Sparse Code Multiple Access-Based Edge Computing for IoT Systems

Ali Alnoman, Serhat Erkucuk, *Senior Member, IEEE*, and Alagan Anpalagan, *Senior Member, IEEE*

***Abstract*—In this paper, a sparse code multiple access (SCMA)-based edge computing scheme is proposed for Internet-of-Things (IoT) systems. The aim of implementing SCMA, which is a nonorthogonal multiple access resource allocation technique, is to improve network connectivity and maximize data rate provision. The proposed edge-IoT system is investigated under different SCMA configurations to explore the various performance aspects such as connectivity, throughput, task completion time, and complexity. First, the problem is formulated as a data rate maximization problem for SCMA-based heterogeneous networks under power constraints. Then, the problem is subdivided into a power allocation problem, which is solved using the water filling approach, and a codebook allocation problem that is solved using a heuristic algorithm. The results show that the SCMA scheme can significantly improve the IoT performance compared to the conventional orthogonal frequency-division multiple access resource allocation scheme in terms of connectivity, throughput, and task completion time provided that SCMA configurations are suitable with IoT processing capabilities to avoid undesired detection latency.**

***Index Terms*—Edge computing, heterogeneous networks (HetNets), Internet-of-Things (IoT), nonorthogonal multiple access (NOMA), orthogonal frequency-division multiple access (OFDMA), sparse code multiple access (SCMA).**

I. Introduction

THE INTERNET of Things (IoT) is expected to remarkably change the way we are living into a smarter, safer, and easier lifestyle. With the current trend toward IoT, it can be realized that IoT is confidently dominating the future of information and communication technologies. However, many challenges still exist regarding the establishment of efficient IoT systems, in particular, device connectivity and service latency. To meet the massive connectivity demands of IoT devices, the sparse code multiple access (SCMA) scheme, which is a nonorthogonal multiple access (NOMA)-based scheme is envisioned as a promising solution to cope with the connectivity challenge [1], [2]. Unlike orthogonal frequency division techniques, NOMA-based schemes allow multiple users to share the same subcarriers to increase the number of users served. In contrast with other NOMA techniques, SCMA provides improved link-level performance and block error rate as introduced in [3] and [4]. Furthermore, comparing SCMA with code division multiple access (CDMA), which is a code-domain multiple access scheme, SCMA allows a multidimensional design of constellation points that in turn enhances system flexibility compared to the 1-D constellation in CDMA [5].

Despite the aforementioned advantages, SCMA detection requires complicated algorithms to decode transmitted signals especially when the number of users sharing one subcarrier is increased. Along with the decoding process, implementing robust interference cancellation techniques is inevitable to maintain satisfactory signal quality at IoT receivers. In addition, the computing capability of IoT devices might be inadequate for fast SCMA detection, thus more delay will be experienced, and that is a serious issue for delay-sensitive applications such as e-health and vehicular communications that can only tolerate few milliseconds of delay [6]. Therefore, it is essential to consider the computing capabilities of IoT systems such as the microprocessor speed [7], when designing SCMA-based systems.

Rather than performing computing tasks using the on-device processors, IoT devices have the opportunity to offload their tasks to edge devices (fog nodes) in the vicinity, benefiting from the reduced end-to-end latency. The physical proximity of edge devices with end-users also supports IoT applications that require location awareness, low latency, and high quality of service (QoS) [8]. Edge devices are equipped with the necessary hardware to enable small-scale cloud-like functions such as authentication, computing, and storage. Performing cloud functions at the network edge helps to save backhaul resources and alleviate the burdens on cloud servers. Instead of sending all data to the distant cloud, the operations of data aggregation, filtration, and analysis can be achieved by edge devices leaving only abstracted data to be further processed by the cloud. In the same context, edge devices can carry out machine learning techniques to harness the big IoT data for achieving accurate content caching and provide timely responses to end-users [9].

To take the full advantage of edge computing, it is necessary to coordinate edge devices with the central cloud on one hand, and with the heterogeneous cloud radio access networks (H-CRANs) on the other hand [10]. With the help of software-defined networking (SDN) technology, efficient coordination of computing and communication nodes can be

Manuscript received December 19, 2018; revised March 5, 2019 and April 8, 2019; accepted April 29, 2019. Date of publication May 2, 2019; date of current version July 31, 2019. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and in part by the Scientific and Technological Research Council of Turkey (TUBITAK). *(Corresponding author: Alagan Anpalagan.)*

A. Alnoman and A. Anpalagan are with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: ali.alnoman@ryerson.ca; alagan@ee.ryerson.ca).

S. Erkucuk is with the Department of Electrical-Electronics Engineering, Kadir Has University, 34083 Istanbul, Turkey (e-mail: serkucuk@khas.edu.tr).

Digital Object Identifier 10.1109/JIOT.2019.2914570

achieved with less complexity. For instance, when an IoT device roams between different edge devices, SDN can assist in evaluating the task offloading profit for that IoT device in regard with experienced delay based on the network traffic, and accordingly decides on whether to initiate the virtual machine (VM) migration process or to keep the VM in its original location [11].

From the communication perspective, the diversity of radio access technologies (RATs), such as LTE, Wi-Fi, and fiber cables utilized in 5G heterogeneous networks (HetNets) requires sophisticated network management strategies to optimally allocate radio and computing resources. However, with the traditional network control techniques, which require manual human intervention in many aspects, it can be quite challenging. To this end, network data aggregation and processing in the centralized baseband unit (BBU) pool, in the well-known architecture of H-CRANs, can achieve huge success in this direction [12]. Small base stations (SBSs) and remote radio heads (RRHs) in H-CRANs belong to the data plane in which they are responsible for providing high data rates by exploiting the spatial frequency reuse. Meanwhile, macro base stations (MBSs) belong to the control plane by providing cross-tier control and management, such as user association, traffic flow, handover management, and network-wide coverage.

In this paper, we conduct a comprehensive investigation on the feasibility of SCMA for distributed IoT computing systems. By selecting different SCMA parameters, the system performance is significantly affected especially with regard to connectivity and computing delay. Scalable SCMA codebook configuration is also proposed, carried out through simulations, and is shown to improve system performance compared to the conventional orthogonal frequency-division multiple access (OFDMA) scheme.

A. Related Work

In [1], an NOMA-based radio and computing resource allocation scheme was proposed to reduce energy consumption in mobile edge computing. To enable an interactive communication among sensors and actuators, a power and channel allocation framework was proposed in [17] for 5G IoT networks. Furthermore, a hierarchical computing resource allocation scheme was proposed in [18] to optimally allocate the limited resources of fog nodes in IoT services. In [19], a comparison study showed that SCMA can provide better throughput in HetNets in contrast with other NOMA schemes at the cost of extra detection complexity.

In terms of SCMA encoding, several papers in the literature considered optimal codebook design as in [5], where the system capacity and outage probability were derived for minimizing outage probability for SCMA users using power allocation (PA). In the same context, the work in [4] aimed to reduce the detection complexity in codebook design, namely, the constellation design and codebook assignment. The detection complexity in SCMA has been investigated in [20], wherein the conventional message passing algorithm (MPA) was enhanced using sphere decoding to reduce the number of superimposed constellation points in SCMA codebooks. Furthermore, decomposing high-order SCMA systems into smaller low-order systems using mapping modules was proposed in [21] to simplify the decoding process. In [2], a learning-based codebook generation and decoding strategy was proposed to adaptively construct codebooks with enhanced bit error rate.

From the computing perspective, different edge computing models have been used in the literature to investigate the computing performance in IoT systems. In [1], the computing capacity of edge devices are divided into resource blocks with certain CPU cycles. Each of these resource blocks is then allocated to a cluster of users that share the same frequency resources in an NOMA-based system. The work in [13], considered associating IoT devices with different fog nodes depending on application requirements and resource availability. Afterward, each associated IoT device is allocated one VM with constant CPU speed. Likewise, the study in [14] considered associating IoT devices with suitable fog nodes; however, the computing resources of each fog node were considered to be shared equally among all IoT devices within that node. Table I summarizes recent related works considering different computing models, multiple access schemes, and objectives. As seen in the table, some studies considered edge computing using NOMA techniques (other than SCMA) in homogeneous networks, whereas other studies considered SCMA in homogeneous networks without incorporating edge computing. However, SCMA has not been considered neither for edge computing in homogeneous or HetNets in general, nor in the context of IoT device connectivity and time latency in particular.

B. Contribution

The major contributions of this paper are as follows.

1) An SCMA-based scheme is proposed for edge computing to improve IoT system connectivity, throughput, and reduce task completion time in HetNets compared to orthogonal multiple access schemes.
2) An optimization problem is formulated to maximize data rate provisioning under the maximum power constraint of base stations. The problem is subdivided into a PA problem which is solved using the water-filling technique, and a codebook allocation algorithm which aims to assign users the codebooks with the highest signal-to-interference-plus-noise-ratio (SINR).
3) SCMA parameters are investigated to fulfill the high quality of experience (QoE) requirements in IoT systems. Since each IoT application has different processing requirements, CPU cycles are allocated considering the total computing capacity of the fog node. Moreover, each IoT device is assumed to have a particular processor speed to consider the detection time with total experienced delay.

C. Organization

The rest of this paper is organized as follows. In Section II, the system model is introduced where network, computing, and SCMA models are presented. The problem formulation

TABLE I
RELATED WORKS

Reference	Channel Allocation	Computing Model	Network Model	Objective
[1]	Non-orthogonal	Edge computing	Homogeneous	Energy consumption
[13]	Orthogonal	Edge computing	Homogeneous	Response time
[14]	Orthogonal	Edge computing	Homogeneous	System revenue
[15]	SCMA (non-orthogonal)	N/A	Homogeneous	Network utility
[16]	SCMA (non-orthogonal)	N/A	Homogeneous	Energy efficiency
[5]	SCMA (non-orthogonal)	N/A	Homogeneous	Outage probability and power allocation
[3]	SCMA (non-orthogonal)	N/A	Homogeneous	Energy efficiency and detection complexity
Our work	SCMA (non-orthogonal)	Edge computing	Heterogeneous	Device connectivity, sum rate, and task completion time

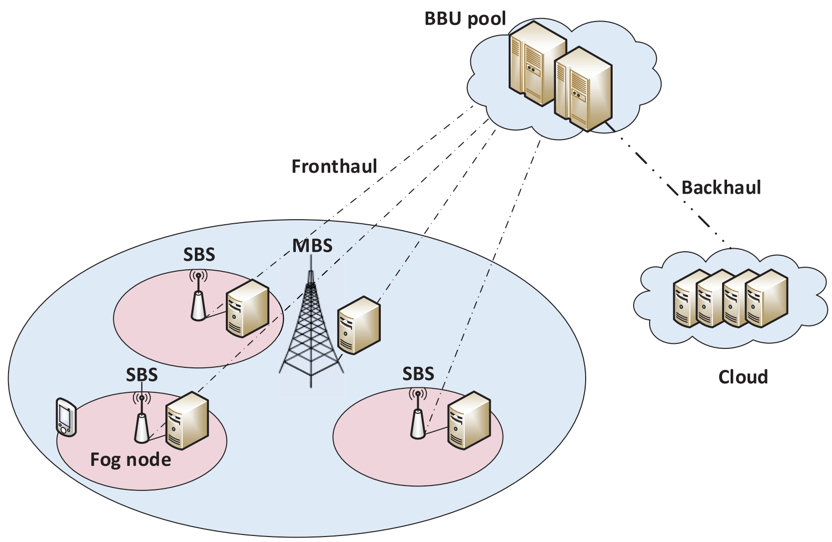

Fig. 1. Proposed system layout.

and solution approach are given in Section III. Simulation results and discussions are presented in Section IV. Finally, Section V provides concluding remarks.

II. SYSTEM MODEL

In this section, the network and computing models are presented. Fig. 1 depicts the proposed joint communication-computing system layout.

A. Network Model

We consider an SCMA-based HetNet consisting of one MBS and a set of pico (small) base stations denoted by $\mathcal{N} = \{1, \ldots, N_p\}$. A set of IoT users $\mathcal{U} = \{1, \ldots, N_u\}$ are served using a bandwidth B that is divided into a set of subcarriers $\mathcal{S} = \{1, \ldots, N_s\}$ which are later mapped into a set of codebooks denoted by $\mathcal{C} = \{1, \ldots, N_c\}$. The sum rate of user $k \in \mathcal{U}$ is given by

$$R_k = \sum_{n=1}^{N_p} \sum_{s=1}^{N_s} a_{k,s}^n \log_2\left(1 + \frac{p_{k,s}^n |h_{k,s}^n|^2}{I_{k,s}^n + N_0}\right) \tag{1}$$

where $a_{k,s}^n$, $p_{k,s}^n$, $h_{k,s}^n$, and $I_{k,s}^n$ denote respectively the user association, PA, channel gain, and interchannel interference of user k over subcarrier s at base station n. N_0 is the noise power spectral density. It should be noted that the second term in (1) represents the SINR offered to user k from base station n over subcarrier s. The interchannel interference can be expressed as

$$I_{k,s}^n = \sum\nolimits_{\{i:|h_{i,s}^n|^2 > |h_{k,s}^n|^2\}} p_{i,s}^n |h_{k,s}^n|^2 \quad \forall n \in \mathcal{N}. \tag{2}$$

Thus, the sum data rate obtained by all IoT devices can be calculated as

$$R_T = \sum_{k=1}^{N_u} R_k. \tag{3}$$

It should be noted that n refers to both the fog node and the SBS since they are considered functioning on the same site and serving the same users.

B. Computing Model

The diversity of IoT devices and applications such as e-health, smart transportation, and smart homes imposes different computing and delay requirements. Therefore, it is essential to address the specific needs of each particular device to maintain satisfactory QoE in terms of both computing and radio resource allocation. In the proposed system, each incoming user (IoT device) k is assumed to have a specific data size D_k bits, and a task completion deadline T_k seconds. Unlike the centralized cloud computing model where all computing tasks are processed in the distant cloud servers, fog nodes in the proposed edge computing model are responsible for handling computing tasks at the vicinity of IoT users within the small-cell tier. The computing (CPU) resources in cycle/s allocated

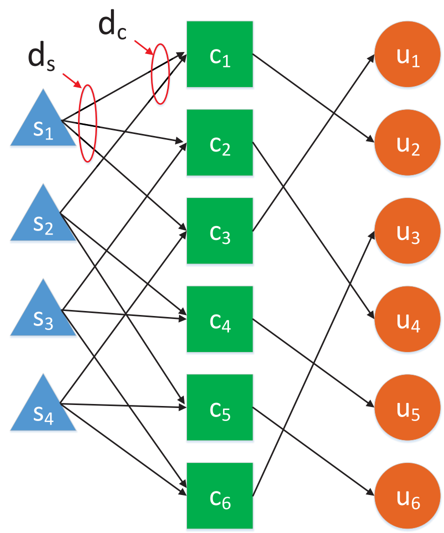

Fig. 2. Factor graph of SCMA with $N_s = 4$, $N_c = 6$, $N_u = 6$, $d_s = 3$, and $d_c = 2$.

by fog node n to user k can be expressed as

$$c_k^n = \frac{\frac{D_k}{T_k}}{\sum_{k \in \mathcal{U}_n} \frac{D_k}{T_k}} \times C_n \tag{4}$$

where C_n and $\mathcal{U}_n$ denote the total computing capacity in cycles/sec and the set of users being served by fog node n, respectively. As shown in (4), the computing resources of a fog node n are shared among all associated users ($k \in \mathcal{U}_n$). Moreover, the amount of resources allocated to a user k depends on the ratio (D_k/T_k) such that a user with larger data size or more strict deadline will be allocated more CPU resources. Assuming that each bit of data requires one cycle for processing (i.e., 1 cycle/s is equivalent to 1 bit/s), the task completion time of user k can be calculated as follows:

$$t_k = \frac{D_k}{R_k} + \frac{D_k}{c_k^n} \tag{5}$$

where (D_k/R_k) and (D_k/c_k^n) denote the delay incurred by wireless transmission and fog node processing, respectively. Each task k is considered satisfied if the task completion time t_k remains below the task completion deadline T_k (i.e., $t_k < T_k$).

C. SCMA Model

1) Codebook Structure: We consider an SCMA system that allows N_s subcarriers to be shared by N_c codebooks which are later allocated to N_u IoT devices. Each individual subcarrier can be used simultaneously by d_s codebooks; whereas each codebook is assigned d_c subcarriers. Fig. 2 demonstrates the mapping relationship of subcarriers, codebooksm and users for an SCMA system with $N_s = 4$, $N_c = 6$, $N_u = 6$, $d_s = 3$, and $d_c = 2$.

Codebook design has been investigated in several studies, including [4], [5], and [20], and is considered beyond the scope of this paper. Nevertheless, the design process implicates that $\log_2 M$ binary information bits are first mapped by the SCMA encoder into a d_c-dimensional constellation points, these constellation points are then zero-padded to spread over N_c codebooks. In this paper, the conventional user-subcarrier (or codebook-subcarrier) association matrix is followed, where this sparse association matrix is also referred to as the factor graph matrix $\mathbf{F}$ [22]

$$\begin{array}{cc} & \begin{array}{cccccc} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 \end{array} \\ \mathbf{F} = & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} . \end{array}$$

Lemma 1: To scale the SCMA system without violating the d_s and d_c constraints when using larger number of subcarriers, the factor graph matrix $\mathbf{F}$ can be used as a block in a diagonal matrix. For instance, the conventional factor graph $\mathbf{F}$ which is a 4×6 matrix can be expanded to a $4m \times 6m$ matrix as

$$\mathbb{F} = \begin{bmatrix} \mathbf{F} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{F} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{F} \end{bmatrix}$$

where $\mathbf{0}$ is a 4×6 zero matrix and $m \in \mathbb{Z}$ indicates the number of diagonal blocks in $\mathbb{F}$.

Proof: Since $\mathbb{F}_{4m\times 6m}$ is a diagonal matrix, this implies that $\sum_j \mathbb{F}_{i,j} = \mathbf{F}$, $\forall i$, and $\sum_i \mathbb{F}_{i,j} = \mathbf{F}$, $\forall j$. Since $\sum_j \mathbf{F}_{i,j} = d_s$, $\forall i$, and $\sum_i \mathbf{F}_{i,j} = d_c$, $\forall j$, then $\mathbb{F}$ maintains the same properties of $\mathbf{F}$ regarding d_s and d_c. ■

2) SCMA Capacity: The motivation behind using SCMA for IoT systems originates from the scarcity of frequency resources to accommodate the massive numbers of IoT devices. Allowing one subcarrier to be shared by multiple users helps improving device connectivity. The capacity of SCMA system is determined by three factors, namely, the number of subcarriers N_s, the number of users (codebooks) sharing one subcarrier (d_s), and the number of subcarriers per codebook (d_c).

Lemma 2: The total number of obtainable codebooks can be expressed by $N_c = \lfloor N_s(d_s/d_c) \rfloor$.

Proof: Since $N_c \sum_j \mathbf{F}_{i,j} = N_s \sum_i \mathbf{F}_{i,j} \equiv$ total number of ones in $\mathbf{F}$, which can also be expressed as $N_c d_c = N_s d_s$, it implies that $N_c = N_s(d_s/d_c)$. For noninteger values of $N_s(d_s/d_c)$ the latter formula can be expressed as $N_c = \lfloor N_s(d_s/d_c) \rfloor$. ■

3) Detection Complexity: The detection complexity in SCMA receivers increases substantially with increasing d_s [21]. As a consequence, more delay will be incurred especially when IoT devices have relatively low computing capabilities. Since IoT devices, such as sensors, actuators, and wearable body sensors are inherently heterogeneous in their computing capabilities, allowing the same subcarrier to be reused by large number of devices could be impractical for IoT systems. Thus, the detection complexity in IoT receivers using SCMA transmission needs to be investigated. For instance, the conventional MPA, which is a common low-complexity decoding technique for SCMA devices based on iterative propagation of messages between resource and user nodes, imposes exponential increase in complexity when the number of users and the codebook size increase. As introduced in [20], the complexity of the addition and multiplication operations required to decode SCMA signals can be expressed as

$$C_{\text{Add}} = d_s N_s M^{d_s} + l_{\max} d_s \Big(N_s M^{d_s} - N_s M \Big) \tag{6}$$

$$\begin{aligned} C_{\text{Mult}} &= N_s M^{d_s}(d_s + 4) + l_{\max} d_s N_s M^{d_s}(d_s - 1) \\ &\quad + l_{\max} N_u d_c M(d_c - 2) \end{aligned} \tag{7}$$

where M denotes the cardinality of the multidimensional constellation points. $l_{\max}$ is the maximum number of message passing update iterations. Assuming that each IoT device has a particular processing capability CPU_k, then t_k in (5) can be rewritten by adding an extra term related to MPA detection time as follows:

$$t_k = \frac{D_k}{R_k} + \frac{D_k}{c_k^n} + \frac{C_{\text{Add}} + C_{\text{Mult}}}{\text{CPU}_k}. \tag{8}$$

III. Problem Formulation

Abiding by the aim of the work, which is improving data rate provision to satisfy the delay requirements of IoT devices, the problem is formulated as a data rate maximization problem as follows:

$$\begin{aligned} \textbf{P1:}\ &\underset{a_{k,s}^n, p_{k,s}^n}{\text{Maximize:}} \quad R_T \\ &\text{Subject to:}\ C1: \quad \sum_{k=1}^{N_u} a_{k,s}^n = d_s \quad \forall n \in \mathcal{N}, \forall s \in \mathcal{S} \\ &C2: \quad \sum_{s=1}^{N_s} a_{k,s}^n = d_c \quad \forall n \in \mathcal{N}, \forall k \in \mathcal{U} \\ &C3: \quad \sum_{k=1}^{N_u} \sum_{s=1}^{N_s} p_{k,s}^n \le P_{\max}^n \quad \forall n \in \mathcal{N} \\ &C4: \quad p_{k,s}^n > 0 \quad \forall n \in \mathcal{N}, \forall k \in \mathcal{U}, \forall s \in \mathcal{S} \\ &C5: \quad a_{k,s}^n \in \{0, 1\}. \end{aligned} \tag{9}$$

The constraint $C1$ ensures that each subcarrier is allocated to d_s codebooks (users), whereas $C2$ ensures that each user is allocated d_c subcarriers. $C3$ sets the upper limit for transmit power at each base station. $C4$ indicates that each subcarrier associated with a user is allocated a nonzero power, while $C5$ is the binary association variable of user k over subcarrier s. It is worth to mention that the task completion deadline T_k is not considered as a constraint but is used to evaluate the system performance by comparing the number of satisfied users under both SCMA and OFDMA schemes as will be seen in the results section.

The aforementioned problem involves both PA at each base station, and subcarrier allocation that is, dependent on the availability of codebooks among all base stations under consideration. It is notable that **P1** involves real, integer, and binary variables which turn the problem into a mixed-integer nonlinear programming (MINLP) problem that is NP-hard and computationally intractable [1], [14], [23]. However, the problem can be solved with less difficulty when subdivided into two consecutive subproblems: 1) codebook allocation in which every user is allocated a codebook that provides the highest data rate (i.e., highest SINR) considering the combined effect of all subcarriers within that codebook and 2) PA whereby each base station undertakes power optimization for associated users. It should be noted that equal PA is carried out in the first subproblem (codebook allocation) since the subcarrier power is one of the parameters required to calculate the data rate according to (1).

Algorithm 1: Codebook Allocation

```
Define:
C: Set of codebooks;
Initialize:
k ← 1;
Set c_k and R_k to zero ∀k ∈ U;
Codebook allocation:
while k ≤ N_u do
    Find c* satisfying R*_k (highest SINR), ∀k ∈ U;
    c_k ← c*;
    C = C\c_k;
    Update R_k;
    k ← k + 1;
end
```

A. Codebook Allocation

The codebook allocation process has two phases: first, subcarriers are mapped onto N_c codebooks taking into account the d_s and d_c constraints. Second, each codebook is matched with an IoT device aiming at providing the highest SINR to IoT devices on a first-come-first-serve basis as shown in Fig. 2. Algorithm 1 illustrates the codebook allocation mechanism.

Lemma 3: The proposed one-to-one matching mechanism is a set-wise stable matching; that is, all users are guaranteed to be associated with a codebook.

Proof: A matching function is considered stable if two conditions hold true. First, no individual element in both sets prefers being single (i.e., with no peer from the other set). Second, no pair prefers other elements on their current outcome (i.e., each of the pair elements does not prefer the matched element). The proposed codebook-user matching algorithm guarantees stability due to the following.

1) All $(\mathcal{C}, \mathcal{U})$ elements are considered rational; that is to say, no user prefers being without a codebook and vice versa.
2) Since codebooks are allocated based on a first-come-first-serve basis, codebooks always prefer their associated users. On the other hand, users might prefer other subcarriers that are already allocated to other users. Nevertheless, that would not violate the second condition. Moreover, codebooks are considered strongly substitutable meaning that a user requesting an already occupied codebook can be allocated the next best codebook (with the next highest SINR). Accordingly,

the codebook allocation mechanism achieves stability. ■

B. Power Allocation

For a given codebook allocation, **P1** can be reduced to the PA problem that aims to maximize the system data rate (or minimize its negative) under power constraints as follows:

$$\begin{aligned}\textbf{P2:}\ \underset{p^n_{k,s}}{\text{Minimize:}}\quad & -\sum_{n=1}^{N_p}\sum_{k=1}^{N_u}\sum_{s=1}^{N_s} a^n_{k,s}\log_2\left(1+\frac{p^n_{k,s}|h^n_{k,s}|^2}{I^n_{k,s}+N_0}\right)\\ \text{Subject to:}\quad C1:\quad & \sum_{k=1}^{N_u}\sum_{s=1}^{N_s} p^n_{k,s}\le P^n_{\max}\quad \forall n\in\mathcal{N}\\ C2:\quad & p^n_{k,s}\ge 0\quad \forall n\in\mathcal{N}, k\in\mathcal{U}, s\in\mathcal{S}. \quad (10)\end{aligned}$$

Assuming that subcarriers have good channel conditions (i.e., SINR >> 1), then the logarithmic function $\log_2$(1+SINR) can provide an accurate approximation of $\log_2$(1+SINR) in **P2** [24]. As a result, the objective function in P2 is a negative concave (convex) function since logarithmic functions are convex on positive real numbers [25]. Therefore, the optimization problem is convex and can be solved using the Lagrange multipliers method

$$\begin{aligned}L\left(p^n_{k,s},\lambda_n,\nu_n\right) = & -\sum_{n=1}^{N_p}\sum_{k=1}^{N_u}\sum_{s=1}^{N_s} a^n_{k,s}\log_2\left(1+\frac{p^n_{k,s}|h^n_{k,s}|^2}{I^n_{k,s}+N_0}\right)\\ & -\lambda_n p^n_{k,s}+\nu_n\left[\sum_{s=1}^{N_s} p^n_s - P^n_{\max}\right] \quad (11)\end{aligned}$$

where λ_n and ν_n are the optimal Lagrange multipliers related to base station n. Note that the interference is considered as additive white Gaussian noise for simplicity. By finding $(\partial L/\partial p^n_{k,s}) = 0$ and fulfilling the Karush–Kuhn–Tucker (KKT) conditions [25], the optimal PA can be calculated by

$$p^n_{k,s} = a^n_{k,s}\left(\frac{1}{\lambda_n}-\frac{N_0}{|h^n_{k,s}|^2}\right)^+ \quad (12)$$

where $(x)^+ = \max\{0, x\}$ and λ_n satisfies the following power constraint:

$$\sum_{k=1}^{N_u}\sum_{s=1}^{N_s} a^n_{k,s}\left(\frac{1}{\lambda_n}-\frac{N_0}{|h^n_{k,s}|^2}\right)^+ = P^n_{\max} \quad (13)$$

which can be rewritten as

$$\frac{1}{\lambda_n} = \frac{1}{\sum_{k=1}^{N_u}\sum_{s=1}^{N_s} a^n_{k,s}}\left(P^n_{\max}+\sum_{k=1}^{N_u}\sum_{s=1}^{N_s}\frac{a^n_{k,s}N_0}{|h^n_{k,s}|^2}\right). \quad (14)$$

where $(1/\lambda_n)$ represents the power (water) level at base station n. Therefore, the power allocated to each subcarrier is determined by $(1/\lambda_n)$ and the subcarrier's channel gain as seen in (12) where higher power is allocated to subcarriers with higher channel gain to maximize the data rate. It is worth mentioning that implementing water-filling in both OFDMA and SCMA systems is fundamentally the same; however, since SCMA allows subcarriers to be shared by multiple users and within different base stations, users that have a common subcarrier in their codebooks can be allocated different amounts of power over that subcarrier depending on the channel conditions. In other words, one subcarrier can have different amounts of power when assigned to different codebooks, unlike OFDMA where each subcarrier can be allocated to at most one user in a timeslot [15], [23]. Algorithm 2 illustrates the PA process in the SCMA scheme.

Algorithm 2: PA in SCMA

Calculate $\frac{1}{\lambda_n}$ using (14), $\forall n\in\mathcal{N}$;
For each user $k\in\mathcal{U}$ within base station n:
$s\leftarrow 1$;
while $s\le d_c$ **do**
  Calculate $p^n_{k,s}$ using (12);
  $s\leftarrow s+1$;
end

IV. Simulation Setup and Results

In this section, different metrics are used to investigate the SCMA performance for edge IoT computing in terms of data rate, connectivity, detection complexity, and computing performance. To ease tracking simulation parameters, Table II presents the list of variables and corresponding values used in simulations.

A. Investigating System Performance Using Different SCMA Settings

The advantage of SCMA over OFDMA stems from the capability of SCMA to accommodate more IoT devices in order to enhance the system capacity. On the one hand, allowing more users to share the same subcarrier (increasing d_s) helps to accommodate more users at the expense of higher incurred interference. On the other hand, allocating more subcarriers to users (increasing d_c) leads to higher data rate provisioning for users but degrades the system connectivity. Hence, choosing d_s and d_c in SCMA depends on the particular demands of IoT systems. Figs. 3 and 4 demonstrate the effect of d_s and d_c on the number of obtainable codebooks and sum data rate, respectively. The number of obtainable codebooks in Fig. 3 exhibits an increasing trend with both increasing d_s and decreasing d_c, and with 16 subcarriers available in the system, up to 48 codebooks can be obtained. Fig. 4 shows that increasing d_s leads to better system throughput; however, the incurred interference could deteriorate the system performance when d_c is not carefully chosen. One downside of SCMA is the interference encountering IoT devices, which leads to large throughput gaps among devices due to the variable channel conditions. It is thus motivating to statistically investigate the per-user throughput difference in SCMA systems. To this end, the standard deviation of the per-user rate can be calculated as follows:

$$\sigma = \sqrt{\frac{1}{N_u}\sum_{k=1}^{N_u}|R_k-\bar{R}|^2} \quad (15)$$

TABLE II
SIMULATION PARAMETERS

Description	Value
Bandwidth (B)	10 MHz
Number of subcarriers (N_s)	6
Maximum number of users per subcarrier (d_s)	3
Maximum number of subcarriers per user (d_c)	2
Cardinality of the constellation points (M)	4
Processing speed at each fog node (C)	10 GHz
Processing speed of IoT devices (CPU_k)	Uniform distribution $[20 - 1000]$ MHz
Data size of user k (D_k)	Uniform distribution $[2 - 8]$ Mb
Coverage of macro BS	1 km
Coverage of small BS	0.1 km
Maximum transmit power (macro BS)	40 W
Maximum transmit power (small BS)	1 W
Path loss (macro BS)	$131.1 + 42.8\ \log_{10}(D)$ dB, D in km
Path loss (small BS)	$145.4 + 37.5\ \log_{10}(D)$ dB, D in km
Shadowing standard deviation (macro BS)	10 dB
Shadowing standard deviation (small BS)	6 dB
Multipath fading (both macro and small BS)	Rayleigh distribution with unit variance
Noise power spectral density (N_0)	-173 dBm/Hz

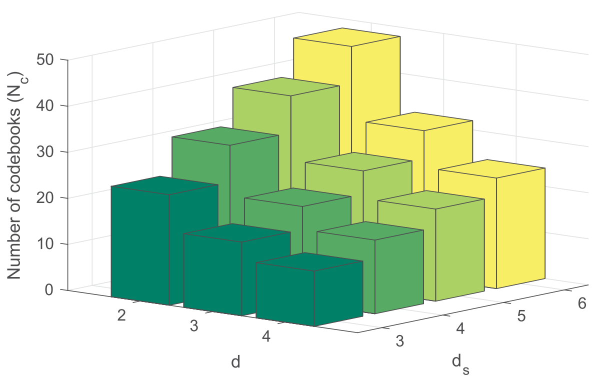

Fig. 3. Obtainable codebooks using different values of d_s and d_c.

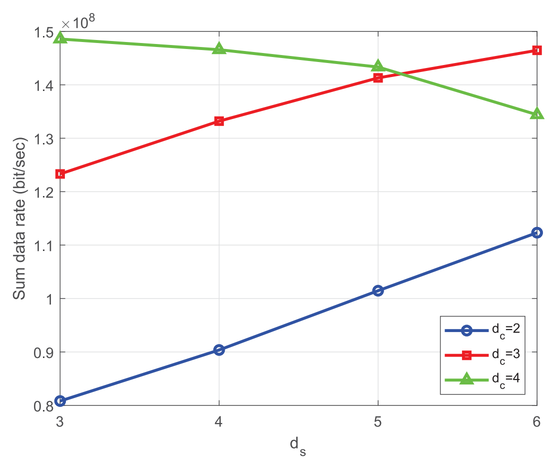

Fig. 4. Effect of d_s and d_c on the sum data rate.

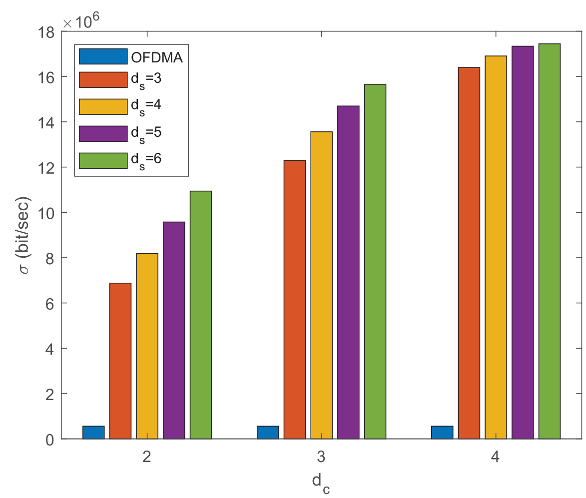

Fig. 5. Effect of d_s and d_c on the standard deviation of data rates experienced by users.

where $\bar{R}$ is the mean data rate of all $k \in \mathcal{U}$. As seen in Fig. 5, the variations in per-user throughput show remarkable increase when the values of d_s and d_c increase.

To support the massive connectivity of IoT devices, the ratio (N_c/N_s) (also referred to as the overloading factor) is desired to be much greater than one [26]. The latter condition can be satisfied by increasing d_s (refer to Fig. 3) which also leads to higher data rate provisioning as depicted in Fig. 4. However, the interference incurred by increasing d_s can also lead to higher variations among users in regard with the experienced data rate as shown in Fig. 5. On the other hand, increasing d_c has a negative impact on IoT connectivity; that is to say, less codebooks will be obtained when d_c is increased. Nevertheless, increasing d_c still shows an increase in the sum data rate since more subcarriers (higher data rates) are allocated to fewer number of users. Furthermore, since increasing d_c allows users to utilize more subcarriers, more interference will be encountered especially at high values of d_s, and that bounds the increase in the sum data rate and incurs more variations in data rate provisioning as seen in Figs. 4 and 5, respectively.

Since SCMA allows each individual user to transmit/receive over multiple (d_c) subcarriers simultaneously, and allow the same subcarrier to be reused by d_s users, the sum throughput obtained shows an obvious increase compared to OFDMA.

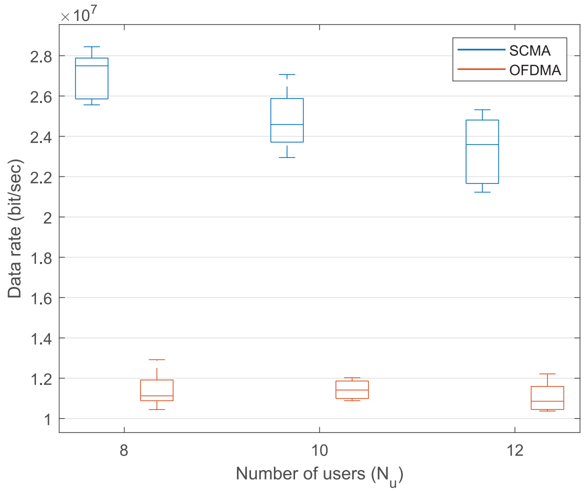

Fig. 6. Comparison of the per-user data rate between SCMA and OFDMA systems at $N_u = 8, 10,$ and 12.

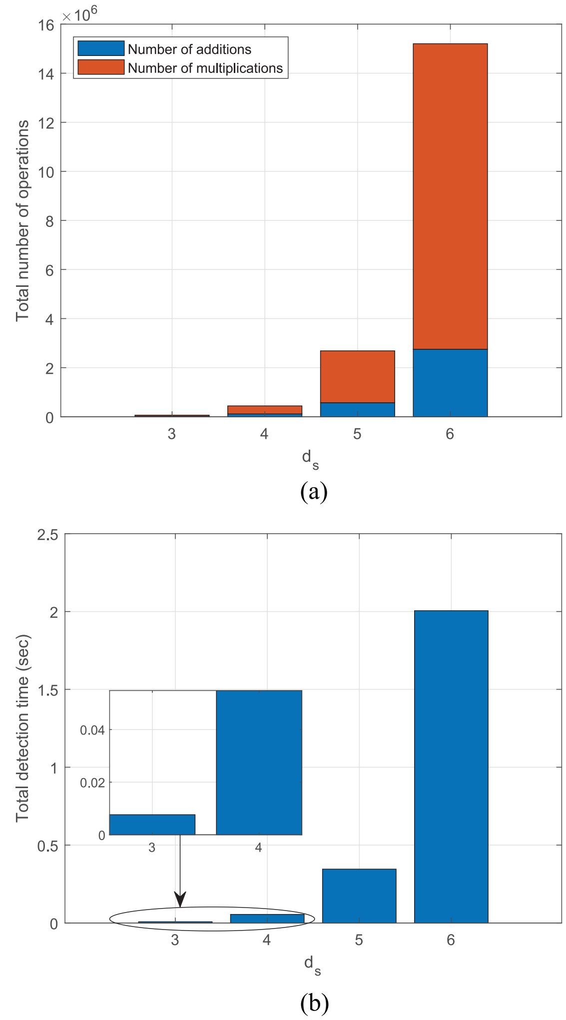

Fig. 7. Detection complexity experienced by SCMA receivers with $d_c = 2$. (a) Total operations required for SCMA detection. (b) Time required by an IoT device with 20 MHz processing capacity.

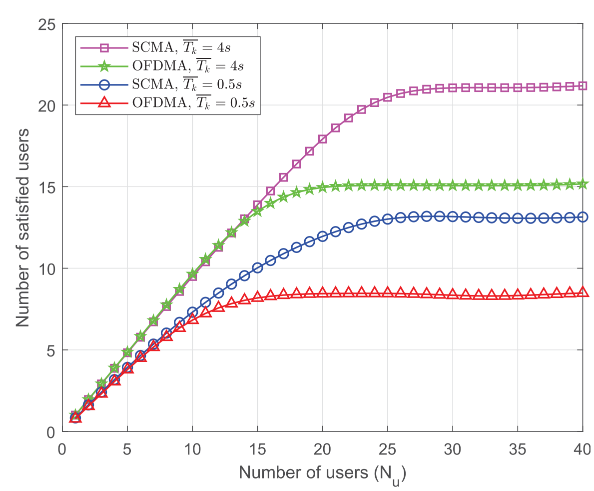

Fig. 8. Computing performance comparison between SCMA and OFDMA under different deadline requirements.

Nevertheless, the per-user data rate provision shows a considerable variation among SCMA users due to interference especially when the number of users increases. Fig. 6 demonstrates that feature of SCMA, where the central line indicates the median data rate, the box edges indicate the 25th and 75th percentiles, while the bottom- and top-most lines indicate the extreme data rate provided to IoT devices. From this figure, it can be seen that the per-user data rate of SCMA is higher than that of OFDMA, but declines with increasing number of users due to the incurred interference. In contrast, OFDMA users experience lower data rates that are relatively fixed with respect to number of users due to channel orthogonality. In addition to their effects on connectivity and throughput, d_s and d_c have significant impact on the detection complexity at the receiver side. From (6) and (7), the total number of operations required to achieve signal detection is depicted in Fig. 7(a), whereas Fig. 7(b) shows the time required to execute these operations by an IoT device with a 20-MHz processor assuming that each multiplication operation requires three clock cycles for processing, and each addition requires one cycle, while the total number of users in the system (N_u) is 12. It is worth mentioning that the effect of d_c on detection complexity is minor according to (6) and (7) compared to d_s which acts as an exponent and thus has significant impact. For this reason, Fig. 7 presents results using different values of d_s while d_c is considered constant.

B. Proposed SCMA System Performance

To quantify the performance gain of SCMA over OFDMA, we consider an SCMA system with $d_s = 3$ and $d_c = 2$ in order to improve connectivity and to satisfy more IoT devices while maintaining the detection complexity within satisfactory limits. The improved system connectivity of SCMA is evident compared to that of OFDMA as manifested by the number of satisfied users in Fig. 8 where users are assumed to have different task completion deadlines that follow a uniform distribution with mean $\overline{T_k}$. It can be noticed that the number of satisfied users saturates when the connectivity limit (i.e., 24 users in SCMA and 16 users in OFDMA) is reached; however, enforcing more stringent delay requirements by IoT devices

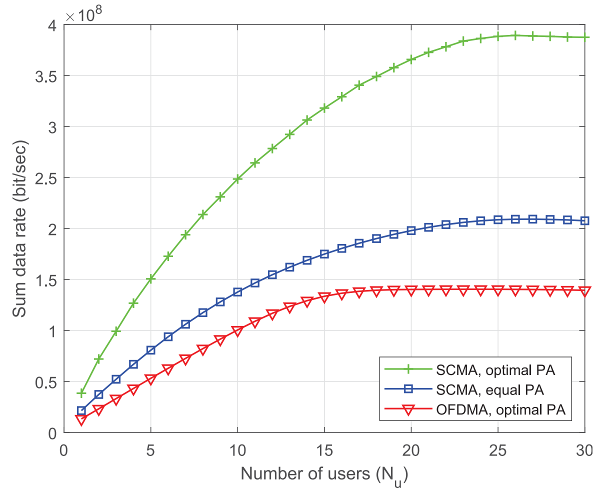

Fig. 9. Sum data rate using different schemes.

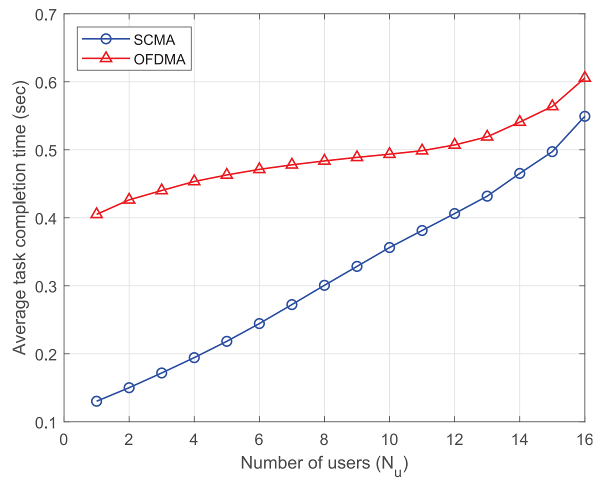

Fig. 10. Average task completion time for IoT devices.

results in less satisfied users since the task completion time will exceed the task completion deadline more easily.

Adopting optimized PA techniques such as water-filling leads to significant performance improvement compared to other strategies as shown in Fig. 9. It can also be observed that the sum data rate shows a continuous increase in SCMA until reaching the maximum connectivity limit which is 24 users at about 3.8 $\times 10^8$ bits/s, and that surpasses OFDMA which saturates at 1.4 $\times 10^8$ bits/s when the maximum connectivity limit is attained at 16 users.

A comparison between SCMA- and OFDMA-based edge-IoT computing is presented in Fig. 10, where SCMA shows a clear advantage over OFDMA regarding the task completion time. This is due to the higher data rate offered by SCMA which reduces the transmission delay of computing tasks. However, when the number of users increases, the advantage of SCMA declines due to the encountered interference. In addition, the competition among users on the limited computing resources of fog nodes tends to increase when more users coexist in the system; as a result, the task completion time of both SCMA and OFDMA schemes increases with the number of users.

V. Conclusion

In this paper, an SCMA-based edge computing scheme for IoT systems was proposed. Different SCMA parameters have been investigated to showcase the applicability of SCMA for IoT systems in comparison with traditional OFDMA-based schemes. The effects of these SCMA parameters, namely, the number of subcarriers allocated to one user and the number of users sharing the same subcarrier have been extensively studied and their effects on the system performance have been presented in detail. An optimization problem was also formulated to maximize system throughput under the power constraint and solved using the water filling approach. The results show the significance of implementing SCMA in improving network connectivity and maximizing data rate provision for better QoE performance in IoT systems.

References

[1] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1299–1306, Apr. 2018.

[2] M. Kim, N.-I. Kim, W. Lee, and D.-H. Cho, "Deep learning-aided SCMA," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 720–723, Apr. 2018.

[3] S. Zhang *et al.*, "Sparse code multiple access: An energy efficient uplink approach for 5G wireless systems," in *Proc. IEEE Glob. Commun. Conf.*, Austin, TX, USA, 2014, pp. 4782–4787.

[4] D. Zhai, "Adaptive codebook design and assignment for energy saving in SCMA networks," *IEEE Access*, vol. 5, pp. 23550–23562, 2017.

[5] J. Chen, Z. Wang, W. Xiang, and S. Chen, "Outage probability region and optimal power allocation for uplink SCMA systems," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4965–4980, Oct. 2018.

[6] Y.-Y. Shih, W.-H. Chung, A.-C. Pang, T.-C. Chiu, and H.-Y. Wei, "Enabling low-latency applications in fog-radio access networks," *IEEE Netw.*, vol. 31, no. 1, pp. 52–58, Jan./Feb. 2017.

[7] T. Adegbija, A. Rogacs, C. Patel, and A. Gordon-Ross, "Microprocessor optimizations for the Internet of Things: A survey," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 1, pp. 7–20, Jan. 2018.

[8] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.

[9] E. Zeydan *et al.*, "Big data caching for networking: Moving from cloud to edge," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 36–42, Sep. 2016.

[10] S.-C. Hung, H. Hsu, S.-Y. Lien, and K.-C. Chen, "Architecture harmonization between cloud radio access networks and fog networks," *IEEE Access*, vol. 3, pp. 3019–3034, 2015.

[11] X. Sun and N. Ansari, "EdgeIoT: Mobile edge computing for the Internet of Things," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 22–29, Dec. 2016.

[12] A. Alnoman, G. H. S. Carvalho, A. Anpalagan, and I. Woungang, "Energy efficiency on fully cloudified mobile networks: Survey, challenges, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1271–1291, 2nd Quart. 2018.

[13] Q. Fan and N. Ansari, "Application aware workload allocation for edge computing based IoT," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2146–2153, Jun. 2018.

[14] Y. Gu, Z. Chang, M. Pan, L. Song, and Z. Han, "Joint radio and computational resource allocation in IoT fog computing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7475–7484, Aug. 2018.

[15] W. Zhu, L. Qiu, and Z. Chen, "Joint subcarrier assignment and power allocation in downlink SCMA systems," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–5.

[16] Y. Li *et al.*, "Cost-efficient codebook assignment and power allocation for energy efficiency maximization in SCMA networks," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.

[17] S. Li, Q. Ni, Y. Sun, G. Min, and S. Al-Rubaye, "Energy-efficient resource allocation for industrial cyber-physical IoT systems in 5G era," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2618–2628, Jun. 2018.

[18] H. Zhang *et al.*, "Computing resource allocation in three-tier IoT fog networks: A joint optimization approach combining Stackelberg game and matching," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1204–1215, Oct. 2017.

[19] M. Moltafet, N. M. Yamchi, M. R. Javan, and P. Azmi, "Comparison study between PD-NOMA and SCMA," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1830–1834, Feb. 2018.

[20] L. Yang, X. Ma, and Y. Siu, "Low complexity MPA detector based on sphere decoding for SCMA," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1855–1858, Aug. 2017.

[21] Y. Han, W. Zhou, M. Zhao, and S. Zhou, "Enabling high order SCMA systems in downlink scenarios with a serial coding scheme," *IEEE Access*, vol. 6, pp. 33796–33809, 2018.

[22] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, London, U.K., Sep. 2013, pp. 332–336.

[23] X. Sun and S. Wang, "Resource allocation scheme for energy saving in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4407–4416, Aug. 2015.

[24] G. Tychogiorgos, A. Gkelias, and K. K. Leung, "A non-convex distributed optimization framework and its application to wireless ad-hoc networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4286–4296, Sep. 2013.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[26] K. Au *et al.*, "Uplink contention based SCMA for 5G radio access," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, Dec. 2014, pp. 900–905.

Ali Alnoman received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Baghdad, Baghdad, Iraq, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada.

From 2012 to 2014, he was a faculty member with Ishik University, Erbil, Iraq. His current research interests include energy efficiency and resource allocation in HetNets, Internet of Things, and cloud computing.

Mr. Alnoman also served as a Technical Program Committee member of the IEEE Vehicular Technology Conference VTC2017-Fall in Toronto.

Serhat Erkucuk (S'99–M'08–SM'19) received the B.Sc. degree in electrical engineering from Middle East Technical University, Ankara, Turkey, in 2001, the M.Sc. degree in electrical engineering from Ryerson University, Toronto, ON, Canada, in 2003, and the Ph.D. degree in engineering science from Simon Fraser University, Burnaby, BC, Canada, in 2007.

He was an NSERC Post-Doctoral Fellow with the University of British Columbia, Vancouver, BC, Canada, in 2008. He then joined Kadir Has University, Istanbul, Turkey, where he is currently an Associate Professor. In 2018, he was a Visiting Professor with Ryerson University, where he conducted research on the design of small cells for 5G networks. His current research interests include physical layer design of emerging communication systems, wireless sensor networks, and communication theory.

Dr. Erkucuk was a recipient of the Governor's General Gold Medal. He is a Marie Curie Fellow.

Alagan Anpalagan (S'98–M'01–SM'04) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada.

He joined the Department of Electrical and Computer Engineering, Ryerson University, Toronto, in 2001, where he became a Full Professor in 2010, and served as the Graduate Program Director from 2004 to 2009 and the Interim Electrical Engineering Program Director from 2009 to 2010. During his sabbatical, he was a Visiting Professor with the Asian Institute of Technology, Khlong Nueng, Thailand, from 2010 to 2011 and a Visiting Researcher with Kyoto University, Kyoto, Japan. His industrial experience includes working for three years with Bell Mobility, Montreal, QC, Canada; Nortel Networks, Ottawa, ON, Canada; and IBM, Armonk, NY, USA. He directs a research group involved with radio resource management and radio access and networking areas within the WINCORE Laboratory. He also completed a course on project management for scientist and engineers at the University of Oxford CPD Center, Oxford, U.K. He coauthored *Design and Deployment of Small Cell Networks* (Cambridge Univ. Press, 2016), *Routing in Opportunistic Networks* (Springer, 2013), *Handbook on Green Information and Communication Systems* (Academic, 2012) and *Game-Theoretic Interference Coordination Approaches for Dynamic Spectrum Access* (Springer, 2016). His current research interests include 5G wireless systems, energy harvesting and green communications technologies, cognitive radio resource management, wireless cross layer design and optimization, cooperative communication, M2M and sensor communication, small cell, and heterogeneous networks.

Dr. Anpalagan was a recipient of the Deans Teaching Award in 2011, the Faculty Scholastic, Research, and Creativity Award in 2010 and 2014, the Faculty Service Award from Ryerson University in 2011 and 2013, the IEEE M. B. Broughton Central Canada Service Award in 2016, the Exemplary Editor Award from IEEE ComSoc in 2013, the Editor-in-Chief Top10 Choice Award in the *Transactions on Emerging Telecommunications Technology* in 2012, and the IEEE SP-S Young Author Best Paper Award in 2015 for his coauthored paper. He served as an Editor for IEEE COMMUNICATIONS SURVEYS AND TUTORIALS from 2012 to 2014, IEEE COMMUNICATIONS LETTERS from 2010 to 2013, and the *EURASIP Journal of Wireless Communications and Networking* from 2004 to 2009. He also served as a Guest Editor for six special issues: IEEE WIRELESS COMMUNICATIONS: "Sustainable Green Networking and Computing in 5G Systems," IEEE ACCESS on "Internet of Things in 5G Systems, *IET Communications* on Evolution and Development of 5G Wireless Communication Systems," *EURASIP* on "Radio Resource Management in 3G+ Systems and on Fairness in Radio Resource Management for Wireless Networks," and *Mobile Networks and Applications* on "Green Cognitive and Cooperative Communication and Networking." He served as the TPC Co-Chair of IEEE VTC Fall 2017, IEEE INFOCOM'16: First International Workshop on Green and Sustainable Networking and Computing, IEEE Globecom15: SAC Green Communication and Computing, and IEEE PIMRC11: Cognitive Radio and Spectrum Management. He served as the Vice Chair of IEEE SIG on Green and Sustainable Networking and Computing with Cognition and Cooperation from 2015 to 2016, the IEEE Canada Central Area Chair from 2012 to 2014, the IEEE Toronto Section Chair from 2006 to 2007, the ComSoc Toronto Chapter Chair from 2004 to 2005, and the IEEE Canada Professional Activities Committee Chair from 2009 to 2011. He is a Registered Professional Engineer in the Province of Ontario, Canada, and a Fellow of the Institution of Engineering and Technology.