

# Energy Efficient Secure Computation Offloading in NOMA-based mMTC Networks for IoT

Shujun Han, Xiaodong Xu, *Senior Member, IEEE*, Sisai Fang, Yan Sun, Yue Cao, *Member, IEEE*, Xiaofeng Tao, *Senior Member, IEEE*, and Ping Zhang, *Fellow, IEEE*

**Abstract**—In the era of Internet of Everything, massive connectivity and various demands of latency for Internet of Thing (IoT) devices will be supported by the massive Machine Type Communication (mMTC). Non-Orthogonal Multiple Access (NOMA) and Mobile Edge Computing (MEC) have the advantages of improving network capacity, reducing MTC devices' (MTCs) latency and enhancing Quality of Service. Exploiting these benefits, we focus on the energy efficient secure computation offloading in NOMA based mMTC networks for IoT, where the relay equipped with an MEC server and a passive malicious eavesdropper are presented. We optimize the joint computation and communication resource allocation to maximize the secrecy energy efficiency of computation offloading while guaranteeing the delay requirements of MTCs. Furthermore, we model the subchannels allocation problem as MTC-to-Subchannel matching. Exploiting difference of convex programming and successive convex approximation, we formulate the Dinkelbach-based SEE optimization algorithm and obtain the closed-form expression of power allocation for MTCs' on each subchannel. Based on the communication resources allocation schemes, we propose the Knapsack algorithm to solve the problem of computation resource allocation. Furthermore, we formulate the joint computation and communication resource allocation algorithm for secure computation offloading. Simulation results demonstrate the effectiveness of proposed algorithm for supporting IoT devices energy efficient secure computation offloading.

**Index Terms**—IoT, mMTC, NOMA, Mobile Edge Computing, secure energy efficiency, joint computation and communication resource allocation

## I. INTRODUCTION

**I**N the coming of the Internet of Everything era, hundreds of billions Internet of Things (IoT) devices will be connected to networks. Since the wide-area coverage capability, service

cost and scalability of cellular network, it fits well for large-scale IoT deployments that requires broader interconnection capabilities [1]. Especially, as the main foundation of IoT, massive machine type communication (mMTC) is considered to serve billions of IoT devices in 5G scenarios [2]–[6]. The mMTC refers to certain IoT scenarios, where a large amount of static sensors are deployed and report sporadically to an application server in the cloud [7], such as Vehicle Networks, Smart Meter Reading, E-health monitoring and Smart surveillance. Ericsson predicts that around 3.5 billion cellular IoT devices, i.e. MTC devices (MTCs), will be widely deployed by 2023 [8].

Because of limited resources in computing, communication, storage and energy, it is hard for MTCs to run sophisticated processing [9] and satisfy the quality of computation experience in longtime working [10]. Soaring computing capabilities are required to improve the security, analytic and connectivity of MTCs. However, the linear growth of cloud computing capabilities cannot match the explosive increasing in massive data that are collected from underlying MTCs. Furthermore, the massive MTCs data will stumble the cloud computing due to its shortcoming, such as high latency, security vulnerability, low coverage, lagged data transmission and limited available bandwidth [11]. These noticeable challenges drive the IoT data to be processed close to where it is produced without being placed in a public cloud.

As a natural development in the evolution of mobile base stations and the convergence of information technology and telecommunications networking, mobile edge computing (MEC) was proposed by European Telecommunications Standards Institute (ETSI) [12], which will be the important enabler to inspire the development of myriads of MTC applications and services. It aims at pushing traffic, cloud computing capabilities and network functions towards the edge of cellular networks. The International Data Corporation predicts that 40% initial IoT data need to be analyzed, processed and stored at the edge of the network by 2019, and IoT will generate 95% of the real-time data by 2025 [13]. With characteristics of MEC in low latency, proximity, high bandwidth, location awareness and real-time insight into radio networks information, massive data generated by MTCs can be efficiently executed by MEC server nearby the data sources. This will reduce the congestions in core networks, improve resource efficiency, increase the security of sensitive data and significantly simplify the functionality of MTCs to make MTCs cost efficiency.

However, limited available computation [14], [15] and com-

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This paper is supported by National Key R&D Program of China No. 2017YFB0801702 and the National Natural Science Foundation of China No. 61871045 and 111 Project of China B16006. This work is also supported by BUPT Excellent Ph.D. Students Foundation CX2019308 and the China Scholarship Council. (Corresponding author: Xiaodong Xu)

Shujun Han, Xiaodong Xu, Sisai Fang, Xiaofeng Tao are with the National Engineering Laboratory for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China (email: hanshujun@bupt.edu.cn, xuxiaodong@bupt.edu.cn, fangsisai@bupt.edu.cn, taoxf@bupt.edu.cn).

Yan Sun is with the Department of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (email: yan.sun@qmul.ac.uk).

Yue Cao is with the School of Computing and Communications, Lancaster University, UK. (email: yue.cao@lancaster.ac.uk).

P. Zhang is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: pzhang@bupt.edu.cn).

munication resources [16], [17] will be a potential bottleneck for supporting massive MTCs in MEC. The MTCs are normally with low power and small size, which is traded off by sacrificing their computing performance, while the computation tasks are generally intensive and latency-critical [18]. Although there are ultra-dense edge devices in 5G wireless systems, including small-cell base stations, wireless access points, laptops, tablets and smartphones, each of them are with limited computation resource. It may lead to Quality of Service degradation when many MTCs computation offloading to one edge devices. In this case, the execution time in MEC server is non-negligible [19], because the less computation resource allocated to the MTCs may cause the violation of their delay requirements. Furthermore, the limited available spectrum resources also cannot accommodate massive MTCs to simultaneously connect the cellular networks for computation offloading or communication, due to the traditional cellular network is mainly designed for traditional human to human communication, rather not mMTC.

Security is another critical requirements of supporting MTCs in MEC [20], [21]. Although MEC offers a more secure infrastructure than cloud computing, but it still has its specific security and privacy challenges due to its unique features. Due to the innate heterogeneity of MEC system, the conventional trust and authentication mechanisms are inapplicable [22]. Moreover, the traditional cryptographic techniques and associated protocols are usually not computationally affordable for MTCs (e.g., sensors in smart grid) [23]. Unlike the wired connections in cloud computing, the computation tasks offloaded from MTCs to MEC servers via wireless channel are more vulnerable to potential attacks due to the broadcast natures of wireless communications. In worst case, the MEC server may be a potential eavesdropping or jamming attacker at the physical layer. For example, the malicious attackers with tremendous processing power can decipher the data overheard from MTCs for launching security attacks, especially, the data contains sensitive and private information.

Instead of orthogonal multiple access, non-orthogonal multiple access (NOMA) has been adopted into 5G New Radio by 3rd Generation Partnership Project (3GPP) [24] and investigated by researchers in [25]–[28]. The NOMA not only supports multiple MTCs computation offloading in the same resource concurrently [25], [26], it is also a diversity-assisted security approaches of improving the security for MTCs computation offloading by exploiting the physical characteristics of wireless channels [27], [28]. In order to effectively protect computation offloading against eavesdropper attacks, the finite computation-communication resources should be well allocated in NOMA-based MEC to minimizing the computation offloading rate of the wiretap channel, while maximizing the computation offloading rate of the main channel. However, none of studies have investigated the energy efficient secure computation offloading with consideration of joint computation-communication resource allocation in NOMA-based mMTC networks.

Due to MTCs generally have strictly limited hardware and signal processing capabilities, limited storage capacity and significant energy constraints, they may not have the ability

TABLE I  
NOTATIONS USED IN THIS PAPER

Notions	Meanings
$\mathcal{M}, \mathcal{N}$	set of MTCs and subchannels
$M, N, f_T$	MTCs/subchannels/computing capacity number
$N_n$	maximum number of MTCs in subchannel $n$
$M_n$	the number of MTCs in subchannel $n$
$h_{m,n}^d, h_{m,n}^e, g_n^R, g_n^e$	channel gain
$\widetilde{g}_n^e, \widetilde{h}_{m,n}^e$	estimated CSI of $g_n^e$ and $h_{m,n}^e$
$\Delta g_n^e, \Delta h_{m,n}^e$	estimated error
$I_{m,n,d}, I_{m,n,e}$	interference and noise
$P_{m,n}, q_n$	transmission power in subchannel $n$
$P_{m,\max}, P_{B,\max}$	maximum transmission power
$x_{m,n}$	subchannel allocation index
$t_m^T, t_m^C, t_m^R$	offloading time/execute time/receiving time
$L_m, K_m$	data of computation offloading task/workload
$D_m$	maximum tolerable delay
$\mathbf{P}_M^*, \mathbf{X}_M^*, \mathbf{Y}_M^*$	optimal power/subchannel allocation matrix
$\mathbf{Y}_M^*$	optimal computing resource allocation matrix
$\mathbf{q}_B^*$	optimal power allocation matrix of R-MECS
$\Psi^*$	optimal secrecy energy efficiency
$\lambda, \theta, \varsigma$	Lagrange multiplier, power amplifier coefficient of residual self-interference power
$\rho$	gradient of F of x, partial differential of F
$\nabla_x F, \partial F$	gradient of F of x, partial differential of F
mMTC	massive Machine Type Communication
MTCs	MTC devices
NOMA	Non-Orthogonal Multiple Access
MEC	Mobile Edge Computing
R-MECS	relay equipped with an MEC server
CRB	computation resource block
SEE, SR	secrecy energy efficiency, secrecy rate
DC	difference of convex

to implement traditional cryptographic protocols. Physical-layer security [29] aims at reinforcing the security (i.e., data confidentiality) of communication systems by exploiting differences in channel conditions to prevent decoding of messages from unauthorized users. Based on the secrecy capacity defined by Wyner [30], the data confidentiality can be guarantee if the achieved secrecy capacity is a positive value. Therefore, by comparing the security requirements of MTCs and the potential benefits of physical-layer security, we see that physical-layer security methods can complement or even replace conventional cryptographic protocols as promising solutions for mMTC.

In this paper, we study the physical layer security enhancement for MTCs computation offloading under eavesdropping attacks in the NOMA-based mMTC networks, which supports massive connectivity for MTCs secure computation offloading. Exploiting the various delay requirements of MTCs (cellular IoT devices), we joint allocate the finite computation-communication resources to MTCs for achieving energy efficient secure computation offloading of IoT in NOMA based mMTC networks. Our main contributions are summarized as follows:

- 1) We propose a NOMA-based mMTC networks for supporting IoT, where a relay with an MEC server integrated (R-MECS) assisting the MTCs secure computation offloading and results forwarding. A passive malicious eavesdropper with imperfect channel state information is present. Based on the secrecy capacity, we define the total number of secure computation offloading bits per Joule to measure the secrecy energy efficiency (SEE) of system

computation offloading.

- 2) Taking the various delay requirements of MTCs and finite computation-communication resources into consideration, we jointly optimize the computation resources, MTCs' transmission power and subchannel allocations to maximize the SEE of system computation offloading.
- 3) We formulate the MTC-to-Subchannel matching algorithm and Dinkelbach-based SEE optimization algorithm to solve the communication resource allocation problem, and obtain the closed-form expressions of power allocation for MTCs on each subchannel. Furthermore, we solve the computation resource allocation problem by Knapsack algorithm.

The rest of this paper is organized as follows. In section II, we review the related works. In section III, we introduce the IoT secure computation offloading with R-MECS assisted in NOMA based mMTC networks. In section IV, we formulate the SEE maximization problem for IoT computation offloading. Joint resources allocation for maximizing SEE is formulated in section V. Furthermore, solution of power allocation scheme is given in VI. Simulation results and conclusions are presented in section VII and VIII, respectively. Notations used in this paper are listed in Table I.

## II. RELATED WORKS

The myriads of interconnected MTCs has a significant surge in demand for computation resources. MEC and IoT facilitate each other with mutual advantages [31]. IoT expands MEC services from smart meters to smart manufacture. Similarly, MEC enhances the computational capacity of IoT devices while saving their battery energies. MEC servers performed as a gateway aggregate, classify and process the data collected by IoT devices through computation offloading, rather than directly transmitting them to the core networks [32]. However, because of the synergistic nature of MEC based on both communication and computation, computing offloading is not beneficial for all IoT devices. Its limited resources may lead the computation offloading with a large execution delay due to the waiting execution time at the servers or the communication delay from MTCs to servers. Therefore, it is important to balance the delay and energy consumption in computation offloading with limited resources constrains. The tradeoff between shortening execution time and extending battery life of devices is investigated in [33]. Furthermore, based on Lyapunov optimization, the authors proposed an energy-efficient offloading-decision algorithm to balance the energy-delay tradeoff based on the criteria of minimum response time or energy consumption [34].

Because separately optimizing the allocation of either computation or communication resource is highly suboptimal, joint optimization of radio and compute resources are studied in [35]. Taking infinite and finite computation capacity into consideration, You *et al.* [35] proved that the optimal resource allocation policy have a threshold-based structure with respect to a derived offloading priority function, which yields priorities for users according to their channel gains and local computing energy consumption. Since massive MTCs are

equipped with certain local computation and communication resources, enable user cooperation among these devices in both computation and communication is an efficient solution to improve the MEC performance. The multi-user cooperative MEC system are investigated in [36], where the limited resources MTCs can offload intensive computation tasks to multiple nearby edge devices serving as helpers for remote execution. Exploiting the benefits of NOMA based MEC in supporting more MTCs computation offloading with lower latency and energy consumption, a group of MTCs form a NOMA cluster and simultaneously offload their computation tasks to MEC server over the same subchannel [37].

Achieving security between IoT devices and MEC servers is challenging, and both physical-layer and higher-layer security mechanisms can be applied to protect the offloading process [38]–[43]. Considered the eavesdropping attack when IoT devices offloading computation tasks to MEC servers over radio frequency channels, resource management for secrecy capacity enhancement in OFDMA and NOMA is investigated [38] [39]. Due to the limited attraction of traditional spectrum-expansion-based schemes for the spectrum-scarce IoT devices, the collaborative information sharing of IoT devices for online security-aware computation offloading is studied in [40] to resist the jamming attacks. Furthermore, exploiting the location information related to the attack, the [42] formulate a deep-learning based unsupervised learning model to automate detect the security threats at the edge of cellular networks. To overcome the high computation costs, low flexibility in key management, and low compatibility in deploying new security algorithms in IoT, the near-user edge device is employed as a security agent to simplify key management and offload the computational costs of security algorithms at IoT devices [43].

Ensuring security and privacy in IoT is particularly complicated, especially for the resource-constrained IoT devices. Therefore, integration of security mechanisms should be better included in the MEC architecture, which has not been addressed adequately in above researches. Although joint computation and communication resources allocation [33]–[36] and NOMA-based MEC system have been studied [25], [26], [37], they have rarely considered the security problem in computation offloading procedure, nor investigated the joint resources allocation problem to guarantee the energy efficient secure computation offloading in NOMA-based mMTC networks. Therefore, to fully gain the benefits of MEC and NOMA in mMTC networks, it is essential to study the maximization energy efficiency technology while ensuring the secure computation offloading of IoT devices.

## III. R-MECS ASSISTED UPLINK NOMA-BASED MMTC NETWORK MODEL

As illustrated in Fig.1, we consider the Relay with a R-MECS assisted computing mMTC network, where a set  $\mathcal{M} \triangleq \{1, \dots, M\}$  of MTCs, an eavesdropper and a legal receive user are presented. The MTCs with limited computation and energy resources want to energy efficient secure offload their all computation-intensive and latency critical tasks to a R-MECS for executing, and the R-MECS forwards the

computation results to the legal user. The procedure is named as cooperative computing and forwarding. The eavesdropper is passive and intercepts the information from MTCDs. All MTCDs are uniformly distributed within a cell range and equipped with a single antenna. Moreover, we consider NOMA as an access method. Although multiple MTCDs can share the same subchannel at the same time, they have to suffer from co-channel interference from the shared MTCDs.

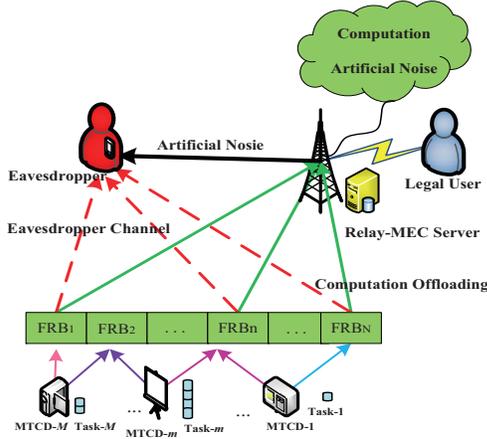


Fig. 1. IoT devices secure computation offloading with R-MECS assisted in NOMA-based mMTC networks

The cooperative computing and forwarding procedure has two stages: secure computation tasks offloading and secure computation results forwarding. In the stage of secure computation tasks offloading, the R-MECS works in full-duplex mode, which has the capability not only to receive the original data from MTCDs through wireless channels and to execute the computation tasks on behalf of the MTCDs, but also to suppress the eavesdropping by transmitting artificial noise to eavesdropper. In the stage of secure computation results forwarding, the R-MECS has enough capability to encrypt the computation results and secure forward them to the destination, which is beyond of the focus in this paper and do not considered.

The computation tasks at  $m$ -th MTCD are characterized by  $(L_m, K_m, D_m)$ .  $L_m$  is the data size of computation offloading task.  $K_m$  is the workload of  $m$ -th MTCD and measured by the number of CPU cycles required to complete the execution of the task, which is application-dependent.  $D_m$  is the tolerable delay of  $m$ -th MTCD, that is the data should be processed within this time. In practical, the available computation capacity of R-MECS is limited with small-scale [14], [15], [35]. Hence, we assume that the computation capacity of R-MECS is divided into  $N_C$  different computation resource blocks (CRBs), and each CRB is equal to  $C$  CPU cycles per second [37]. The computation resources made available by the R-MECS and to be shared among the MTCDs are quantified by the computational capacity  $f_T$ , expressed as numbers of CPU cycles per second.

The available system bandwidth is shared by the MTCDs using NOMA to support more MTCDs computation offloading at the same time. The NOMA channel composes of  $N$

subchannels, denoted by  $\mathcal{N} = \{1, \dots, N\}$ , and each has a bandwidth  $B$ . The noise power at the receiver of the R-MECS is denoted as  $\sigma^2$ . The R-MECS exploits its full-duplex capability to conduct artificial noise to eavesdropper. However, it also negatively influences itself in receiving information from MTCDs, known as self-interference. We denote the self-interference channel gain in  $n$ -th subchannel as  $g_n^R$ , and the channel gain from R-MECS to eavesdropper in  $n$ -th subchannel is  $g_n^e$ . The channel gains from  $m$ -th MTCD to R-MECS (destination) and eavesdropper in subchannel  $n$  are denoted by  $h_{m,n}^d$  and  $h_{m,n}^e$ , respectively. We assume that the R-MECS perfectly knows the channel state information and the computation information of all MTCDs, but only partially knows that of  $g_n^e$  and  $h_{m,n}^e$ .

As commonly adopted in the physical-layer security literatures [38], we consider the deterministic CSI uncertainty model for  $g_n^e$  and  $h_{m,n}^e$ , where  $\widetilde{g}_n^e = g_n^e + \Delta g_n^e$  and  $\widetilde{h}_{m,n}^e = h_{m,n}^e + \Delta h_{m,n}^e$ .  $\widetilde{g}_n^e$  and  $\widetilde{h}_{m,n}^e$  denote the estimated CSI of  $g_n^e$  and  $h_{m,n}^e$  at the R-MECS, in addition,  $\Delta g_n^e$  and  $\Delta h_{m,n}^e$  denote the estimated error that is bounded by a possible value  $\varepsilon \geq 0$  as  $|\Delta g_n^e| \leq \varepsilon$ ,  $|\Delta h_{m,n}^e| \leq \varepsilon$ .  $g_n^e$  and  $h_{m,n}^e$  are the best possible channel gain of the eavesdropper known by the R-MECS, and  $\varepsilon$  is also known by R-MECS. Moreover, both distance dependent path loss and Rayleigh fading on each of the subchannel are considered, such that  $h_{m,n}^d = Dd_{m,d}^{-\alpha/2} h_{m,n}^d$  and  $\widetilde{h}_{m,n}^e = Dd_{m,e}^{-\alpha/2} \widetilde{h}_{m,n}^e$ , where  $Dd_{m,d}$  and  $Dd_{m,e}$  are the distances from MTCD  $m$  to R-MECS (destination) and eavesdropper.  $\alpha$  denotes the path-loss exponent and  $h_{m,n}^d, \widetilde{h}_{m,n}^e \sim \mathcal{CN}(0, 1)$  denote the normalized Rayleigh fading. The block fading model is adopted, where all the subchannels are invariable during a complete transmission cycle and the co-channel interference among MTCDs on each subchannel is considered.

#### IV. SECURE COMPUTATION OFFLOADING OVER NOMA-BASED mMTC NETWORKS

##### A. Secure Computation Offloading Energy Efficiency

In R-MECS assisted computing mMTC system, each MTCD offload their computation tasks to R-MECS by using different subchannels, and each subchannel can be shared by more than one MTCD simultaneously with actively introducing interferences. To reduce the complexity of receiver, an upper limitation of the number of sharing same subchannel is given  $\sum_{m=1}^M x_{m,n} \leq N_n, m \in \mathcal{M}, n \in \mathcal{N}$  where  $x_{m,n} = 1$  denote the  $n$ -th subchannel is assigned to the  $m$ -th MTCD, otherwise  $x_{m,n} = 0$ .

When R-MECS works in full-duplex mode, it will bring self-interference to its receiver and cannot be canceled completely due to the hardware limitation. We describe the residual self-interference with the self-interference channel gain  $g_n^R$  as a linear coefficient  $\rho$  [44]. The residual self-interference in  $n$ -th subchannel is denoted as  $\rho q_n |g_n^R|^2$ , and  $q_n$  is the transmission power from R-MECS to eavesdropper in subchannel  $n$ . Specially, for the self-interference, we assume  $E \left\{ |g_n^R|^2 \right\} = 1, n \in \mathcal{N}$ .

Without loss of generality, we assume channel gains of MTCDs on subchannel are perfectly known with descending

order in destination channel and eavesdropping channel. The achievable secrecy rate and lavage information rate in eavesdropper of the  $m$ -th MTCD in the  $n$ -th subchannel are given by,

$$\begin{aligned} r_{m,n,d} &= x_{m,n} B \log_2 (1 + \gamma_{m,n,d}) \\ r_{m,n,e} &= x_{m,n} B \log_2 (1 + \gamma_{m,n,e}) \end{aligned} \quad (1)$$

$\gamma_{m,n,d}$  and  $\gamma_{m,n,e}$  are the signal-to-noise-plus-interference ratio (SINR), particularly, they are evaluated by

$$\begin{aligned} \gamma_{m,n,d} &= \frac{p_{m,n} |h_{m,n}^d|^2}{\sigma^2 + \sum_{i=m+1}^M p_{i,n} |h_{i,n}^d|^2 + \rho q_n |g_n^R|^2} \\ \gamma_{m,n,e} &= \frac{p_{m,n} |h_{m,n}^e|^2}{\sigma^2 + \sum_{i'=m+1}^M p_{i',n} |h_{i',n}^e|^2 + q_n |g_n^e|^2} \end{aligned} \quad (2)$$

The computation offloading rate in the  $n$ -th subchannel of MTCD-to-R-MECS and that of the  $n$ -th subchannel in MTCD-to-eavesdropper are given by

$$r_{n,d} = \sum_{m=1}^{M_n} r_{m,n,d}, r_{n,e} = \sum_{m=1}^{M_n} r_{m,n,e} \quad (3)$$

Furthermore, the maximum achievable computation offloading secrecy rate (SR) of each MTCD is given by the difference of the MTCD-to-R-MECS computation offloading rate and the MTCD-to-eavesdropper computation offloading rate, which can be expressed as

$$U_{SC} = \sum_{n=1}^N [r_{n,d} - r_{n,e}]^+ \quad (4)$$

Due to the low-cost and power-limited features of MTCDs, a major concern is to secure computation offloading with limited power and energy as much as possible. The performance metric of secrecy energy efficiency (SEE) for computation offloading needs to be maximized, which is defined as the total number of secure computation offloading bits per Joule from the MTCDs to R-MECS. The SEE of all uplink system is defined as

$$SEE = \frac{U_{SC}}{U_{TP}} \quad (5)$$

which is the ratio of the total SR to the total energy consumption.  $U_{TP}$  is the total power consumption of all MTCDs, and can be expressed as

$$U_{TP} = \sum_{n=1}^N \left( \sum_{m=1}^{M_n} \frac{1}{\varsigma} x_{m,n} p_{m,n} + P_C \right) \quad (6)$$

The transmission power consumption of  $m$ -th MTCD is  $\sum_{n=1}^N x_{m,n} p_{m,n}$ , which is not the only component to calculate SEE for computation offloading, static circuit power consumption  $P_C$  also plays a significant role of SEE and includes circuit power and processing power.  $\varsigma \in [0, 1]$  is a constant value which accounts for the power amplifier efficiency.

### B. MEC Server Scheduling

The maximum tolerable delay  $D_m$  of  $m$ -th MTCD incorporates the time of offloading tasks to MEC server for executing, which are the time for computation tasks offloading from  $m$ -th MTCD to MEC server  $t_m^T$ , the time necessary for MEC server executing  $t_m^C$  and the time for legal user receiving

computed results  $t_m^R$ . In other word,  $t_m^T + t_m^C + t_m^R \leq D_m$ , where  $t_m^T = \frac{L_m}{R_{SC}^m}$ ,  $t_m^C = \frac{L_m K_m}{y_m}$ .  $y_m$  represents computing resources assigned to  $m$ -th MTCD, and they are subject to the computational budget constraint  $\sum_{m=1}^M y_m \leq f_T$ .

In general, R-MECS has enough energy to transmit computation results with small size to legal users, we assume that the time for legal users to receive the computed results is negligible [26], i.e.,  $t_m^R = 0$ . Thus, the total secure computation offloading time and execution time of  $m$ -th MTCD's task should be no larger than its maximum tolerable delay  $D_m$

$$t_m^T + t_m^C \leq D_m \quad (7)$$

The maximum tolerable delay constraint in (7) is what couples computation and communication optimization variables, which motivates a joint optimization of computation resource and communication resource. The total execution time in (7) is equivalent to the following minimum computation offloading SR requirement  $U_{SC}^m \geq \frac{L_m}{D_m - t_m^C}$  where  $t_m^C < D_m$ . Due to the constraints of the minimum computation offloading SR, the overall network performance in terms of SEE may be degraded.

### C. Optimization Problem Formulation

The power allocation scheme of MTCDs  $\mathbf{P}_M^*$ , subchannel allocation scheme  $\mathbf{X}_M^*$ , computation resource allocation  $\mathbf{Y}_M^*$  and the transmission power allocation of R-MECS  $\mathbf{q}_B^*$  can be obtained by solving

$$\begin{aligned} \mathbf{P1} \quad & \max_{x_{m,n}, y_m, p_{m,n}, q_n} \text{SEE} \\ \text{s.t.} \quad & C1 : R_{SC}^m \geq \frac{L_m}{D_m - t_m^C}, m \in \mathcal{M} \\ & C2 : t_m^C < D_m, m \in \mathcal{M} \\ & C3 : \sum_{m=1}^M x_{m,n} \leq N_n, m \in \mathcal{M}, n \in \mathcal{N} \\ & C4 : x_{m,n} (x_{m,n} - 1) = 0, m \in \mathcal{M}, n \in \mathcal{N} \\ & C5 : \sum_{m=1}^M y_m \leq f_T, m \in \mathcal{M} \\ & C6 : \sum_{n=1}^N q_n \leq P_{B,\max}, n \in \mathcal{N} \\ & C7 : \sum_{n=1}^N x_{m,n} p_{m,n} \leq P_{m,\max}, m \in \mathcal{M} \\ & C8 : q_n \geq 0, n \in \mathcal{N} \\ & C9 : p_{m,n} \geq 0, m \in \mathcal{M}, n \in \mathcal{N} \end{aligned} \quad (8)$$

where C1 represents the minimum computation offloading SR requirements for each MTCD to meet the maximum tolerable delay constraint; C2 imposes a limitation on the task execution time  $t_m^C$  with the delay limitation  $D_m$ ; C3 ensures the total number of MTCDs sharing the same subchannel is not exceed  $N_n$  and C4 restricts  $x_{m,n}$  is a binary variable; C5 is the limitation of the computation resource in MEC server available to the MTCDs, C6 and C7 are the power budget constraints of R-MECS and MTCDs, respectively; Moreover, C8 and C9 guarantee the transmission power of R-MECS and MTCDs to be positive values.

In general, depending on the MTCDs' requests, channel state and computation capability, not all requests can be

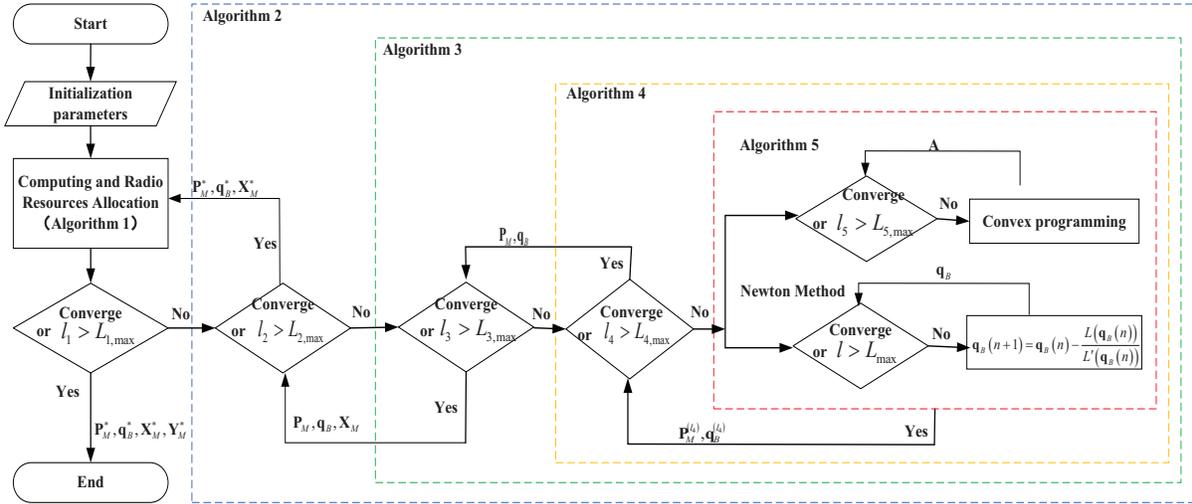


Fig. 2. The Relationship between the Following Proposed Algorithm

accommodated. The C1-C9 can be enforced by a proper admission control scheme through solving **P1**. However, the introduced co-channel interference in NOMA system leads a coupling among all MTCDs, which makes the objective function in **P1** is a ratio between a nonconvex function and an affine function. In addition, the subchannel indicator is binary variable. Thus, **P1** is classified into a non-convex mixed integer nonlinear fractional programming problem, which is non-polynomial problem. The optimal SEE depends now on both computation and communication parameters, such as maximum tolerable delay, input size of MTCDs, CPU cycle budget of R-MECS, etc. Different from the classical secure transmission in **P1**, each MTCD's secure computation offloading rate has to meet the latency constraint in (7), which means that a lowest secrecy computation offloading rate is no longer optimal. Therefore, we develop computationally efficient algorithms for SEE maximization problem suitable to uplink NOMA-based mMTC systems. To clearly express the relationship among the following proposed algorithms, we make a high-level illustrative representation as shown in Fig.2.

## V. JOINT RESOURCES ALLOCATION FOR MAXIMIZING SEE

In this section, we will give a overview of the SEE maximization scheme, which is obtained upon a two stages resource allocation algorithm. In the first stage, SEE with communication resource allocation is solved by an iterative algorithm of joint subchannel and power allocation, where subchannel allocation is solved by Match Game and power allocation algorithm is obtained from fractional programming and DC programming. The iterative radio resources allocation for maximizing SEE is given in **Algorithm 2**, which is to solve **P2**. In the second stage, to take full use of computation resource and maximize SEE, the optimization problem of computation resource allocation shown in **P3** is modeled as a dynamic programming and solved by Knapsack algorithm, which is illustrated in **Algorithm 1**.

### A. Knapsack-based Algorithm to Maximize SEE

Because MTCDs are energy-critical devices in mMTC networks, it is meaningful to discuss the relationship between SEE and secure computation offloading time, and in order to exploit the various delay requirements of MTCDs to prolong the lifetime of MTCDs.

**Theorem 1.** *The SEE of  $m$ -th MTCD is monotonically increasing and convex in its secure computation offloading time  $t_m^T$ .*

*Proof:* It can be seen in Appendix A. ■

Therefore, making full use of the various delay requirements of MTCDs can effectively improve the SEE of uplink computation offloading mMTC networks. Since the offloading tasks must be offloaded to R-MECS within given amount of time and little execution time left cannot complete the computation tasks in R-MECS, the secure computation offloading time of any computation task cannot be arbitrarily long. However, constrained by limited computing resources in R-MECS, each MTCD wants to be assigned as more computing resources as possible, which will reduce the execution time and increase secure computation offloading time under delay requirement. It is essential to reasonable allocate the computing resources to maximize SEE.

We take the confidential of computation tasks into consideration in our proposed Knapsack problem. Moreover, the joint computation and communication resource allocation for maximizing SEE problem is reformulated as a Knapsack problem, which optimizes the communication resource allocation scheme with computation resource constraints taken into consideration,

$$\begin{aligned}
 & \mathbf{P2} \quad \max_{x_{m,n}, p_{m,n}, q_n} \quad \text{SEE} \\
 & \text{s.t.} \quad \text{C3-C4, C6-C9} \\
 & \quad \widetilde{\text{C1}} : SR_m \geq \frac{L_m}{D_m} = R_{m,\min}
 \end{aligned} \tag{9}$$

and

$$\begin{aligned} \mathbf{P3} \quad & \max_{y_m} \quad \mathbf{Y}_M \\ \text{s.t.} \quad & y_m \geq \frac{L_m K_m}{D_m - \frac{L_m}{SR_m}}, \forall m \in \mathcal{M} \quad (10) \\ & \text{C2, C5} \end{aligned}$$

where  $SR_m = \sum_{n=1}^N [r_{m,n,d} - r_{m,n,e}]^+$ ,  $\eta(\mathbf{Y}_M) = \text{SEE}$ .

In the proposed Knapsack problem, the R-MECS is seen as a knapsack and the computation capacity in C5 is the volume of knapsack. The computation resource  $y_m$  required by the  $m$ -th MTCD represents the volume goods to be packed into the knapsack and the computation offloading SEE achieved by the  $m$ -th MTCD is the value of corresponding goods  $y_m$ . Different from the Knapsack algorithm in existing researches to solve the problem of computation resource allocation, the goods have its value only when the required MTCD can guarantee the confidential of computation tasks, i.e., the computation offloading SEE achieved by the MTCD has a positive value.

---

**Algorithm 1:** Joint Resource Allocation for Maximizing SEE

---

**Input:**  $\mathbf{P}_M, \mathbf{X}_M, \mathbf{q}_B, \{D_m, L_m\}, N_n, N_c$

**Output:**  $SEE^*$

- 1: **While**  $|SEE(l_1) - SEE(l_1 - 1)| > \varepsilon_1$  or  $l_2 \leq L_{1,\max}$
  - 2: Calculate SEE based on the iterative communication resource allocation for maximizing SEE algorithm (in **Algorithm 2**).
  - 3: Based on  $\bar{C1}$ , calculate the required computing resources  $y_m$
  - 4: Perform computing resources allocation based on the **Knapsack algorithm**.
  - 5: **End While**
- 

The **P2** in (9) is a non-convex mixed integer non-linear programming (MINLP) problem due the product term  $x_{m,n} p_{m,n}$  in objective and constraints of **P2**, where  $x_{m,n}$  is a binary variable. The MINLP problem is known to belong to NP-hard problems and no efficient solutions exists because the complexity may increase exponentially with problem size. This type of problems is among the most challenging problems, and obtaining the globally optimal solution is intractable for mMTC networks. In this section, instead of directly solving the MINLP problem, we transform the optimization problem into MTCD-to-subchannel matching subproblem and power allocation subproblem.

---

**Algorithm 2:** Iterative Communication Resource Allocation for Maximizing SEE

---

**Input:** maximum tolerance  $\varepsilon_2$  and maximum iterations  $L_{2,\max}$

**Output:** optimal secure energy efficient  $SEE^*$

1. Initialization : the power  $p_{m,n}$  for each MTCD and  $q_n$  for R-MECS
  - 2: **While**  $|SEE(l_2) - SEE(l_2 - 1)| > \varepsilon_2$  or  $l_2 \leq L_{2,\max}$
  - 3: Given the power allocation, update  $x_{m,n}$  based on **Algorithm 3**
  - 4: Reallocate  $p_{m,n}$  and  $q_n$  according to **Algorithm 4**.
  - 5: Set  $l_2 = l_2 + 1$  and calculate  $SEE(l_2 + 1)$ .
  - 6: **End While**
- 

By iteratively solving the two sub-problems, the system SEE slightly improves at each iteration and will converge at last, the detail procedure is presented in **Algorithm 2**. It finds at least a locally optimal solution with polynomial time, which has the potential to be a global optimum.

### B. MTCD-to-Subchannel Matching Problem

The optimal solution  $\mathbf{X}_M$  of **P2** can only be obtained by performing an exhaustive search over all subchannel  $\mathbf{X}_M =$

$[x_{m,n}]$  that respect the C3 and C4 in **P2**. There are at least  $M^N$  different subchannel allocation matrices to be evaluated, which greatly increased the processing time in MTCN and the computing complexities in MEC server grow fast with the number of MTCNs. In the subchannel allocation process, the MTCNs prefers to access the subchannel with good quality to achieve the best service, while the subchannel allocation aims at maximizing the system SEE by arranging right MTCNs using the subchannels. Thus, in this section, we model the subchannel allocation problem as a matching process where MTCNs and subchannels match with each other to maximize the system SEE.

---

**Algorithm 3:** MTCD-to-Subchannel Matching Algorithm

---

**Input:**  $p_{m,n}, y_m, q_n, N_n$

**Output:** optimal subchannel allocation matrix  $\mathbf{X}_M^*$

**1: Initialization Phase** (Subchannel Selection)

Initialize the set of unmatched subchannels  $\Omega^u$

For each MTCD

Calculate  $SEE_{m,n}$  of each unmatched subchannel

Construct the preferable list  $\mathbf{PMD}(m)$  based on  $SEE_{m,n}$

End

**2: Matching Phase** (Subchannel Allocation)

For each MTCD  $m$

If the proposal set of  $m$ -th MTCD  $\mathbf{PMD}(m) \neq \emptyset$

$m$ -th MTCD propose itself to its preferable unmatched subchannel according to  $\mathbf{PMD}(m)$

For each unmatched  $n$ -th subchannel

If  $n$ -th subchannel has received proposal from MTCD  $m$ -th

Accept the proposal from  $m$ -th MTCD

End If

For each MTCD favors to  $n$ -th subchannel

Construct the preferable list  $\mathbf{PSC}(n)$  based on SEE of the accepted MTCNs

Assign subchannel  $n$ -th to its favorite MTCD with maximum SEE and set  $x_{i,n} = 1$

End For

**3: Updating Phase** (Unmatched Subchannels Updating)

For each unmatched  $n$ -th subchannel

If  $\sum_{m=1}^M x_{m,n} = N_n$

$\Omega^u = \Omega^u - \{n\}$

End If

End For

---

Let  $(\text{MTCD}_m, \text{SC}_n)$  denote a matching pair if subchannel  $n$  is assigned to  $m$ -th MTCD. The unmatched MTCD set is  $\Omega^u$ . The preferable matrix  $\mathbf{PMD}$  of MTCNs record their preferable subchannel sequence ranked by affordable SEE in each subchannel, i.e.  $SEE_{m,n}$ . The preferable matrix  $\mathbf{PMD}$  is constructed at the beginning of the matching, and will not be changed throughout the matching process. It guarantees the matching algorithm is low complexity and operability, and is same as the conventional Gale-Shapley algorithm. However, the preferable matrix  $\mathbf{PSC}$  based on the achieved secure energy efficiency  $SEE_n$  on itself are unreducible in static matching iterations, because there is interdependence relationship between the MTCNs who share the same subchannel in matching algorithm. To further reduce the complexity, the preferable matrix  $\mathbf{PSC}$  of subchannels is not constructed in advance.

In this matching model, each MTCD makes their decisions first according to their own preferable matrix  $\mathbf{PMD}(m)$ . In each round of proposals from source nodes, each MTCD proposes itself to at most one subchannel and then wait for the response from the unmatched subchannels. The unmatched subchannels accept all proposals from MTCNs and construct

their preferable matrix  $\mathbf{PSC}$  based on the SEE of each MTCD. The subchannel will assign its favorite MTCD with maximum SEE to sharing itself, and the unassigned MTCD on this subchannel will reselect the new subchannel. The detailed entire process of MTCD-Subchannel matching algorithm are described in **Algorithm 3**, which consist of initialization phase, matching phase and updating phase.

### C. Power Allocation Algorithm for Maximizing SEE

Without loss of generality, we define the maximum SEE, of  $\mathbf{P2}$  in (13) as  $\Psi^*$ . Based on the fractional programming, the primary problem is associated with a parametric program problem stated as follows,

$$\Psi^* = \frac{U_{SC}(\mathbf{P}_M^*, \mathbf{q}_B^*)}{U_{TP}(\mathbf{P}_M^*, \mathbf{q}_B^*)} = \max_{\{p_{m,n}, q_n\}} \frac{U_{SC}(\mathbf{P}_M, \mathbf{q}_B)}{U_{TP}(\mathbf{P}_M, \mathbf{q}_B)} \quad (11)$$

**Theorem 2.** *The maximum SEE can be achieved only when  $\Psi_m^*$  and the optimal resource allocation policies  $\{\mathbf{P}_m^*, \mathbf{q}_B^*\}$  satisfy,*

$$\begin{aligned} \max_{\{p_{m,n}, q_n\}} U_{SC}(\mathbf{P}_M, \mathbf{q}_B) - \Psi^* U_{TP}(\mathbf{P}_M, \mathbf{q}_B) \\ = U_{SC}(\mathbf{P}_M^*, \mathbf{q}_B^*) - \Psi^* U_{TP}(\mathbf{P}_M^*, \mathbf{q}_B^*) \\ = 0 \end{aligned} \quad (12)$$

*Proof:* It can be seen in Appendix B. ■

To solving (12) with an equivalent objective function, we propose a Dinkelbach-based SEE optimization algorithm, which is shown in **Algorithm 4**. The iterative approach presented in the Dinkelbach-based SEE optimization algorithm provides closer approximations to the optimal SEE at each iteration, stopping when a fixed convergence error  $\varepsilon_4$  or the maximum of iterations  $L_{4,SEE}$  is achieved. The objective function at the process of iteration of the Dinkelbach algorithm is written as

$$\begin{aligned} \mathbf{P4} \quad \max_{\{p_{m,n}, q_n\}} U_{SC}(\mathbf{P}_M, \mathbf{q}_B) - \Psi^{(l)} U_{TP}(\mathbf{P}_M, \mathbf{q}_B) \\ \text{s.t.} \quad \mathbf{C6-C9} \\ \widetilde{\mathbf{C1}} : SR_m \geq R_{m,\min}, m \in \mathcal{M} \end{aligned} \quad (13)$$

where  $\Psi^{(l)}$  is the temporary parameter used in the  $l^{th}$  iteration of Dinkelbach algorithm.

---

#### Algorithm 4: Dinkelbach-based SEE Optimization Algorithm

---

**Input:**  $\varepsilon_4$  and  $L_{4,\max}$  and  $l_4 = 0$

**Output:**  $\Psi^*$  and  $\mathbf{P}_m^*(\Psi^*)$ ,  $\mathbf{q}_B^*(\Psi^*)$

1. Initialization :  $L_{4,SEE}$ ,  $\varepsilon_4$ ,  $\Psi^{(0)}=0$ ,

2: Repeat

3: Obtain  $\{\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)}\}$  by **Algorithm 5**

4: If  $U_{SC}(\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)}) - \Psi^{(l_4-1)} U_{TP}(\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)}) < \varepsilon_4$

5: Convergence=true

6: Return  $\{\mathbf{P}_M^*, \mathbf{q}_B^*\} = \{\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)}\}$  and  $\Psi^* = \frac{U_{SC}(\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)})}{U_{TP}(\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)})}$ ;

7: Else

8: Convergence=false

9: Set  $\Psi^{(l_4+1)} = \frac{U_{SC}(\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)})}{U_{TP}(\mathbf{P}_M^{(l_4)}, \mathbf{q}_B^{(l_4)})}$  and  $l_4 = l_4 + 1$ ;

10: End if

11: Until Convergence=true or  $l_4 = L_{4,\max}$

---

## VI. SOLUTIONS TO POWER ALLOCATION SCHEME

Although given the MTCD-to-Subchannel Matching scheme,  $U_{SC}(\mathbf{P}_M, \mathbf{q}_B)$  in optimization problem (18), the sum of multiple rate function, is still non-convex optimization problem with respect to  $p_{m,n}$ . In order to reduce the complexity of solving the power allocation problem, SCA and DC programming are exploited to solve this problem. We first establish a concave lower bound of user rate, which is parameterized by a given power allocation  $\widetilde{\mathbf{P}}_M$  [45],

$$\log_2(1+z) \geq \alpha \log_2(z) + \beta \quad (14)$$

where  $\alpha = \frac{\hat{z}}{1+\hat{z}}$ ,  $\beta = \log_2(1+\hat{z}) - \frac{\hat{z}}{1+\hat{z}} \log_2(\hat{z})$ , and the equality holds when  $z = \hat{z}$ . By exploiting the rate lower bound, the lower bound of data rate of the  $m$ -th MTCD-to-R-MECS and that of the  $m$ -th MTCD-to-eavesdropper on the  $n$ -th subchannel can be written as

$$\begin{aligned} \tilde{r}_{m,n,d} &= B\alpha_{m,n,d} \log_2(\tilde{\gamma}_{m,n,d}(A_{m,n}, \mathbf{q}_B)) + \beta_{m,n,d} \\ \tilde{r}_{m,n,e} &= B\alpha_{m,n,e} \log_2(\tilde{\gamma}_{m,n,e}(A_{m,n}, \mathbf{q}_B)) + \beta_{m,n,e} \end{aligned} \quad (15)$$

where  $\mathbf{A} = \mathbf{P}_M^*$  and  $\tilde{r}_{m,n,d}(A_{m,n}, \mathbf{q}_B)$  is a concave function of  $A_{m,n}$ , which is proved in [46]. We rewrite the **P4** as

$$\begin{aligned} \mathbf{P5} \quad \max_{\{p_{m,n}, q_n\}} \widetilde{U}_{SC}(\mathbf{A}, \mathbf{q}_B) - \Psi U_{TP}(\mathbf{A}, \mathbf{q}_B) \\ \text{s.t.} \quad \widetilde{\mathbf{C6}} : \sum_{n=1}^{N_f} q_n \leq P_{B,\max}, n \in \mathcal{N}_f \\ \widetilde{\mathbf{C7}} : \sum_{n=1}^{N_f} A_{m,n} \leq P_{m,\max}, m \in \mathcal{M} \\ \widetilde{\mathbf{C8}} : q_n \geq 0, n \in \mathcal{N}_f \\ \widetilde{\mathbf{C9}} : A_{m,n} \geq 0, m \in \mathcal{M}, n \in \mathcal{N}_f \\ \mathbf{Col} : \widetilde{SR}_m \geq R_{m,\min}, m \in \mathcal{M} \end{aligned} \quad (16)$$

where the  $\widetilde{U}_{SC}(\mathbf{A}, \mathbf{q}_B) = \sum_{m=1}^M \sum_{n=1}^{N_f} [\tilde{r}_{m,n,d} - \tilde{r}_{m,n,e}]^+$  and  $\widetilde{SR}_m(\mathbf{A}, \mathbf{q}_B) = \sum_{n=1}^N [\tilde{r}_{m,n,d} - \tilde{r}_{m,n,e}]^+$ .

### A. Dual Decomposition with the Fixed SEE $\Psi$

Due to the nonconvex objective function and nonconvex feasible set of **P5**, it is difficult to obtain the optimal solution. We intend to solve **P5** by exploiting dual decomposition. In this way, the nonconvex constraint **Col** can be incorporated into the objective function and the feasible domain is turn into a convex set. Therefore, the DC programming can be exploited in the next subsection.

If the constraint **Col** is eliminated, the feasible domain of **P4** will be convex and the problem will be decoupled. Thus, we construct the lagrangian function of **P5** as

$$\begin{aligned} L(\theta, \mathbf{A}, \mathbf{q}_B) &= \widetilde{U}_{SC}(\mathbf{A}, \mathbf{q}_B) - \Psi U_{TP}(\mathbf{A}, \mathbf{q}_B) \\ &+ \sum_{m=1}^M \theta_m (\widetilde{SR}_m - R_{m,\min}) \end{aligned} \quad (17)$$

The dual problem of **P5** is written as

$$\min_{\theta \geq 0} \max_{\mathbf{A}, \mathbf{q}_B \in \mathcal{S}_1} L(\theta, \mathbf{A}, \mathbf{q}_B) \quad (18)$$

where  $\lambda \geq 0$  is the lagrange multiplier. Then, the feasible domain of  $\mathbf{A}, \mathbf{q}_B$  is turn into the convex set  $\mathcal{S}_1 = \{\mathbf{A} : \widetilde{\mathbf{C6}}, \widetilde{\mathbf{C8}}, \mathbf{q}_B : \widetilde{\mathbf{C7}}, \widetilde{\mathbf{C9}}\}$ .

The sub-gradient method can be applied to solve the higher level subproblem  $\min_{\theta \geq 0} L(\theta, \mathbf{A}, \mathbf{q}_B)$  of (24) with fixed  $\mathbf{A}, \mathbf{q}_B$ . The gradient of  $L(\theta, \mathbf{A}, \mathbf{q}_B)$  with respect to  $\theta_m$  is  $\sum_{n=1}^{N_f} (\tilde{r}_{m,n,d} - \tilde{r}_{m,n,e}) - R_{m,\min}$ , thus the dual variables  $\theta_m$  can be updated by

$$\theta_m(s+1) = \theta_m(s) + \phi_s \left[ \sum_{n=1}^{N_f} (\tilde{r}_{m,n,d} - \tilde{r}_{m,n,e}) - R_{m,\min} \right] \quad (19)$$

where  $\phi_s$  is the positive step size at the  $s$ -th iteration.

The lower level subproblem  $\max_{\mathbf{A}, \mathbf{q}_B \in \mathcal{S}_1} L(\theta, \mathbf{A}, \mathbf{q}_B)$  can be solved when  $\theta$  is fixed. However, due to the nonconvexity of  $L(\theta, \mathbf{A}, \mathbf{q}_B)$  of both  $\mathbf{A}$  and  $\mathbf{q}_B$  simultaneously, the lower level subproblem  $\max_{\mathbf{A}, \mathbf{q}_B \in \mathcal{S}_1} L(\theta, \mathbf{A}, \mathbf{q}_B)$  is still difficult to solve. Therefore, with fixed  $\mathbf{A}$  or  $\mathbf{q}_B$ , the alternate search strategy can be adopted to solve the lower problem. In alternate search, there is only one of  $\mathbf{A}, \mathbf{q}_B$  optimized in every step while others keeps fixed. In other words, we can find the solution by iteratively solving following problems

$$\mathbf{P6} \quad \min_{q_n} L(\theta, \mathbf{A}, \mathbf{q}_B) \quad (20a)$$

$$s.t. \quad \widetilde{C6}, \widetilde{C8}$$

$$\min_{A_{m,n}} L(\theta, \mathbf{A}, \mathbf{q}_B) \quad (20b)$$

$$s.t. \quad \widetilde{C7}, \widetilde{C9}$$

With fixed  $\theta$ ,  $L(\theta, \mathbf{A}, \mathbf{q}_B)$  is a concave function of  $\mathbf{q}_B$ , the proof of which is shown in **Appendix C**, but it is still not convex of  $\mathbf{A}$ . Thus, the optimal value  $\mathbf{q}_B$  in (20a) can be found by Newton method, while the optimal value  $\mathbf{A}$  in (20b) can be obtained uniquely by DC programming in the next subsection.

### B. DC Programming for Power Allocation

Although  $L(\theta, \mathbf{A}, \mathbf{q}_B)$  is not convex with respect to  $\mathbf{A}$ , the key idea of DC programming is to convert a non-convex problem to convex sub-problems by SCA and obtain a locally optimal solution. By exploiting the DC structures, we convert  $L(\theta, \mathbf{A}, \mathbf{q}_B)$  as

$$L(\theta, \mathbf{A}, \mathbf{q}_B) = F_1(\mathbf{A}) - F_2(\mathbf{A}) \quad (21)$$

where the function  $F_1(\mathbf{A})$  and  $F_2(\mathbf{A})$  can be expressed as

$$F_1(\mathbf{A}) = \sum_{n=1}^N \sum_{m=1}^{M_n} (\theta_{m,n} + 1) \tilde{r}_{m,n,d} - \Psi \sum_{n=1}^N \left( \sum_{m=1}^{M_n} \frac{1}{\zeta} p_{m,n} + p_{m,c} \right) \quad (22)$$

$$F_2(\mathbf{A}) = \sum_{n=1}^N \sum_{m=1}^{M_n} (\theta_{m,n} + 1) \tilde{r}_{m,n,e}$$

The objective  $F_1(\mathbf{A})$  and  $F_2(\mathbf{A})$  are continuous concave function [46] on compact and convex set  $\widetilde{C7}$  and  $\widetilde{C9}$ , it belongs to DC function programming. Thus, (27b) can be solved uniquely by DC programming. Since  $F_2(\mathbf{A})$  is DC function for any feasible point  $\mathbf{A}^{(k)}$ , its convex majorant can be expressed as  $F_2(\mathbf{A}) \approx F_2(\mathbf{A}^{(k)}) + \nabla_{\mathbf{A}} F_2^T(\mathbf{A}^{(k)}) (\mathbf{A} - \mathbf{A}^{(k)})$ , which is affine functions representing the global underestimation of  $F_2(\mathbf{A})$ .

Then, we can solve the **P6** by

$$\mathbf{P7} \quad \min_{A_{m,n}} L(\mathbf{A}^{(k)}) \quad (23)$$

$$s.t. \quad \widetilde{C7}, \widetilde{C9}$$

The solution of DC programming problem in (23) is concluded in Algorithm 5. For differentiable  $F_1(\mathbf{A})$  and  $F_2(\mathbf{A})$ , Algorithm 5 returns a stationary point of  $L(\mathbf{A})$ , which is proved in **Appendix D**.

---

#### Algorithm 5: Successive Convex Approximation for DC problems

---

**Input:**  $F_1(\mathbf{A}), F_2(\mathbf{A}^{(k)}), \varepsilon_5, L_5, \lambda_{\max}$

**Output:**  $\mathbf{A}^*$

- 1: Initialization :  $k = 1, \mathbf{A}^{(k)}$
  - 2: Repeat
  - 3: update the optimization problem with  $\{\mathbf{A}^{k+1}\}$  and obtain  $\{\mathbf{A}^{k+1}\} = \arg \min_{A_{m,n}, q_n} L^{(k)}(\mathbf{A}^k)$
  - 4:  $k \leftarrow k + 1$
  - 5: Until  $|L(\mathbf{A}^{(k+1)}) - L(\mathbf{A}^{(k)})| \leq \varepsilon_5$  or  $k > L_5, \lambda_{\max}$
- 

Although CVX can solve this problem, it will suffer from the curse-of-dimensionality problem as the number of both MTCs and subchannels growing vastly. A Lagrangian Primal-Dual based algorithm is formulated to solve this problem. Given the Lagrange multiplier  $\lambda$  and SEE  $\Psi$ , the optimal power allocation for the  $m$ -th MTC on subchannel  $n$  can be derived as

$$A_{m,n} = \frac{C_{m,n} \alpha_{m,n,d} A_{m,n}^{(k)}}{C_{m,n} \alpha_{m,n,e} + \ln 2 \left( \Psi \frac{1}{\zeta} + \lambda_{m,n} + \sum_{i=1}^{m-1} B_{m,n}^i \right) A_{m,n}^{(k)}} \quad (24)$$

where

$$B_{m,n}^i = \frac{C_{i,n} \alpha_{i,n,d} |h_{m,n}^d|^2}{\ln 2 I_{i,n,d}} - \frac{C_{i,n} \alpha_{i,n,e} |h_{m,n}^e|^2}{\ln 2 I_{i,n,e}}$$

$$C_{i,n} = (\theta_{i,n} + 1) W$$

*Proof:* The proof can be seen in **Appendix E**. ■

### C. Computation Complexity of The Proposed Algorithm,

The computational complexity of Algorithm 1 is the total computational complexity of the proposed scheme, which including the computational complexity of Algorithm 2 and that of Knapsack algorithm. The computational complexity of knapsack algorithm is  $O(Mf_T)$ . The computational complexity of Algorithm 3 is  $O(l_3 MN)$ . The computational complexity of finding optimal  $\mathbf{q}_B^*$  in Algorithm 4 is  $O(lN)$ , where  $l$  is the iterations of Newton method. The computational complexity of finding the optimal  $\mathbf{P}_M^*$  in Algorithm 5 is  $O(MNl_5)$ . Thus, the total computational complexity of Algorithm 1 can be computed as  $O(l_1 l_2 l_3 MN + l_1 l_2 l_4 l_5 MN + l_1 l_2 l_4 lN + l_1 Mf_T)$ , where  $l_1, l_2, l_3, l_4, l_5$  are the number of iterations in each algorithm. The computational complexities of the proposed schemes increase with the scale of MTCs increasing.

TABLE II  
SIMULATION PARAMETERS

Parameter	Value Range
Transmission bandwidth	10MHz
Number of subchannels	50
Bandwidth of one subchannel	180kHz
Distance dependent path loss	$128.1+37.6\log_{10}(d)$ (dB), $d$ (km)
Shadowing standard deviation	8dB
Fading	Rayleigh fading
Max MTCs in same subchannel $N_n$	2
Max transmission power of R-MECS $P_{B,max}$	46dBm
Max transmission power of MTC $P_{m,max}$	23dBm
Static circuit power $P_C$	5dBm
Power amplifier efficiency $\varsigma$	0.38
coefficient of residual self-interference $\rho$	-110dB [44]
Noise $N_o$	-90dBm [44]
Total computational Resource Blocks $N_C$	30 [37]
Computing capacity of each RB $C$	10 Giga cycles/s [37]
Workload of each MTC $K_m$	[0.5,1]Giga cycles [37]
Data size of computation tasks for each MTC $L_m$	[5000,7000] bits [37]
Deadline of each MTC $D_m$	[400,500] ms [37]

## VII. PERFORMANCE EVALUATIONS

In this section, we evaluate the SEE performance in MEC assisted NOMA-based mMTC networks through Monte Carlo simulations. The system parameters used in simulations are given in Table II unless specified, where the wireless parameters are referred to 3GPP standard [47]. As the R-MECS works in full-duplex mode, its effective distance is 150m, we consider one R-MECS located in the cell center and  $M$  users are uniformly distributed on the circular range with radius of 150 m. In the simulations, we consider a worse case and set the distance between base station and eavesdropper as 30m. In Orthogonal Frequency Division Multiple Access (OFDMA) systems, each user can only be assigned to one subchannel. In NOMA systems, the maximum users can be multiplexed on the same subchannel is  $N_n$ . Considering the characteristic of MTCs, we set the static circuit power of MTCs as 5dBm.

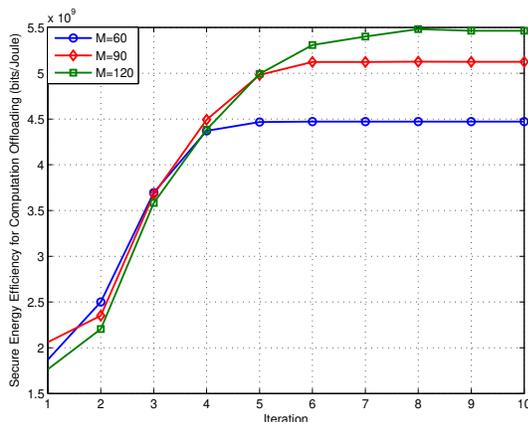


Fig. 3. SEE for Computation Offloading vs. number of iteration

The convergence behavior of proposed joint resources allocation for SEE algorithm is illustrated in Fig. 3. The proposed algorithm converges after 5 iteration when  $M = 60$ , via the update of computation resource allocation based on Knapsack algorithm and iterative communication resource allocation for maximizing SEE algorithm. With the increasing number of MTCs, although more MTCs will join the procedure of MTC-to-Subchannel matching to maximize the SEE, the increased iterations of the proposed algorithm are limited and the total required iterations are not high.

We name the proposed joint resource allocation for maximizing SEE algorithm applied in NOMA system as SEE-NOMA in Fig.3 and Fig.4. For convenience, we give the explanation of the compared schemes in [35] [37] [38] and other baseline schemes:

- 1) We take the offloading priority function scheme for users in each subchannel [35] as a compared scheme, which is based on the OFDMA system and called as SEE-OPF in the simulation results.
- 2) We take the proposed schemes in [37] as SEE-MENOMA, which aims to minimize the total energy consumption in NOMA based MEC computation offloading scenario.
- 3) Furthermore, we compare our proposed algorithm with the schemes in [38] and name it as SEE-MENOMA, which minimizes the energy consumption in OFDMA based system while guaranteeing the secure computation offloading.
- 4) To study the effect of subchannel allocation, we name the proposed joint resource allocation for maximizing SEE applied in a conventional OFDMA system as SEE-OFDMA.
- 5) Moreover, we name the scheme of substituting random subchannel allocation algorithm for MTC to subchannel matching algorithm in the proposed SEE-NOMA as SEE-RRB.
- 6) To study the influence of computation resource allocation, we name the scheme of replacing Knapsack algorithm with the fractional computation resource (FCR) allocation method  $y_m = \frac{K_m}{\sum_{i=1}^M K_i} f_T$  in the proposed SEE-NOMA as SEE-FCR, where  $f_T = 300$  Giga cycles/s is the total computing resource.
- 7) We also compared our proposed SEE-NOMA algorithm with the joint resource allocation algorithm for maximizing conventional computation offloading SR in NOMA system, which is shorted as SR-NOMA.

The performance of SEE versus the number of MTCs is evaluated in Fig.4. The total number of MTCs ranges from 30 to 150 and the upper limit number of MTCs per subchannel is  $N_n = 2$ . As it can be seen from Fig.4, the proposed joint algorithm, i.e. SEE-NOMA, achieves higher SEE than other compared algorithms. The SEE of SEE-OFDMA algorithm is higher than that of SR-NOMA, which means that performing proper power allocation can improve SEE. Although SEE-RRB is the scheme to maximize SEE, its performance is the worst due to the random subchannel allocation without schedul-

ing. The performance of SEE-OPF [35] has lower SEE than that of SEE-NOMA and SEE-FCR, which means that offloading priority function scheme is not applicable for guarantee the secure energy efficient computation offloading. Furthermore, the SEE-MENOMA [37] and SEE-MEOFDMA [38] have worst performances in SEE, because they are not mainly designed for secure energy efficient for computation offloading in NOMA-based mMTC networks.

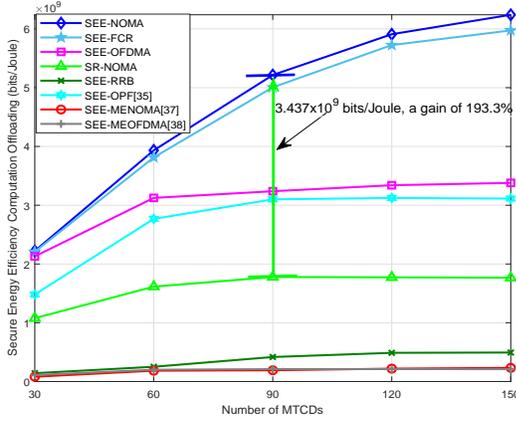


Fig. 4. Computation Offloading SEE vs. number of MTCDs

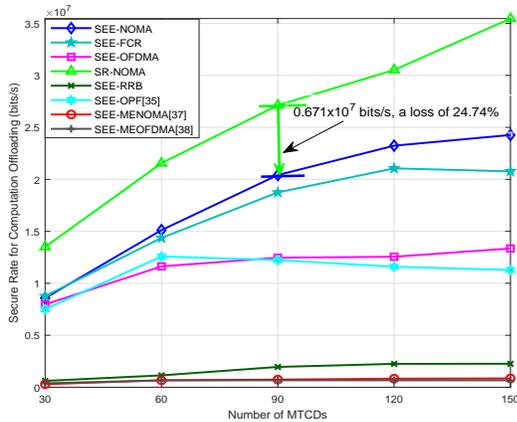


Fig. 5. Computation Offloading SR vs. number of MTCDs

The achieved SR of different schemes is also compared in Fig.5, where the simulation conditions are same with in Fig.4. The optimization goals of SR-NOMA is maximizing the SR while other schemes are maximizing SEE. Compared with the achieved SR of the SR maximization problem, the proposed SEE maximization leads to a certain loss of SR, especially the SR of proposed SEE-NOMA has a loss of 24.74%. However, the SEE of SEE-NOMA scheme has a gain of 193.30% compared with that of SR-NOMA (shown in Fig.4), which means that the SEE-NOMA will securely transmit more bit than other schemes with unit Joule. The phenomenon is caused by the tradeoff between SEE and SR. Because the computation resources strategy is the main

influencing factor when the number of MTCDs is smaller, the SR of SEE-OPF [35] is larger than that of SEE-OFDMA. While the proposed SEE-OFDMA outperforms than SEE-OPF when the number of MTCDs is large, as the power allocation strategy is the main influencing factor. As the performance in Fig.4, the SEE-MENOMA [37] and SEE-MEOFDMA [38] also have worst performances in SR. In the IoT era, it is necessary to establish differentiated service mechanisms for various requirements of IoT devices. The SR is applicable for the rate-sensitive applications. While the SEE is more suitable for energy-limited applications than SR, especially for massive MTCDs who cannot be charged once their energy drained.

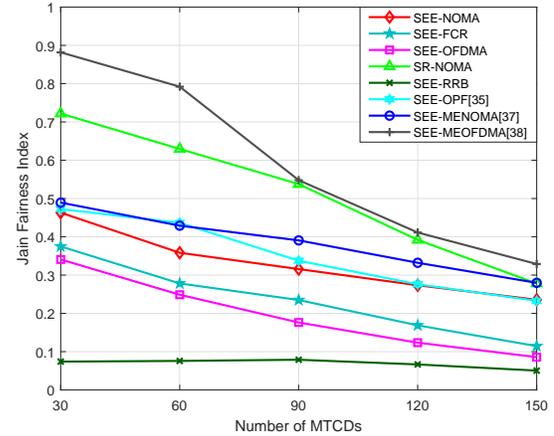


Fig. 6. Fairness index vs. number of MTCDs

With the same simulation conditions as in Fig.4, the fairness performance versus the number of MTCDs of is shown in Fig.6. Because fairness is an important metric for resource allocation (with respect to both computation and communication resource), we redefine the Jain's fairness index [48] with computation resource taken into considering as Fairness index =  $\frac{(\sum_{m=1}^M SEE_s^m \Theta(SR_m > R_{m,min}))^2}{M \sum_{m=1}^M (SEE_s^m \Theta(SR_m > R_{m,min}))^2}$ . If  $SR_m > R_{m,min}$ ,  $\Theta(SR_m > R_{m,min}) = 1$ , it means  $m$ -th MTCD has probability to perform computing offloading. Otherwise,  $\Theta(SR_m > R_{m,min}) = 0$ , the  $m$ -th MTCD will be refused to computation offloading. The fairness decreases with the total number of MTCDs increasing, because more MTCDs will compete for the limited resources. Different with the worst performance in both SEE and SR, the SEE-MENOMA [37] and SEE-MEOFDMA [38] have better performance in Jain Fairness Index. The Jain Fairness Index of SEE-MEOFDMA is highest, because it actually maximizes the transmission rate without taking security into consideration when obtaining the optimal power allocation in equation (10) of reference [38]. Furthermore, the SEE-OPF and SEE-MENOMA has also have better performance in Jain Fairness Index, because they also did not consider the restrictions of security. Due to SR-NOMA aims to maximizing the secrecy rate, the number of MTCDs that meet the requirement of  $\Theta(SR_m > R_{m,min})$  is bigger than that of algorithms maximizing SEE, which lead to a fairer state than other algorithms. Since the algorithms maximizing SEE have same characteristic in  $\Theta(SR_m > R_{m,min})$ , the

trend of fairness index is similar with Fig.4, i.e. SEE-NOMA > SEE-FCR > SEE-OFDMA > SEE-RRB.

The SEE with respect to the maximum transmission power  $P_{\text{MTCD,max}}$  of each MTCD is demonstrated in Fig.7. Due to the MTCDs are energy-critical devices, we set  $P_{\text{MTCD,max}}$  as 10 dBm, 23 dBm and 30 dBm in order not to exceed that of H2H users. With the number of MTCDs is fixed, the SEE increases when the maximum transmission power of MTCDs increases, which illustrate that the system secrecy computation offloading rate grows faster than the total power consumption in our proposed algorithm. Although the larger transmission power will enhance the secrecy performance, it also aggravates the co-channel interference. Thus, the increased SEE between  $P_{m,\text{max}} = 23$  dBm and  $P_{m,\text{max}} = 10$  dBm is larger than that between  $P_{m,\text{max}} = 23$  dBm and  $P_{m,\text{max}} = 30$  dBm.

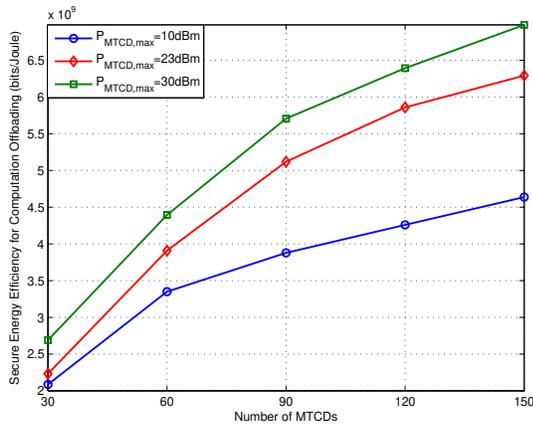


Fig. 7. Computation Offloading SEE vs. maximum transmission power of MTCDs

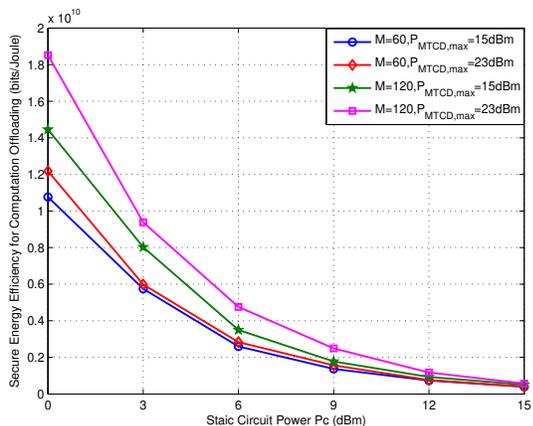


Fig. 8. Computation Offloading SEE vs. static circuit power of MTCDs

The performance of SEE versus the static circuit power  $P_c$  of each MTCD is shown in Fig.8. The static circuit power of MTCDs varies from 0dBm to 15dBm. With the number of MTCDs and maximum transmission power fixed, the SEE decreases with the static circuit power increasing. In addition, with the increase of static circuit power, the influence of

MTCDs' number and maximum transmission power on SEE is getting smaller and smaller, and static circuit power becomes the main factor affecting SEE. Thus, the design of MTCDs with low cost and low energy consumption is urgent.

The performance of SEE versus the total number of computation resource blocks in MEC server is evaluated in Fig.9. The number of CRB  $N_C$  varies from 1 to 29. With the number of MTCDs fixed, there exists a critical value of the number of CRB. Particularly, the critical value of CRB in x-axis is the number of CRB that can meet the computation offloading requirements of MTCDs. The SEE increases when the number of CRB is smaller than the critical value, and it will keep constant if the number of CRB beyond the critical value. In addition, the critical value of requiring CRB increases with the total number of MTCDs increasing. The results imply that the CRB assigning to each MEC server should be according to the total number of served MTCDs.

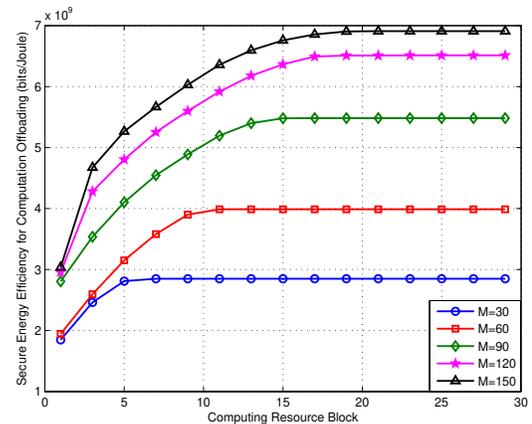


Fig. 9. Computation Offloading SEE vs. number of CRB

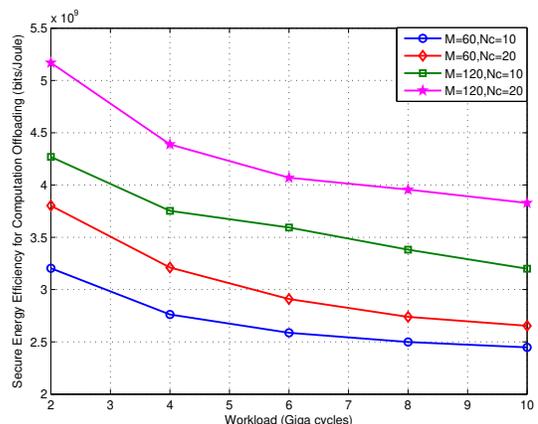


Fig. 10. Computation Offloading SEE vs. workload of MTCDs

The performance of SEE versus the workload  $K$  of each MTCD is evaluated in Fig.10. The workload of each MTCD varies from 0.2 Giga cycles to 1 Giga cycles. With given the number of MTCDs and total computation resource, because the increasing workload will need more computation resource,

which lead to the system SEE decreasing. Nevertheless, the SEE can be improved by allocating more computation resource to the MEC server, which need to be execute large high concurrent computing workload.

The performance of SEE versus the input data size  $L$  of each MTCd is evaluated in Fig.11. The input data size of each MTCd ranges from 2 Kbits to 10 Kbits. To fulfill the constraint of maximum delay, the minimum secure computation offloading rate increases with the increasing of input data size, which lead to more transmission power to be needed. Thus, SEE decreases with the increasing of input data size  $L$  with given the number of MTCds and total computing resources. Furthermore, it can be seen from Fig.11 that the influence of input data size on secrecy energy efficiency is smaller than that of MTCds' number and total computing resources. As the number of MTCds increases, the total computing resources has an increasing influence on SEE.

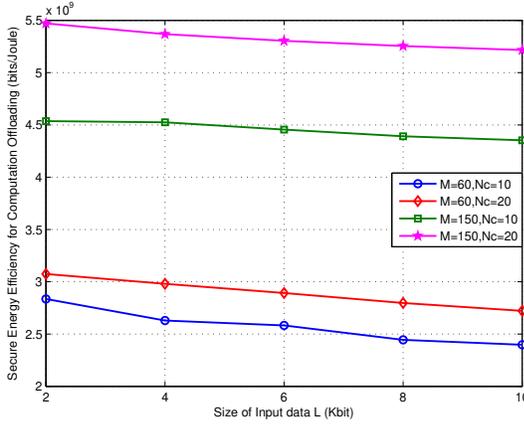


Fig. 11. Computation Offloading SEE vs. input data size of MTCds

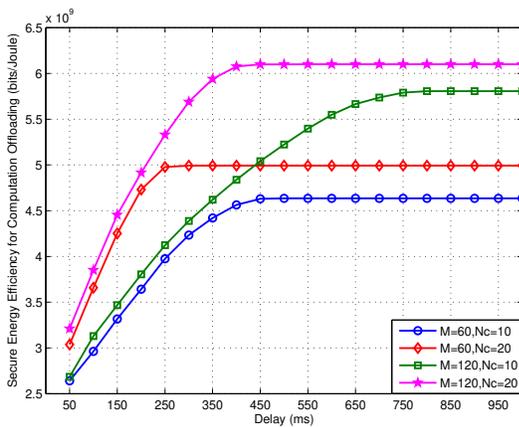


Fig. 12. Computation Offloading SEE vs. delay of MTCds

The performance of SEE versus the maximum delay  $D$  of each MTCd is evaluated in Fig.12. The maximum delay of each MTCd varies from 50 ms to 1s. With the number of MTCds and total computation resource fixed, there exists a critical value of the maximum delay, which represents the

minimum SR that all MTCds can reach to, i.e.,  $\frac{L}{D}$ . The minimum SR decreases with the number of MTCds increasing with total computing resources given, because limited communication resource also have influence on SR. When maximum delay is smaller than the critical value, the SEE increases with the increasing of maximum delay, while it keeps stable when maximum delay exceeds the critical value and maximum delay is no longer the main factor to constrain the SEE. In addition, when the total computation resource increasing and the number of MTCds fixed, more computation resource will be allocated to MTCds and the execution time will decrease, which will prolong the computation offloading time and increase the SR and SEE, thus the minimum SR decreases with the increasing of total computation resource.

## VIII. CONCLUSION

With the supporting for IoT in 5G scenarios, the R-MECS was proposed to assist the NOMA based mMTC networks for IoT secure computation offloading. By sharing computation and communication resource among MTCds, R-MECS can support many MTCds computation tasks execution at the same time and meet the MTCds' various delay requirements. Specifically, based on the defined SEE, the joint computation and communication resource allocation algorithm was formulated to guarantee the procedure of IoT computation offloading securely and energy efficient. A MTCd-to-Subchannel matching algorithm was proposed to maximize the SEE in solving subchannel allocation problem. Furthermore, a Dinkelbach-based SEE optimization algorithm was proposed by exploiting SCA and DC programming, and the closed-form expressions of power allocation for MTCds' on each subchannel were obtained. Given the communication resource allocation schemes, the computation resource allocation problem was solved by the Knapsack algorithm. Finally, the effectiveness of the proposed algorithms and the influence of computation-communication parameters on SEE has been evaluated via simulations.

## APPENDIX A PROOF OF THEOREM 1

*Proof:* Here, we investigate the relationship between secrecy energy efficiency and secure offloading transmission time, where  $m$ -th MTCd with  $L_{m,n}$  bits computation task need to be securely offloading to MEC through subchannel  $n$ . The secrecy computing offloading rate is

$$\begin{aligned} SR_{m,n} &= (r_{m,n,d} - r_{m,n,e})^+ \\ &= \left[ W \log_2 \left( \frac{1 + p_{m,n} |H_{m,n}^d|^2}{1 + p_{m,n} |H_{m,n}^e|^2} \right) \right]^+ \end{aligned} \quad (25)$$

where  $|H_{m,n}^d|^2 = \frac{|h_{m,n}^d|^2}{I_{m,n,d}}$  and  $|H_{m,n}^e|^2 = \frac{|h_{m,n}^e|^2}{I_{m,n,e}}$ . When  $|H_{m,n}^d|^2 \geq |H_{m,n}^e|^2$ ,  $SR_{m,n} = W \log_2 \left( \frac{1 + p_{m,n} |H_{m,n}^d|^2}{1 + p_{m,n} |H_{m,n}^e|^2} \right)$ . Furthermore,

$$\frac{L_{m,n}}{t_m^T} = SR_{m,n} \quad (26)$$

Then, the transmission power is

$$p_{m,n} = \frac{2^{\frac{SR_{m,n}}{W}} - 1}{|H_{m,n}^d|^2 - |H_{m,n}^e|^2 2^{\frac{SR_{m,n}}{W}}} \quad (27)$$

Based on (26) and (27), the secrecy energy efficiency can be expressed as

$$SEE_{m,n} = \frac{SR_{m,n}}{p_{m,n}} = \frac{L_{m,n} \left( |H_{m,n}^d|^2 - |H_{m,n}^e|^2 2^{\frac{SR_{m,n}}{W}} \right)}{t_m^T \left( 2^{\frac{L_{m,n}}{t_m^T W}} - 1 \right)} \quad (28)$$

we can obtain from (28) as following

$$SEE_{m,n}(t_m^T) \geq 0, SEE'_{m,n}(t_m^T) \geq 0, SEE''_{m,n}(t_m^T) \leq 0. \quad (29)$$

It is easy to see  $SEE_{m,n}$  is monotonically increasing and concave in transmission time  $t_m^T$ , and  $SEE_{m,n}$  can be increased by increasing transmission time and correspondingly decreasing power. ■

#### APPENDIX B PROOF OF THEOREM 2

*Proof:* According to (11), we have

$$\Psi^* = \frac{U_{SC}(\mathbf{P}_M^*, \mathbf{q}_B^*)}{U_{TP}(\mathbf{P}_M^*, \mathbf{q}_B^*)} = \max_{\{p_{m,n}, q_n\}} \frac{U_{SC}(\mathbf{P}_M, \mathbf{q}_B)}{U_{TP}(\mathbf{P}_M, \mathbf{q}_B)} \quad (30)$$

where  $\mathbf{P}_M^*$  and  $\mathbf{q}_B^*$  are the optimal power allocation strategies with respect to optimal SEE  $\Psi^*$ . Therefore, we have

$$\Psi^* = \frac{U_{SC}(\mathbf{P}_M^*, \mathbf{q}_B^*)}{U_{TP}(\mathbf{P}_M^*, \mathbf{q}_B^*)} \geq \frac{U_{SC}(\mathbf{P}_M, \mathbf{q}_B)}{U_{TP}(\mathbf{P}_M, \mathbf{q}_B)} \quad (31)$$

and

$$\begin{aligned} U_{SC}(\mathbf{P}_M, \mathbf{q}_B) - \Psi^* U_{TP}(\mathbf{P}_M, \mathbf{q}_B) &\leq 0 \\ U_{SC}(\mathbf{P}_M^*, \mathbf{q}_B^*) - \Psi^* U_{TP}(\mathbf{P}_M^*, \mathbf{q}_B^*) &= 0 \end{aligned} \quad (32)$$

We can obtain that  $\max_{\{p_{m,n}, q_n\}} \{U_{SC}(\mathbf{P}_M, \mathbf{q}_B) - \Psi^* U_{TP}(\mathbf{P}_M, \mathbf{q}_B)\} = 0$ . Thus, the sufficient condition of **Theorem 2** has been proved.

Then, the necessary condition should be proved. If  $\{\mathbf{P}_M^*, \mathbf{q}_B^*\}$  is the optimal power allocation policy, we have  $U_{SC}(\mathbf{P}_M^*, \mathbf{q}_B^*) - \Psi^* U_{TP}(\mathbf{P}_M^*, \mathbf{q}_B^*) = 0$ . Then, for any feasible  $\{\mathbf{P}_M, \mathbf{q}_B\}$ , we can obtain

$$\begin{aligned} U_{SC}(\mathbf{P}_M, \mathbf{q}_B) - \Psi^* U_{TP}(\mathbf{P}_M, \mathbf{q}_B) \\ \leq U_{SC}(\mathbf{P}_M^*, \mathbf{q}_B^*) - \Psi^* U_{TP}(\mathbf{P}_M^*, \mathbf{q}_B^*) = 0 \end{aligned} \quad (33)$$

Hence,

$$\frac{U_{SC}(\mathbf{P}_M, \mathbf{q}_B)}{U_{TP}(\mathbf{P}_M, \mathbf{q}_B)} = \Psi^* \frac{U_{SC}(\mathbf{P}_M, \mathbf{q}_B)}{U_{TP}(\mathbf{P}_M, \mathbf{q}_B)} \leq \Psi^* \quad (34)$$

Therefore, the optimal power allocation strategies  $\{\mathbf{P}_M^*, \mathbf{q}_B^*\}$  of the transformed objective function is also the optimal ones of the original objective function. ■

#### APPENDIX C

##### PROOF OF THE CONCAVITY $L(\theta, \mathbf{A}, \mathbf{q}_B)$ IN $\mathbf{q}_B$

*Proof:* When  $\mathbf{A} = \mathbf{P}_M^*$ ,  $L(\theta, \mathbf{A}, \mathbf{q}_B) = U_{SC}(\mathbf{A}, \mathbf{q}_B) - \Psi U_{TP}(\mathbf{A}, \mathbf{q}_B) + \sum_{m=1}^M \theta_m (SR_m - R_{m,\min})$ , then we have

$$\begin{aligned} L(\theta, \mathbf{A}, \mathbf{q}_B) &= W \sum_{n=1}^N \sum_{m=1}^{M_n} (\theta_m + 1) [f_1 - f_2 - f_3 + f_4] \\ &\quad - \Psi \sum_{n=1}^N \left( \sum_{m=1}^{M_n} \frac{1}{\zeta} p_{m,n} + p_{m,c} \right) \end{aligned} \quad (35)$$

where

$$\begin{aligned} f_1 &= \log_2 \left( \sigma^2 + \sum_{i=m}^M p_{i,n} |h_{i,n}^d|^2 + \rho q_n |g_n^R|^2 \right) \\ f_2 &= \log_2 \left( \sigma^2 + \sum_{i=m+1}^M p_{i,n} |h_{i,n}^d|^2 + \rho q_n |g_n^R|^2 \right) \\ f_3 &= \log_2 \left( \sigma^2 + \sum_{i'=m}^M p_{i',n} |h_{i',n}^e|^2 + q_n |g_n^e|^2 \right) \\ f_4 &= \log_2 \left( \sigma^2 + \sum_{i'=m+1}^M p_{i',n} |h_{i',n}^e|^2 + q_n |g_n^e|^2 \right) \end{aligned}$$

Furthermore the second derivative of  $f_1, f_2, f_3, f_4$  with respect to  $q_n$  are

$$\begin{aligned} \frac{\partial^2 f_1}{\partial (q_n)^2} &= - \frac{(\rho |g_n^R|^2)^2}{\left( \sigma^2 + \sum_{i=m}^M p_{i,n} |h_{i,n}^d|^2 + \rho q_n |g_n^R|^2 \right)^2} \\ \frac{\partial^2 f_2}{\partial (q_n)^2} &= - \frac{(\rho |g_n^R|^2)^2}{\left( \sigma^2 + \sum_{i=m+1}^M p_{i,n} |h_{i,n}^d|^2 + \rho q_n |g_n^R|^2 \right)^2} \\ \frac{\partial^2 f_3}{\partial (q_n)^2} &= - \frac{(|g_n^e|^2)^2}{\left( \sigma^2 + \sum_{i'=m}^M p_{i',n} |h_{i',n}^e|^2 + q_n |g_n^e|^2 \right)^2} \\ \frac{\partial^2 f_4}{\partial (q_n)^2} &= - \frac{(|g_n^e|^2)^2}{\left( \sigma^2 + \sum_{i'=m+1}^M p_{i',n} |h_{i',n}^e|^2 + q_n |g_n^e|^2 \right)^2} \end{aligned} \quad (36)$$

where  $\frac{\partial^2 f_1}{\partial (q_n)^2} \approx 0$  and  $\frac{\partial^2 f_2}{\partial (q_n)^2} \approx 0$  if  $\rho$  is small enough, for example  $\rho \leq 10^{-5}$ . Then, we have

$$\begin{aligned} &\frac{\partial^2 L(\mathbf{P}_M, \mathbf{q}_B)}{\partial (q_n)^2} \\ &= \sum_{m=1}^{M_n} \frac{W(\theta_m + 1) (|g_n^e|^2)^2}{\left( \sigma^2 + \sum_{i'=m}^{M_n} p_{i',n} |h_{i',n}^e|^2 + q_n |g_n^e|^2 \right)^2} \\ &\quad - \sum_{m=1}^{M_n} \frac{W(\theta_m + 1) (|g_n^e|^2)^2}{\left( \sigma^2 + \sum_{i'=m+1}^M p_{i',n} |h_{i',n}^e|^2 + q_n |g_n^e|^2 \right)^2} \\ &\leq 0 \end{aligned} \quad (37)$$

Hence,  $L(\theta, \mathbf{A}, \mathbf{q}_B)$  is a concave function of  $\mathbf{q}_B$ . ■

## APPENDIX D

## PROOF OF THE CONVERGENCE OF ALGORITHM 5

*Proof:* It can be easily obtained that,

$$L(\mathbf{A}^{(k)}) = L_1^{(k)}(\mathbf{A}^{(k)}) \leq L_1^{(k)}(\mathbf{A}^{(k+1)}) \leq L(\mathbf{A}^{(k+1)}) \quad (38)$$

where  $\mathbf{A}^{(k)}$  is the optimal value of the (23) after the  $k$ -th iteration,  $L(\mathbf{A}^{(k)}) = L_1^{(k)}(\mathbf{A}^{(k)})$  is the  $k$ -th iteration. Furthermore,  $L_1^{(k)}(\mathbf{A}^{(k)}) \leq L_1^{(k)}(\mathbf{A}^{(k+1)})$  can be obtained by the concave sub-problem. Thus,  $L(\mathbf{A}^{(k)})$  is monotonically increases when  $k$  increases. Because  $L(\mathbf{A})$  is optimize over constrained resources set, there exists a limited iteration step until it stops decreasing. Hence, the **Algorithm 5** converges to a stationary point and outputs optimal power allocation. ■

## APPENDIX E

## DERIVATION OF OPTIMAL POWER ALLOCATION STRATEGIES IN EQUATION (24)

*Proof:* For each subchannel  $n$ , the channel gains are sorted as  $|h_{1,n}^d|^2 \geq \dots |h_{m,n}^d|^2 \geq \dots \geq |h_{M,n}^d|^2$  and  $|h_{1',n}^e|^2 \geq \dots |h_{m',n}^e|^2 \geq \dots \geq |h_{M',n}^e|^2$ , when  $m = 1$ , the user with highest channel gain is decoded first and

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial A_{1,n}} &= \frac{(\theta_{1,n} + 1) W \alpha_{1,n,d}}{\ln 2 A_{1,n}} - \frac{(\theta_{1,n} + 1) W \alpha_{1,n,e}}{\ln 2 A_{1,n}^{(k)}} \\ &\quad - \Psi \frac{1}{\varsigma} - \lambda_{1,n} \end{aligned} \quad (39)$$

then

$$A_{1,n} = \frac{(\theta_{1,n} + 1) W \alpha_{1,n,d} A_{1,n}^{(k)}}{(\theta_{1,n} + 1) W \alpha_{1,n,e} + \ln 2 (\Psi \frac{1}{\varsigma} + \lambda_{1,n}) A_{1,n}^{(k)}} \quad (40)$$

When  $m = 2$ , the user with second highest channel gain is decoded and

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial A_{2,n}} &= \frac{(\theta_{2,n} + 1) W \alpha_{2,n,d}}{\ln 2 A_{2,n}} - \frac{(\theta_{2,n} + 1) W \alpha_{2,n,e}}{\ln 2 A_{2,n}^{(k)}} \\ &\quad - \Psi \frac{1}{\varsigma} - \lambda_{2,n} + \frac{\partial B \alpha_{1,n,d} \log_2(\tilde{\gamma}_{1,n,d}(A_{1,n}))}{\partial A_{2,n}} \\ &\quad - \frac{\partial B \alpha_{m,n,e} \log_2(\tilde{\gamma}_{1,n,e}(A_{1,n}^{(k)}))}{\partial A_{2,n}} \\ &= \frac{(\theta_{2,n} + 1) W \alpha_{2,n,d}}{\ln 2 A_{2,n}} - \frac{(\theta_{2,n} + 1) W \alpha_{2,n,e}}{\ln 2 A_{2,n}^{(k)}} \\ &\quad - \Psi \frac{1}{\varsigma} - \lambda_{2,n} - B_{2,n}^1 \end{aligned} \quad (41)$$

$$\begin{aligned} B_{2,n}^1 &= \frac{(\theta_{1,n} + 1) W \alpha_{1,n,d} |h_{2,n}^d|^2}{\ln 2 I_{1,n,d}} - \frac{(\theta_{1,n} + 1) W \alpha_{1,n,e} |h_{2,n}^e|^2}{\ln 2 I_{1,n,e}} \end{aligned}$$

Then we have

$$A_{2,n} = \frac{(\theta_{2,n} + 1) W \alpha_{2,n,d} A_{2,n}^{(k)}}{(\theta_{2,n} + 1) W \alpha_{2,n,e} + \ln 2 (\Psi \frac{1}{\varsigma} + \lambda_{2,n} + B_{2,n}^1) A_{2,n}^{(k)}} \quad (42)$$

When  $m = 3$ , the user with third highest channel gain is decoded and

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial A_{3,n}} &= \frac{(\theta_{3,n} + 1) W \alpha_{3,n,d}}{\ln 2 A_{3,n}} - \frac{(\theta_{3,n} + 1) W \alpha_{3,n,e}}{\ln 2 A_{3,n}^{(k)}} \\ &\quad - \Psi \frac{1}{\varsigma} - \lambda_{3,n} + \frac{\partial B \alpha_{1,n,d} \log_2(\tilde{\gamma}_{1,n,d}(A_{1,n}))}{\partial A_{3,n}} \\ &\quad - \frac{\partial B \alpha_{m,n,e} \log_2(\tilde{\gamma}_{1,n,e}(A_{1,n}^{(k)}))}{\partial A_{3,n}} \\ &\quad + \frac{\partial B \alpha_{2,n,d} \log_2(\tilde{\gamma}_{2,n,d}(A_{2,n}))}{\partial A_{3,n}} \\ &\quad - \frac{\partial B \alpha_{m,n,e} \log_2(\tilde{\gamma}_{2,n,e}(A_{2,n}^{(k)}))}{\partial A_{3,n}} \\ &= \frac{(\theta_{3,n} + 1) W \alpha_{3,n,d}}{\ln 2 A_{3,n}} - \frac{(\theta_{3,n} + 1) W \alpha_{3,n,e}}{\ln 2 A_{3,n}^{(k)}} \\ &\quad - \Psi \frac{1}{\varsigma} - \lambda_{3,n} - B_{3,n}^1 - B_{3,n}^2 \end{aligned} \quad (43)$$

$$\begin{aligned} B_{3,n}^1 &= \frac{(\theta_{1,n} + 1) W \alpha_{1,n,d} |h_{3,n}^d|^2}{\ln 2 I_{1,n,d}} - \frac{(\theta_{1,n} + 1) W \alpha_{1,n,e} |h_{3,n}^e|^2}{\ln 2 I_{1,n,e}} \\ B_{3,n}^2 &= \frac{(\theta_{2,n} + 1) W \alpha_{2,n,d} |h_{3,n}^d|^2}{\ln 2 I_{2,n,d}} - \frac{(\theta_{2,n} + 1) W \alpha_{2,n,e} |h_{3,n}^e|^2}{\ln 2 I_{2,n,e}} \end{aligned}$$

Then, we obtain

$$\begin{aligned} A_{3,n} &= \frac{(\theta_{3,n} + 1) W \alpha_{3,n,d} A_{3,n}^{(k)}}{(\theta_{3,n} + 1) W \alpha_{3,n,e} + \ln 2 (\Psi \frac{1}{\varsigma} + \lambda_{3,n} + B_{3,n}^1 + B_{3,n}^2) A_{3,n}^{(k)}} \end{aligned} \quad (44)$$

Therefore, by deduction, we have

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial A_{m,n}} &= (\theta_{m,n} + 1) \frac{W \alpha_{m,n,d}}{\ln 2 A_{m,n}} - (\theta_{m,n} + 1) \frac{W \alpha_{m,n,e}}{\ln 2 A_{m,n}^{(k)}} \\ &\quad - \Psi \frac{1}{\varsigma} - \lambda_{m,n} - \sum_{i=1}^{m-1} B_{m,n}^i \end{aligned} \quad (45)$$

$A_{m,n}$

$$= \frac{C_{m,n} \alpha_{m,n,d} A_{m,n}^{(k)}}{C_{m,n} \alpha_{m,n,e} + \ln 2 (\Psi \frac{1}{\varsigma} + \lambda_{m,n} + \sum_{i=1}^{m-1} B_{m,n}^i) A_{m,n}^{(k)}} \quad (46)$$

where

$$\begin{aligned} B_{m,n}^i &= \frac{C_{i,n} \alpha_{i,n,d} |h_{m,n}^d|^2}{\ln 2 I_{i,n,d}} - \frac{C_{i,n} \alpha_{i,n,e} |h_{m,n}^e|^2}{\ln 2 I_{i,n,e}} \\ C_{i,n} &= (\theta_{i,n} + 1) W \end{aligned}$$

■

## REFERENCES

- [1] M. Elsaadany, A. Ali, and W. Hamouda, "Cellular lte-a technologies for the future internet-of-things: Physical layer features and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2544–2572, Fourthquarter 2017.
- [2] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, Feb. 2017.
- [3] L. Liang, L. Xu, B. Cao, and Y. Jia, "A cluster-based congestion-mitigating access scheme for massive m2m communications in internet of things," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2200–2211, Jun. 2018.
- [4] Z. Yang, W. Xu, Y. Pan, C. Pan, and M. Chen, "Energy efficient resource allocation in machine-to-machine communications with multiple access and energy harvesting for iot," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 229–245, Feb. 2018.
- [5] T. Lv, Y. Ma, J. Zeng, and P. T. Mathiopoulos, "Millimeter-wave noma transmission in cellular m2m communications for internet of things," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1989–2000, Jun. 2018.
- [6] J. Huang, C. Xing, S. Y. Shin, F. Hou, and C. Hsu, "Optimizing m2m communications and quality of services in the iot for sustainable smart cities," *IEEE Trans. on Sustainable Computing*, vol. 3, no. 1, pp. 4–15, Jan. 2018.
- [7] J. Lianghai, B. Han, M. Liu, and H. D. Schotten, "Applying device-to-device communication to enhance iot services," *IEEE Commun. Standards Mag.*, vol. 1, no. 2, pp. 85–91, Jun. 2017.
- [8] Ericsson, "Ericsson mobility report," <https://www.ericsson.com/en/mobility-report>, Jun. 2018.
- [9] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. on Signal and Inform. Process. over Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.
- [10] N. Li, J.-F. Martinez-Ortega, and G. Rubio, "Distributed joint offloading decision and resource allocation for multi-user mobile edge computing: A game theory approach," *arXiv preprint arXiv:1805.02182*, 2018.
- [11] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, Feb. 2018.
- [12] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing: A key technology towards 5g," *ETSI white paper*, vol. 11, no. 11, pp. 1–16, 2015.
- [13] S. C. Carrie MacGillivray, "Worldwide iot market update," IDC, <https://www.idc.com/getdoc.jsp?containerId=US44244318&pageType=PRINTFRIENDLY>, Sep. 2018.
- [14] H. Zhang, F. Guo, H. Ji, and C. Zhu, "Combinational auction-based service provider selection in mobile edge computing networks," *IEEE Access*, vol. 5, pp. 13 455–13 464, Jul. 2017.
- [15] T. Zhao, S. Zhou, X. Guo, and Z. Niu, "Tasks scheduling and resource allocation in heterogeneous cloud for delay-bounded mobile edge computing," in *Proc. IEEE Intern. Conf. on Commun. (ICC)*, Paris, France, May 2017, pp. 1–7.
- [16] L. Song, Y. Li, Z. Ding, and H. V. Poor, "Resource management in non-orthogonal multiple access networks for 5g and beyond," *IEEE Network*, vol. 31, no. 4, pp. 8–14, Jul. 2017.
- [17] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular iot: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.
- [18] F. Wang, J. Xu, and Z. Ding, "Optimized multiuser computation offloading with multi-antenna noma," in *Proc. IEEE Global Commun. Conf. Workshops (GC Wkshps)*, Singapore, Singapore, Dec. 2017, pp. 1–7.
- [19] Y. Yu, J. Zhang, and K. B. Letaief, "Joint subcarrier and cpu time allocation for mobile edge computing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [20] I. Ahmad, T. Kumar, M. Liyanage, J. Okwuibe, M. Ylianttila, and A. Gurtov, "Overview of 5g security challenges and solutions," *IEEE Commun. Standards Mag.*, vol. 2, no. 1, pp. 36–43, Mar. 2018.
- [21] D. He, S. Chan, and M. Guizani, "Security in the internet of things supported by mobile edge computing," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 56–61, Aug. 2018.
- [22] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, Fourthquarter 2017.
- [23] J. Ni, K. Zhang, X. Lin, and X. S. Shen, "Securing fog computing for internet of things applications: Challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 601–628, Firstquarter 2018.
- [24] 3GPP, "3rd generation partnership project; technical specification group radio access network; study on non-orthogonal multiple access (noma) for nr," TR 38.812, Release 15, Aug. 2018.
- [25] Z. Ding, P. Fan, and H. V. Poor, "Impact of non-orthogonal multiple access on the offloading of mobile edge computing," *arXiv preprint arXiv:1804.06712*, 2018.
- [26] F. Wang, J. Xu, and Z. Ding, "Multi-antenna noma for computation offloading in multiuser mobile edge computing systems," *arXiv preprint arXiv:1707.02486*, 2017.
- [27] B. He, A. Liu, N. Yang, and V. K. N. Lau, "On the design of secure non-orthogonal multiple access systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2196–2206, Oct 2017.
- [28] S. Han, X. Xu, X. Tao, and P. Zhang, "Joint power and sub-channel allocation for secure transmission in noma-based mmhc networks," *IEEE Systems Journal*, pp. 1–12, 2019.
- [29] Y. Shiu, S. Y. Chang, H. Wu, S. C. . Huang, and H. Chen, "Physical layer security in wireless networks: a tutorial," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 66–74, Apr. 2011.
- [30] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [31] P. Porabage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on multi-access edge computing for internet of things realization," *arXiv preprint arXiv:1805.06695*, 2018.
- [32] X. Sun and N. Ansari, "Edgeiot: Mobile edge computing for the internet of things," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 22–29, Dec. 2016.
- [33] H. Wu, Q. Wang, and K. Wolter, "Tradeoff between performance improvement and energy saving in mobile cloud offloading systems," in *2013 IEEE International Conference on Communications Workshops (ICC)*, Budapest, Hungary, Jun. 2013, pp. 728–732.
- [34] H. Wu, Y. Sun, and K. Wolter, "Energy-efficient decision making for mobile cloud offloading," *IEEE Transactions on Cloud Computing*, pp. 1–1, Jan. 2018.
- [35] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.
- [36] H. Xing, L. Liu, J. Xu, and A. Nallanathan, "Joint task assignment and wireless resource allocation for cooperative mobile-edge computing," *arXiv preprint arXiv:1802.06862*, 2018.
- [37] A. Kiani and N. Ansari, "Edge computing aware noma for 5g networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, Apr. 2018.
- [38] J. Xu and J. Yao, "Exploiting physical-layer security for multiuser multicarrier computation offloading," *IEEE Wireless Commun. Letters*, vol. X, no. X, pp. 1–1, Jun. 2018.
- [39] Y. Wu, L. P. Qian, H. Mao, X. Yang, H. Zhou, X. Tan, and D. H. Tsang, "Secrecy-driven resource management for vehicular computation offloading networks," *IEEE Network*, vol. 32, no. 3, pp. 84–91, May 2018.
- [40] B. Li, T. Chen, and G. B. Giannakis, "Secure mobile edge computing in iot via collaborative online learning," *arXiv preprint arXiv:1805.03591*, 2018.
- [41] L. Xiao, C. Xie, T. Chen, H. Dai, and H. V. Poor, "A mobile offloading game against smart attacks," *IEEE Access*, vol. 4, pp. 2281–2291, May 2016.
- [42] Y. Chen, Y. Zhang, and S. Maharjan, "Deep learning for secure mobile edge computing," *arXiv preprint arXiv:1709.08025*, 2017.
- [43] R. Hsu, J. Lee, T. Q. S. Quek, and J. Chen, "Reconfigurable security: Edge-computing-based framework for iot," *IEEE Network*, vol. 32, no. 5, pp. 92–99, Sep. 2018.
- [44] X. Zhang, T.-H. Chang, Y.-F. Liu, C. Shen, and G. Zhu, "Max-min fairness user scheduling and power allocation in full-duplex ofdma systems," *arXiv preprint arXiv:1803.07395*, Apr. 2018.
- [45] J. Papandriopoulos and J. S. Evans, "Scale: A low-complexity distributed protocol for spectrum balancing in multiuser dsl networks," *IEEE Trans. on Inf. Theory*, vol. 55, no. 8, pp. 3711–3724, Aug. 2009.
- [46] T. Wang and L. Vandendorpe, "Iterative resource allocation for maximizing weighted sum min-rate in downlink cellular ofdma systems," *IEEE Trans. on Sig. Process.*, vol. 59, no. 1, pp. 223–234, Jan. 2011.
- [47] 3GPP, "3rd generation partnership project; technical specification group radio access network; evolved universal terrestrial radio access (e-utra); further advancements for e-utra physical layer aspects," TR 36.814, Release 9, Mar. 2017.

- [48] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, 1984.



**Shujun Han** is a Ph.D. in the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. She received the B.E. and M.S. degrees from Zhengzhou University in 2013 and 2016. Her main research interests are in the area in cognitive radio networks, machine type communications and networks, Internet of Things, with currently focusing on radio resource management for enhancing the physical layer security.



**Xiaodong Xu** (S'06-M'07-SM'18) received his B.S degree in Information and Communication Engineering and Master's Degree in Communication and Information System both from Shandong University in 2001 and 2004 separately. He received his Ph.D. degrees of Circuit and System in Beijing University of Posts and Telecommunications (BUPT) in 2007. He is currently a professor of BUPT. He has co-authored seven books and more than 120 journal and conference papers. He is also the inventor or co-inventor of 37 granted patents. His research interests cover network architecture, moving network, coordinated multi-point, mobile network visualization and IoT. He is supported by Beijing Nova Programme on mobile networking.



**Sisai Fang** is a M.S. in the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China. He received the B.E. degree from University of Science and Technology Beijing in 2016. His main research interests include Radio Resource Management, Cell Range Expansion under multi-frequency networks and deployment density optimization in 5G.



**Yan Sun** received the B.Eng. degree in telecommunications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2001, and the M.Sc. and Ph.D. degrees in electronic engineering from the Queen Mary University of London, London, U.K., in 2003 and 2009, respectively. In 2001, she joined Siemens Ltd., Beijing, as a Network Optimization Engineer, which she then rejoined as a System Engineer in research and development in 2003 and a Product Manager for five years. In 2009, she joined the Queen Mary University of London, as a Lecturer. Her current research interests include ad hoc networks, energy saving for modern mobile networks, software defined networks, and mobile healthcare networks.



**Yue Cao** (M'16) received the PhD degree from the Institute for Communication Systems (ICS), University of Surrey, UK in 2013. He was the Research Fellow at University of Surrey, UK; Lecturer and Senior Lecturer at Department of Computer and Information Sciences, Northumbria University, UK; and has been the International Lecturer at School of Computing and Communications, Lancaster University, UK. His research interests focus on Intelligent Transport Systems. He is the Associate Editor of IEEE Access, KSII Transactions on Internet and Information Systems, IGI Global International Journal of Vehicular Telematics and Infotainment Systems.



**Xiaofeng Tao** (S'99-A'02-M'03-SM'13) received the B.S. degree in electrical engineering from Xian Jiaotong University, Xi'an, China, in 1993, and the M.S.E.E. and Ph.D. degrees in telecommunication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1999 and 2002, respectively. He was a Visiting Professor with Stanford University, Stanford, CA, USA, from 2010 to 2011; the Chief Architect with the Chinese National FUTURE Fourth-Generation (4G) TDD Working Group from 2003 to 2006; and established the 4G TDD CoMP Trial Network in 2006. He is currently a Professor with the BUPT and a fellow of the Institution of Engineering and Technology. He is the inventor or co-inventor of 50 patents and the author or co-author of 120 papers in 4G and beyond 4G. He is currently involved in fifth-generation networking technology and mobile network security.



**Ping Zhang** (M'07-SM'15-F'18) received his Ph.D. degree from BUPT in 1990. He is currently a professor at Beijing University of Posts and Telecommunications. He has published eight books and more than 400 papers, and he holds approximately 170 patents. He is Executive Associate Editor-in-chief on information sciences of Chinese Science Bulletin, a Member of next-generation broadband wireless communication network in National Science and Technology Major Project committee, a Member of the 5th Advisory Committee of NSFC(National Natural Science Foundation of China), the Chief Scientist of "973" National Basic Research Program of China, member of MOST(The Ministry of science and technology) 863 Program Expert team, member of MOST IMT-Advanced 5G Expert team. His current research interests include mobile communications, ubiquitous networking, and service provisioning, especially in the key techniques of the 5G systems.