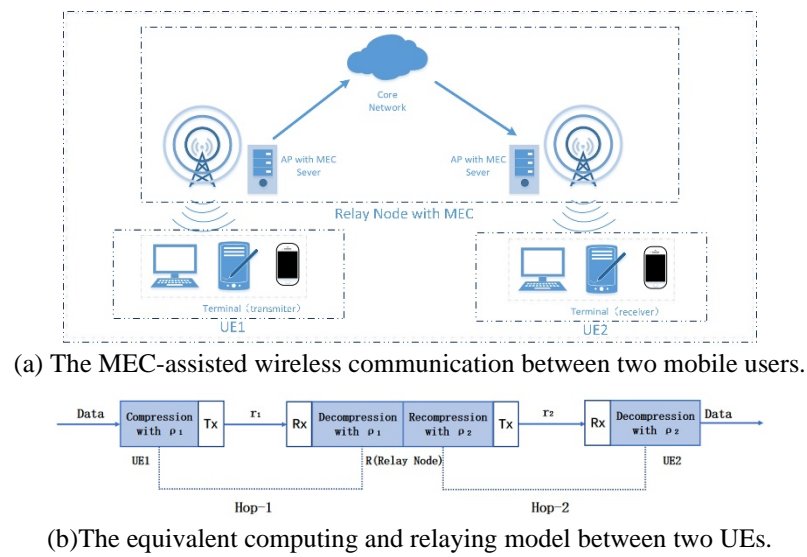
(b)The equivalent computing and relaying model between two UEs.
Fig. 1. The system model

The P2P wireless communication scenario with MEC is shown in **Fig. 1(a)** as two mobile devices exchanging application data, i.e., photo, video and other multi-media content, with each other via a wireless communication network in which mobile devices access to the wireless network through APs that are connected with the core networks through optical cables. The data is first captured and compressed by the mobile transmitter, namely UE1, then uploaded to a local AP via a wireless channel. Next, the received data is transmitted to another AP which is adjacent to the mobile receiver, namely UE2, through the core networks. Finally, the data is downloaded by the receiver from the adjacent AP through a wireless channel and decompressed into the original application data. In this scenario we assume that the APs are adopted with MEC servers, so that the computing capability of the MEC server can be used to accelerate the wireless transmission between the APs and the mobile devices. For this purpose, the data received from UE1 is first decompressed and then re-compressed with another ratio by the APs before it is transmitted to UE2.

In order to facilitate the analysis of the above P2P communication, we modeled the wireless communication network including the APs as a relay node, R. Therefore, the process of the P2P communication can be described as a 2-hops relaying system with communication and computation phases, which is shown in **Fig. 1(b)**. Firstly, D bits of original application data is collected and compressed with a ratio of ρ_1 by UE1, and then the compressed data is transmitted from UE1 to R with a data rate of r_1. With the help of MEC servers, R

decompresses the received data and re-compresses it with a ratio of ρ_2, and then the re-compressed data is transmitted to UE2 with a data rate of r_2. Finally, UE2 receives the data and decompresses it to obtain the original application data. In this model, the wireless channel between UE1 and R, namely h_1, and that between R and UE2, namely h_2, are assumed to obey Rayleigh distribution, i.e., $h_1 \sim \mathcal{CN}(0, \sigma_1^2)$ and $h_2 \sim \mathcal{CN}(0, \sigma_2^2)$.

In specific, the computation and communication parameters in the relaying system are given as follows. As for computation, we describe the computing capability of the UEs and the relay node in terms of CPU cycles/s. According to [28], to compress or decompress 1-bit data will consume approximately the computing capability as

$$U(\rho_i) = \xi_i(e^{\varepsilon \rho_i} - e^\varepsilon), \qquad (1)$$

where ε and ξ_i denote constants depending on the compression method, and $\rho_i \in \left[1, \rho_{\max,i}\right]$ denotes the compression or decompression ratio of the UEs and relay node, $\forall i = 1,2$. Meanwhile, it is assumed that in the tow-hops relaying system the wireless channels of two hops are perfectly known by the receivers. By referring to the transmit power of hop-i as p_i, the outage probability of hop-i, which is defined as

$$P_i = P\{p_i g_i < f(r_i)\} = 1 - \exp\left(-\frac{f(r_i)}{p_i g_i}\right), i = 1,2, \qquad (2)$$

where $f(x) = 2^{x/B} - 1$ with the bandwidth B and $g_i = \sigma_i^2 / N_0$ with the power spectral density of the noise N_0, $i = 1,2$. Since the power of relay node is supplied by the power grid, which is nearly unconstrained compared to mobile devices, we set p_2 to fixed value.

2.2. Performance Metric

In P2P wireless communications, two key performance metrics are usually discussed: the energy consumption and the time of delay. To further investigate the system performance on the tradeoff between energy consumption and time delay, we define a cost function that covers the two key metrics as follows.

Definition 1: The cost of transmitting D bits application data from UE1 to UE2 through this considered relay system is defined as $L_{MEC} = \alpha_1 E_1 + \alpha_2 E_2 + \gamma T$, where E_1 and E_2 denote the energy consumptions of UE1 and UE2, respectively, T denotes the total time delay of the entire process of communication and computation and $\alpha_1, \alpha_2, \gamma \in [0,1]$ are scalar weights.

Note that $\alpha_1, \alpha_2, \gamma$ are all variable weights in the definition of the cost function, and the tradeoff between the energy consumption and time delay can be modified when we vary them. Therefore as long as a particular series of weights are provided, the optimal transmission and compression strategy for this tradeoff can be derived according to the cost function.

Then the expressions of energy consumption and delay are given based on the computation and communication process to calculate the cost function.

Energy consumption: According to the computation and communication scheme, the energy cost consists of two main parts: the energy consumption for the compression or decompression of data, and that for the transmission or reception of data. As the relay node is supplied by the power grid and thus is insensitive to energy consumption, its energy consumption is not taken into account in the optimal problem of this model. Hence only the energy consumption of the mobile devices are counted in this model.

Firstly, according to (1), the energy consumption of UE1 to compress D bits of data with a ratio of ρ_1 is given as $E_c = q_c D \cdot U(\rho_1) = q_c D \xi_1 (e^{\varepsilon \rho_1} - e^\varepsilon)$, where q_c (in Joule/cycle) denotes the energy consumption for each CPU cycle of UE 1. Similarly, the energy consumption of UE2 to decompress D/ρ_2 bits of data with a ratio of ρ_2 is given by $E_d = q_d D \xi_2 (e^{\varepsilon \rho_2} - e^\varepsilon)$, where q_d (in Joule/cycle) denotes the energy consumption for each CPU

cycle of UE2.

Secondly, with the outage probability of hop-i, P_i, $i = 1,2$, the energy consumption of UE1 in data transmission is given by $E_{tx} = p_1 D / \{\rho_1 r_1 (1 - P_1)\}$. In a similar way, the energy consumption in UE2 in data reception is given by $E_{rx} = q_{rx} D / \{\rho_2 r_2 (1 - P_2)\}$, where q_{rx} is the energy consumption of UE2 for receiving each bit.

As above, we obtained roughly the amount of energy consumption. Then we define other energy consumption of circuits in the UEs, which are usually regarded as constant values. We denoted them as C_1 and C_2 for UE1 and UE2, respectively. Therefore, the total energy consumptions of the UEs are expressed as

$$E_1 = E_c + E_{tx} + C_1, \tag{3}$$
$$E_2 = E_d + E_{rx} + C_2. \tag{4}$$

Delay Time: As for the entire process of computation and communication, the time delay includes mainly two parts: one is when data is being compressed or decompressed, and the other is when data is being transmitted or received. Noting that the time delay both at the UEs and at the relay node is taken into account.

UE1 for compression of D bits with compression ratio ρ_1 is given by $T_c = D\xi_1(e^{\varepsilon \rho_1} - e^{\varepsilon})/f_c$, where f_c (in cycles/s) denotes the CPU frequency of UE1. Similarly, the computing delay in UE 2 for decompression of D/ρ_2 bits with a ratio of ρ_2 is given by $T_d = D\xi_2(e^{\varepsilon \rho_2} - e^{\varepsilon})/f_d$, where f_d (in cycles/s) is the CPU frequency of UE2.

Secondly, the computing delay within the relay node is described as $T_r = T_{r1} + T_{r2}$, where T_{r1} and T_{r2} represent the delay of decompressing D/ρ_1 bits data with ratio ρ_1 and compressing D bits data with ratio ρ_2, respectively, i.e., $T_{r1} = D\xi_1(e^{\varepsilon \rho_1} - e^{\varepsilon})/f_m$, $T_{r2} = D\xi_2(e^{\varepsilon \rho_2} - e^{\varepsilon})/f_m$, with f_m (in cycles/s) denoting the equivalent CPU frequency of the relay node. Moreover, the transmission delay in hop-1 and hop-2 are given by $T_{tx} = D/\{\rho_1 r_1 (1 - P_1)\}$ and $T_{rx} = D/\{\rho_2 r_2 (1 - P_2)\}$, respectively

To conclude, the total time delay of this system is represented as $T = T_c + T_{r1} + T_{tx} + T_{rx} + T_d + T_{r2}$.

Cost Function: Based on Definition 1, the total cost function for MEC-assisted relay is expressed as

$$L_{MEC} = A_1(e^{\varepsilon \rho_1} - e^{\varepsilon}) + A_2(e^{\varepsilon \rho_2} - e^{\varepsilon}) + \frac{(\alpha'_1 p_1 + \gamma')}{\rho_1} e^{(B_1/p_1)} + \frac{A_3}{\rho_2} + A_4, \tag{5}$$

where $A_1 = \alpha_1 q_c D\xi_1 + \gamma D\xi_1(f_c^{-1} + f_m^{-1})$, $A_2 = \alpha_2 q_d D\xi_2 + \gamma D\xi_2(f_d^{-1} + f_m^{-1})$, $\alpha'_1 = \alpha_1 D/r_1$, $\gamma' = \gamma D/r_1$, $A_3 = (\gamma + \alpha_2 q_{rx})e^{(B_2/\rho_2)} \cdot D/r_2$, $A_4 = \alpha_1 C_1 + \alpha_2 C_2 + \gamma T_0$, $B_i = f(r_i)/g_i$, $i = 1,2$.

The main parameters in (5) are explained as follows. A_1 is related to the cost consisting of energy and delay when the original data are compressed at UE1 and then decompressed at the relay node, both with the ratio of ρ_1. A_2 is related to the cost when the original data are compressed at the relay node and then decompressed at UE2, both with the ratio of ρ_2. A_3 is related to the cost of the transmission in hop-2, noting that p_2 is constant and the energy consumption of the relay node is not counted. A_4 denotes the other cost consisting of the constant energy consumption of UEs, i.e., C_1, C_2, and the constant transmitting delay within the relay node, i.e., T_0.

In the following expressions, the computing and relaying model adopting a conventional relay node without the capability of MEC is considered to compare with the model with the MEC-assisted relay node. The total cost function of this model can be defined as a special case of $L(\rho_1, \rho_2, p_1)$ when $\rho_1 = \rho_2$, except that the delay within relay node is exactly the transmission delay between the APs, i.e., T_0, without the delay of data computing for the MEC-assisted relay. Therefore, the total cost function for the conventional relay model is given by

$$L_{non} = (A_1 + A_2 - A_0)(e^{\varepsilon\rho} - e^{\varepsilon}) + \frac{(\alpha'_1 p_1 + \gamma')}{\rho} e^{(B_1/p_1)} + \frac{A_3}{\rho} + A_4, \quad (6)$$

where $A_0 = \gamma D(\xi_1 + \xi_2)/f_m$ which is related to the time delay of data computation of MEC-assisted relay node.

3. Optimal Strategies

In this section, the transmission and compression strategies are firstly optimized for the MEC-assisted relaying model to minimize the total cost of energy and delay of the system. It comes out as a non-convex problem and we decompose this problem into two subproblems of which the optimal solutions are obtained with closed-form expressions. Then an iterative algorithm is proposed to tackle this problem by using the block coordinate descent (BCD) method. In the sequel, the cost of the non-MEC relaying model is also optimized with a similar optimal strategy used in the MEC-assisted relaying model.

3.1. MEC-Assisted Scenario

To begin with, we regard L_{MEC} as a function of five variables of the computation and communication process, i.e., $\rho_1, \rho_2, r_1, r_2, p_1$ (p_2 is set as a fixed value because the relay node is supplied by the power grid.), and our target is to find the optimal solution to minimize L_{MEC}. The optimization problem is formulated as follows.

$$\mathcal{P}1: \min_{\substack{\rho_1,\rho_2\in[1,+\infty) \\ r_1,r_2\in(0,+\infty) \\ p_1\in[p_{1,min},p_{1,max}]}} L_{MEC}(\rho_1,\rho_2,r_1,r_2,p_1). \quad (7)$$

Unfortunately, this problem is non-convex, which means it is hard to find an optimal solution of (7). To tackle this problem, we firstly decompose P1 into two subproblems, i.e., P2 and P3 as follows, and manage to prove the convexity or equivalent convexity of them. Then we propose an algorithm using the BCD method to optimize P2 and P3 iteratively. This algorithm can provide a local optimum for P1 with linear complexity and global convergence, which is illustrated as follows.

To begin with, we decompose P1 into

$$\mathcal{P}2: \min_{r_1,r_2\in(0,+\infty)} L_{MEC}(r_1,r_2) \quad (8)$$

$$\text{and } \mathcal{P}3: \min_{\substack{\rho_1,\rho_2\in[1,+\infty) \\ p_1\in[p_{1,min},p_{1,max}]}} L_{MEC}(\rho_1,\rho_2,p_1). \quad (9)$$

Then, we will show that $\mathcal{P}2$ is convex, and that $\mathcal{P}3$ is equivalently convex Firstly, it is proved in Theorem 1 that $\mathcal{P}2$ is a convex problem of both r_1 and r_2, and the optimal solution of $\mathcal{P}2$ is also given by closed-form expressions.

Theorem 1: *Assuming that p_1 is fixed, the optimal transmission rates of the MEC-assisted relaying system is given by*

$$r_i^* = \frac{BW(\eta_i)}{\ln 2}, \quad (10)$$

$W(\cdot)$ is Lambert W function and $\eta_i = \frac{p_i\sigma_i^2}{N_0}, i = 1,2.$

Proof: With considering $L_{MEC}(r_1,r_2)$ as a function of r_1 and r_2, it is easily proved that $L_{MEC}(r_1,r_2)$ is convex over r_1 and r_2, respectively, and the optimal solution is given by (10).

Theorem 1 indicates that for each hop, the optimal data rate depends on the SNR. This is because when p_1 and p_2 are given, the cost of the transmission phase is proportional to the

transmission time. Then $\mathcal{P}2$ is simplified as a minimization problem of the transmission time of two hops. Therefore, r_i^* in Theorem 1 represents the optimal data rate determined by the SNR, which leads to the minimal transmission time for each hop.

However, the subproblem $\mathcal{P}3$ is non-convex. To cope with this, we transform $\mathcal{P}3$ into

$$\mathcal{P}4: \min_{\substack{\rho_1,\rho_2\in[1,+\infty] \\ t\in[p_{1,max}^{-1},p_{1,min}^{-1}]}} L_{MEC}(\rho_1,\rho_2,t). \tag{11}$$

equivalently by replacing p_1 with $1/t$. We can prove that $\mathcal{P}4$ is jointly convex in (ρ_1,ρ_2,t) by (See Appendix A), and the optimal solution of $\mathcal{P}3$ is given by the following theorem .

Theorem 2: *The optimum for P3 is given by*

$$\rho_1^* = \max\left\{1,\frac{2}{\varepsilon}W\left(\frac{\sqrt{\varepsilon}}{2}\sqrt{\frac{(\alpha\prime_1/t^*+\gamma\prime)e^{B_1t^*}}{A_1}}\right)\right\}, \tag{12}$$

$$\rho_2^* = \max\left\{1,\frac{2}{\varepsilon}W\left(\frac{\sqrt{\varepsilon}}{2}\sqrt{\frac{A_3}{A_2}}\right)\right\}, \tag{13}$$

$$p_1^* = \begin{cases} p_{1,min}, & \frac{1}{t^*} < p_{1,min} \\ \frac{1}{t^*}, p_{1,min} \leq \frac{1}{t^*} \leq p_{1,max} \\ p_{1,max}, & \frac{1}{t^*} > p_{1,max} \end{cases} \tag{14}$$

where $t^* = \left(\sqrt{B_1\alpha_1(B_1\alpha_1+4\gamma)}-B_1\alpha_1\right)/(2B_1\gamma)$ *and* $W(\cdot)$ *is Lambert W function.*

Proof: See Appendix A.

Theorem 2 presents the optimal transmission and compression strategy of the MEC-assisted relaying system. The concise solution reveals that the optimal solution of hop-2, ρ_2^* depends on r_2 only and that hop-1, i.e., p_1^* and ρ_1^* depends on r_1 only, which means the optimization of two hops are thoroughly decoupled. Compared with the conventional relaying system, the MEC-assisted relaying system breaks through the limit caused by the asymmetry of communication and computing resources of UEs by optimizing the transmission and computation strategy for each hop independently, so that an advantage of cost performance may be achieved.

Remark 1: *The variation of ρ_2^* with respect to r_1 is shown as follows. It is derived from (13) that ρ_2^* is a monotonically increasing function of* $\exp\left[\left(2^{r_2/B}-1\right)/\eta_2\right]/r_2$ *, which is convex over r_2 and the optimal solution is $r_2^* = BW(\eta_2)/ln2.$. Therefore, ρ_2^* is convex over r_2 and it is minimized when $r_2 = r_2^*$. As has been discussed in Theorem 1, the transmission time of hop-2 is also minimized when $r_2 = r_2^*$ as well. This indicates that for hop-2, the cost of the transmission phase and computation phase are reduced simultaneously with the increase of r_2 when $r_2 < r_2^*$ and both of them are minimized when $r_2 = r_2^*$.*

Remark 2: *The variation of ρ_1^* and p_1^* with respect to r_1 is shown as follows. Firstly, ρ_1^* is convex over r_1 and achieves the minimum value when $r_1 = r_1^*$. Meanwhile, it is proved that p_1^* is positively correlated with r_1 from (14), which makes the correlation between ρ_1^* and r_1 intricate because ρ_1^* is also affected by p_1^* and the effect is uncertain. This indicates that when r_1 grows, the corresponding value of ρ_1^* needs to increase along with it, in case that the cost of transmission phase rises rapidly due to the increase of outage errors. Moreover, it is proved that the value of ρ_1^* is influenced by p_1^* and r_1, but not vice versa .*

So far, we have proved that as for a non-convex optimal problem, $\mathcal{P}1$ can be decomposed into two subproblems, $\mathcal{P}2$ and $\mathcal{P}3$, which are convex or equivalently convex, having closed-forme solutions both. Also, we notice that the optimal solution of $\mathcal{P}2$ and $\mathcal{P}3$ interacts with each other. Therefore the problem $\mathcal{P}1$ is solved by optimizing the two subproblems iteratively. With a BCD method, an algorithm (JETMRS) for $\mathcal{P}1$ is proposed as follows, and the

complexity and convergence of which will be analyzed in the next subsection.

   Algorithm 1 Joint optimization algorithm for Energy consumption and Time delay of MEC-assisted Relaying System (JETMRS)

Initialization:

Define the tolerance of accuracy δ and the maximum number of iterations N_{max};

Initialize the scalar weights of the cost function, i.e., α_1, α_2, γ, and the parameters of the MEC-assisted system, p_2, q_c, q_d, f_m, f_c, f_d, ect.;

Set the iteration number $i = 0$;

1: Calculate $r_2 \rho_1^{(i)}$ according to (10), set $r_1^{(i)} = r_2$;

2: Calculate ρ_2 according to (13);

3: **repeat**

4:      Calculate $p_1^{(i)}$ and $\rho_1^{(i)}$ according to (14) and (12), respectively;

5:      Calculate $r_1^{(i+1)}$ according to (10);

6:      Update the iteration number: $i \leftarrow i + 1$;

7: **until** $\left| L_{MEC}\left(\rho_1^{(i)}, p_1^{(i)}, r_1^{(i+1)}\right) - L_{MEC}\left(\rho_1^{(i-1)}, p_1^{(i-1)}, r_1^{(i)}\right) \right| \leq \delta$ or reaching the maximum iteration number.

3.2. Non-MEC Scenario Scenarios

In this subsection, we investigate the conventional relaying model without adopting MEC servers, of which the total cost function is optimized similarly with the optimization of the MEC-assisted relaying system, and then the performance of the two systems are compared.

Firstly, the optimization problem for the non-MEC relaying system is formulated as

$$\mathcal{P}5: \min_{\substack{\rho, \in[1,+\infty) \\ r_1, r_2 \in (0,+\infty) \\ p_1 \in [p_{1,min}, p_{1,max}]}} L_{non}(\rho, r_1, r_2, p_1). \tag{15}$$

It can be proved that $\mathcal{P}5$ can be optimized with a similar method with that of $\mathcal{P}1$, therefore the optima of r_i are given by Theorem 1, and ρ^* and p_1^* are derived by the following theorem.

Theorem 3: *The optimum of $\mathcal{P}7$ is given by*

$$\rho^* = \max\left\{1, \frac{2}{\varepsilon} W\left(\frac{\sqrt{\varepsilon}}{2} \sqrt{\frac{\left(\alpha'_1/t^* + \gamma'\right) e^{B_1 t^*} + A_3}{(A_1 + A_2 - A_0)}}\right)\right\}, \tag{16}$$

$$p_1^* = \begin{cases} p_{1,min}, & \frac{1}{t^*} < p_{1,min} \\ \frac{1}{t^*}, p_{1,min} \leq \frac{1}{t^*} \leq p_{1,max}, \\ p_{1,max}, & \frac{1}{t^*} > p_{1,max} \end{cases} \tag{17}$$

where $t^* = \left(\sqrt{B_1 \alpha_1 (B_1 \alpha_1 + 4\gamma)} - B_1 \alpha_1\right)/(2B_1 \gamma)$.

Proof: Refer to the proof of Theorem 2.

Based on the results of Theorems 1 and 3, a BCD-based algorithm is also proposed to optimize $\mathcal{P}5$ iteratively. This algorithm, namely JETCRS, is given as follows.

   Algorithm 2 Joint optimization algorithm for Energy consumption and Time delay of Conventional Relaying System (JETCRS)

Initialization:

Define the tolerance of accuracy δ and the maximum number of iterations N_{max};

Initialize the scalar weights of the cost function, i.e., α_1, α_2, γ, and the parameters of the MEC-assisted system, p_2, q_c, q_d, f_m, f_c, f_d, ect.;

Set the iteration number $i = 0$;

1: Calculate $r_2 \rho^{(i)}$ according to (10), set $r_1^{(i)} = r_2$;

2: **repeat**

3:　　　Calculate $p_1^{(i)}$ and $\rho^{(i)}$ according to (20) and (19), respectively;

4:　　　Calculate $r_1^{(i+1)}$ according to (10);

5:　　　Update the iteration number: $i \leftarrow i + 1$;

6: **until** $\left| L_{non}\left(\rho_1^{(i)}, p_1^{(i)}, r_1^{(i+1)}\right) - L_{non}\left(\rho_1^{(i-1)}, p_1^{(i-1)}, r_1^{(i)}\right) \right| \leq \delta$ or reaching the maximum iteration number.

Remark 3: *It is observed that the optimal solution of p_1 and r_1 in the non-MEC relaying system are the same as that in the MEC relaying system under the identical fading channel of hop-1. This means that the optimal transmit power for UE 1 and the optimal value of transmission data rate of hop-1 are only influenced by the channel fading, whether the relay node has the capability of MEC or not. Besides, the factor of the optimal strategy influenced by the relay type is the compression ratio of data, which makes a difference not only on the cost during the computation phase but also on that in the transmission phase because of the changing of transmitting time. Different from the MEC relaying system, the conventional relaying system has only one compression ratio, ρ, through the two hops, which narrows the space for optimization of the system performance and this is shown by the following corollary.*

Corollary 1. *Assuming $f_m \gg \max\{f_c, f_d\}$, we have two conclusions from Theorem 2 and 3 as follows.*

1. ρ_1^* and ρ_2^* are positively correlated to f_c and f_d, respectively.

2. $\min\{\rho_1^*, \rho_2^*\} \leq \rho^* \leq \max\{\rho_1^*, \rho_2^*\}$, and the equality holds only when $\rho^* = \rho_1^* = \rho_2^* = 1$.

　　Proof: First, it is easily observed that ρ_1^* and ρ_2^* are negatively correlated with A_1 and A_2, respectively, as $W(x)$ monotonically increases with x in domain $\mathcal{R}_+$. Then, when $f_m \gg \max\{f_c, f_d\}$, A_1 and A_2 are positively correlated with f_c^{-1} and f_d^{-1}, respectively, and thus the first conclusion is obtained.

　　Next, with comparing the optimal compression ratios of the two system which are given by Theorem 2 and 3, it is easily observed that $\rho^* = \rho_1^* = \rho_2^*$ holds only if $\rho^* = \rho_1^* = \rho_2^* = 1$ based on the monotonically increasing property of $W(x)$. Moreover, if $f_m \gg \max\{f_c, f_d\}$, we obtain $A_1 + A_2 \gg A_0$ and therefore

$$\rho^* \approx \max\left\{1, \frac{2}{\varepsilon}\left(\frac{\sqrt{\varepsilon}}{2}\sqrt{\frac{(\alpha'_1/t+\gamma')e^{B_1 t^*}+A_3}{(A_1+A_2)}}\right)\right\}. \tag{18}$$

In this case, it is observed that

$$\min\left\{\frac{(\alpha'_1/t+\gamma')e^{B_1 t^*}}{A_1}, \frac{A_3}{A_2}\right\} \leq \frac{(\alpha'_1/t+\gamma')e^{B_1 t^*}+A_3}{(A_1+A_2)} \leq \max\left\{\frac{(\alpha'_1/t+\gamma')e^{B_1 t^*}}{A_1}, \frac{A_3}{A_2}\right\}. \tag{19}$$

　　Again, with using the monotonically increasing property of $W(x)$ in domain $\mathcal{R}_+$, the second conclusion of Corollary 1 is proved.

Remark 4: *Corollary* 1 *considers the scenarios when the MEC servers is far more capable of computation than the UEs, i.e.,* $f_m \gg \max\{f_c, f_d\}$, *which are the most cases in practical scenarios. For these scenarios, the physical meaning of the two conclusions can be explained as follows. The first conclusion of Corollary 1 indicates that the optimal compression rate in each hop is bottlenecked by the computing capability of the corresponding UE. It is easily understood that the shortage of the computing resource in each hop decides the upper limit of maximal compression rate that the optimal strategy can achieve.*

The second conclusion of Corollary 1 reveals that ρ^* *of non-MEC system that optimize the performance of two hops as a whole is approximately between* ρ_1^* *and* ρ_2^* *of MEC-assisted system that optimize them separately. Thus, it is presumed that the gap between* ρ_1^* *and* ρ_2^* *being greater, the advantage of the MEC-assisted system over the non-MEC system will be more significant and vice versa.*

3.3. The Convergency and Complexity of Two Algorithms

Note that $\mathcal{P}1$ is non-convex, therefore there is no guarantee to obtain a global optimum of $\mathcal{P}1$. However, it is proved that $\mathcal{P}1$ is equivalent to a block multi-convex problem, of which a local optimal solution (a critical point) can be obtained with the BCD method algorithm, namely JETMRS. Moreover, the JETMRS algorithm is proved to be globally convergent to a local optimum. Similarly, it is proved that $\mathcal{P}5$ is equivalent to a block multi-convex problem. Therefore, the JETCRS algorithm can guarantee a local optimum of $\mathcal{P}5$ with global convergence .

The overall computational complexity of algorithms of JETMRS and JETCRS are both determined by the number and complexity of iterations which are derived as follows. The JETMRS algorithm consists of two iterative procedures, optimizing r_1 when fixing ρ_1 and p_1 and vice versa. Since the optimal solution of r_1, ρ_1 and p_1 are all given in closed-form expressions. The complexity of calculating them in a single loop is $O(1)$. Hence the computation complexity of the JETMRS algorithm is $O(T_1)$, where T_1 denotes the iterations of the JETMRS algorithm. Similarly, the complexity of the JETCRS algorithm is $O(T2)$, where $T2$ denotes the number of iterations of it. Therefore, these two algorithms both have linear computation complexity, which is efficient to be applied in practical scenarios.

4. Performance Analysis

In this section, we further investigate the performance of the MEC-assisted system including comparison with the non-MEC system. Firstly, we evaluate the superiority of MEC-assisted system from the cost difference between the two systems to obtain the condition of when MEC-assisted relay is more preferable than conventional relay in saving the total cost. In addition, some special cases are shown to give some physical insight for practical scenarios.

4.1 Cost Difference

As has been discussed, the MEC-assisted system achieve better performance than the non-MEC system by decoupling the compression data rates of the hops with the assistance of MEC servers. However, the cost of the MEC-assisted system including the time delay of data processing and transmitting within MEC servers should be also counted. To carefully examine the advantage of the MEC-assisted system, we define the cost difference between the non-MEC and MEC-assisted system as $G_{diff}^* = L_{non}(\rho^*, p_1^*) - L(\rho_1^*, \rho_2^*, p_1^*)$, where $L_{non}(\rho^*, p_1^*)$ and $L(\rho_1^*, \rho_2^*, p_1^*)$ are both optimized by the proposed algorithms under the same fading

channels and energy status.

It is easily observed that G^*_{diff} indicates the performance gain of the MEC-assisted system over the non-MEC system with their optimal solutions. status. When $G^*_{diff} > 0$, it means that the MEC-assisted system saves more energy-delay cost in this communication-and-computing process than the non-MEC system does and vice versa. Moreover, if $G^*_{diff} = 0$, it means that the two relaying systems are equal in terms of the total cost in their optimal status. However, it is obvious that conventional relaying is preferable to the MEC-assisted relaying when $G^*_{diff} = 0$, as the latter consumes extra computing resources of MEC servers. To compare the two relaying systems in detail with G^*_{diff}, we show the following theorem.

Theorem 4: *The condition of whether the non-MEC or MEC-assisted relaying system is better for saving the total cost is given by*

$$\begin{cases} G^*_{diff} > 0, & \text{when } \rho^* < f_d(\rho^*_1, \rho^*_2) \\ G^*_{diff} = 0, & \text{when } \rho^* = f_d(\rho^*_1, \rho^*_2), \\ G^*_{diff} < 0, & \text{when } \rho^* > f_d(\rho^*_1, \rho^*_2) \end{cases} \tag{20}$$

where $f_d(\rho^*_1, \rho^*_2) = \dfrac{\varepsilon b_1 + \sqrt{b_1\left((c_1(2+\varepsilon\rho^*_1)^2 - 4A_0 e^\varepsilon \varepsilon(\rho^*_1)^2)(\rho^*_2)^2 + A_3(\rho^*_1)^2(2+\varepsilon\rho^*_2)^2\right)}}{2\left((c_1 + c_1\varepsilon\rho^*_1 - A_0 e^\varepsilon \varepsilon(\rho^*_1)^2)(\rho^*_2)^2 + A_3(\rho^*_1)^2(1+\varepsilon\rho^*_2)\right)}$ (21)

with $b_1 = (A_3 + c_1)(\rho^*_1)^2(\rho^*_2)^2$ *and* $c_1 = (\alpha'_1 p^*_1 + \gamma')e^{(B_1/p^*_1)}$.

Proof: To begin with, the expression of cost difference is given as

$$G^*_{diff} = A_1\left(e^{\varepsilon\rho^*} - e^{\varepsilon\rho^*_1}\right) + A_2\left(e^{\varepsilon\rho^*} - e^{\varepsilon\rho^*_2}\right) + c_1\left(\frac{1}{\rho^*} - \frac{1}{\rho^*_1}\right) + A_3\left(\frac{1}{\rho^*} - \frac{1}{\rho^*_2}\right) - A_0\left(e^{\varepsilon\rho^*} - e^\varepsilon\right), \tag{22}$$

where $c_1 = (\alpha'_1 p^*_1 + \gamma')e^{(B_1/p^*_1)}$. Without loss of generality, we assume $\rho^*, \rho^*_1, \rho^*_2 > 1$ and therefore obtain $\rho^* = \frac{2}{\varepsilon}W\left(\frac{\sqrt{\varepsilon}}{2}\sqrt{\frac{c_1 + A_3}{(A_1 + A_2 - A_0)}}\right)$, $\rho^*_1 = \frac{2}{\varepsilon}W\left(\frac{\sqrt{\varepsilon}}{2}\sqrt{\frac{c_1}{A_1}}\right)$ and $\rho^*_2 = \frac{2}{\varepsilon}W\left(\frac{\sqrt{\varepsilon}}{2}\sqrt{\frac{A_3}{A_2}}\right)$.

Then with using the property of Lambert W function, i.e., $x = W(x)e^{W(x)}$, we obtain $e^{\varepsilon\rho^*_1} = \frac{c_1}{\varepsilon A_1}\cdot(\rho^*_1)^{-2}$, $e^{\varepsilon\rho^*_2} = \frac{A_3}{\varepsilon A_2}\cdot(\rho^*_2)^{-2}$ and $e^{\varepsilon\rho^*} = \frac{c_1 + A_3}{\varepsilon(A_1 + A_2 - A_0)}\cdot(\rho^*)^{-2}$. We substitute the above equations into (22) and therefore obtain

$$G^*_{diff} = c_1\left(\frac{1}{\rho^*} - \frac{1}{\rho^*_1}\right)\left(1 + \frac{1}{\varepsilon\rho^*} + \frac{1}{\varepsilon\rho^*_1}\right) + A_3\left(\frac{1}{\rho^*} - \frac{1}{\rho^*_2}\right)\left(1 + \frac{1}{\varepsilon\rho^*} + \frac{1}{\varepsilon\rho^*_2}\right) + A_0 e^\varepsilon. \tag{23}$$

Next, by solving the equation of $G^*_{diff} = 0$, we obtain $\rho^* = f_d(\rho^*_1, \rho^*_2)$. Regarding G^*_{diff} in (23) as a function of ρ^*, which monotonically decreases with ρ^*, therefore Theorem 4 is proved.

This theorem reveals the condition of whether the MEC-assisted relaying system is superior to the conventional relaying system or not in terms of the energy-delay cost. We note that the closed-form expression of the condition is influenced by many factors, i.e., $\rho^*, \rho^*_1, \rho^*_2, A_0, A_3$ and c_1, and it is too complicated to analyze the effect of any single factor because these factors interact with each other constantly. However, by calculating the numerical results of (21), Theorem 4 can provide a criterion for the choice between the two systems in practical scenarios.

As has been discussed, it is hard to track the effect of any single factor on the cost difference because all factors affect one another mutually. Therefore we need to consider some special cases and further investigations on the cost difference are shown in the following subsection.

4.2 Special Cases

It is obvious that if the computing resources of UEs were sufficient and the channels in the relaying model were ideal, there would be no need for MEC servers. Hence we should consider what condition will allow the MEC-assisted system to outperform the non-MEC system in the special case when the computing or communication resources are poor. For the facility of analysis several extreme conditions are assumed in the following special cases.

- *Poor Computing in UE1 or UE2*

Consider the case when the computing capability of UE1 or UE2 is extremely insufficient, which forces the compression or decompression rate of the user to be 1, i.e., $\rho^* = \rho_1^* = 1$ or $\rho^* = \rho_2^* = 1$. For these extreme cases, we obtain the following corollary.

Corollary 2: *In the case of the poor computing capability of UE*1 *or UE*2, *i.e.,* $\rho^* = \rho_1^* = 1$ *or* $\rho^* = \rho_2^* = 1$, *the MEC-assisted system is always superior to the non-MEC system when* $\rho_2^* > 1$ *or* $\rho_1^* > 1$ *respectively, which is*

$$\begin{cases} G_{diff}^* > 0, \text{ when } \rho^* = \rho_1^* = 1 \text{ and } \rho_2^* > 1 \\ G_{diff}^* > 0, \text{ when } \rho^* = \rho_2^* = 1 \text{ and } \rho_1^* > 1 \end{cases} \tag{24}$$

Moreover, it is observed that $\rho_2^* > 1$ *is equivalent to* $A_3/A_2 > \varepsilon e^\varepsilon$ *and* $\rho_1^* > 1$ *is equivalent to* $c_1/A_1 > \varepsilon e^\varepsilon$.

Proof: Firstly, we rewrite the cost different of the special cases as

$$G_{diff}^* = \begin{cases} (A_2 e^\varepsilon + A_3) - (A_2 e^{\varepsilon \rho_2^*} + A_3/\rho_2^*), \rho^* = \rho_1^* = 1 \\ (A_1 e^\varepsilon + c_1) - (A_1 e^{\varepsilon \rho_1^*} + c_1/\rho_1^*), \rho^* = \rho_2^* = 1 \end{cases} \tag{25}$$

Note that ρ_1^* is the optimum of the convex function $A_1 e^{\varepsilon \rho} + c_1/\rho$ when $\rho_1^* > 1$, and it is the same with ρ_2^* to $A_2 e^{\varepsilon \rho} + A_3/\rho$ when $\rho_2^* > 1$. Therefore it is easily obtained that $G_{diff}^* > 0$ for the cases of $\rho^* = \rho_2^* = 1$ or $\rho^* = \rho_2^* = 1$.

Remark 5: *It is easily understood that the MEC-assisted system has notable advantages over the non-MEC system when one UE is in lack of computing capability. This is because the cost gain from resource optimizing of the non-MEC system is restricted to* 0 *due to the poor computing capability of one UE whilst the MEC-assisted system only loses them in one hop. Therefore the MEC-assisted system is a better choice, i.e.,* $G_{diff}^* > 0$ *in these special cases, as long as the optimal compression rate of another hop is greater than* 1 *under the condition given by Corollary* 2. *As has been discussed before,* A_1 *and* A_2 *are related to the energy-delay cost of the computation phase while* c_1 *and* A_3 *are related to that of the transmission phase. Then* $A_3/A_2 > \varepsilon e^\varepsilon$ *and* $c_1/A_1 > \varepsilon e^\varepsilon$ *indicate that when the cost ratio of the transmission phase over the computation phase in one hop is larger than a given threshold, i.e.* $\varepsilon e^\varepsilon$, *it is feasible to use the computing resources of the MEC servers to alleviate the cost of communication. From these special cases, we propose a reasonable deduction that the advantage of the MEC-assisted system is enhanced when the disparity between the computing source of two UEs enlarges, which will be verified by the simulation results.*

- *Poor Communication in hop-1 or hop-2*

Consider the case when the wireless channel of hop-1 or hop-2 experiences a deep fade so that the transmission data of the corresponding hop is compressed/decompressed with the maximum ratio, i.e., $\rho^* = \rho_1^* = \rho_{1,\max}$ or $\rho^* = \rho_2^* = \rho_{2,\max}$. For these extreme cases, we obtain the following corollary.

Corollary 4: *In the case of the channel of hop-*1 *or hop-*2 *in deep fading, the condition of whether the MEC-assisted system or the non-MEC system is better in the total cost is given as follows. When* $\rho^* = \rho_1^* = \rho_{1,max}$, *we obtain*

$$\begin{cases} G_{diff}^* > 0, \text{ when } \rho_2^* > \left[A_3\varepsilon + \sqrt{A_3\varepsilon(4a_2 + A_3\varepsilon)}\right]/(2a_2\varepsilon) \\ G_{diff}^* = 0, \text{ when } \rho_2^* = \left[A_3\varepsilon + \sqrt{A_3\varepsilon(4a_2 + A_3\varepsilon)}\right]/(2a_2\varepsilon), \\ G_{diff}^* < 0, \text{ when } \rho_2^* < \left[A_3\varepsilon + \sqrt{A_3\varepsilon(4a_2 + A_3\varepsilon)}\right]/(2a_2\varepsilon) \end{cases} \quad (26)$$

where $a_2 = A_2 e^{\varepsilon\rho_{2,max}} + A_3/\rho_{2,max} - A_0(e^{\varepsilon\rho_{2,max}} - e^\varepsilon)$.

When $\rho^* = \rho_2^* = \rho_{2,max}$, we obtain

$$\begin{cases} G_{diff}^* > 0, \text{ when } \rho_1^* > \left[c_1\varepsilon + \sqrt{c_1\varepsilon(4a_1 + c_1\varepsilon)}\right]/(2a_1\varepsilon) \\ G_{diff}^* = 0, \text{ when } \rho_1^* = \left[c_1\varepsilon + \sqrt{c_1\varepsilon(4a_1 + c_1\varepsilon)}\right]/(2a_1\varepsilon), \\ G_{diff}^* < 0, \text{ when } \rho_1^* < \left[c_1\varepsilon + \sqrt{c_1\varepsilon(4a_1 + c_1\varepsilon)}\right]/(2a_1\varepsilon) \end{cases} \quad (27)$$

where $a_1 = A_1 e^{\varepsilon\rho_{1,max}} + c_1/\rho_{1,max} - A_0(e^{\varepsilon\rho_{1,max}} - e^\varepsilon)$.

Proof: Firstly, we represent the cost difference as

$$G_{diff}^* = A_2\left(e^{\varepsilon\rho_{1,max}} - e^{\varepsilon\rho_2^*}\right) + A_3\left(\frac{1}{\rho_{1,max}} - \frac{1}{\rho_2^*}\right) - A_0(e^{\varepsilon\rho_{1,max}} - e^\varepsilon), \ \rho^* = \rho_1^* = \rho_{1,max}, \quad (28)$$

or $G_{diff}^* = A_1\left(e^{\varepsilon\rho_{2,max}} - e^{\varepsilon\rho_1^*}\right) + c_1\left(\frac{1}{\rho_{2,max}} - \frac{1}{\rho_1^*}\right) - A_0(e^{\varepsilon\rho_{2,max}} - e^\varepsilon), \ \rho^* = \rho_2^* = \rho_{2,max}. \quad (29)$

Then, we substitute $e^{\varepsilon\rho_2^*} = A_3/[\varepsilon A_2(\rho_2^*)^2]$ into (28) and obtain $G_{diff}^* = -\frac{A_3}{\varepsilon(\rho_2^*)^2} - \frac{A_3}{\rho_2^*} + a_2$ with $a_2 = A_2 e^{\varepsilon\rho_{2,max}} + A_3/\rho_{2,max} - A_0(e^{\varepsilon\rho_{2,max}} - e^\varepsilon)$. Noting that G_{diff}^* increases monotonically with ρ_2^*. Then we solve the equation of $G_{diff}^* = 0$ and obtain (26). Next, we substitute $e^{\varepsilon\rho_1^*} = c_1/[\varepsilon A_1(\rho_1^*)^2]$ into (29) and obtain $G_{diff}^* = -\frac{c_1}{\varepsilon(\rho_1^*)^2} - \frac{c_1}{\rho_1^*} + a_1$ with $a_1 = A_1 e^{\varepsilon\rho_{1,max}} + c_1/\rho_{1,max} - A_0(e^{\varepsilon\rho_{1,max}} - e^\varepsilon)$. Noting that G_{diff}^* increases monotonically with ρ_1^*. Similarly, we solve the equation of $G_{diff}^* = 0$ and obtain (27).

Remark 6: *Firstly, it is observed that G_{diff}^* increases with the increase of f_m when $f_m \gg \max\{f_c, f_d\}$.). This means when the computing capability of MEC is enhanced, the advantage of the MEC-assisted system over the non-MEC system will increase to some extent.*

Secondly, we note that G_{diff}^* is also influenced by the disparity between $\rho_{1,max}$ and ρ_2^* or between $\rho_{2,max}$ and ρ_1^*, respectively. This means that the advantage of the MEC-assisted system over the non-MEC system will diminish when the disparity of them becomes small. Typically, the MEC-assisted system is inferior to the non-MEC system when $\rho_2^* = \rho_{1,max}$ or $\rho_1^* = \rho_{2,max}$. For example, we consider a practical scenario when two hops both suffer from deep fading and both UEs have the same level of maximum compression/decompression rate, i.e., $\rho_{1,max} = \rho_{2,max}$. For this special case, there is no need to adopt MEC-assisted servers because no performance gain can be achieved compared with the non-MEC system.

To sum up, for the aforementioned special cases, the advantage of the MEC-assisted system holds when one of the UEs is short of computing capability whereas it is not the case when the channel of one hop suffers from deep fading. In particular, there is no advantage for adopting the MEC-assisted system in comparison with the conventional relaying system when channels of both hops are in deep fading.

5. Numerical Results and Discussion

In this section, we further evaluate the performance of the MEC-assisted system by simulation results with the comparison of the non-MEC system. Some of the relevant parameters in simulations are given as follows. The transmit rate for hop-1 and hop-2 are

preset to $r_1 = r_2 = 2$ Mbit/s while the transmit power of UE1 and the relay node are preset as $p_1 = p_2 = 1$ W. The amount of data D that UE1 transmits to UE2 through the relay node is 10 M bits and the utility weights of the cost function are $\gamma = 0.5$/Sec, $\alpha_1 = 0.25$/Joul, and $\alpha_2 = 0.25$/Joul. The energy consumption for receiving each bit in the receiver is $q_{Rx} = 0.42 \times 10^{-6}$ Joul/bits. For the data processing, the required numbers of CPU cycles for compression and decompression are set to $\xi_1 = 1500$ cycles/bit and $\xi_2 = 500$ cycles/bit, respectively. The CPU-cycle frequency of the UEs is set to $f_{c(d)} \in [10,1000]$ MHz and the energy consumption per cycle is $q_{c(d)} = 1 \times 10^{-13}$ Joul/bit. The equivalent CPU-cycle frequency of the MEC server is $f_m = 100$GHz and $\varepsilon = 0.5$. For simplicity, we set other constant energy consumptions, i.e., C_1, C_2, to 0 and denote $g_i = \sigma_i^2/N_0$, $i = 1,2$. Besides, we set the tolerance of accuracy $\delta = 10^{-4}$ and the maximum number of iterations $N_{max} = 500$ for the running of Algorithms 1 and 2.

Fig. 2(a) verifies the effectiveness of Algorithms 1 and 2 in this paper. A very fast convergence rate is shown in global sub-optimal solutions of the MEC-assisted and non-MEC systems with 15 iterations for most cases. **Fig. 2(a)** also compares the total cost of the MEC-assisted system with the non-MEC system under different channel conditions of $g_1 = g_2 = -20$ dB and $g_1 = g_2 = 0$ dB while $f_c = 200$ MHz and $f_d = 500$ MHz. It is observed that the cost difference of the system, i.e., G_{diff}^*, varies within a large range under different fading channels. Therefore we define an indicator of cost gain, $G_L = L/L_{non}$, for a better comparison of the system performance between the MEC-assisted system and the non-MEC system.
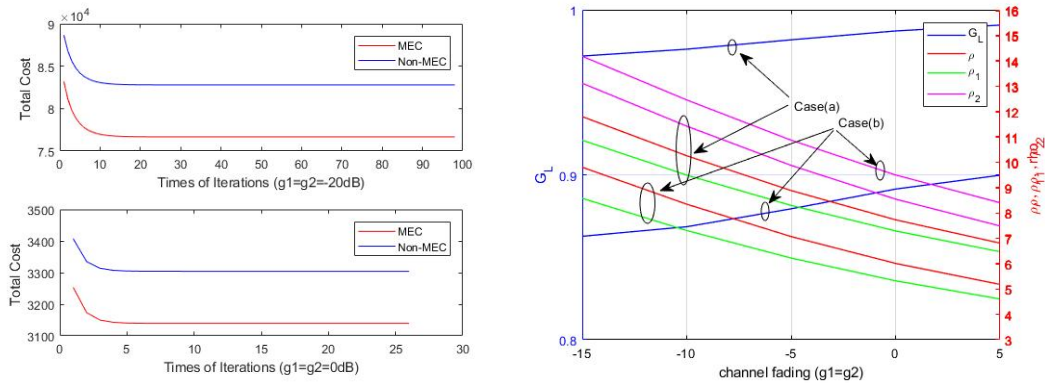
(a) Comparison of total cost of the two systems. (b) The cost gain and optimal compression rates.
Fig. 2. The performance comparison with different channel gains and different CPU frequencies.

Fig. 2(b) illustrates the effect of different CPU frequencies of the UEs on the two systems, where figures are given for Case (a): $f_c = f_d = 500$ MHz and Case (b): $f_c = 100$ MHz, $f_d = 1000$ MHz. It is observed in **Fig. 2(a)** that the MEC-assisted system is superior to the non-MEC system in Case (b) by acquiring smaller G_L than in Case (a). Furthermore, **Fig. 2(a)** shows that the gap between ρ_1^* and ρ_2^* of Case (b) is much greater than that of Case (a), which verified the conclusion in Remark 5 that a greater gap of $|\rho_1^* - \rho_2^*|$ leads to better performance for the MEC-assisted system.

For the following figures, we consider more practical situations where $\rho_{max} = \rho_{1,max} = \rho_{2,max} = 15$, $p_{1,max} = 1W$ and $p_2 = 1W$ unless otherwise noted (Therefore g_2 is equivalent to the SNR of hop-2, and g_1 is equivalent to the SNR of hop-1 approximately.). **Fig. 3** illustrates the performance gain of the proposed systems in terms of the SNR of hop-1 or hop-2 where $f_c = f_d = 500$ MHz. For **Fig. 3(a)**, the channel gain of hop-2 varies between $[-15,5]$ dB while the channel gain of hop-1 is set to $g_1 = -10$ dB. For **Fig. 3(b)**, the channel

gain of hop-1 varies between $[-15,5]$ dB while the channel gain of hop-2 is set to $g_2 = -10$ dB.

It is observed in **Fig. 3** that G_L is generally smaller than 1 during the variation of the SNR of hop-2 or hop-1 with low SNR regime of the other hop (below -10 dB). This means that the MEC-assisted system is superior to the non-MEC system when the access channel of either UE is not in good condition. It is also shown that the advantage of the MEC-assisted system relies on the SNR quality of hop-1 much more than on hop-2. This is validated from **Fig. 3** that G_L decreases rapidly when g_1 is larger than g_2, especially where g_1 is above 0 dB. This is because when the SNR of hop-1 is much better than hop-2, the cost of the hop-1 will drop off significantly, therefore the problem of minimizing the cost of hop-2 is highlighted. Then the advantage of MEC-assisted relaying over conventional relaying is fully utilized whereas it is further impaired for the same reason when SNR of hop-1 is much lower than hop-2. Besides, it is observed from **Fig. 3** that ρ_1^* and ρ_2^* decrease respectively with the increase of g_1 and g_2, while ρ^* falls with each of them. This is easily understood that when the channel quality is improved, the cost in the communication phases will decline and more emphasis will be put on reducing that for the computing phases to optimize the total cost, which leads to the reduction of the compression rates of the corresponding hop.
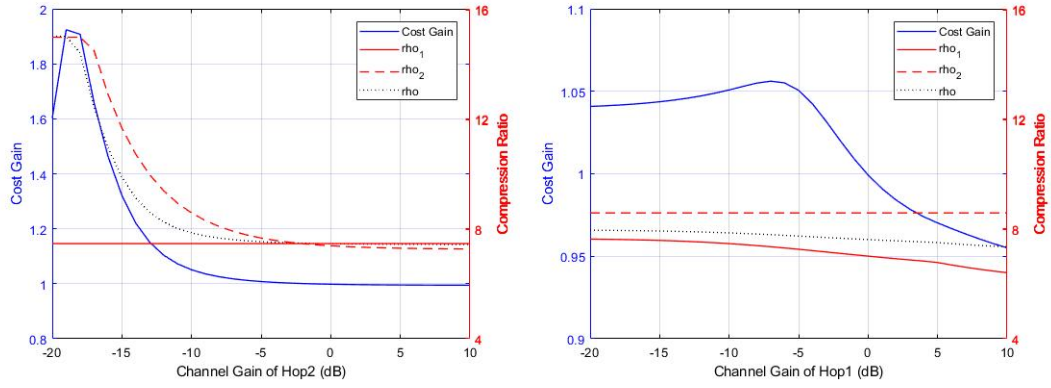
(a) The cost gain and compression rates with g_2.   (b) The cost gain and compression rates with g_1.
Fig. 3. The performance comparison of two relaying systems under different fading channels.
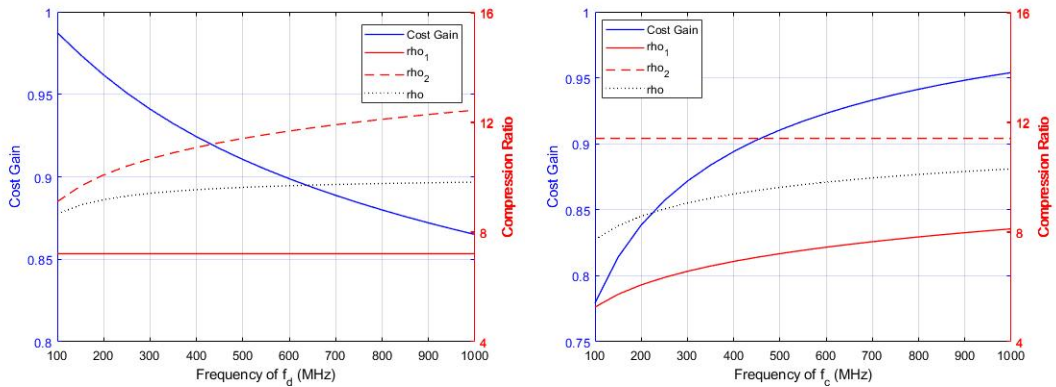
(a) The cost gain and compression rates with f_d.   (b)The cost gain and compression rates with f_c.
Fig. 4. The performance comparison of two systems under different computing capability of UEs.

Fig. 4 illustrates the performance gain of the MEC-assisted system with different CPU frequencies of the UEs, where $f_c = 500$ MHz for **Fig. 4(a)** and $f_d = 500$ MHz for **Fig. 4(b)**.

As we discussed previously, the condition of the SNR of hop-1 being better than hop-2 benefits for improving the advantage of MEC servers. Therefore we set $g_1 = 0$ dB and $g_2 = -10$ dB in **Fig. 4** for the simplicity of investigation. It is observed that the performance of the MEC-assisted system outweighs that of the non-MEC system in both **Figs. 4(a) and 4(b)**. Besides, it is observed in **Fig. 4(a)** that the advantage of the MEC-assisted system weakens with the increase of computing capability of UE1 but strengthens with the increase of computing capability of UE2. There are two reasons for this phenomenon: Firstly, as we mentioned before the advantage of MEC relaying is enhanced when the gap between ρ_1^* and ρ_2^* increases and vice versa. Secondly, given the situation that the SNR of hop-1 is much better than that of hop-2, it is obvious that for the optimal strategy ρ_2^* will be greater than ρ_1^* to combat the worse channel fading. Hence when ρ^* and ρ_2^* increase with the rise of f_d, the gap between ρ_1^* and ρ_2^* is expanded, which leads to the increase of G_L in **Fig. 4(a)**. On the other hand, as ρ^* and ρ_1^* increase with the rise of f_c, the gap between ρ_1 and ρ_2 is narrowed, which leads to the decrease of G_L in **Fig. 4(b)**. To sum up, the advantage of the MEC-assisted system is notably strengthened when the channel condition of hop-1 is better than hop-2 and it is further enhanced while the computing capability of UE2 is better than UE1.

As shown in the analytical and numerical results of this paper, this computing and relaying system was analyzed on to what extent MEC servers at the APs can accelerate the P2P communication between the mobile devices in saving the cost of energy consumption and time delay. The advantage of the MEC- assisted relaying system was discussed in detail, which provides valuable physical insights for the practical scenarios. But in addition, certain issues still need to be considered for the extension of this work. Firstly, this model is based on the P2P communication between one pair of mobile devices, which can be further extended to the multi-pairs P2P communication scenario. Secondly the task offloading scheduling of MEC servers should be considered as well as the cost minimization and a cost-aware offloading scheme needs to be designed to jointly optimize the offloading policy and cost performance for the extension model. Then, as the energy consumption of mobile devices is strictly constrained by the limited battery capacity, for this reason we concentrated on them and neglected the energy consumption of the APs and MEC servers which are supplied by the grid. Furthermore, the energy efficiency of the APs and MEC servers needs further investigation if a conversion factor is elaborately defined to depict correlation between the improvement of the communication and the energy consumption of the APs and MEC servers. Lastly, more investigations of the system performance on other wireless channels, e.g., millimeter wave channels are advisable for physical insights in the practical scenarios of 6G communication networks.

6. Conclusion

In this paper, we studied the P2P wireless relaying system where two mobile users communicate with each other through the aid of adjacent APs adopting MEC servers. In this MEC-assisted communication system the APs can adjust the compression ratios of transmission data to accommodate the computation and communication resources of each hop with the help of MEC servers, which leads to less time delay and energy consumption. The optimal strategy of computing and transmitting was investigated for this system with consideration of the variation of wireless channels and the computing capacity of the MEC servers. Because this problem is non-convex, a BCD-method algorithm was proposed to obtain a local optimal solution with linear complexity and global convergency. It is shown that the optima of the transmit power and data rates are decided by the variation of the fading channels

only whilst the optima of compression ratios are influenced by the state of the fading channels and computation resources. Then the performance of the MEC-assisted communication system was evaluated through comparison with that of the conventional communication system which was also optimized with a BCD-method algorithm. From the optimal solution of the problem and the Monte-Carlo simulations, the advantage of the MEC-assisted system over the conventional system are elaborated as follows. Firstly, by decoupling the two hops and optimizing the compression ratios of transmission data separately, the MEC-assisted system can save more cost than the non-MEC system in the most cases with consideration of the extra processing cost at MEC servers. Specifically, the superiority of the MEC-assisted system is positively correlated to the gap between the optimal compression ratios of the two hops, i.e., $|\rho_1^* - \rho_2^*|$. Secondly, a closed-form condition of when the MEC-assisted system gains an advantage over the non-MEC system was derived for the general case, which shows that the advantages hold for the special case when the computing capability of one of the mobile users is extremely poor. Lastly, the performance of the MEC- assisted system in the practical scenarios was further studied. It is revealed that the advantage of the MEC-assisted system is strengthened by the increase of the computing capability of UE2 but weakened by the increase of the computing capability of UE1. Meanwhile, the advantage of the MEC-assisted system shows a remarkable improvement when the SNR of hop-1 is greater than that of hop-2.

Appendix A

Proof of Theorem 2

The proof of Theorem 2 relies on the following lemma.

Lemma 1: *The optimization problem $\mathcal{P}4$ is jointly convex in (ρ_1, ρ_2, t).*

Proof: By replacing p_1 with $1/t$, we reformulate the total cost function as

$$L_{MEC}(\rho_1, \rho_2, t) = A_1(e^{\varepsilon \rho_1} - e^\varepsilon) + A_2(e^{\varepsilon \rho_2} - e^\varepsilon) + \left(\alpha'_1/t + \gamma'\right) e^{B_1 t}/\rho_1 + A_3/\rho_2 + A_4. \quad (30)$$

It is easily derived that $A_1(e^{\varepsilon \rho_1} - e^\varepsilon) + A_2(e^{\varepsilon \rho_2} - e^\varepsilon) + A_3/\rho_2 + A_4$ is convex in ρ_1 and ρ_2, then the only remain part is to prove that $f(\rho_1, t) = \left(\alpha'_1/t + \gamma'\right) e^{B_1 t}/\rho_1$ is jointly convex function of ρ_1 and t, which is given as follows. To prove the convexity of $f(\rho_1, t)$, we define the Hessian Matrix of $f(\rho_1, t)$ as

$$\boldsymbol{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial t^2} & \frac{\partial^2 f}{\partial t \partial \rho_1} \\ \frac{\partial^2 f}{\partial \rho_1 \partial t} & \frac{\partial^2 f}{\partial \rho_1^2} \end{bmatrix}. \quad (31)$$

Then we have $H_{11} = \left[\left(\sqrt{\alpha'_1} - \sqrt{\alpha'_1} B_1 t\right)^2 + \alpha'_1 + \gamma' B_1^2 t^3\right] e^{B_1 t}/(t^3 \rho_1) > 0$ and $|\boldsymbol{H}| = \left[\left(B_1 t \left(\alpha'_1 + \gamma' t\right) - \alpha'_1\right)^2 + \alpha'_1 \left(2\alpha'_1 + 4\gamma' t\right)\right] e^{2B_1 t}/(\rho_1^4 t^4) > 0$, which fulfills the necessary and sufficient condition of a jointly convex function. Therefore we prove that $L_{MEC}(\rho_1, \rho_2, t)$ is a jointly convex function, and hence $\mathcal{P}4$ is jointly convex in (ρ_1, ρ_2, t).

Since we proved that $\mathcal{P}4$ is convex by Lemma 1, there is an optimal point, $(\rho_1^*, \rho_2^*, t^*)$ to achieve the minimum cost of this system. According to Karush-Kuhn-Tucker (KKT) conditions, the optimal point of $\mathcal{P}4$ is given by the following equations.

$$\frac{\partial L}{\partial \rho_1} = A_1 \varepsilon e^{\varepsilon \rho_1} - \frac{(\alpha'_1/t + \gamma')}{\rho_1^2} e^{B_1 t} = 0, \quad (32)$$

$$\frac{\partial L}{\partial \rho_2} = A_2 \varepsilon e^{\varepsilon \rho_2} - \frac{A_3}{\rho_2^2} = 0, \tag{33}$$

$$\frac{\partial L}{\partial t} = \frac{B_1 \gamma' t^2 + B_1 \alpha'_1 t - \alpha'_1}{\rho_1 t^2} e^{B_1 t} = 0. \tag{34}$$

Since $\mathcal{P}4$ is convex, we can solve the KKT conditions of $\mathcal{P}4$ to obtain the optimal solution of $\mathcal{P}4$, and transform them into the optimal solution of $\mathcal{P}3$. The KKT conditions of $\mathcal{P}4$ are solved as follows. First, we note that equation (33) is independent to other equations and only related to ρ_2. Then we reformulate (33) as $\varepsilon \rho_2 / 2 \cdot e^{\varepsilon \rho_2 / 2} = \sqrt{\varepsilon A_3 / A_2}/2$, which has the form of $x = W(x)e^{W(x)}$. We can solve this equation and obtain the optimum of ρ_2 as $\rho_2^* = 2/\varepsilon \cdot W(\sqrt{\varepsilon A_3 / A_2}/2)$, where $W(\cdot)$ is Lambert W function. Next, we find that the solution of (34) is equivalent to the solution of $B_1 \gamma' t^2 + B_1 \alpha'_1 t - \alpha'_1 = 0$, and obtain that $t^* = \left[\sqrt{B_1 \alpha_1 (B_1 \alpha_1 + 4\gamma)} - B_1 \alpha_1\right]/(2B_1 \gamma)$. Then, we substituting t^* into (32), and hence we obtain ρ_1^* in (12). Finally, as $\mathcal{P}4$ is equivalent to $\mathcal{P}3$, we replace $1/t^*$ with p_1^* and take into account the constraints of ρ_1, ρ_2, p_1, then the optimal solution of problem $\mathcal{P}3$ is obtained. Since P4 has been proved to be convex, we first solve the KKT conditions of $\mathcal{P}4$ to obtain the optimal solution of $\mathcal{P}4$, and then transfer them into the optimal solution of $\mathcal{P}3$.

References

[1]   V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," *White Paper*, Nov. 2018. Article (CrossRef Link)

[2]   S. Abolfazli, Z. Sanaei, E. Ahmed, A. Gani, and R. Buyya, "Cloud-based augmentation for mobile devices: Motivation, taxonomies, and open challenges," *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 337–368, 1st Quart. 2014. Article (CrossRef Link)

[3]   H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: Architecture, applications, and approaches," *Wireless Communications and Mobile Computing*, vol. 13, no. 18, pp. 1587–1611, Sep. 2013. Article (CrossRef Link)

[4]   A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile internet: Haptic communications over next generation 5G cellular networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 82–89, Apr. 2017. Article (CrossRef Link).

[5]   Y. Xu, G. Gui, H. Gacanin and F. Adachi, "A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 668-695, 2021. Article (CrossRef Link)..

[6]   X. Wang, Z. Ning, and L. Wang, "Offloading in internet of vehicles: A fog-enabled real-time traffic management system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4568–4578, Otc. 2018. Article (CrossRef Link).

[7]   F. Bonomi, R. A. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proc. of IEEE International Conference on Cloud Computing Technology and Science (2012)*, pp. 13–16, Apr. 2012. Article (CrossRef Link).

[8]   M. P. et al, "Mobile-edge computing introductory technical white paper," *White Paper, Mobile-edge Comput. Ind. Initiative*, Sep. 2014. Article (CrossRef Link).

[9]   J. Ahn, J. Lee, S. Yoon, and J. K. Choi, "A novel resolution and power control scheme for energy-efficient mobile augmented reality applications in mobile edge computing," *IEEE Commun. Lett.*, vol. 9, no. 6, pp. 750–754, Jun. 2020. Article (CrossRef Link).

[10] S.-R. Yang, Y.-J. Tseng, C.-C. Huang, and W.-C. Lin, "Multi-access edge computing enhanced video streaming: Proof- of-concept implementation and prediction/QoE models," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1888–1902, Feb. 2019. Article (CrossRef Link).

[11] A. Faro, D. Giordano, and M. Venticinque, "Internetworked wrist sensing devices for pervasive and M-Connected eldercare," in *Proc. of IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, Nara, Japan, pp. 454–456, Mar. 2021. Article (CrossRef Link).

[12] W. Jin, R. Xu, T. You, Y.-G. Hong, and D. Kim, "Secure edge computing management based on independent microservices providers for gateway-centric iot networks," *IEEE Access*, vol. 8, pp. 187975–187990, Oct. 2020. Article (CrossRef Link).

[13] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3590–3605, Dec. 2016. Article (CrossRef Link).

[14] M. Kamoun, W. Labidi, and M. Sarkiss, "Joint resource allocation and offloading strategies in cloud enabled cellular networks," in *Proc. of IEEE International Conference on Communications (ICC)*, pp. 5529–5534, Jun. 2015. Article (CrossRef Link).

[15] W. Labidi, M. Sarkiss, and M. Kamoun, "Joint multi-user resource scheduling and computation offloading in small cell networks," in *Proc. of IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 794–801, Oct. 2015. Article (CrossRef Link).

[16] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile cloud computing," in *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 354–358, Jun. 2014. Article (CrossRef Link).

[17] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, Oct. 2016. Article (CrossRef Link).

[18] S. Cao, X. Tao, Y. Hou, and Q. Cui, "An energy-optimal offloading algorithm of mobile computing based on HetNets," in *Proc. of International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 254–258, Oct. 2015. Article (CrossRef Link).

[19] Y. Zhao, S. Zhou, T. Zhao, and Z. Niu, "Energy-efficient task offloading for multiuser mobile cloud computing," in *Proc. of IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 1–5, Nov. 2015. Article (CrossRef Link).

[20] O. Mun˜oz, A. Pascual-Iserte, and J. Vidal, "Joint allocation of radio and computational resources in wireless application offloading," in *Proc. of Future Network Mobile Summit*, pp. 1–10, Jul. 2013. Article (CrossRef Link).

[21] O. Mun˜oz, A. Pascual Iserte, J. Vidal, and M. Molina, "Energy-latency trade-off for multiuser wireless computation offloading," in *Proc. of IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 29–33, Apr. 2014. Article (CrossRef Link).

[22] C. You, K. Huang, H. Chae, and B. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017. Article (CrossRef Link).

[23] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–6, Dec. 2016. Article (CrossRef Link).

[24] W. Shi et al., "Joint UL/DL Resource Allocation for UAV-Aided Full-Duplex NOMA Communications," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8474-8487, Sept. 2021. Article (CrossRef Link).

[25] W. Sun, H. Zhang, R. Wang, and Y. Zhang, "Reducing offloading latency for digital twin edge networks in 6G," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12240–12251, Oct. 2020. Article (CrossRef Link).

[26] J. Tang, H. Tang, X. Zhang, K. Cumanan, G. Chen, K.-K. Wong, and J. A. Chambers, "Energy minimization in D2D-assisted cache-enabled internet of things: A deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5412–5423, Aug. 2020. Article (CrossRef Link).

[27] G. Gui, M. Liu, F. Tang, N. Kato and F. Adachi, "6G: Opening New Horizons for Integration of Comfort, Security, and Intelligence," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 126-132, Oct. 2020. Article (CrossRef Link).

[28] K. Barr and K. Asanovic, "Energy-aware lossless data compression," *ACM Transactions on Computer Systems*, vol. 24, no. 3, pp. 250–291, Aug. 2006. Article (CrossRef Link).

[29] M. Qin, L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Computing and relaying: Utilizing mobile edge computing for P2P communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1582–1594, Feb. 2020. Article (CrossRef Link).

Xuesong Liang received the Ph.D. degree in communications and information systems from the Southeast University, Nanjing, in 2013. He is currently with School of Communication Engineering Department, Hangzhou Dianzi University, Zhejiang, China. His current research interests include non-orthogonal multiple access (NOMA), mobile edge computing (MEC), and relaying systems for wireless communications.
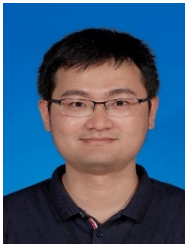
Yongpeng Wu received the B.S. degree in telecommunication engineering from Wuhan University, Wuhan, China, in July 2007, the Ph.D. degree in communication and signal processing with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in November 2013. Dr. Wu is currently a Tenure-Track Associate Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Previously, he was senior research fellow with Institute for Communications Engineering, Technical University of Munich, Germany and the Humboldt research fellow and the senior research fellow with Institute for Digital Communications, University Erlangen- Nurnberg, Germany. During his doctoral studies, he conducted cooperative research at the Department of Electrical Engineering, Missouri University of Science and Technology, USA. His research interests include massive MIMO/MIMO systems, massive machine type communication, physical layer security, and signal processing for wireless communications. Dr. Wu was awarded the IEEE Student Travel Grants for IEEE International Conference on Communications (ICC) 2010, the Alexander von Humboldt Fellowship in 2014, the Travel Grants for IEEE Communication Theory Workshop 2016, the Excellent Doctoral Thesis Awards of China Communications Society 2016，the Exemplary Editor Award of IEEE Communication Letters 2017，Young Elite Scientist Sponsorship Program by CAST 2017, and Excellent Youth Science Fund Project of National Natural Science Foundation of China 2021.

Yujin Huang, enrolled in 2018. She is now majoring in communication engineering at the School of Communication Engineering, Hangzhou Dianzi University.

Derrick Wing Kwan Ng received the bachelor degree with first-class honors and the Master of Philosophy (M.Phil.) degree in electronic engineering from the Hong Kong University of Science and Technology (HKUST) in 2006 and 2008, respectively. He received his Ph.D. degree from the University of British Columbia (UBC) in Nov. 2012. He was a senior postdoctoral fellow at the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany. He is now working as Scientia Associate Professor at the University of New South Wales, Sydney, Australia. His research interests include convex and non-convex optimization, physical layer security, IRS-assisted communication, UAV-assisted communication, wireless information and power transfer, and green (energy-efficient) wireless communications. Dr. Ng has been listed as a Highly Cited Researcher by Clarivate Analytics since 2018. He received the Australian Research Council (ARC) Discovery Early Career Researcher Award 2017, the Best Paper Awards at the WCSP 2020, 2021, IEEE TCGCC Best Journal Paper Award 2018, INISCOM 2018, IEEE International Conference on Communications (ICC) 2018, 2021, IEEE International Conference on Computing, Networking and Communications (ICNC) 2016, IEEE Wireless Communications and Networking Conference (WCNC) 2012, the IEEE Global Telecommunication Conference (Globecom) 2011, 2021 and the IEEE Third International Conference on Communications and Networking in China 2008.

Li Pei received the M. S. degree in Operational Research and Cybernetics from Yanshan University, Hebei, China, in 2012, and the Ph. D degree in Information and Communication Engineering at the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in 2019. She is currently a Lecturer with the College of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China. Her research interests include delay analysis, energy efficient design, queue theory and integrated sensing and communication networks.

Yingbiao Yao received Ph.D. degree in information and communication engineering from the Zhejiang University, Hangzhou, China, in 2006. From 2008 to 2016, he was an Associate Professor with the School of Communication Engineering, Hangzhou Dianzi University, China. In 2011, he was a visiting scholar at RPI University, NY, USA. Since 2017, he has been a Professor with the School of communication engineering, Hangzhou Dianzi University, China.