Random-Access NOMA in URLL Energy-Harvesting IoT Networks with Short Packet and Diversity Transmissions

Article *in* IEEE Access · December 2020

DOI: 10.1109/ACCESS.2020.3042744

CITATIONS
5

READS
48

2 authors, including:

Mohammed W. Baidas
Kuwait University
87 PUBLICATIONS **343** CITATIONS

SEE PROFILE

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Random-Access NOMA in URLL Energy-Harvesting IoT Networks with Short Packet and Diversity Transmissions

MOHAMMAD REZA AMINI[1], *Senior Member, IEEE*, and MOHAMMED W. BAIDAS[2], *Senior Member, IEEE*

[1]Department of Electrical Engineering, Borujerd Branch, Islamic Azad University, Borujerd, Iran (email: mr.amini@iaub.ac.ir)
[2]Department of Electrical Engineering, College of Engineering and Petroleum, Kuwait University, Kuwait City 13060, Kuwait (e-mail: m.baidas@ku.edu.kw)

Corresponding author: Mohammed W. Baidas (e-mail: m.baidas@ku.edu.kw).

This work was partially supported by the Kuwait Foundation for the Advancement of Sciences (KFAS), under project code PN17-15EE-02.

ABSTRACT Non-orthogonal multiple-access (NOMA) has recently been proposed to improve throughput and spectrum-efficiency of 5G cellular networks and beyond. It is also a key enabler for ultra-reliable and low-latency (URLL) communications. Moreover, the Internet-of-Things (IoT) paradigm has emerged to practically provide massive connectivity for smart devices and systems, which entails spectrum- and energy-efficient transmission schemes. To this aim, NOMA and energy-harvesting (EH) solutions have been put forth to address such demands, making the combination of such technologies inevitable in NOMA-based URLL-EH-IoT networks. On the other hand, random-access (RA) techniques are the key solution to enable massive URLL-EH-IoT networks, since they reduce signaling overhead, packet latency, and energy consumption when massive numbers of clustered IoT devices with sporadic traffic behavior are considered. In this paper, uplink RA-NOMA in URLL-EH-IoT networks with short packet and diversity transmissions is studied and analyzed. Network metrics—such as average packet latency, reliability, and GoodPut—are derived for the RA-NOMA scenario and compared with its RA-OMA counterpart to explore the advantages of RA-NOMA over RA-OMA. Moreover, the analytical derivations have been validated and shown to coincide with the simulation results. Furthermore, the effect of the transmission diversity and number of data bits per blocklength on the different network metrics are extensively evaluated. Lastly, the analytical derivations are utilized to find the optimum values of the IoT nodes' transmit power, number of packet replicas, and number of transmitted data bits per blocklength, such that the maximum sum-GoodPut of a RA-NOMA IoT cluster is achieved, subject to URLL constraints.

INDEX TERMS Energy-harvesting, Internet-of-Things, low-latency, NOMA, random-access, ultra-reliability.

I. INTRODUCTION

The explosive and exponential growth in the number of smart mobile devices, applications, and services in the current 5G cellular and emerging Internet-of-Things (IoT) networks calls for intelligent spectrum- and energy-efficient transmission techniques to meet the demands for massive connectivity, high data rates, and low-latency [1], [2]. The advent of non-orthogonal multiple access (NOMA) has made spectrum usage more efficient by allowing multiple users to share the same resource blocks (RBs) simultaneously, while employing successive interference cancellation (SIC) for multi-user detection [3]. Particularly, NOMA has been shown to be superior to the conventional OMA schemes in terms of capacity and transmission latency [4]. Thus, NOMA has been proposed for new emerging applications, such as LiFi access-points [5], device-to-device communications [6], wireless power transfer [7], mobile edge computing [8], and many others [9]. In turn, incorporating NOMA into IoT networks is envisioned to be a promising solution for high spectral-efficiency and massive connectivity requirements [10]. Energy-harvesting (EH) techniques have also been proposed to scavenge environmental energy, and thus prolong the lifetime of wireless devices and systems [11], [12]. On the other hand, ultra-reliable and low-latency (URLL) transmissions are considered to be the main features that distinguish 5G from previous generations [13], [14],

and serve as foundations for many 5G and IoT applications, such as tactile Internet, factory automation, Industrial IoT (IIoT), and those under the Industry 4.0 paradigm [15]–[18]. To achieve URLL communication, several techniques can be adopted, such as diversity transmission, which can be used to enhance transmission reliability by sending multiple replicas of a data packet. For low-latency, short packet transmissions via finite blocklength (FBL) codes can be utilized [19]. The combination of the aforementioned transmission techniques is extremely attractive to realize URLL transmissions in NOMA-based EH-IoT networks, which are the focus of this paper. In general, IoT networks incorporate massive numbers of smart devices and systems, which are hungry for spectrum and energy resources. Moreover, such devices and systems are characterized by sporadic traffic behavior and spectrum access, which motivates the use of random-access (RA) techniques. However, the conventional RA techniques (e.g. the current RA-LTE standard) cannot used for low-latency transmissions due to the excessive signaling overhead [10]. Hence, proposing and analyzing RA-NOMA for EH-IoT networks with URLL requirements is of the essence.

A. RELATED WORKS

Several research works have recently considered NOMA-based communications. Despite the importance of URLL communications, only few studies have focussed on the analysis and characterization of NOMA-based transmission with URLL requirements. In general, NOMA transmission can be classified as grant-based (GB) and grant-free (GF) access schemes. As opposed to the conventional GB access schemes, GF access has been identified as a key medium access control technique for providing massive connectivity, minimizing signaling overhead, and reducing latency, especially in URLL IoT networks [20], [21].

1) GB-NOMA Access

The packet error probability in uplink (UL) NOMA networks with URLL requirements has been addressed in [22]. Particularly, a hybrid automatic repeat request (HARQ) mechanism has been augmented with NOMA, namely NOMA-HARQ. Moreover, the number of re-transmissions and transmit power for each user have been optimized to meet the reliability constraint. The proposed NOMA-HARQ has been shown to be superior to the classical HARQ mechanism as well as OMA. Performance analysis of a two-users downlink (DL) NOMA system with finite blocklength codes is provided in [23]. Specifically, transmission error and queuing-delay probabilities have been derived based on the effective-bandwidth analysis, and shown to improve the reliability for time-critical applications. In [24], cross-layer optimization for delay-optimal UL NOMA transmission in delay-sensitive applications has been presented. Using a Constraint Markov Decision Process (CMPD), the authors devised a framework based on a linear program, where the long-term transmission delay is minimized under power consumption and reliability constraints. Also, the optimal joint scheduling and super-position coding policy has been derived, and illustrated to minimize transmission latency in practical NOMA systems. In [25], a simple two-users time-slotted UL NOMA scenario has been studied, with the aim of characterizing the stable throughput region and average delay based on two types of SIC receivers. Particularly, the authors consider the effect of unsaturated traffic, which gives rise to discontinuous interference. It is has been demonstrated that NOMA is superior to OMA, due to the larger stable throughput region, and the lower average delay. The performance of DL NOMA in short-packet communications has been studied in [26]. In particular, closed-form expressions for the block error rates of a two-users NOMA network have been derived, while demonstrating the superior performance of NOMA over OMA in reducing transmission latency. Closed-form expressions for the rate and outage probability in an UL NOMA scenario with two users have been derived in [27] for the Gaussian and Rayleigh fading channels. Furthermore, the authors investigated the trade-off between reliability and throughput in the ARQ regime, and demonstrated that NOMA is always superior to OMA in terms of throughput, reliability, and latency.

2) GF-NOMA Access

The authors in [28] propose an uncoordinated NOMA scheme for machine-to-machine (M2M) communications, where the devices have strict latencies and no retransmission opportunities. In particular, devices randomly choose pilot sequences from a predefined set, and use them to transmit their data simultaneously. The average system throughput under joint decoding and massive access over Rayleigh fading channels is derived, where the uncoordinated NOMA scheme has been shown to be primarily dominated by the collision probability, ultimately requiring a collision resolution scheme. An novel framework for massive grant-free NOMA is proposed in [29]. To be specific, this work extends the NOMA scheme in [28] and derives simplified expressions for the outage probability and throughput of the system for both successive joint decoding (SJD) and SIC. It has been shown that for low data-rate transmissions, the outage probability performance of the SIC is close to the SJD; however, SJD has been shown to achieve almost double the throughput of SIC. In [30], a novel UL NOMA scheme—based on orthogonal frequency division multiplexing (OFDM), and OFDM with index modulation (OFDM-IM)—is proposed to achieve URLL communications. Particularly, the OFDM-IM is utilized for latency-tolerant applications, since wideband interference can be converted into narrowband interference by fractional subcarrier activation. It has been shown that the proposed NOMA scheme based on OFDM-IM can significantly reduce latency in comparison to the classical NOMA based on pure OFDM.

Another line of work considered NOMA-based RA for the analysis of various network metrics. For example, random-access with multichannel ALOHA in UL NOMA systems is considered in [31]. Specifically, the proposed NOMA scheme

uses a set of predefined power levels for multiple-access, which improves the throughput of multichannel ALOHA without any bandwidth expansion. This is particularly attractive for machine-type communications (MTC), which are bandwidth-limited. A closed-form expression for a lower-bounded throughput has been derived, and the proposed scheme has been shown to be superior to the classical multichannel ALOHA. Additionally, channel-dependent selection for subchannel and power level has also been proposed to reduce the transmit power. S-ALOHA (SA) with UL NOMA has also been proposed in [32], where two detection techniques have been proposed to mitigate interference, namely SIC with optimal decoding order, and joint decoding (JD). The outage probability and average throughput have been derived, and the SA-NOMA-JD has been shown to outperform SA-NOMA-SIC. Based on NOMA, a non-orthogonal random-access (NORA) scheme has been proposed in [33] to alleviate the access congestion problem. Particularly, NORA is based on the difference of time of arrival to identify multiple user equipments (UEs) with identical preamble, and enables power-domain multiplexing of collided UEs. Moreover, the base-station (BS) applies SIC based on the channel conditions obtained through preamble detection. The proposed scheme has been shown to outperform the conventional orthogonal random-access (ORA) scheme in terms of preamble collision probability, access success probability, and throughput. The authors in [34] proposed UL NORA techniques for 5G mobile communication networks. To be specific, by utilizing the channel inversion technique, the UEs can adjust their transmit power such that their received power at the BS is one of two target values. This enables the BS to decode two packets simultaneously via SIC. The proposed NORA techniques have been analyzed in terms of access delay, throughput and energy-efficiency, where it has been shown that the maximum achievable throughput can exceed 0.7, in comparison to 0.368 of the conventional S-ALOHA. A throughput-oriented NORA scheme is proposed in [35] for massive MTC (mMTC) networks. In particular, tagged preambles are employed by multiple MTC devices (MTCDs), forming power-domain NOMA groups. After that, an optimization problem is formulated for throughput maximization, subject to constraints on the power back-off factor, number of MTCDs, and successful transmission probability. A particle swarm optimization (PSO) algorithm is devised, along with a low-complexity suboptimal algorithm. The proposed scheme has been shown to improve the network throughput performance in comparison to existing schemes.

B. MOTIVATION AND CONTRIBUTIONS

Most of the current 5G communication protocols and access schemes are grant-based, with RA strategies being mostly used for setting up a connection[1]. Moreover, most of the published works on NOMA-based random-access networks do not consider EH for sustainable and green IoT networks. Additionally, the massive numbers of IoT nodes are characterized with sporadic traffic patterns; however, most studies ignore the data- and energy-causality aspects in the analysis and design of EH NOMA-based URLL communications. More importantly, the interplay and trade-offs between critical network performance metrics are often overlooked.

Since various IoT applications require URLL transmissions along with the ever-increasing sustainability demands, an URLL-EH-IoT network with UL RA-NOMA transmissions and clustered IoT nodes is considered in this work. Specifically, the IoT nodes harvest environmental energy to cater for their data transmissions, and the IoT NOMA clusters randomly select RBs for their data transmissions. To achieve the target ultra-reliability and realize low-latency, short packet and diversity transmissions are adopted. Moreover, network metrics such as average packet latency, reliability, and GoodPut are mathematically derived, while accounting for IoT nodes' sporadic traffic behavior and energy arrival rate. In deriving packet latency, transmission delay and queue waiting time are considered as the two crucial delay components. Two kinds of impairments are considered in deriving the reliability, namely inter-cluster collisions occurring between the IoT nodes, and decoding errors. Also, the effect of transmission diversity, number of transmitted data bits per blocklength, and nodes' transmit powers on the RA-NOMA network performance is investigated and compared to its RA-OMA counterpart. The main contributions of this work are summarized as follows:

- Analyzed the performance of UL NOMA-based URLL-EH-IoT networks for the first time. Specifically, an URLL-IoT network with clustered energy-harvesting IoT nodes utilizing RA-NOMA transmissions is explored, with URLL transmissions achieved via diversity transmission in the FBL regime.
- Derived analytical expressions for critical IoT network performance metrics (i.e. average packet latency, reliability, and GoodPut), where energy- and data-causality are considered simultaneously in deriving the network metrics. The analytical derivations have been numerically validated and shown to coincide with the simulation results.
- Derived the equivalent RA-OMA network metrics, and highlighted the advantages of RA-NOMA over RA-OMA. The trade-offs between the different network metrics have also been extensively evaluated.
- Utilized the analytical derivations to maximize the sum-GoodPut of a RA-NOMA cluster, subject to constraints on reliability, average packet latency, SIC decoding order, and transmit power. Specifically, IoT nodes' transmit power, number of data bits, and number of replicas for each data packet are optimized such that sum-GoodPut is maximized under URLL requirements.

To the best of our knowledge, no prior work has analyzed UL

[1]Establishing a connection requires about 30-50% of the payload size in the FBL regime [36]. Additionally, the grant-based access procedure in LTE-A takes around 5-8 ms, excluding the time required for data transmission, and assuming error- and collision-free signaling [36].

RA-NOMA URLL-EH-IoT networks, and provided analytical expressions and extensive comparative results for various IoT network performance metrics.

The rest of this paper is organized as follows. Section II introduces the system model. The analytical derivations of the different IoT network metrics are given in Section III, while the numerical results are presented in Section IV. In Section V, the sum-GoodPut maximization problem is formulated and solved. A summary of the findings is provided in Section VI. Finally, conclusions are drawn in Section VII.

II. SYSTEM MODEL

The studied system model is based on a single-cell UL IoT network consisting of a BS, and N EH transmitting nodes with URLL transmission requirements[2]. Also, the IoT nodes transmit their sporadic data packets to the BS in the FBL regime via RA-NOMA. The IoT nodes scavenge environmental energy (e.g. solar and/or wind), and store it in a finite-capacity rechargeable battery for data communication purposes. The IoT nodes are paired in clusters of two nodes, and hence, there are $M = N/2$ clusters in the IoT network[3]. Additionally, there are R orthogonal RBs, each of bandwidth W, which are utilized by the nodes within each cluster for frame-based RA-NOMA transmission. Let $U_{i,m}$ denote IoT node $i \in \{1, 2\}$ in the m^{th} cluster, for $m \in \{1, \ldots, M\}$. Furthermore, $U_{i,m}$ transmits its data packets over the selected RB with transmit power of $P_{i,m}$, such that $P_{i,m} \leq P_{\max}$, $\forall i \in \{1, 2\}$, $\forall m \in \{1, \cdots, M\}$, where $P_{\max}$ is the maximum transmit power per IoT node.

There are N communication links, which originate from the N nodes to the BS. All the links in the network experience independent but not necessarily identically distributed Rayleigh block fading. Moreover, all the links are assumed to be constant during each transmission frame, which is due to the short packet transmission. The channel coefficient between $U_{i,m}$ and the BS is denoted $h_{i,m}$. Therefore, the corresponding channel gain $|h_{i,m}|^2$ follows an exponential distribution with mean $d_{i,m}^{-\nu}$, where $d_{i,m}$ is the distance between the IoT node $U_{i,m}$ and the BS, while ν is the path-loss exponent. Furthermore, The background noise in all links is assumed to be independent and identically distributed (i.i.d.) zero-mean additive white Gaussian noise with variance $\sigma^2 = WN_0$, where N_0 is the noise spectral density. For notational convenience, let $U_{1,m}$ and $U_{2,m}$ be the nodes with the higher and lower channel gains over the selected RB, respectively. That is, $|h_{1,m}|^2 > |h_{2,m}|^2$, and hence $P_{1,m} \geq P_{2,m}$ [43]. In turn, nodes $U_{1,m}$ and $U_{2,m}$ are referred to as the stronger and weaker nodes, respectively.

Remark 1: According to the principle of UL NOMA, the signal of node $U_{1,m}$ must be decoded first at the BS, and thus experiences interference from $U_{2,m}$. In turn, node $U_{2,m}$ enjoys interference-free transmission upon applying SIC [44]. Thus, assuming perfect SIC, the received signal-to-interference-plus-noise ratio (SINR) of node $U_{1,m}$ at the BS is given by

$$\gamma_{1,m} = \frac{P_{1,m}|h_{1,m}|^2}{P_{2,m}|h_{2,m}|^2 + \sigma^2}, \tag{1}$$

while that for node $U_{2,m}$ is obtained as

$$\gamma_{2,m} = \frac{P_{2,m}|h_{2,m}|^2}{\sigma^2}. \tag{2}$$

A. FRAME STRUCTURE AND CHANNEL ACCESS

The IoT nodes transmit their data packets in a frame-based structure. The IoT nodes in the same cluster access the same RB by exploiting power-domain NOMA [3], [4], while different clusters access the RBs randomly[4]. Each frame has two transmission phases, namely preamble transmission, and short data packet transmission of durations T_p and T_d, respectively. Hence, the whole frame duration is $T_f = T_p + T_d$. In the first phase, each IoT node in a cluster transmits its orthogonal preamble[5] sequence on the randomly selected RB r (for $r \in \{1, \ldots, R\}$) to the BS[6]. In the second phase of the frame, each IoT node transmits its data packets over the selected RB.

Remark 2: Note that a typical IoT node in a cluster does not transmit any preamble sequence if it does not have any data in its buffer or enough energy to transmit its data packet. When the BS detects only one preamble sequence on a RB, it starts decoding the only received signal without applying SIC.

B. LINK SPECIFICATIONS

To achieve UR communications, transmission diversity is adopted, in which multiple replicas of each frame, say K, are transmitted successively. To realize LL transmissions, the FBL codes are adopted [19], [48]. In such a case, Shannon's capacity is no longer applicable, since the decoding error cannot be ignored. Thus, given a blocklength of $n_b > 100$ with n_d data bits per data packet, the instantaneous block error rate of decoding the signal of node $U_{i,m}$ (for $i \in \{1, 2\}$ and $m \in \{1, \ldots, M\}$) at the BS can be approximated as [49]

$$\Upsilon\left(\gamma_{i,m}, n_b, n_d\right) = Q\left(\sqrt{\frac{n_b}{\chi\left(\gamma_{i,m}\right)}}\left(C\left(\gamma_{i,m}\right) - \frac{n_d}{n_b}\right)\right). \tag{3}$$

where $C\left(\gamma_{i,m}\right) = \log_2(1 + \gamma_{i,m})$ is the Shannon capacity, while $\chi\left(\gamma_{i,m}\right) = \left(1 - \frac{1}{1+\gamma_{i,m}^2}\right)\left(\log_2 e\right)^2$ represents the channel dispersion.

[2]For simplicity, and without loss of generality, N is assumed to even.

[3]The two-node NOMA communication is an elementary block of NOMA, addressed in the 3GPP-LTE Advanced [37]. Also, in IoT networks, the wireless devices may be clustered according to their geographical locations, traffic similarity/activity, task allocation, or resource requirements [38]–[42].

[4]From a practical perspective, this can be performed by feeding the random number generators with the same seed [45].

[5]Orthogonal preamble transmissions are used in RA-LTE, and defined in 3GPP RA [46], [47].

[6]The preamble phase enables the BS to detect the stronger/weaker nodes within each cluster in order to determine the SIC decoding order. Furthermore, by receiving only one preamble from a typical cluster, the BS knows the node that is not transmitting any data packet in the transmission phase.

Remark 3: The inter-cluster interference occurs when nodes in at least two clusters transmit data over the same RB. In such a case, a collision occurs and all data packets transmitted by all the nodes on that RB are lost.

C. DATA AND ENERGY ARRIVAL MODELS

Since the IoT nodes are assumed to scavenge environmental energy, random amounts of energy are harvested in each frame. The energy arrival process $\mathcal{E}_n$ at each IoT node $U_{i,m}$ in each frame n (for $n = 1, 2, \cdots$) is modeled as an i.i.d. stationary random process. Specifically, the number of harvested energy packets is modeled as a homogeneous Poisson counting process with rate λ_e. That is, the probability of the arrival of ε energy packets—each of ξ amount of energy—during a typical frame equals $P_E^{pkt}(\varepsilon) = \frac{(\lambda_e T_f)^\varepsilon e^{-\lambda_e T_f}}{\varepsilon!}$, for $\varepsilon \in \{0, 1, \ldots\}$. Thus, the expected amount of harvested energy during a frame is given by $\lambda_e T_f \xi$. The harvested energy is then stored in a battery of finite-capacity $B_{\max}$. In turn, the battery energy level at the beginning of the n^{th} frame is a random variable, which is denoted by $\mathcal{B}_{i,m}^n$. Specifically, the battery energy level is quantized into a set of finite discrete states, i.e., $\mathcal{B}_{i,m}^n \in \{0, 1, \ldots, B_{\max}\}$. Hence, the battery capacity equals $B_{\max}\beta$, where β is the minimum quantized energy level of the battery.

The minimum battery energy level for transmitting a frame by node $U_{i,m}$ is given by

$$\Delta_{i,m} = \left\lfloor \frac{P_{i,m}\left(T_p + T_d\right)}{\beta} \right\rfloor. \tag{4}$$

The packet arrival process of IoT node $U_{i,m}$ is independent of the other nodes, and is modeled as a homogeneous Bernoulli process with probability of arrival—in each frame—of $\alpha_{i,m}$[7]. It is further assumed that each data packet can be sent over one RB within T_d seconds. Note that each IoT node has a buffer, in which the packets are stored and queued to be sent. The number of data packet in a typical IoT node's buffer at the beginning of the n^{th} frame is a random variable, which is denoted $\mathcal{D}_{i,m}^n$. Furthermore, the steady-state probability that an IoT node $U_{i,m}$ has d data packets in its buffer with the battery being in state b at the beginning of a transmission frame is denoted by $\pi_{i,m}(d, b)$, which is derived in subsection III-B.

Before analyzing the URLL-EH-IoT network metrics, the main symbols used in this paper as well as their descriptions are given in Table 1.

III. DERIVATION OF PERFORMANCE METRICS

A. DEFINITIONS

Definition 1 (Transmission Cycle): A transmission cycle $\mathcal{T}$ is the time period during which a typical data packet and all its replicas are transmitted, which also incorporates all the waiting frames due to insufficient harvested energy. Thus, this cycle lasts at least $K \times T_f$ seconds if there is always sufficient energy to send all K successive replicas of a packet.

[7]Such a model complies with the sporadic traffic flow assumption that occurs within short time transmissions.

TABLE 1. Notations

Symbol	Description
$P_{i,m}$	Transmit power of node $U_{i,m}$
T_f	Frame duration
T_p	Preamble duration
T_d	Data packet transmission duration
K	Number of replicas of a packet
M	Number of clusters
N	Number of nodes
R	Number of resource blocks
n_d	Number of data bits of a packet
n_b	Blocklength
$\gamma_{i,m}$	SINR of $U_{i,m}$'s signal at BS
λ_e	Energy packets arrival rate
ξ	Amount of energy for each quantized energy packets
β	Minimum quantized battery energy level
ν	Path-loss exponent
$\mathcal{B}_{i,m}^n$	$U_{i,m}$'s battery energy level at the beginning of the n^{th} frame
$\mathcal{D}_{i,m}^n$	Number of data packets in $U_{i,m}$'s buffer at the beginning of the n^{th} frame
$\Delta_{i,m}$	Minimum energy level at $U_{i,m}$ for transmitting a frame
$\alpha_{i,m}$	Probability of data packet arrival at $U_{i,m}$'s buffer in a frame
$\pi_{i,m}(d,b)$	Steady-state probability that $U_{i,m}$ has d data packets with battery in state b
$\mathcal{T}_{i,m}$	$U_{i,m}$'s transmission cycle
$\mathcal{L}_{i,m}$	$U_{i,m}$'s latency
$\mathcal{R}_{i,m}$	Reliability of IoT node $U_{i,m}$
$\mathcal{G}_{i,m}$	$U_{i,m}$'s GoodPut

Definition 2 (Average Packet Latency): The latency $\mathcal{L}$ is defined as the average delay of delivering all replicas of a data packet to the BS, which includes the transmission delay and the queuing (buffer) waiting time.

Definition 3 (Reliability): The transmission reliability $\mathcal{R}$ is defined as the probability that a typical packet transmitted from an IoT node is received successfully at the BS (i.e. without any inter-cluster collisions or decoding errors).

Definition 4 (GoodPut): The GoodPut $\mathcal{G}$ of an IoT node is defined as the average effective rate (i.e. non-redundant data bits per time unit) received successfully at the BS.

In what follows, the RA-NOMA scenario refers to the case of a cluster of IoT nodes randomly selecting a RB from the R RBs; while in the RA-OMA scenario, each IoT node randomly selects one RB for data transmission.

B. STEADY-STATE PROBABILITIES

To derive the steady-state probabilities $\pi_{i,m}(d, b)$, define the stochastic process $X_n = \left(\mathcal{D}_{i,m}^n, \mathcal{B}_{i,m}^n\right)$, and consider the following lemma.

Lemma 1: X_n is a time-homogeneous *quasi-birth-and-death* (QBD) process [50], with $\mathcal{D}_{i,m}^n$ and $\mathcal{B}_{i,m}^n$ being the level and phase variables, respectively.

Proof: See Appendix A. ■

The transition probabilities of the process X_n are given by **Lemma 2**. For notational convenience, define $\Theta_{i,m}^{d_2|d_1,b_2|b_1} \triangleq \Pr(\mathcal{D}_{i,m}^n = d_2, \mathcal{B}_{i,m}^n = b_2 \mid \mathcal{D}_{i,m}^{n-1} = d_1, \mathcal{B}_{i,m}^{n-1} = b_1)$, which is the transition probability from state $(\mathcal{D}_{i,m}^{n-1} = d_1, \mathcal{B}_{i,m}^{n-1} = b_1)$ to state $(\mathcal{D}_{i,m}^n = d_2, \mathcal{B}_{i,m}^n = b_2)$.

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = \begin{cases} \mathbb{J}_{\alpha_{i,m}}(d_2-d_1)P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1)\beta}{\xi}\right\rfloor\right)\mathbb{I}(b_2-b_1), & b_1<\Delta_{i,m}, b_2<B_{\max} \text{ or } d_1=0, b_2<B_{\max}, \\ \mathbb{J}_{\alpha_{i,m}}(d_2-d_1)\sum_{k=B_{\max}}^{\infty}P_E^{pkt}\left(\left\lfloor\frac{(k-b_1)\beta}{\xi}\right\rfloor\right), & b_1<\Delta_{i,m}, b_2=B_{\max} \text{ or } d_1=0, b_2=B_{\max}, \\ \mathbb{J}_{\alpha_{i,m}}(d_2-d_1+1)P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right)\mathbb{I}(b_2-b_1+\Delta_{i,m}), & b_1\geq\Delta_{i,m}, d_1\neq 0, b_2<B_{\max}, \\ \mathbb{J}_{\alpha_{i,m}}(d_2-d_1+1)\sum_{k=B_{\max}}^{\infty}P_E^{pkt}\left(\left\lfloor\frac{(k-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right), & b_1\geq\Delta_{i,m}, d_1\neq 0, b_2=B_{\max} \end{cases} \tag{5}$$

Lemma 2: The transition probabilities of X_n are obtained as given by (5), in which $\mathbb{J}_{\alpha_{i,m}}(d)$ is defined as

$$\mathbb{J}_{\alpha_{i,m}}(d) = \begin{cases} \alpha_{i,m}, & d=1, \\ 1-\alpha_{i,m}, & d=0, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

and $\mathbb{I}(b)$ is a binary indicator function, defined as

$$\mathbb{I}(b) = \begin{cases} 1, & b\geq 0, \\ 0, & \text{otherwise}. \end{cases} \tag{7}$$

Proof: See Appendix B. ∎

Now that the transition probabilities have been obtained, $\pi_{i,m}(d,b)$ can be derived by solving the linear system of equations $\boldsymbol{\pi}_{i,m}\boldsymbol{\Theta} = \boldsymbol{\pi}_{i,m}$, in which $\boldsymbol{\Theta}$ is the transition probability matrix with elements of $\Theta_{i,m}^{d_2|d_1,b_2|b_1}$. Moreover, $\boldsymbol{\pi}_{i,m}$ is the vector form of the steady-state probabilities, given by $\boldsymbol{\pi}_{i,m} = \left(\boldsymbol{\pi}_{i,m}(0), \boldsymbol{\pi}_{i,m}(1), \boldsymbol{\pi}_{i,m}(2), \cdots\right)$, where $\boldsymbol{\pi}_{i,m}(d) = \left(\pi_{i,m}(d,0), \cdots, \pi_{i,m}(d,B_{\max})\right)$. To solve $\boldsymbol{\pi}_{i,m}\boldsymbol{\Theta} = \boldsymbol{\pi}_{i,m}$, the matrix geometric approach is used [51], in which $\boldsymbol{\Theta}$ is structured into blocks as

$$\boldsymbol{\Theta} = \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{A}_{0,1} & & & \\ \mathbf{A}_{1,0} & \mathbf{A}_1 & \mathbf{A}_0 & & \\ & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{8}$$

where $\mathbf{A}_{0,0}$, $\mathbf{A}_{0,1}$, $\mathbf{A}_{1,0}$, $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{A}_2$ are the corresponding $(B_{\max}+1)\times(B_{\max}+1)$ matrices. By defining the rate matrix[8] $\boldsymbol{\Omega}$, the steady-state probabilities are given by $\boldsymbol{\pi}_d = \boldsymbol{\pi}_1\boldsymbol{\Omega}^{d-1}$, for $d = 0,1,2,\ldots$ [51]. Moreover, the vectors $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_1$ are the unique positive solutions to the linear system of

$$\boldsymbol{\pi}_0 = \boldsymbol{\pi}_0\mathbf{A}_{0,0} + \boldsymbol{\pi}_0\mathbf{A}_{0,0}, \tag{9a}$$

and

$$\begin{aligned} \boldsymbol{\pi}_1 &= \boldsymbol{\pi}_0\mathbf{A}_{0,1} + \boldsymbol{\pi}_1\mathbf{A}_1 + \boldsymbol{\pi}_0 \\ &= \boldsymbol{\pi}_0\mathbf{A}_{0,0} + \boldsymbol{\pi}_2\mathbf{A}_2. \end{aligned} \tag{9b}$$

Remark 4: Deriving $\pi_i(d,b)$ for the RA-OMA scenario is exactly the same as $\pi_{i,m}(d,b)$, since the QBD structure is the same for each IoT node in the RA-OMA and RA-NOMA.

[8]The rate matrix $\boldsymbol{\Omega}$ is the minimal nonnegative solution to the nonlinear equation $\boldsymbol{\Omega} = \mathbf{A}_0 + \boldsymbol{\Omega}\mathbf{A}_1 + \boldsymbol{\Omega}^2\mathbf{A}_2$. Note that in general, there is no explicit solution to the matrix $\boldsymbol{\Omega}$. However, some algorithms have been proposed to compute it recursively [50].

C. TRANSMISSION CYCLE AND AVERAGE PACKET LATENCY

To derive the average packet latency, the probability density function (PDF) of the transmission cycle for IoT node $U_{i,m}$'s data packet is firstly derived, as per **Lemma 3**.

Lemma 3: The PDF of IoT node $U_{i,m}$'s packet transmission cycle $\mathcal{T}$ is obtained as

$$f_{i,m}^{\mathcal{T}}(t) = \sum_{k=K}^{\infty}\binom{k-1}{K-1}\Pi_{i,m}^{K}\left(1-\Pi_{i,m}\right)^{k-K}\delta(t-kT_f), \tag{10}$$

in which $\delta(.)$ is the Dirac delta function, and $\Pi_{i,m} = \sum_{d=1}^{\infty}\sum_{b=\Delta_{i,m}}^{B_{\max}}\pi_{i,m}(d,b)$ is the probability that $U_{i,m}$ has a battery level of at least $\Delta_{i,m}$ upon transmitting a typical data packet. Furthermore, the expected value of the transmission cycle is obtained as

$$\begin{aligned} \mu_{\mathcal{T}_{i,m}} &\triangleq \mathbb{E}\left[\mathcal{T}_{i,m}\right] \\ &= \frac{K}{\Pi_{i,m}}T_f. \end{aligned} \tag{11}$$

Proof: See Appendix C. ∎

In addition to the packet transmission time, another important factor affecting the average packet latency is the IoT node's packet waiting time. Considering the steady-state probability $\pi_{i,m}(d,b)$, the expected number of data packets in IoT node $U_{i,m}$'s buffer is obtained as $\mathbb{E}\left[\mathcal{D}_{i,m}^{n}\right] = \sum_{d=0}^{\infty}d\left(\sum_{b=0}^{B_{\max}}\pi_{i,m}(d,b)\right)$. Thus, by exploiting Little's formula, the average packet latency is obtained as [52]

$$\begin{aligned} \mathcal{L}_{i,m}^{NOMA} &= \mu_{\mathcal{T}_{i,m}}\left(\mathbb{E}\left[\mathcal{D}_{i,m}^{n}\right]+1\right) \\ &= \mu_{\mathcal{T}_{i,m}}\left(\sum_{d=0}^{\infty}d\left(\sum_{b=0}^{B_{\max}}\pi_{i,m}(d,b)\right)+1\right). \end{aligned} \tag{12}$$

Furthermore, let $\rho_{i,m}$ represent $U_{i,m}$'s link utilization, which is given by $\rho_{i,m} \triangleq \frac{\alpha_{i,m}}{T_f}\mathbb{E}\left[\mathcal{T}_{i,m}\right]$, with $\frac{\alpha_{i,m}}{T_f}$ being the average arrival packet rate. In addition, the probability that $U_{i,m}$ does not transmit data in a frame (e.g. possibly due to not having any data packets in the buffer or not having sufficient energy) is determined as

$$\begin{aligned} \Psi_{i,m}^{NT} &= 1-\rho_{i,m} \\ &= 1-\frac{\alpha_{i,m}}{T_f}\mathbb{E}\left[\mathcal{T}_{i,m}\right] \\ &= 1-\frac{\alpha_{i,m}}{\Pi_{i,m}}K. \end{aligned} \tag{13}$$

Remark 5: The steady-state probabilities of the defined QDB process X_n for both RA-NOMA and RA-OMA are the same as long as the frame duration, transmit power, data and energy arrival patterns are the same.
Remark 6: The packet transmission time and packet waiting time in an IoT node's buffer in the RA-OMA scenario are identical to those of the RA-NOMA scenario, for the same number of transmitted replicas K, frame duration T_f, and steady-state probabilities $\pi_{i,m}(d,b)$ and $\pi_i(d,b)$.

Remark 6 can be justified by noting that the waiting time in the buffer and the packet transmission time for each IoT node's packet in the underlying RA-NOMA and RA-OMA scenarios are only dependent on the packet and energy arrival rates, and not on the transmission strategies. Hence, the packet latency of an IoT node in the RA-OMA scenario is given as

$$\mathcal{L}_i^{OMA} = \mu_{\mathcal{T}_i}\left(\sum_{d=0}^{\infty} d\left(\sum_{b=0}^{B_{\max}} \pi_i(d,b)\right)+1\right), \tag{14}$$

where $\mu_{\mathcal{T}_i} \triangleq \frac{K}{\Pi_i}T_f$. Also, $\Pi_i = \sum_{d=1}^{\infty}\sum_{b=\Delta_i}^{B_{\max}}\pi_i(d,b)$, where $\Delta_i = \left\lfloor \frac{P_i(T_p+T_d)}{\beta} \right\rfloor$ is the minimum battery energy level for transmitting a frame by the node U_i in OMA scenario.

D. RELIABILITY

In this subsection, the reliability for RA-NOMA and RA-OMA scenarios is derived.

1) Reliability in the RA-NOMA Scenario

Each IoT node in a cluster may experience a different reliability, depending on the transmit power and SIC decoding order. Specifically, the reliability is affected by two different impairments in the underlying RA-NOMA scenario; RA collisions due to inter-cluster transmissions, and the decoding errors at the BS. **Lemma 4** provides the reliability for an IoT node under the RA-NOMA scenario.
Lemma 4: The reliability of IoT node $U_{i,m}$ in a RA-NOMA cluster m is derived as

$$\mathcal{R}_{i,m}^{\text{NOMA}} = 1-\left(1-\Pr\left(E_m^{NIC}\right)\Pr\left(E_{i,m}^{NDE}\right)\right)^K. \tag{15}$$

In particular, $\Pr\left(E_m^{NIC}\right)$ is the probability of no inter-cluster collisions (NIC) experienced by cluster m over the selected RB, as given by (16), where ω_l is defined as

$$\omega_l = \begin{cases} 1, & l=0, \\ \left(\frac{R-1}{R}\right)^l, & \text{otherwise}, \end{cases} \tag{17}$$

and $\phi_{l,m}$ is a set containing indices all possible combinations of l IoT clusters from the total M clusters excluding the cluster of interest (i.e. cluster m). Mathematically, $\phi_{l,m} = \{(G,G') \mid G=(g_1,\cdots,g_l), G'=(g_{l+1},\cdots,g_{M-1}), g_i \in \{1,\cdots,M\}, g_i \neq g_j, \forall i,j, 1\leq i,j\leq l, g_i \neq m\}$, where g_i represents a cluster index, G is a set including l clusters from all existing M clusters excluding cluster m, and G' is the complement of G. On the other hand, $\Pr\left(E_{i,m}^{NDE}\right)$ is the probability of no decoding error (NDE), which is as given in (18). Furthermore, in (18), $\overline{\Upsilon}_{i|i,m}(n_b,n_d)$ is given as (19), while $\overline{\Upsilon}_{i|j,m}(n_b,n_d)$ is given by

$$\overline{\Upsilon}_{i|j,m}(n_b,n_d) = \frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\int_0^{\infty}\Upsilon(\gamma,n_b,n_d)e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\gamma}d\gamma. \tag{20}$$

Proof: See Appendix D. ■

2) Reliability in the RA-OMA Scenario

The reliability in the RA-OMA scenario is given by **Lemma 5**.
Lemma 5: The reliability of IoT node U_i (for $i \in \{1,\cdots,N\}$) is derived as

$$\mathcal{R}_i^{\text{OMA}} = 1-\left(1-\Pr\left(E_i^{NIC}\right)\Pr\left(E_i^{NDE}\right)\right)^K. \tag{21}$$

in which $\Pr\left(E_i^{NIC}\right)$ and $\Pr\left(E_i^{NDE}\right)$ are the probabilities of no inter-node collisions and no decoding error, respectively. Particularly, $\Pr\left(E_i^{NIC}\right)$ is determined as

$$\Pr\left(E_m^{NIC}\right) = \sum_{l=0}^{M-1}\omega_l\left(\sum_{\forall(g_1,\cdots,g_l)\in\phi_{l,m}}\left(\prod_{g=g_1}^{g_l}\left(1-\Psi_{1,g}^{NT}\Psi_{2,g}^{NT}\right)\prod_{\substack{g'=g_{l+1}\\ g'\neq m}}^{g_M}\Psi_{1,g'}^{NT}\Psi_{1,g'}^{NT}\right)\right) \tag{16}$$

$$\Pr\left(E_{i,m}^{NDE}\right) = \frac{P_{i,m}d_{j,m}^{\nu}\left(1-\overline{\Upsilon}_{i|i,m}(n_b,n_d)\right)+P_{j,m}d_{i,m}^{\nu}\left(1-\overline{\Upsilon}_{i|j,m}(n_b,n_d)\right)\left(1-\overline{\Upsilon}_{j|j,m}(n_b,n_d)\right)}{P_{j,m}d_{i,m}^{\nu}+P_{i,m}d_{j,m}^{\nu}}\left(1-\Psi_{j,m}^{NT}\right)+\left(1-\overline{\Upsilon}_{i|j,m}(n_b,n_d)\right)\Psi_{j,m}^{NT} \tag{18}$$

$$\overline{\Upsilon}_{i|i,m}(n_b,n_d) = \int_0^{\infty}\Upsilon(\gamma,n_b,n_d)\frac{d_{i,m}^{\nu}d_{j,m}^{\nu}\sigma^2\left(P_{i,m}d_{j,m}^{\nu}+P_{j,m}d_{i,m}^{\nu}\gamma\right)+\left(P_{i,m}P_{j,m}d_{i,m}^{\nu}d_{j,m}^{\nu}\right)}{\left(P_{i,m}d_{j,m}^{\nu}+P_{j,m}d_{i,m}^{\nu}\gamma\right)^2}e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_i}\gamma}d\gamma \tag{19}$$

$$\Pr\left(E_i^{NIC}\right) = \sum_{l=0}^{N-1} \omega_l \left(\sum_{\forall (f_1,\cdots,f_l)\in\phi_l} \left(\prod_{f=f_1}^{f_l} \left(1-\Psi_f^{NT}\right) \prod_{\substack{f'=f_{l+1}\\ f'\neq i}}^{f_N} \Psi_{f'}^{NT} \right) \right), \tag{22}$$

where ϕ_l is a set containing indices all possible combinations of l IoT nodes excluding that of interest (i.e. U_i). Mathematically, $\phi_l = \{(F,F') \mid F = (f_1,\cdots,f_l), F' = (f_{l+1},\cdots,f_{N-1}),\ f_n \in \{1,\cdots,N\},\ f_n \neq f_j, \forall n,j, 1 \leq n,j \leq l, f_n \neq i\}$, where f_n represents the node index, F is a set including l nodes from all existing N nodes excluding node i, and F' is the complement of F. In turn, ϕ_l is a set of all combinations of l nodes and their complement. Furthermore, $\Pr\left(E_i^{NDE}\right)$ is obtained as

$$\Pr\left(E_i^{NDE}\right) = 1 - \overline{\Upsilon}_i(n_b,n_d), \tag{23}$$

where $\overline{\Upsilon}_i(n_b,n_d)$ is given as

$$\overline{\Upsilon}_i(n_b,n_d) = \int_0^\infty \Upsilon(\gamma,n_b,n_d) \frac{d_i^\nu \sigma^2}{P_i} e^{-\frac{d_i^\nu \sigma^2}{P_i}\gamma} d\gamma. \tag{24}$$

Proof: See Appendix E. ■

E. GOODPUT

1) GoodPut in the RA-NOMA Scenario

According to the definition provided in subsection III-A, the GoodPut for IoT node $U_{i,m}$ (namely $\mathcal{G}_{i,m}$) is defined as the average effective rate received at the BS due to $U_{i,m}$'s transmission. To derive such a metric, the successfully transferred "non-redundant" data bits per time unit from $U_{i,m}$ to the BS must be considered. Note that during a transmission cycle $\mathcal{T}$, the number of effective successfully delivered bits is $n_d \mathcal{R}_{i,m}^{\text{NOMA}}$. Correspondingly, the GoodPut of node $U_{i,m}$ is determined as

$$\mathcal{G}_{i,m}^{\text{NOMA}} = \frac{n_d \mathcal{R}_{i,m}^{\text{NOMA}}}{\mathbb{E}[\mathcal{T}_{i,m}]} = \frac{n_d}{KT_f/\Pi_{i,m}} \mathcal{R}_{i,m}^{\text{NOMA}}. \tag{25}$$

2) GoodPut in the RA-OMA Scenario

Similar to the derivation of GoodPut in the RA-NOMA scenario and considering that there is no clustering in RA-OMA, the GoodPut $\mathcal{G}_i^{\text{OMA}}$ for node U_i (for $i = 1,\ldots,N$) is obtained as

$$\mathcal{G}_i^{\text{OMA}} = \frac{n_d \mathcal{R}_i^{\text{OMA}}}{\mathbb{E}[\mathcal{T}_i]} = \frac{n_d}{KT_f/\Pi_i} \mathcal{R}_i^{\text{OMA}}. \tag{26}$$

in which $\mathbb{E}[\mathcal{T}_i] = \frac{K}{\Pi_i} T_f$.

IV. NUMERICAL RESULTS

This section numerically evaluates the effect of the number of replicas K of an IoT node's packet, number of data bits n_d in each packet per blocklength, and the transmit power of the IoT nodes on the GoodPut, reliability, and packet latency for the RA-NOMA and RA-OMA scenarios. Table 2 summarizes the network parameters [53][9]. In the simulations, and for simplicity, the $N = 50$ nodes are assumed to be split into two sets of nodes; near and far [55]. The BS is assumed to be located at the center of the network area. Moreover, the near (far) nodes with better (worse) channel conditions are assumed to be located at a distance of $d_{1,m} = 0.7$ ($d_{2,m} = 1$) Km for the RA-NOMA scenario ($\forall m \in \{1,\ldots,25\}$)[10]. Similarly, for the RA-OMA scenario, $d_1 = 0.7$ Km and $d_2 = 1$ Km are for users U_1 and U_2, respectively.

TABLE 2. Simulation Parameters

Parameter	Value	Parameter	Value
N	50	R	5 RBs
ν	3	$\lambda_e \xi$	4 J/s
n_b	400	β	5×10^{-4} J
P_1, P_2	0.2 W	$B_{\max}$	20 J
T_d, T_p	0.2 ms	α_1, α_2	0.02
W	1 MHz	N_0	-174 dBm/Hz

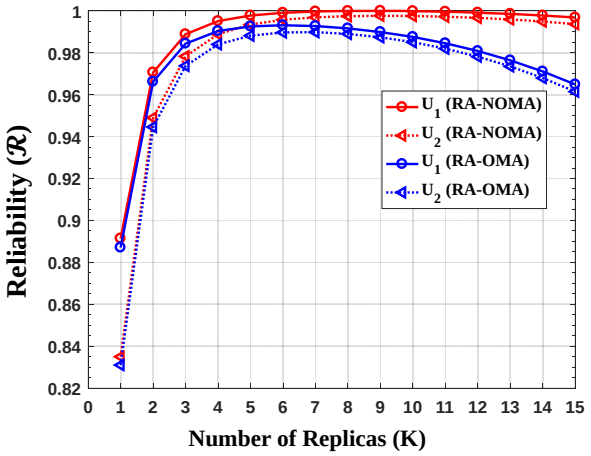


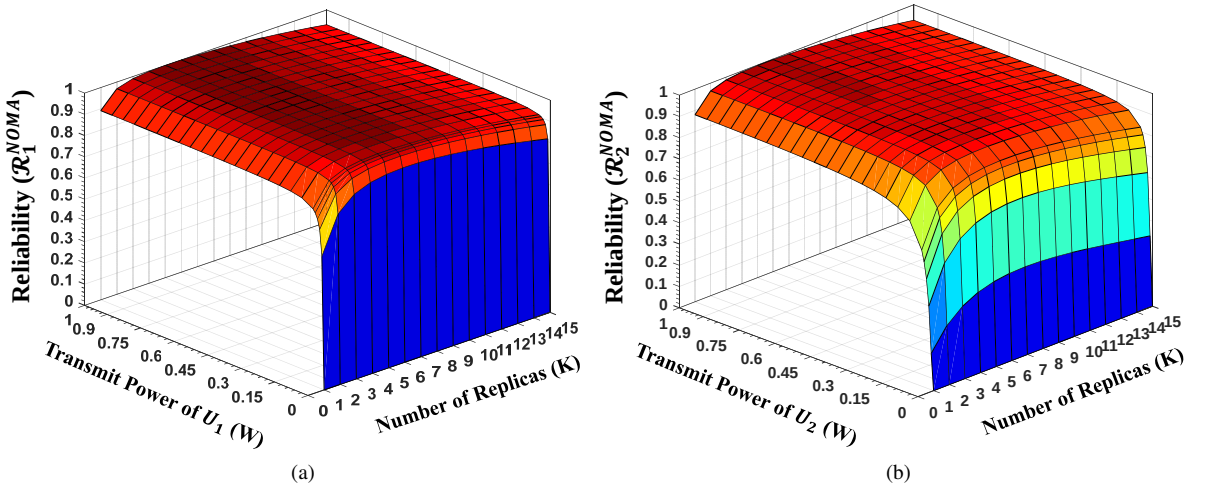
FIGURE 1. Reliability vs. number of replicas K - $n_d = 500$.

A. EFFECT OF NUMBER OF REPLICAS

Fig. 1 illustrates the nodes' reliability as a function of K for the stronger and weaker RA-NOMA and RA-OMA nodes. As can be seen, the reliability for both RA-NOMA and RA-OMA nodes increases sharply as K increases, with the RA-NOMA scenario being superior to the RA-OMA scenario. This is expected since by transmitting multiple replicas of a packet, the number of successful transferred data bits increases. Particularly, for RA-NOMA, the reliability is maximized when $K = 8$; while for RA-OMA, it is $K = 5$. However, the reliability starts to decrease when the number of retransmitted packets increases. This is due to the effect of the non-saturated data traffic and the resulting congestion, since the higher the number of transmitted replicas is, the more packets are queued in the nodes' buffer, and the higher the possibility of RA collisions during retransmission. Notably,

[9]Note that the preamble time T_p is that adopted by RA-LTE as PRACH preamble format #4 for UL-TX [54].

[10]From this point onwards, the subscript referring to the cluster index is dropped for the RA-NOMA scenario. Hence, the stronger and weaker nodes in each cluster will be referred to as U_1 and U_2, respectively.



(a) (b)

FIGURE 2. Reliability vs. transmit power and number of replicas K for RA-NOMA nodes: (a) Stronger node U_1, and (b) Weaker node U_2 - $n_d = 500$

the RA-NOMA nodes suffer less reduction in the reliability than their RA-OMA counterparts, since the collisions are lower in RA-NOMA than in RA-OMA, which is due to the clustering of nodes in RA-NOMA. Thus, the optimal number of replicas must be carefully chosen according to the URLL requirement, the number of nodes N in the network, and their data traffic pattern (i.e. packet arrival probability in each frame). Furthermore, it is observed that the reliability for the stronger and weaker RA-NOMA nodes is higher than that their RA-OMA counterparts, for all values of K, as clustering in RA-NOMA helps reduce the collisions. Moreover, the reliability of the stronger nodes is higher than the weaker nodes under both scenarios, as would be expected. Another point is that no value for K can meet the minimum general requirement of 99.999% reliability in the RA-OMA scenario; however, this is not the case for RA-NOMA.

To investigate the effect of RA-NOMA nodes' transmit power and transmission diversity K on the reliability, Figs. 2a and 2b are plotted. As can be seen from Fig. 2a, the reliability for the stronger node U_1 (i.e. $\mathcal{R}_1^{NOMA}$) increases when its transmit power increases, while fixing the transmit power of user U_2 at $P_2 = 0.2$ W. This is expected, as increasing the transmit power of the stronger node in a cluster helps increase its SINR at the BS, which decreases the decoding error probability. The same observation can be made for the weaker node U_2 when fixing $P_1 = 0.2$ W, as shown in Fig. 2b. Evidently, the reliability of the stronger node can be verified to be superior to the weaker node for all transmit power and K values, since the decoding error probability for the stronger node is lower than the weaker node, which is due to its higher SINR at the BS.

The effect of transmit diversity on the GoodPut for both the RA-NOMA and RA-OMA nodes is plotted in Fig. 3. It can be observed that the GoodPut decreases with the increase in K. This is due to the fact that the higher the number of transmitted replicas is, the longer the transmission cycle, and consequently, the lower the number of effective transferred data bits per time unit.

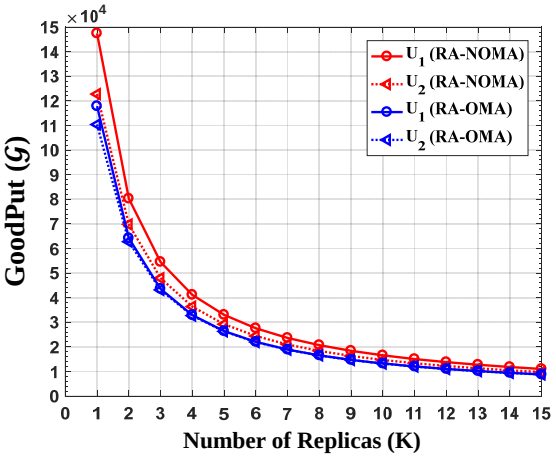
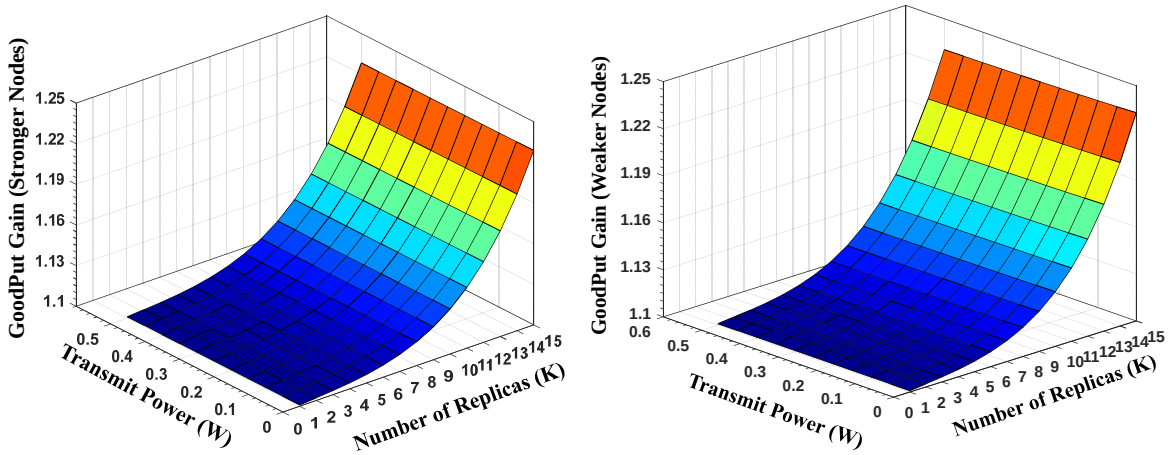


FIGURE 3. GoodPut vs. number of replicas $\mathbf{K}$ - $\mathbf{n_d = 500}$.

To quantify the GoodPut gain of RA-NOMA over RA-OMA, Figs. 4a and 4b are respectively plotted for the stronger and weaker nodes with respect to their RA-OMA counterparts. It is observed that GoodPut for RA-NOMA outweighs that for RA-OMA up to about 20% for higher transmission diversity values, since the reliability of RA-OMA decreases at a faster rate with the increase in K, as explained in Fig. 1.

The average packet latency as a function of the transmit power and number of replicas for both the stronger and weaker nodes in the RA-NOMA and RA-OMA scenarios is depicted in Fig. 5[11]. Clearly, the packet latency increases with the increase in K, as the transmission cycle also increases. Remarkably, the packet latency increases when the nodes' transmit power increases as well. This reflects the interplay between data- and energy-causality, which indicates that the higher the transmit power is, the more time they must wait to harvest sufficient energy from the environment, and hence,

[11]Note that according to **Remark 6**, all the nodes under both scenarios have the same packet latency when α and K are identical for all nodes.



(a) (b)

FIGURE 4. GoodPut Gain of RA-NOMA over RA-OMA vs. transmit power and number of replicas $\mathbf{K}$: (a) Stronger nodes, and (b) Weaker nodes - $\mathbf{n_d = 500}$.

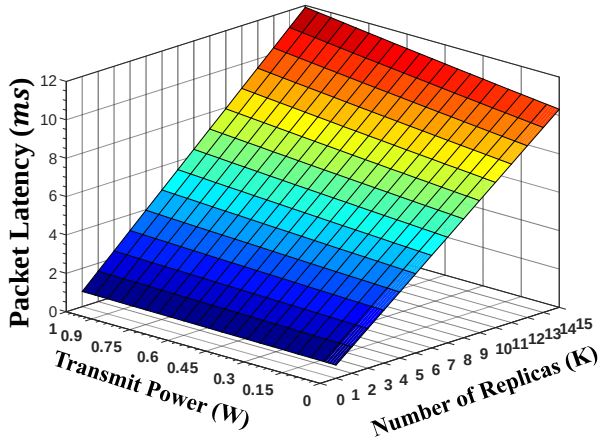


FIGURE 5. Packet Latency vs. transmit power and number of replicas $\mathbf{K}$ - $\mathbf{n_d = 500}$.

the higher the packet latency is. Note that increasing the transmit power improves the reliability (as shown earlier), but exacerbates the average packet latency in EH systems. To achieve the latency requirement of the URLLC in IIoT belonging to the emerging Industry 4.0 paradigm[12], the transmit power and transmission diversity should be considered as critical network operation parameters. Another point here is that when ignoring the reliability, both the RA-OMA and RA-NOMA scenarios have the same packet latency[13]. However, when the same target reliability for both scenarios is considered, the RA-OMA nodes have to send more replicas than RA-NOMA nodes. This makes the average packet latency longer, and the GoodPut lower in RA-OMA than in RA-NOMA.

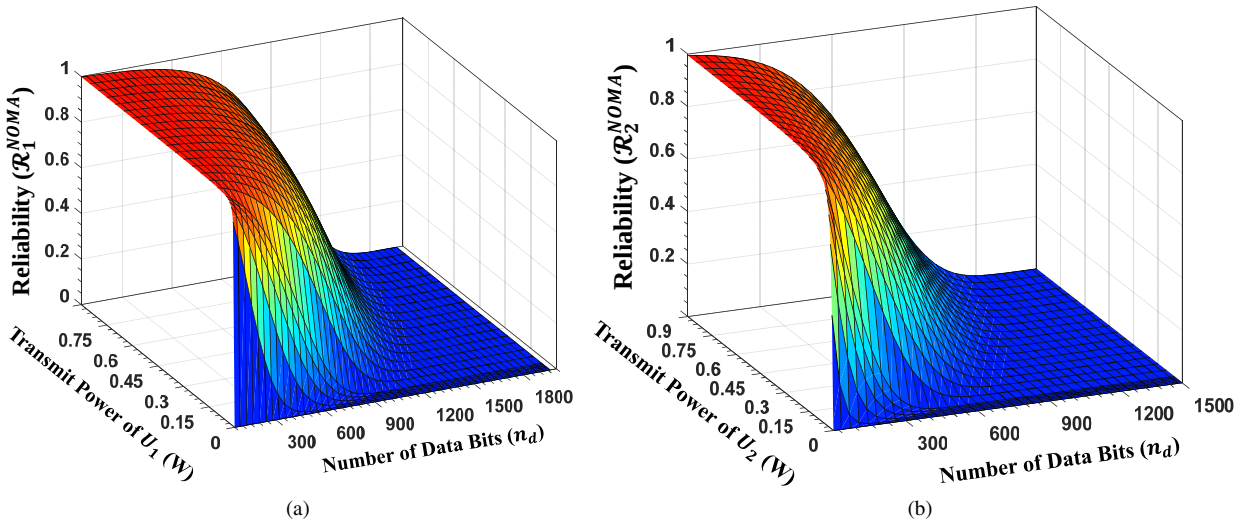
[12]Example applications include motion control, factory automation, and process automation with latency of 1, 10, and 50 ms, respectively [56], [57].

[13]Note that the RA-OMA transmission considered in this study is different from other conventional OMA schemes, such as TDMA or FDMA. Also, such OMA schemes have higher packet latency than the conventional NOMA scheme [4], [58].

B. EFFECT OF NUMBER OF DATA BITS

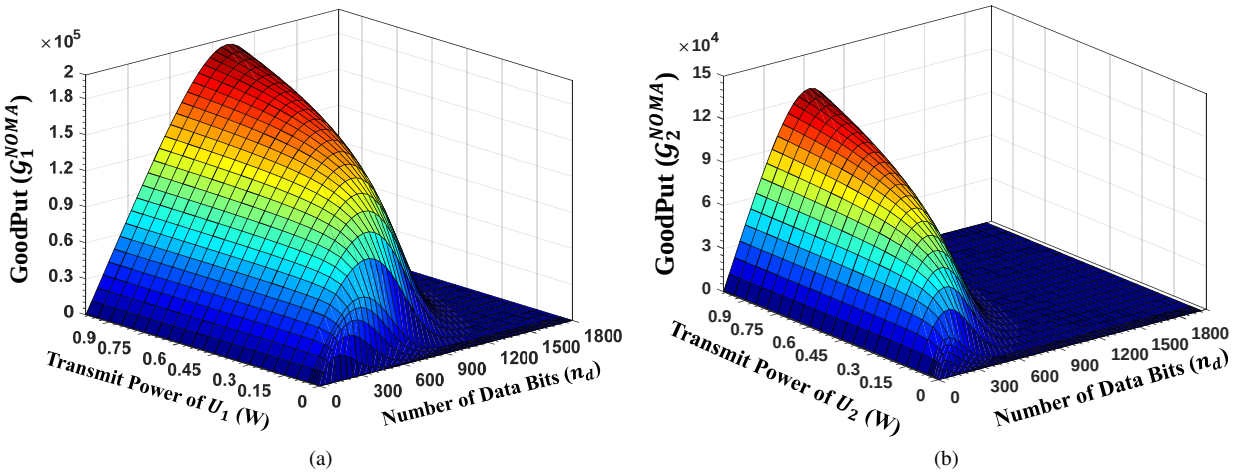
To explore the effect of finite blocklength regime, n_b is assumed to be fixed, and the number of data bits n_d is varied to have similar time-slot duration for both the stronger and weaker nodes. Figs. 6a and 6b illustrate the reliability for RA-NOMA nodes U_1 and U_2, respectively. As can be seen and also explained before, the reliability improves with the increase in the transmit power; however, it decreases when n_d increases. This is because the higher coding rate in fixed blocklength results in higher decoding error probability when short packet lengths are adopted. Also, note that the reliability decays slowly when the nodes use higher transmit powers, since for higher SINR/SNR values, more data bits can be transmitted successfully for a target decoding error. Furthermore, $\mathcal{R}_1^{NOMA}$ is verified to be superior to $\mathcal{R}_2^{NOMA}$, since the stronger node has higher SINR at the BS than the SNR of the weaker one.

Figs. 7a and 7b show the GoodPut for nodes U_1 and U_2—under the RA-NOMA scenario—as a function of the transmit power and n_d, respectively. Clearly, the higher the nodes' transmit power is, the greater the GoodPut, which is due to the higher the reliability. On the other hand, the GoodPut for both RA-NOMA nodes increases as n_d increases, peaks at some values, and then starts to decrease. For instance, for U_1, the GoodPut peaks at $n_d = 642$ for transmit power of 0.2 W, while for U_2, it peaks at $n_d = 322$ for the same transmit power. The reason for this phenomenon is that the higher the number of data bits is, the higher GoodPut, in agreement with (25). However, as explained earlier, increasing n_d increases the decoding error probability, which lowers the reliability. Thus, the GoodPut starts to decrease simultaneously with the decrease in the reliability, since the excessive increase in n_d outweighs the improvement in the GoodPut. Moreover, and in general, $\mathcal{G}_1^{NOMA}$ is higher than $\mathcal{G}_2^{NOMA}$, which is due to the higher U_1's SINR at the BS than U_2's SNR. It is noteworthy that at the same transmit power for nodes U_1 and U_2, the peak value in $\mathcal{G}_1^{NOMA}$ is greater than $\mathcal{G}_2^{NOMA}$,



(a) (b)

FIGURE 6. Reliability vs. transmit power and number of data bits $\mathrm{n_d}$ for RA-NOMA nodes: (a) Stronger node U_1, and (b) Weaker node U_2 - $\mathrm{K} = 2$.



(a) (b)

FIGURE 7. GoodPut vs. transmit power and number of data bits $\mathrm{n_d}$ for RA-NOMA nodes: (a) Stronger node U_1, and (b) Weaker node U_2 - $\mathrm{K} = 2$.

indicating that the stronger node can convey higher number of data bits (i.e. has higher GoodPut) per time-slot than the weaker node in FBL regime[14].

Figs. 8 shows the average packet latency for both RA-NOMA nodes as a function of the transmit power and n_d, respectively. As mentioned for Fig. 5, increasing the transmit power increases the packet latency due to the effect of energy and data causality. However, the higher the number of data bits per blocklength included into the time-slot is, the lower the average packet latency. This is because fewer bits are queued in the nodes' buffer and consequently, the buffer waiting time becomes shorter.

Finally, the reliability of the RA-NOMA nodes U_1 and U_2 along with the SIC decision plane (SDP) is illustrated in Figs. 9a and 9b, respectively. Plotted in transparent red, SDP is the plane where the received power of the stronger

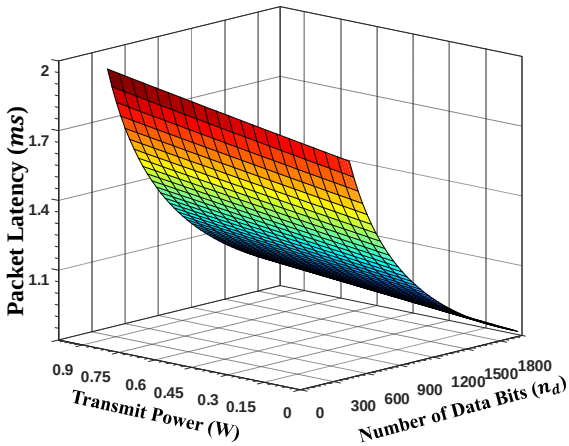


FIGURE 8. Packet Latency vs. transmit power and number of data bits $\mathrm{n_d}$ - $\mathrm{K} = 2$.

[14]Note that the same observations for the RA-OMA nodes over the same range of n_d values can be made, as the decoding error for short packet transmissions is identical for RA-NOMA and RA-OMA nodes. Therefore, the corresponding figures are not plotted to avoid unnecessary repetition. Nevertheless, the GoodPut of the RA-NOMA nodes is better than its RA-OMA counterpart, in agreement with Fig. 3.

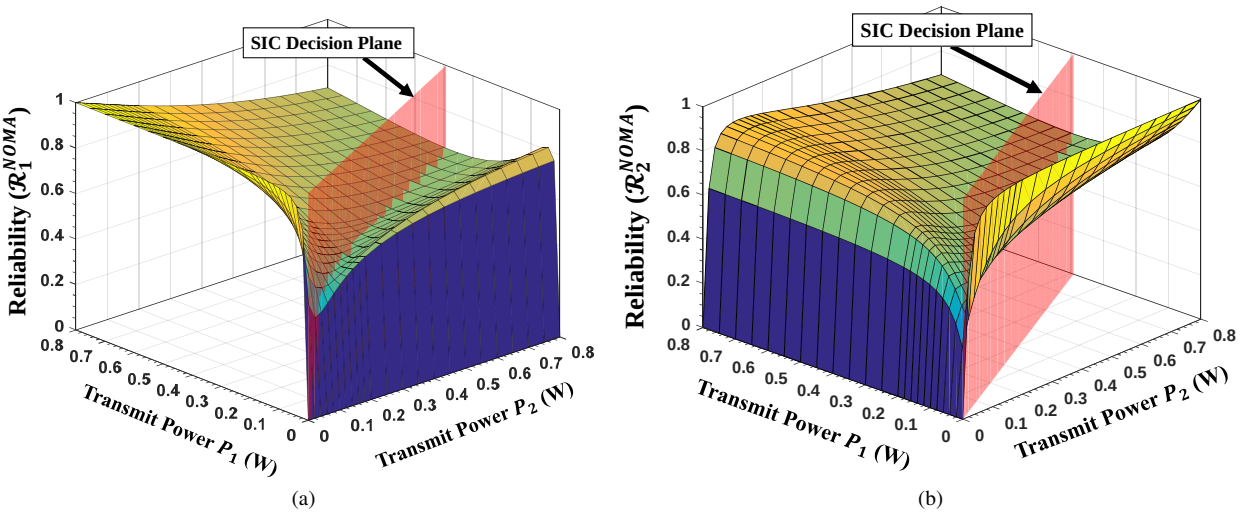


FIGURE 9. Reliability vs. transmit powers for RA-NOMA nodes: (a) Stronger node U_1, and (b) Weaker node U_2 - $\mathbf{K = 2}$ and $\mathbf{n_d = 500}$.

node U_1 at BS equals that of the stronger node U_2[15]. This is a threshold plane for the BS to determine the SIC decoding order. As can be seen from Fig. 9a, on the left side of the SDP—where the U_1's received power is higher than U_2's (i.e. $P_1 > P_2$)—$\mathcal{R}_1^{NOMA}$ increases as P_1 increases. However, it decreases with the increase in P_2. This is because in the left side of the SDP, the BS firstly decodes U_1's signal in the presence of U_2's signal as interference, and then decodes U_2's signal by applying SIC. On the right side of the SDP in Fig. 9a, U_1's received power is lower than U_2's and hence, the BS firstly decodes U_2's signal and then applies SIC to decode U_1's signal[16]. Thus, U_1 experiences interference-free decoding. Now, going back to the right side of the SDP in Fig. 9a, it is observed that for small values of P_1, $\mathcal{R}_1^{NOMA}$ is also low, since U_1's SNR is low and hence, its decoding error becomes high. By increasing P_1 for fixed P_2, $\mathcal{R}_1^{NOMA}$ improves, since U_1's SNR is increased. However, since $\mathcal{R}_1^{NOMA}$ depends on the correct decoding of U_2's signal, excessive increase in P_1 is counterproductive. This is because it results in decreasing U_2's SINR, and consequently, lowers $\mathcal{R}_1^{NOMA}$. Such a decrease in $\mathcal{R}_1^{NOMA}$ continues toward the SDP on its right side, where U_2's SINR reaches its lowest value. For further increase in P_1 (and thus moving to left side of the SDP), $\mathcal{R}_1^{NOMA}$ increases, since the BS changes the SIC decoding order. The same observations for $\mathcal{R}_2^{NOMA}$ can be made in Fig. 9b by noting that in the left side of SDP, U_2 experiences interference-free signal decoding; while in the right side, its signal is decoded by treating U_1's signal as interference.

C. VALIDATION OF ANALYTICAL DERIVATIONS

Discrete event simulation (DES) is performed to validate the analytical derivations, where all independent network events—including IoT nodes' packet and energy arrival events—are scheduled. To simulate the channel behavior, the exponential distribution is exploited and sampled to generate channel gains for each node. The RBs are selected for each node according to the uniform distribution. The simulation time advances to the next nearest scheduled event, while the next time instant for the corresponding event is generated. Upon reaching the simulation time to the next nearest event time, all the network metrics are then calculated and updated. The simulation end time is set to 5×10^3 seconds to collect a sufficient number of events—say 5×10^4—for each run. The simulation is executed 50 times, and then the network metrics are averaged. Figs. 10a, 10b and 10c show both theoretical and simulated results for RA-NOMA packet latency, reliability and GoodPut, respectively. As can be seen, the simulation results are in good agreement with those obtained via the theoretical analyses.

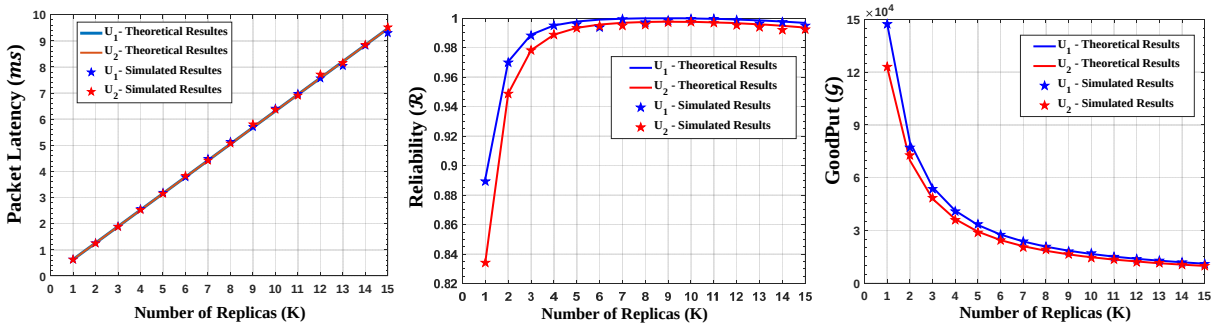
V. SUM-GOODPUT MAXIMIZATION PER RA-NOMA IOT CLUSTER

The obtained analytical derivations can be utilized to optimize the sum-GoodPut of a RA-NOMA cluster m in the URLL-EH-IoT network, subject to constraints on reliability, average packet latency, SIC decoding order, and transmit power. The optimization problem is formulated as[17]

[15]Recall that the preamble sequences are orthogonal and thus, in the preamble phase, the BS is able to measure the received SNR of each node. Then, for the same noise power at the BS, the received power for each signal is measured, and used to determine the SIC decoding order. Also, note that all the nodes transmit preamble sequence with the same predefined power by the BS. Hence, the node with higher received power (or equivalently SNR) is the stronger node, and vice versa for the weaker node [59].

[16]Note that in right side of the SDP in Fig. 9a, U_1 experiences interference-free signal decoding, since $P_1 < P_2$.

[17]Appropriate choice of transmit power and K can achieve the stringent ultra-reliability requirement of 99.999% as per 3GPP and ITU specification for URLL IoT applications [60], [61].



(a) (b) (c)

FIGURE 10. Simulated Results vs. Theoretical Results for RA-NOMA: (a) Packet Latency, (b) Reliability, and (c) GoodPut - $\mathrm{n_d} = 500$.

Sum-GoodPut-Max:

$$\max_{n_d, K, \mathbf{P}_m} \quad \sum_{i=1}^{2} \mathcal{G}_{i,m}^{NOMA} \tag{27a}$$

$$\text{s.t.} \quad \mathcal{L}_{i,m} \leq \delta_{th}^{L}, \qquad \forall i \in \{1,2\}, \tag{27b}$$

$$\mathcal{R}_{i,m}^{NOMA} \geq \delta_{th}^{R}, \qquad \forall i \in \{1,2\}, \tag{27c}$$

$$P_{1,m} \geq P_{2,m}, \tag{27d}$$

$$0 \leq P_{i,m} \leq P_{\max}, \qquad \forall i \in \{1,2\}, \tag{27e}$$

$$n_d, K \in \{1,2,...\}. \tag{27f}$$

where $\mathbf{P}_m = [P_{1,m}, P_{2,m}]$ is the transmit power vector IoT cluster m. Constraint (27b) ensures that the maximum average packet latency δ_{th}^{L} is satisfied, while Constraint (27c) is the reliability requirement, which is at least δ_{th}^{R}. Constraint (27d) enforces the SIC decoding order. The last two constraints define the range of values the decision variables take.

Remark 7: Problem **Sum-GoodPut-Max** is classified as a mixed-integer non-linear programming (MINLP) problem, which is non-convex and NP-hard [62]. This can be verified from the derived non-linear expressions of $\mathcal{G}_{i,m}^{NOMA}$, $\mathcal{L}_{i,m}$, and $\mathcal{R}_{i,m}^{NOMA}$, and the integer-valued decision variables n_d and K. In other words, problem **Sum-GoodPut-Max** is computationally-intensive. Nevertheless, due to the steady-state analysis, the computational delay due to the solution of problem **Sum-GoodPut-Max** becomes irrelevant[18].

Fig. 11 shows the number of iterations to reach the optimal solution of problem **Sum-GoodPut-Max**. The optimal objective function value is 4.336×10^4 bits/s, while the optimal values of the decision variables are $(n_d, K, P_{1,m}, P_{2,m}) = (592, 4, 0.31, 0.17)$ when $P_{\max} = 0.5$ W, $\delta_{th}^{R} = 0.99999$, and $\delta_{th}^{L} = 10$ms. Also, the optimal value of the sum-GoodPut is determined in less than 20 iterations, and hence, the optimal solution can be obtained efficiently.

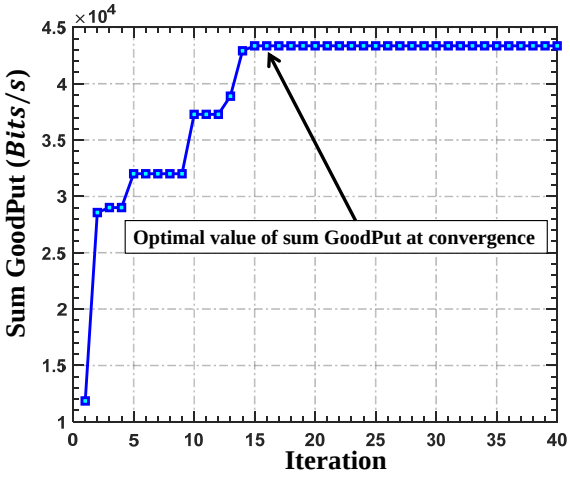


FIGURE 11. Number of iterations to obtain the optimal solution.

VI. SUMMARY OF FINDINGS

The following findings can be stated for the analyzed EH-IoT network metrics. For UR transmissions, the three main factors of the number of replicas K, nodes' transmit power, and number of data bits n_d per blocklength should be finely tuned. Similar to some existing studies, in the analyzed network scenario, the reliability for both RA-NOMA and RA-OMA nodes increases sharply as K increases. Remarkably, the reliability starts to decrease when K excessively increases. This is because in a realistic massive IoT network with nodes characterized by sporadic traffic behavior, the congestion due to excessive number of packet replicas is inevitable. Such effect of diversity transmission has not been discussed in the literature before, since the saturated data traffic behavior is commonly assumed. Particularly, the higher the number of transmitted replicas is, the more packets are queued in the nodes' buffer, and consequently, the adverse effect inter-cluster collisions due to RA surpasses the positive effect of diversity transmission with respect to reliability. Thus, the number of replicas must be carefully chosen. Furthermore, the stronger and weaker RA-NOMA nodes' transmit powers affect their received SINR/SINR at the BS, which in turn influences their reliability. On one hand, increasing the stronger node's transmit power improves the reliability of both nodes in a cluster. On the other hand, increasing the weaker node's transmit power—as long as its received power at BS does not outweigh that of the stronger node—improves its reliability, and decreases that of the stronger node. More importantly, by exploiting orthogonal preamble

[18]In this work, problem **Sum-GoodPut-Max** is solved via the global optimization package MIDACO [63], [64], with tolerance set to 10^{-3}.

transmissions at the beginning of each frame, the BS is able to perform correct SIC based on the nodes' received signal power, and thus avoid reliability degradation due to fixed SIC decoding order. The number of data bits transmitted in a frame also affects the reliability, where for fixed blocklength, increasing n_d for each node increases the error decoding probability at BS, and hence decreases the node's reliability.

Additionally, the aforementioned three factors also affect the average packet latency, and hence the LL requirement. Specifically, the average packet latency increases as K increases. This is due to the fact that the transmission cycle of each packet increases when multiple copies of each data packet are scheduled for transmission. Another interesting finding is that the packet latency increases when the nodes' transmit power increases. This signifies the interplay between data- and energy-causality, where the higher the transmit power is, the more time the nodes must wait to harvest sufficient energy, and hence, the higher the latency. Moreover, n_d also affects the average packet latency such that for fixed blocklength, the higher the number of data bits included into the time-slots is, the lower the average packet latency. This is because fewer bits are queued in the nodes' buffer and consequently, the buffer waiting time becomes shorter.

Finally, from the perspective of reliability-latency trade-off, all the three parameters play important roles in achieving URLL transmissions. In general, increasing K improves the reliability but increases packet latency. Such a trade-off is seen by considering n_d, where the higher the number of transmitted data bits in a time-slot is, the less the reliability but the shorter the packet latency. Such a trade-off still holds irrespective of the nodes' transmit power, as increasing the nodes' transmit power in general improves the reliability but increases packet latency. In summary, for URLL transmissions in EH-IoT networks, optimal tuning of the three aforementioned parameters is a must.

VII. CONCLUSIONS

This paper has considered UL RA-NOMA EH-IoT networks with URLL requirements. To realize URLL, short packets with transmission diversity have been adopted. The IoT nodes have been clustered and SIC has been employed through preamble transmissions. Network metrics such as average packet latency, reliability and GoodPut—considering the FBL regime—have been mathematically derived. Data- and energy-causality have been considered simultaneously throughout the provided analyses, and their interdependency has been incorporated into the average packet latency derivation. Furthermore, the effect of number of packet replicas, number of transmitted data bits per blocklength, and the IoT nodes' transmit power on the network metrics for the RA-NOMA clusters have been explored numerically and compared with their RA-OMA counterparts to highlight the merits of RA-NOMA. Lastly, the derived expressions have then been exploited to maximize the sum-GoodPut for a RA-NOMA IoT cluster under URLL and power constraints.

.

APPENDIX A : PROOF OF LEMMA 1

Proof: Firstly, process X_n must be shown to be a discrete-time Markov chain (DTMC). To this aim, the following property must be proved,

$$\Pr(X_n \mid X_{n-1}, X_{n-2}, \cdots, X_0) = \Pr(X_n \mid X_{n-1}). \quad \text{(A.1)}$$

For the sake of simplicity, the indices i, m in $\mathcal{D}^n_{i,m}$ and $\mathcal{B}^n_{i,m}$ are dropped in the following expressions. Note that the battery energy level is independent of the number of data packets in the buffer. Hence, $\Pr(X_n \mid X_{n-1}, X_{n-2}, \cdots, X_0)$ can be written as

$$\begin{aligned}
&\Pr(X_n = (\mathcal{D}^n, \mathcal{B}^n) \mid X_{n-1}, X_{n-2}, \cdots, X_0) \\
&\quad = \Pr(\mathcal{B}^n \mid \mathcal{D}^n, X_{n-1}, X_{n-2}, \cdots, X_0) \\
&\qquad \times \Pr(\mathcal{D}^n \mid X_{n-1}, X_{n-2}, \cdots, X_0) \qquad \text{(A.2)} \\
&\quad = \Pr(\mathcal{B}^n \mid X_{n-1}, X_{n-2}, \cdots, X_0) \\
&\qquad \times \Pr(\mathcal{D}^n \mid X_{n-1}, X_{n-2}, \cdots, X_0).
\end{aligned}$$

Furthermore, by defining $\mathcal{A}_p^{n-1}$ and $\mathcal{Q}_p^{n-1}$ as the number of arriving and transmitted packets of an IoT node during the $(n-1)^{th}$ frame, the number of data packets in that node's buffer can be written as

$$\mathcal{D}^n = \begin{cases} \mathcal{D}^{n-1} + \mathcal{A}_p^{n-1} - \mathcal{Q}_p^{n-1}, & \mathcal{D}^{n-1} \neq 0, \\ \mathcal{A}_p^{n-1}, & \mathcal{D}^{n-1} = 0. \end{cases} \quad \text{(A.3)}$$

Moreover, note that the packet arrival process in the $(n-1)^{th}$ frame is assumed to be independent of other frames, and also identically distributed among all frames. In turn, the time instance index $(n-1)$ in $\mathcal{A}_p^{n-1}$ can be removed. It is obvious that $\mathcal{Q}_p^{n-1}$ depends only on $\mathcal{D}^{n-1}$ and the energy level of the battery in the $(n-1)^{th}$ frame, since data transmission only happens if there is at least one data packet in the buffer, and sufficient energy in the battery at the beginning of the $(n-1)^{th}$ frame. Thus, $\mathcal{D}^n$ in (A.3) is only dependent on the state in the previous frame, i.e., X_{n-1}. Hence, in (A.2), $\Pr(\mathcal{D}^n \mid X_{n-1}, X_{n-2}, \cdots, X_0) = \Pr(\mathcal{D}^n \mid X_{n-1})$.

On the other hand, by defining $\mathcal{A}_e^{n-1}$ and $\mathcal{Q}_e^{n-1}$ as the amount of harvested and consumed energy during the $(n-1)^{th}$ frame, then the energy level of an IoT node's battery can be written as

$$\mathcal{B}^n = \begin{cases} \mathcal{B}^{n-1} + \mathcal{A}_e^{n-1} - \mathcal{Q}_e^{n-1}, & \mathcal{B}^{n-1} \geq \Delta_{i,m} \\ \mathcal{A}_e^{n-1}, & \text{otherwise.} \end{cases} \quad \text{(A.4)}$$

As can be seen from (A.4), $\mathcal{B}^n$ only depends on the state in the previous frame, i.e., X_{n-1}. Hence, in (A.2), $\Pr(\mathcal{B}^n \mid X_{n-1}, X_{n-2}, \cdots, X_0) = \Pr(\mathcal{B}^n \mid X_{n-1})$. Therefore, (A.2) can be written as

$$\begin{aligned}
&\Pr(X_n = (\mathcal{D}^n, \mathcal{B}^n) \mid X_{n-1}, X_{n-2}, \cdots, X_0) \\
&\quad = \Pr(\mathcal{B}^n \mid X_{n-1}, X_{n-2}, \cdots, X_0) \\
&\qquad \times \Pr(\mathcal{D}^n \mid X_{n-1}, X_{n-2}, \cdots, X_0) \qquad \text{(A.5)} \\
&\quad = \Pr(\mathcal{B}^n \mid X_{n-1}) \times \Pr(\mathcal{D}^n \mid X_{n-1}) \\
&\quad = \Pr(X_n \mid X_{n-1}),
\end{aligned}$$

and this proves that X_n is a DTMC. Now, note that $\mathcal{D}^n$ can change up to one packet in a successive frame. Also, $\mathcal{B}^n$ can take finite values from the battery energy level set. Therefore, the process X_n becomes a time-homogeneous *quasi-birth-and-death* process [51], with $\mathcal{D}^n$ and $\mathcal{B}^n$ being the level and phase variables, respectively. ■

APPENDIX B : PROOF OF LEMMA 2

Proof: To derive the transition probabilities, the initial states are classified according to several cases. Particularly, **Case 1** pertains to the states when there is no packet transmission, which is related to the conditions that either the battery energy level $\mathcal{B}_{i,m}^{n-1}$ is lower than the minimum level for transmission $\Delta_{i,m}$, or the buffer is empty. On the other hand, cases **Cases 2–4** are classified based on the status of a data packet arriving during the frame.

Case 1 ($b_1 < \Delta_{i,m}$ or $d_1 = 0$): In this case, no transmission happens, since node $U_{i,m}$ does not have sufficient energy to send data. Hence, the transitions from $\mathcal{B}_{i,m}^{n-1} = b_1$ to all states $\mathcal{B}_{i,m}^{n} = b_2$ with $b_2 > b_1$ states (and no state with lower energy levels) is possible, which happens with probability of $P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1)\beta}{\xi}\right\rfloor\right)$. Furthermore, the transition from state $\mathcal{D}_{i,m}^{n-1} = d_1$ to states $\mathcal{D}_{i,m}^{n} = d_2$ with $d_2 \geq d_1$ is also possible, because of the sporadic behavior assumption of the data packets, and that it is possible for only one data packet to arrive in the a typical frame, which happens with the probability of $\alpha_{i,m}$. Thus, $\Theta_{i,m}^{d_2|d_1,b_2|b_1}$ can be written as

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = \mathbb{J}_{\alpha_{i,m}}(d_2-d_1)P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1)\beta}{\xi}\right\rfloor\right)\mathbb{I}(b_2-b_1), \tag{B.1}$$

where $\mathbb{J}_{\alpha_{i,m}}(d)$ is the probability mass function of the number of data packets arriving in a frame, defined as

$$\mathbb{J}_{\alpha_{i,m}}(d) = \begin{cases} \alpha_{i,m}, & d = 1, \\ 1-\alpha_{i,m}, & d = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{B.2}$$

Moreover, $\mathbb{I}(b)$ is a binary indicator function, defined as

$$\mathbb{I}(b) = \begin{cases} 1, & b \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{B.3}$$

Note that (B.1) also holds for all the states with $d_1 = 0$, since no transmission occurs when no data packets exist in the IoT node buffer. Furthermore, the maximum battery level is $B_{\max}$. Then, those transitions from state $\mathcal{B}_{m,i}^{n-1} = b_1$ to states with energy level higher than $B_{\max}$ must be directed to $\mathcal{B}_{i,m}^{n} = B_{\max}$, which is due to battery saturation. Thus, for $b_2 = B_{\max}$, $\Theta_{i,m}^{d_2|d_1,b_2|b_1}$ can be written as

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = \mathbb{J}_{\alpha_{i,m}}(d_2-d_1)\sum_{k=B_{\max}}^{\infty} P_E^{pkt}\left(\left\lfloor\frac{(k-b_1)\beta}{\xi}\right\rfloor\right). \tag{B.4}$$

Case 2 ($b_1 \geq \Delta_{i,m}$ and $d_1 \neq 0$ and $d_2 < d_1$): In this case, because there is at least one data packet in the buffer and the IoT node has sufficient energy for transmission, it transmits a data packet. Notably, since $d_2 < d_1$, then no data packet has arrived. Therefore, the transition from state $\mathcal{B}_{i,m}^{n-1} = b_1$ to state $\mathcal{B}_{i,m}^{n} = b_2$ is equivalent to the arrival of $b_2 - b_1 + \Delta_{i,m}$ energy packets, which happens with probability of $P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right)$. Note that transition to states $\mathcal{B}_{i,m}^{n} = b_2$ with $b_2 < b_1 - \Delta_{i,m}$ is impossible, since maximum energy level reduction is $\Delta_{i,m}$. Furthermore, the probability of no data packet arriving is $1-\alpha_{i,m}$. Hence, for $b_2 < B_{\max}$, $\Theta_{i,m}^{d_2|d_1,b_2|b_1}$ can be written as

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = (1-\alpha_{i,m})\times P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right)\mathbb{I}(b_2-b_1+\Delta_{i,m}). \tag{B.5}$$

Note that the maximum battery level is $B_{\max}$, and those transitions from state $\mathcal{B}_{m,i}^{n-1} = b_1$ to states with energy level higher than $B_{\max}$ must be directed to $\mathcal{B}_{i,m}^{n} = B_{\max}$, which is due to battery saturation. Therefore, for $b_2 = B_{\max}$ $\Theta_{i,m}^{d_2|d_1,b_2|b_1}$ can be written as

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = (1-\alpha_{i,m})\sum_{k=B_{\max}}^{\infty} P_E^{pkt}\left(\left\lfloor\frac{(k-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right). \tag{B.6}$$

Case 3 ($b_1 \geq \Delta_{i,m}$ and $d_1 \neq 0$ and $d_2 = d_1$): Similar to **Case 1**, in this case, because there is at least one data packet in the buffer and the IoT node has sufficient energy for transmission, it transmits a data packet. However, since $d_2 = d_1$, a data packet must have also arrived. In turn, the transition from state $\mathcal{B}_{m,i}^{n-1} = b_1$ to state $\mathcal{B}_{i,m}^{n} = b_2$ is equivalent to the arrival of $b_2 - b_1 + \Delta_{i,m}$ energy packets, which happens with the probability of $P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right)$ and transition to states $\mathcal{B}_{i,m}^{n} = b_2$ with $b_2 < b_1 - \Delta_{i,m}$ is impossible due to the maximum energy level reduction. Furthermore, the probability of arriving a data packet is $\alpha_{i,m}$. Hence, for $b_2 < B_{\max}$, $\Theta_{i,m}^{d_2|d_1,b_2|b_1}$ is written as

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = \alpha_{i,m}P_E^{pkt}\left(\left\lfloor\frac{(b_2-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right)\mathbb{I}(b_2-b_1+\Delta_{i,m}). \tag{B.7}$$

Similar to the explanation provided for **Case 2**, for $b_2 = B_{\max}$, $\Theta_{i,m}^{d_2|d_1,b_2|b_1}$ can be written as

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = \alpha_{i,m}\sum_{k=B_{\max}}^{\infty} P_E^{pkt}\left(\left\lfloor\frac{(k-b_1+\Delta_{i,m})\beta}{\xi}\right\rfloor\right). \tag{B.8}$$

Case 4 ($b_1 \geq \Delta_{i,m}$ and $d_1 \neq 0$ and $d_2 > d_1$): This case is impossible, since moving from state $\mathcal{D}_{i,m}^{n-1} = d_1$ to states $\mathcal{D}_{i,m}^{n} = d_2$ with $d_2 > d_1$ entails the arrival of more than one data packet in a frame. Hence, in this case, $\Theta_{i,m}^{d_2|d_1,b_2|b_1} = 0$. To simplify the **Cases 2** to **4**, they can be combined into the following two cases by exploiting the function $\mathbb{J}_{\alpha_{i,m}}(d)$, as

- $\mathbf{b_1 \geq \Delta_{i,m}}$ and $\mathbf{d_1 \neq 0}$ and $\mathbf{b_2 < B_{\max}}$:

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = \mathbb{J}_{\alpha_{i,m}}(d_2 - d_1 + 1)\times P_E^{pkt}\left(\left\lfloor \frac{(b_2 - b_1 + \Delta_{i,m})\beta}{\xi} \right\rfloor\right)\mathbb{I}(b_2 - b_1 + \Delta_{i,m}). \tag{B.9}$$

- $\mathbf{b_1 \geq \Delta_{i,m}}$ and $\mathbf{d_1 \neq 0}$ and $\mathbf{b_2 = B_{\max}}$:

$$\Theta_{i,m}^{d_2|d_1,b_2|b_1} = \mathbb{J}_{\alpha_{i,m}}(d_2 - d_1 + 1)\sum_{k=B_{\max}}^{\infty} P_E^{pkt}\left(\left\lfloor \frac{(k - b_1 + \Delta_{i,m})\beta}{\xi} \right\rfloor\right). \tag{B.10}$$

■

APPENDIX C : PROOF OF LEMMA 3

Proof: To obtain the packet transmission cycle, the PDF of a typical packet transmission time $\Phi_{i,m}$ must first be derived, which is a random variable indicating the time required for a packet to be sent by IoT node $U_{i,m}$. Note that $\Phi_{i,m} = T_f$ provided that the IoT node has sufficient energy at the first attempt of transmission (i.e. if its battery has at least $\Delta_{i,m}$ energy levels when a packet is ready for transmission in the buffer). Accordingly, $\Phi_{i,m} = lT_f$ if IoT node $U_{i,m}$ does not have sufficient energy for the first $l-1$ frames, and harvests sufficient energy in l^{th} frame. In addition, the probability that the battery level is less than $\Delta_{i,m}$ for $l-1$ successive frames, but at least $\Delta_{i,m}$ is the l^{th} frame is geometrically distributed as

$$P_{i,m}(n) = (1 - \Pi_{i,m})^{l-1}\,\Pi_{i,m}, \tag{C.1}$$

in which $\Pi_{i,m}$ is the probability that $U_{i,m}$ has a battery level at least $\Delta_{i,m}$ at the beginning of a typical frame. Furthermore, $\Pi_{i,m}$ can be written as

$$\Pi_{i,m} = \sum_{d=1}^{\infty}\sum_{b=\Delta_{i,m}}^{B_{\max}} \pi_{i,m}(d,b). \tag{C.2}$$

Additionally, note that a transmission cycle $\mathcal{T}$ can be expressed as the sum of K independent and identical geometric random variables, and thus has a negative binomial distribution [65]. Hence, the PDF of $U_{i,m}$'s packet transmission cycle is given as

$$f_{i,m}^{\mathcal{T}}(t) = \sum_{k=K}^{\infty}\binom{k-1}{K-1}\Pi_{i,m}^{K}\,(1-\Pi_{i,m})^{k-K}\,\delta(t - kT_f), \tag{C.3}$$

in which $\delta(.)$ is the Dirac delta function. Lastly, the mean of the transmission cycle can accordingly be written as [65]

$$\mu_{\mathcal{T}_{i,m}} = \frac{K}{\Pi_{i,m}}T_f, \tag{C.4}$$

and this completes the proof. ■

APPENDIX D : PROOF OF LEMMA 4

Proof: By definition, reliability is the probability that a typical packet is delivered successfully to the BS. Since transmission diversity is adopted to enhance the reliability, a data packet is delivered successfully if at least one packet among the K transmitted replica packets is received successfully. Denoting E_{suc} as the event of receiving one replica of a typical packet successfully, and $\bar{E}_{suc}$ as its complement, then the reliability for IoT node $U_{i,m}$ can be written as

$$\begin{aligned}\mathcal{R}_{i,m}^{\text{NOMA}} &= 1 - \left(\Pr\left(\bar{E}_{suc}\right)\right)^{K} \\ &= 1 - \left(1 - \Pr(E_{suc})\right)^{K}.\end{aligned} \tag{D.1}$$

When $U_{i,m}$ transmits a data packet, the packet is delivered successfully if no IoT node from another cluster selects the same RB (i.e. no inter-cluster collision happens), and no decoding error occurs at the BS. Defining E_m^{NIC} as the event of no inter-cluster collision (NIC), and $E_{i,m}^{NDE}$ as that of no decoding error (NDE) at node $U_{i,m}$, then $\Pr(E_{suc})$ in (D.1) can be written as

$$\Pr(E_{suc}) = \Pr\left(E_m^{NIC}\right)\Pr\left(E_{i,m}^{NDE}\right). \tag{D.2}$$

Now, let the number of IoT clusters that are able to transmit data packets (i.e. clusters with at least one node having at least one data packet in its buffer and a sufficient energy level to transmit)—except the underlying cluster m—be denoted $\bar{N}_m$. Then, $\Pr\left(E_m^{NIC}\right)$ can be conditioned on $\bar{N}_m$ as

$$\Pr\left(E_m^{NIC}\right) = \sum_{l=0}^{M-1}\Pr\left(E_m^{NIC} \mid \bar{N}_m = l\right)\Pr\left(\bar{N}_m = l\right). \tag{D.3}$$

To obtain $\Pr\left(\bar{N}_m = l\right)$, define $\phi_{l,m}$ as a set containing indices of all possible combinations of l IoT clusters from the total M clusters excluding the cluster of interest (i.e. cluster m). Mathematically, $\phi_{l,m} = \{(G, G') \mid G = (g_1, \cdots, g_l), G' = (g_{l+1}, \cdots, g_{M-1}), g_i \in \{1, \cdots, M\}, g_i \neq g_j, \forall i, j, 1 \leq i, j \leq l, g_i \neq m\}$, where g_i represents a cluster index, G is a set including l clusters from all existing M clusters excluding cluster m, and G' is the complement of G. In turn, $\phi_{l,m}$ is a set of all combinations of l clusters and their complement. Then, the probability that l IoT clusters, except the cluster of interest m, have data in their buffers can be expressed as

$$\Pr(\bar{N}_m = l) = \sum_{\forall\,(g_1,\cdots,g_l)\in\phi_{l,m}}\left(\prod_{g=g_1}^{g_l}\left(1 - \Psi_{1,g}^{NT}\Psi_{2,g}^{NT}\right)\prod_{\substack{g'=g_{l+1}\\ g'\neq m}}^{g_M}\Psi_{1,g'}^{NT}\Psi_{1,g'}^{NT}\right), \tag{D.4}$$

$$\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}\right) = \Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{i,m}^{S}\right)\Pr\left(E_{i,m}^{S} \mid E_{j,m}^{T}\right) + \Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{j,m}^{S}\right)\Pr\left(E_{j,m}^{S} \mid E_{j,m}^{T}\right) \tag{D.7}$$

$$dF_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma) = \frac{d_{i,m}^{\nu} d_{j,m}^{\nu} \sigma^2 \left(P_{i,m} d_{j,m}^{\nu} + P_{j,m} d_{i,m}^{\nu}\gamma\right) + \left(P_{i,m} P_{j,m} d_{i,m}^{\nu} d_{j,m}^{\nu}\right)}{\left(P_{i,m} d_{j,m}^{\nu} + P_{j,m} d_{i,m}^{\nu}\gamma\right)^2} e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_i}\gamma} \tag{D.10}$$

in which $\Psi_{1,g}^{NT}\Psi_{2,g}^{NT}$ is the probability that none of the two IoT nodes in a cluster g transmit data and $\Psi_{i,g}^{NT}$ ($i \in \{1,2\}$ $g \in \{1,\cdots,M\}$) is defined in (13). On the other hand, $\Pr(E_m^{NIC} \mid \bar{N}_m = l)$ can be obtained as

$$\begin{aligned}\omega_l &\triangleq \Pr\left(E_m^{NIC} \mid \bar{N}_m = l\right) \\ &= \begin{cases} 1, & l = 0, \\ \left(\frac{R-1}{R}\right)^l, & \text{otherwise}, \end{cases}\end{aligned} \tag{D.5}$$

which indicates that there is no collision on the selected RB if all l IoT nodes select their RBs among the remaining $(r-1)$ RBs. For the case of $l = 0$, no collision happens, since no other IoT clusters transmit any data packets. Eventually, by substituting (D.4) and (D.5) into (D.3), the expression of $\Pr(E_m^{NIC})$ is obtained as given in (16).

The other probability metric (i.e. $\Pr\left(E_{i,m}^{NDE}\right)$) must be derived to complete the proof. Now, define $E_{j,m}^{T}$ as the event of data packet transmission by $U_{j,m}$, for $j \in \{1,2\}$ and $j \neq i$, while $\bar{E}_{j,m}^{T}$ as its complement. Then, conditioning on those events, $\Pr(E_{i,m}^{NDE})$ can be written as

$$\begin{aligned}\Pr\left(E_{i,m}^{NDE}\right) = &\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}\right)\Pr\left(E_{j,m}^{T}\right) \\ &+ \Pr\left(E_{i,m}^{NDE} \mid \bar{E}_{j,m}^{T}\right)\Pr\left(\bar{E}_{j,m}^{T}\right).\end{aligned} \tag{D.6}$$

By considering (13), it can be easily verified that $\Pr\left(E_{j,m}^{T}\right) = 1 - \Psi_{j,m}^{NT}$ and $\Pr\left(\bar{E}_{j,m}^{T}\right) = \Psi_{j,m}^{NT}$. Now, note that $\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}\right)$ is different for the stronger and weaker nodes. Hence, it must be determined whether $U_{i,m}$ is the stronger node or not. Recall that in subsection II-A, node $U_{1,m}$ has been defined as the stronger node. Moreover, since the channel gains are stochastic, the stronger node must be determined by the BS in the preamble phase in each frame. Therefore, node $U_{i,m}$ is determined as stronger if its received signal power at the BS in the preamble phase is found to be higher than the other node in the cluster. Defining $E_{U_i}^{S}$ as the event that the BS finds $U_{i,m}$ as the stronger node in the preamble phase which is equivalent to receiving more power from $U_{i,m}$ than $U_{j,m}$ ($i,j \in \{1,2\}, i \neq j$), and by conditioning on the stronger node perceived by the BS, $\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}\right)$ can be written as given in (D.7).

Furthermore, considering (3), $\Pr(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{i,m}^{S})$ can be obtained as

$$\begin{aligned}&\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{i,m}^{S}\right) \\ &= \int_0^{\infty} \Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{i,m}^{S}, \gamma_{i,m} = \gamma\right) dF_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma) \\ &= \int_0^{\infty} \Big(1 - \Upsilon\left(\gamma, n_b, n_d\right)\Big) dF_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma) \\ &= 1 - \int_0^{\infty} \Upsilon\left(\gamma, n_b, n_d\right) dF_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma).\end{aligned} \tag{D.8}$$

Additionally, $F_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma)$ can be expressed as

$$\begin{aligned}&F_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma) \\ &= \Pr\left(\gamma_{i,m} < \gamma \mid E_{i,m}^{S}\right) \\ &\overset{a}{=} \Pr\left(\frac{P_{i,m}|h_{i,m}|^2}{P_{j,m}|h_{j,m}|^2 + \sigma^2} < \gamma \mid E_{i,m}^{S}\right) \\ &= \int_0^{\infty} F_{|h_{i,m}|^2|E_{i,m}^{S}}\left(\frac{\gamma}{P_{i,m}}(P_{j,m}\hbar + \sigma^2)\right) f_{|h_{j,m}|^2|E_{i,m}^{S}}(\hbar) d\hbar,\end{aligned} \tag{D.9}$$

Note that the equality a is from the fact that when $U_{i,m}$ is the stronger node, the BS directly decodes its signal without applying SIC, as per **Remark 1**. Furthermore, $F_{|h_{i,m}|^2|E_{i,m}^{S}}(\gamma) = F_{|h_{i,m}|^2}(\gamma) = 1 - e^{-d_{i,m}^{\nu}\gamma}$ and $f_{|h_{j,m}|^2|E_{i,m}^{S}}(\hbar) = f_{|h_{j,m}|^2}(\hbar) = d_{j,m}^{\nu} e^{-d_{j,m}^{\nu}\hbar}$ (for $i,j \in \{1,2\}, i \neq j$) is the conditional PDF of $U_{j,m}$'s channel gain when BS received stronger power from $U_{i,m}$. Substituting the aforementioned functions into (D.9) and exploiting conditional expectations, $dF_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma)$ is obtained as given in (D.10). In turn, $\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{i,m}^{S}\right)$ can be written as

$$\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{i,m}^{S}\right) = 1 - \overline{\Upsilon}_{i|i,m}(n_b, n_d), \tag{D.11}$$

where $\overline{\Upsilon}_{i|i,m}(n_b, n_d) = \int_0^{\infty} \Upsilon(\gamma, n_b, n_d) dF_{\gamma_{i,m}|E_{i,m}^{S}}(\gamma)$. Deriving $\Pr\left(E_{i,m}^{S} \mid E_{j,m}^{T}\right)$ is straightforward, since

$$\begin{aligned}&\Pr\left(E_{i,m}^{S} \mid E_{j,m}^{T}\right) = \Pr\left(P_{i,m}|h_{i,m}|^2 > P_{j,m}|h_{j,m}|^2\right) \\ &= \int_0^{\infty} \Pr\left(|h_{i,m}|^2 > \frac{P_{j,m}}{P_{i,m}}|h_{j,m}|^2 \,|\, |h_{j,m}|^2 = \hbar\right) f_{|h_{j,m}|^2}(\hbar) d\hbar \\ &= \int_0^{\infty} e^{-\frac{d_{i,m}^{\nu} P_{j,m}}{P_{i,m}}\hbar} d_{j,m}^{\nu} e^{-d_{j,m}^{\nu}\hbar} d\hbar \\ &= \frac{P_{i,m} d_{j,m}^{\nu}}{P_{i,m} d_{j,m}^{\nu} + P_{j,m} d_{i,m}^{\nu}}.\end{aligned} \tag{D.12}$$

Two additional metrics in (D.7) must be derived, which are $\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{j,m}^{S}\right)$ and $\Pr\left(E_{j,m}^{S} \mid E_{j,m}^{T}\right)$. In a similar

$$\begin{aligned}\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{j,m}^{S}\right) &\overset{b}{=} \Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{j,m}^{S}, E_{j,m}^{NDE}\right) \Pr\left(E_{j,m}^{NDE} \mid E_{j,m}^{T}, E_{j,m}^{S}\right) \\ &\overset{c}{=} \int_{0}^{\infty} \Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{S}, E_{j,m}^{NDE}, \gamma_{i,m}=\gamma\right) dF_{\gamma_{i,m}|E_{j,m}^{S},E_{j,m}^{NDE}}(\gamma)\Big(1-\overline{\Upsilon}_{j|j,m}(n_b,n_d)\Big) \\ &= \int_{0}^{\infty} \Big(1-\Upsilon(\gamma,n_b,n_d)\Big) dF_{\gamma_{i,m}|E_{j,m}^{S},E_{j,m}^{NDE}}(\gamma)\Big(1-\overline{\Upsilon}_{j|j,m}(n_b,n_d)\Big)\end{aligned} \tag{D.13}$$

$$\begin{aligned}\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{j,m}^{S}\right) &= \int_{0}^{\infty} \Big(1-\Upsilon(\gamma,n_b,n_d)\Big) dF_{\gamma_{i,m}|E_{j,m}^{S},E_{j,m}^{NDE}}(\gamma)\Big(1-\overline{\Upsilon}_{j|j,m}(n_b,n_d)\Big) \\ &= \Big(1-\overline{\Upsilon}_{i|j,m}(n_b,n_d)\Big)\Big(1-\overline{\Upsilon}_{j|j,m}(n_b,n_d)\Big)\end{aligned} \tag{D.15}$$

$$\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}\right) = \frac{P_{i,m}d_{j,m}^{\nu}\left(1-\overline{\Upsilon}_{i|i,m}(n_b,n_d)\right)}{P_{j,m}d_{i,m}^{\nu}+P_{i,m}d_{j,m}^{\nu}} + \frac{P_{j,m}d_{i,m}^{\nu}\Big(1-\overline{\Upsilon}_{i|j,m}(n_b,n_d)\Big)\Big(1-\overline{\Upsilon}_{j|j,m}(n_b,n_d)\Big)}{P_{j,m}d_{i,m}^{\nu}+P_{i,m}d_{j,m}^{\nu}} \tag{D.17}$$

manner to the derivation of (D.8), $\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{S}\right)$ can be obtained as given in (D.13). Note that equality b is conditioned on the event of successfully decoding of $U_{j,m}$'s signal. This is because in this case, successfully decoding of $U_{i,m}$'s signal is dependent on successfully decoding of $U_{j,m}$'s signal due to applying SIC. It is assumed that erroneously decoding of $U_{j,m}$'s signal results in error on decoding of $U_{i,m}$'s signal. Hence, the remaining part due to conditioning in equality b is ignored. Furthermore, the expression of $\Pr\left(E_{j,m}^{NDE} \mid E_{j,m}^{S}\right)$ can be derived exactly as in (D.7) through (D.10), where $\overline{\Upsilon}_{j|j,m}(n_b,n_d) = \int_0^{\infty} \Upsilon(\gamma,n_b,n_d) dF_{\gamma_{j,m}|E_{j,m}^{S}}(\gamma)$, which is specified by $\Big(1-\overline{\Upsilon}_{j|j,m}(n_b,n_d)\Big)$ in equality c. Note that $F_{\gamma_{i,m}|E_{j,m}^{S},E_{j,m}^{NDE}}(\gamma)$ is easily obtained as

$$\begin{aligned}F_{\gamma_{i,m}|E_{j,m}^{S},E_{j,m}^{NDE}}(\gamma) &= \Pr\left(\gamma_{i,m} < \gamma \mid E_{j,m}^{S}, E_{j,m}^{NDE}\right) \\ &= \Pr\left(\frac{P_{i,m}|h_{i,m}|^2}{\sigma^2} < \gamma\right) \\ &= 1 - e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\gamma},\end{aligned} \tag{D.14}$$

and hence, $dF_{\gamma_{i,m}|E_{j,m}^{S},E_{j,m}^{NDE}}(\gamma) = \frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}} e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\gamma}$. Thus, $\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T}, E_{j,m}^{S}\right)$ from (D.13) is obtained as given in (D.15), where $\overline{\Upsilon}_{i|j,m}(n_b,n_d) = \frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}} \int_0^{\infty} \Upsilon(\gamma,n_b,n_d) e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\gamma} d\gamma$. Also, $\Pr\left(E_{j,m}^{S} \mid E_{j,m}^{T}\right)$ is derived in the same way as (D.11), and equals

$$\Pr\left(E_{j,m}^{S} \mid E_{j,m}^{T}\right) = \frac{P_{j,m}d_{i,m}^{\nu}}{P_{j,m}d_{i,m}^{\nu}+P_{i,m}d_{j,m}^{\nu}}. \tag{D.16}$$

Now, by substituting (D.11), (D.12), (D.15), (D.16) into (D.7), $\Pr(E_{i,m}^{NDE} \mid E_{j,m}^{T})$ can be written as given in (D.17).

To obtain $\Pr\left(E_{i,m}^{NDE} \mid \bar{E}_{j,m}^{T}\right)$ in (D.6), note that if $U_{j,m}$ does not transmit any data, then node $U_{i,m}$ does not experience any interference from it. In this case,

$$\begin{aligned}&\Pr\left(E_{i,m}^{NDE} \mid E_{j,m}^{T'}\right) \\ &= \int_{0}^{\infty} \Pr\left(E_{i}^{D} \mid E_{j,m}^{T'}, \gamma_{i,m}=\gamma\right) f_{\gamma_{i,m}}(\gamma) d\gamma \\ &= \int_{0}^{\infty} \Big(1-\Upsilon(\gamma,n_b,n_d)\Big) f_{\gamma_{i,m}}(\gamma) d\gamma \\ &= 1 - \int_{0}^{\infty} \Upsilon(\gamma,n_b,n_d) \frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}} e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\gamma} d\gamma \\ &= 1 - \overline{\Upsilon}_{i,m}(n_b,n_d),\end{aligned} \tag{D.18}$$

where $\overline{\Upsilon}_{i,m}(n_b,n_d) = \int_0^{\infty} \Upsilon(\gamma,n_b,n_d) \frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}} e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\gamma} d\gamma$ and it is equal to $\overline{\Upsilon}_{i|j,m}(n_b,n_d)$. Note that in the case of no RA collision between IoT nodes, then $\gamma_{i,m} = \frac{P_{i,m}|h_{i,m}|^2}{\sigma^2}$. Hence, $f_{\gamma_{i,m}}(\gamma)$ in (D.18) equals $f_{\gamma_{i,m}}(\gamma) = \frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}} e^{-\frac{d_{i,m}^{\nu}\sigma^2}{P_{i,m}}\gamma}$; otherwise, $\Pr\left(E_{i,m}^{NDE}\right) = 0$. Finally, $\Pr\left(E_{i,m}^{NDE}\right)$ is obtained as given in (18) by substituting (D.17) and (D.18) into (D.6). Finally, to determine $\mathcal{R}_{i,m}^{\text{NOMA}}$ and complete the proof, substitute $\Pr\left(E_{i,m}^{NDE}\right)$ and (D.6) into (D.2) and then using (D.1) to obtain (15). ■

APPENDIX E : PROOF OF LEMMA 5

Proof: The derivation of reliability for the RA-OMA scenario is similar to the RA-NOMA scenario. As in the RA-NOMA scenario, two different impermanent factors should be considered, inter-cluster collisions and decoding errors. Due to employing transmission diversity, the reliability for a typical IoT node U_i (for $i \in \{1,\cdots,N\}$) can be written as

$$\begin{aligned}\mathcal{R}_{i}^{\text{OMA}} &= 1 - \Big(\Pr\left(\bar{E}_{suc}\right)\Big)^{K} \\ &= 1 - \Big(1 - \Pr\left(E_{suc}\right)\Big)^{K},\end{aligned} \tag{E.1}$$

where E_{suc} is the event of receiving one replica of a typical packet successfully, and $\bar{E}_{suc}$ is its complement. Similar

to deriving reliability in the RA-NOMA scenario, define two events, namely E^{NIC} as the event of no inter-cluster collisions, and E_i^{NDE} as that of no decoding error. Then, then $\Pr\left(E_{suc}\right)$ can be written as

$$\Pr\left(E_{suc}\right) = \Pr\left(E_i^{NIC}\right)\Pr\left(E_i^{NDE}\right). \tag{E.2}$$

Then, $\Pr\left(E_i^{NIC}\right)$ can be written as

$$\Pr\left(E_i^{NIC}\right) = \sum_{l=0}^{N-1} \Pr\left(E_i^{NIC} \mid \bar{N}=l\right)\Pr\left(\bar{N}=l\right), \tag{E.3}$$

where $\bar{N}$ is the number of IoT nodes that are able to transmit data, except the underlying node U_i. To obtain $\Pr\left(\bar{N}_m=l\right)$, define ϕ_l as a set containing indices all possible combinations of l IoT nodes excluding that of interest (i.e. U_i). Mathematically, $\phi_l = \{(F,F') \mid F=(f_1,\cdots,f_l), F'=(f_{l+1},\cdots,f_{N-1}),\ f_n \in \{1,\cdots,N\},\ f_n \neq f_j, \forall n,j, 1 \leq n,j \leq l, f_n \neq i\}$, where f_n represents the node index, F is a set including l nodes from all existing N nodes excluding node i, and F' is the complement of F. In turn, ϕ_l is a set of all combinations of l nodes and their complement. Then, the probability that l IoT nodes, except the node of interest U_i, have data in their buffers can be expressed as

$$\Pr(\bar{N}_m=l) = \sum_{\forall(f_1,\cdots,f_l)\in\phi_l}\left(\prod_{f=f_1}^{f_l}\left(1-\Psi_f^{NT}\right)\prod_{\substack{f'=f_{l+1}\\ f'\neq i}}^{f_N}\Psi_{f'}^{NT}\right), \tag{E.4}$$

in which Ψ_s^{NT} is the probability that U_s does not transmit any data packet. Additionally, $\Pr\left(E_i^{NIC} \mid \bar{N}=l\right) = \omega_l$ where ω_l is given in (17). Eventually, $\Pr(E_i^{NIC})$ in (E.3) is obtained as

$$\Pr\left(E_i^{NIC}\right) = \sum_{l=0}^{N-1}\omega_l\left(\sum_{\forall(f_1,\cdots,f_l)\in\phi_l}\left(\prod_{f=f_1}^{f_l}\left(1-\Psi_f^{NT}\right)\prod_{\substack{f'=f_{l+1}\\ f'\neq i}}^{f_N}\Psi_{f'}^{NT}\right)\right). \tag{E.5}$$

Furthermore, $\Pr\left(E_i^{NDE}\right)$ can be derived as

$$\begin{aligned}\Pr\left(E_i^{NDE}\right) &= \int_0^\infty \Pr\left(E_i^{NDE} \mid \gamma_i=\gamma\right) f_{\gamma_i}(\gamma)d\gamma \\ &= \int_0^\infty \Big(1-\Upsilon(\gamma,n_b,n_d)\Big) f_{\gamma_i}(\gamma)d\gamma \\ &= 1-\int_0^\infty \Upsilon(\gamma,n_b,n_d)\frac{d_i^\nu\sigma^2}{P_i}e^{-\frac{d_i^\nu\sigma^2}{P_i}\gamma}d\gamma \\ &= 1-\overline{\Upsilon}_i(n_b,n_d),\end{aligned} \tag{E.6}$$

in which $\overline{\Upsilon}_i(n_b,n_d) = \int_0^\infty \Upsilon(\gamma,n_b,n_d)\frac{d_i^\nu\sigma^2}{P_i}e^{-\frac{d_i^\nu\sigma^2}{P_i}\gamma}d\gamma$. Note that in the case of no RA collision between the IoT nodes, $\gamma_i = \frac{P_i|h_i|^2}{\sigma^2}$, Hence, $f_{\gamma_i}(\gamma)$ in (E.6) equals $f_{\gamma_i}(\gamma) = \frac{d_i^\nu\sigma^2}{P_i}e^{-\frac{d_i^\nu\sigma^2}{P_i}\gamma}$; otherwise, $\Pr\left(E_i^{NDE}\right) = 0$. Therefore, by substituting (E.5) and (E.6) into (E.2) and then into (E.1), $\mathcal{R}_i^{\text{OMA}}$ in (21) is obtained, and this completes the proof. ■

REFERENCES

[1] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the Internet of Things: Communication technologies and challenges," IEEE Access, vol. 6, pp. 3619–3647, 2017.

[2] L. Chettri and R. Bera, "A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems," IEEE Internet of Things Journal, vol. 7, no. 1, pp. 16–32, 2020.

[3] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," IEEE Communications Magazine, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[4] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," IEEE Journal on Selected Areas in Communications, vol. 35, no. 10, pp. 2181–2195, Jul. 2017.

[5] H. Abumarshoud, H. Alshaer, and H. Haas, "Dynamic multiple access configuration in intelligent Lifi attocellular access points," IEEE Access, vol. 7, pp. 62 126–62 141, Feb. 2019.

[6] M. W. Baidas, M. S. Bahbahani, E. Alsusa, K. Hamdi, and Z. Ding, "Joint D2D group association and channel assignment in uplink multi-cell NOMA networks: A matching-theoretic approach," IEEE Transactions on Communications, vol. 67, no. 12, pp. 8771–8785, Dec. 2019.

[7] M. W. Baidas, E. Alsusa, and Y. Shi, "Resource allocation for SWIPT-enabled energy-harvesting downlink/uplink clustered NOMA networks," Computer Networks, vol. 182, Dec. 2020.

[8] Z. Ding, P. Fan, and H. V. Poor, "Impact of non-orthogonal multiple access on the offloading of mobile edge computing," IEEE Transactions on Communications, vol. 67, no. 1, pp. 375–390, Jan. 2019.

[9] Y. Liu, Z. Qin, and Z. Ding, Non-Orthogonal Multiple Access for Massive Connectivity. Springer International Publishing, 2020.

[10] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," IEEE Communications Magazine, vol. 55, no. 9, pp. 55–61, 2017.

[11] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable 5G networks," IEEE Wireless Communications, vol. 24, no. 4, pp. 72–80, Aug. 2017.

[12] M. Imran, L. U. Khan, I. Yaqoob, E. Ahmed, M. A. Qureshi, and A. Ahmed, "Energy harvesting in 5G netowrks: Taxomnomy, requirements, challenges, and future directions," IEEE Wireless Communications, Oct. 2019, arXiv: https://arxiv.org/abs/1910.00785.

[13] J. Sachs, G. Wikstrom, T. Dudda, R. Baldemair, and K. Kittichokechai, "5G radio network design for ultra-reliable low-latency communication," IEEE Network, vol. 32, no. 2, pp. 24–31, Apr. 2019.

[14] S. Henry, A. Alsohaily, and E. S. Sousa, "5G is real: Evaluating the compliance of the 3GPP 5G new radio system with the ITU IMT-2020 requirements," IEEE Access, vol. 8, pp. 42 828–42 840, Mar. 2020.

[15] G. Hampel, C. Li, and J. Li, "5G ultra-reliable low-latency communications in factory automation leveraging licensed and unlicensed bands," IEEE Communications Magazine, vol. 57, no. 5, pp. 117–123, 2019.

[16] Z. Ma, M. Xiao, Y. Xiao, Z. Pang, H. V. Poor, and B. Vucetic, "High-reliability and low-latency wireless communication for Internet of Things: Challenges, fundamentals, and enabling technologies," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 7946–7970, Oct. 2019.

[17] M. A. Siddiqi, H. Yu, and J. Joung, "5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices," Electronics, vol. 8, no. 9, 2019, article No. 981. [Online]. Available: https://www.mdpi.com/2079-9292/8/9/981

[18] M. Fuentes, J. L. Carcel, C. Dietrich, L. Yu, E. Garro, V. Pauli, F. I. Lazarakis, O. Grondalen, O. Bulakci, J. Yu, W. Mohr, and D. Gomez-Barquero, "5G new radio evaluation against IMT-2020 key performance indicators," IEEE Access, vol. 8, pp. 110 880–110 896, Jun. 2020.

[19] M. Shirvanimoghaddam, M. S. Mohammadi, R. Abbas, A. Minja, C. Yue, B. Matuz, G. Han, Z. Lin, W. Liu, Y. Li, S. Johnson, and B. Vucetic, "Short block-length codes for ultra-reliable low latency communications," IEEE Communications Magazine, vol. 57, no. 2, pp. 130–137, 2019.

[20] R. Abbas, T. Huang, B. Shahab, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Grant-free non-orthogonal multiple access: A key enabler for 6G-IoT," ArXiv, 2020, Online: https://arxiv.org/abs/2003.10257.

[21] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1805–1838, May 2020.
[22] R. Kotaba, C. N. Manchon, N. M. K. Pratas, T. Balercia, and P. Popovski, "Improving spectral efficiency in URLLC via NOMA-based retransmissions," in Proc. of IEEE International Conference on Communications (ICC), 2019, pp. 1–7.
[23] M. Amjad and L. Musavian, "Performance analysis of NOMA for ultra-reliable and low-latency communications," in Proc. of IEEE GLOBECOM Workshops (GC Wkshps), 2018, pp. 1–5.
[24] X. Zhao and W. Chen, "Non-orthogonal multiple access for delay-sensitive communications: A cross-layer approach," IEEE Transactions on Communications, vol. 67, no. 7, pp. 5053–5068, 2019.
[25] L. Liu, M. Sheng, J. Liu, Y. Dai, and J. Li, "Stable throughput region and average delay analysis of uplink NOMA systems with unsaturated traffic," IEEE Transactions on Communications, vol. 67, no. 12, pp. 8475–8488, 2019.
[26] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, "On the performance of non-orthogonal multiple access in short-packet communications," IEEE Communications Letters, vol. 22, no. 3, pp. 590–593, 2018.
[27] E. Dosti, M. Shehab, H. Alves, and M. Latva-aho, "On the performance of non-orthogonal multiple access in the finite blocklength regime," Ad Hoc Networks, vol. 84, pp. 148–157, 2019.
[28] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "On the performance of massive grant-free NOMA," Proc. of IEEE International Symposium on Persona, Indoor, and Mobile Radio Communications (PIMRC), Oct. 2017.
[29] ——, "A novel analytical framework for massive grant-free NOMA," IEEE Transactions on Communications, vol. 67, no. 3, pp. 2436–2449, Nov. 2019.
[30] S. Dogan, A. Tusha, and H. Arslan, "NOMA with index modulation for uplink URLLC through grant-free access," IEEE Journal on Selected Topics in Signal Processing, vol. 13, no. 6, pp. 1249–1257, Oct. 2019.
[31] J. Choi, "NOMA-based random access with multichannel ALOHA," IEEE Journal on Selected Areas in Communications, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.
[32] S. A. Tegos, P. D. Diamantoulakis, A. S. Lioumpas, P. G. Sarigiannidis, and G. K. Karagiannidis, "Slotted ALOHA with NOMA for the next generation IoT," IEEE Transactions on Communications, vol. 68, no. 10, pp. 6289–6301, Oct. 2020.
[33] Y. Liang, X. Li, J. Zhang, and Z. Ding, "Non-orthogonal random access for 5G networks," IEEE Transactions on Wireless Communications, vol. 16, no. 7, pp. 4817–4831, Jul. 2017.
[34] J. Seo, B. C. Jung, and H. Jin, "Nonorthogonal random access for 5G mobile communication systems," IEEE Transactions on Vehicular Technology, vol. 67, no. 8, pp. 7867–7871, 2018.
[35] Y. Wang, T. Wang, Z. Yang, D. Wang, and J. Cheng, "Throughput-oriented non-orthogonal random access scheme for massive MTC networks," IEEE Transactions on Communications, vol. 68, no. 3, pp. 1777–1793, Mar. 2020.
[36] Q. Chen, H. Gao, Z. Cai, L. Cheng, and J. Li, "Energy-collision aware data aggregation scheduling for energy harvesting sensor networks," in Proc. of IEEE Conference on Computer Communications (INFOCOM), Apr. 2018, pp. 117–125.
[37] 3GPP TR 36.859 V0.1.0, "3GPP; technical specification group radio access network; study on downlink multiuser superposition transmission (MUST) for LTE," 2015. [Online]. Available: www.portal.3gpp.org
[38] A. M. Ortiz, D. Hussein, S. Park, S. N. Han, and N. Crespi, "The cluster between Internet of Things and social networks: Review and research challenges," IEEE Internet of Things Journal, vol. 1, pp. 206–215, 2014.
[39] M. Nitti, L. Atzori, and I. P. Cvijikj, "Friendship selection in social Internet of Things: Challenges and possible strategies," IEEE Internet of Things Journal, vol. 2, no. 3, pp. 240–247, Jun. 2015.
[40] Q. Ju and Y. Zhang, "Clustered data collection for Internet of Batteryless Things," IEEE Internet of Things Journal, vol. 4, no. 6, pp. 2275–2285, Dec. 2017.
[41] D. Wu, J. Shi, and N. Mamoulis, "Density-based place clustering using geo-social network data," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 5, pp. 838–851, May 2018.
[42] J. Zhang, X. Feng, and Z. Liu, "A grid-based clustering algorithm via load analysis for industrial Internet of Things," IEEE Access, vol. 6, pp. 13 117–13 128, Jan. 2018.
[43] M. D. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," IEEE Access, vol. 4, pp. 6325–6343, 2016.
[44] M. Zeng, A. Yadav, O. A. Dobre, and H. V. Poor, "Energy-efficient joint user-RB association and power allocation for uplink hybrid NOMA-OMA," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5119–5131, Jun. 2019.
[45] H. Tang, T. Qin, Z. Hui, P. Cheng, and W. Bai, "Design and implementation of a configurable and aperiodic pseudo random number generator in FPGA," in Proc. of IEEE 2nd International Conference on Circuits, System and Simulation (ICCSS), 2018, pp. 47–51.
[46] 3GPP TS 36.211 V13.2.0, "3rd generation partnership project; technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation," 2016. [Online]. Available: www.portal.3gpp.org
[47] S. Sesia, I. Toufik, and M. Baker, LTE - The UMTS Long Term Evolution: From Theory to Practice, 2nd Ed. Wiley, 2011.
[48] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2488–2524, 2019.
[49] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.
[50] W. J. Stewart, Probability, Markov Chains, Queues, and Simulation. Princeton University Press, 2009.
[51] H. Ming, Fundamentals of Matrix-Analytic Methods. Springer New York, 2014.
[52] S. M. Ross, Introduction to Probability Models, 10th Ed. Elsevier Academic Press, 2010.
[53] ETSI TR 136 942 V13.0.0, "3rd generation partnership project; LTE;Evolved Universal Terrestrial Radio Access (E-UTRA); radio frequency (RF) system scenarios," 2016. [Online]. Available: www.portal.3gpp.org
[54] ETSI TS 136 211 V11.0.0, "3rd generation partnership project; LTE;Evolved Universal Terrestrial Radio Access (E-UTRA); physical channels and modulation," 2012. [Online]. Available: www.portal.3gpp.org
[55] H. Tabassum, M. S. Ali, E. Hossain, M. J. Hossain, and D. I. Kim, "Uplink vs. downlink NOMA in cellular networks: Challenges and research directions," Proc. of IEEE Vehicular Technology Conference (VTC) - Spring, pp. 1–7, Jun. 2017.
[56] 3GPP TS 22.261 V17.3.0, "3rd generation Partnership Project; Technical Specification Group Services and System Aspects; Service requirements for the 5G system; stage 1," 2020. [Online]. Available: www.portal.3gpp.org
[57] P. Popovski, . Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," IEEE Transactions on Communications, vol. 67, no. 8, pp. 5783–5801, 2019.
[58] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 721–742, 2017.
[59] Y. Gao, B. Xia, K. Xiao, Z. Chen, X. Li, and S. Zhang, "Theoretical analysis of the dynamic decode ordering SIC receiver for uplink NOMA systems," IEEE Communications Letters, vol. 21, no. 10, pp. 2246–2249, 2017.
[60] ITU, "Telecommunication standardization sector of ITU, recommendation ITU-T E.800," in Definitions of terms related to quality of service, Jul. 2009, online: http://www.itu.int/rec/T-REC-E.800-200809-I/en.
[61] 3GPP TR 38.825 V16.0.0, "3rd generation partnership project; technical specification group radio access network; study on NR Industrial Internet of Things (IoT)," 2019. [Online]. Available: www.portal.3gpp.org
[62] G. L. Nemhauser, Integer and Combinatorial Optimization. Wiley, 1988.
[63] M. Schlueter, "MIDACO software performance on interplanetary trajectory benchmarks," Advances in Space Research, vol. 54, no. 4, pp. 744–754, 2014.
[64] MIDACO-Solver, "Mixed integer distributed ant colony optimization," Online: http://www.midaco-solver.com.
[65] A. Leon-Garcia, Probability, Statistics, and Random Processes For Electrical Engineering. Pearson, 2008.

MOHAMMAD REZA AMINI (SM'20) received the B.Eng. and M.Sc. in electrical and communication system engineering from Isfahan University of Technology (IUT) and Malek-Ashtar University of Technology, Iran, respectively. He also received his Ph.D. in Telecommunications from Isfahan University of Technology, Isfahan, Iran in 2018. He is currently an Assistant professor with the Department of Electrical Engineering in Islamic Azad University, Borujerd Branch, Iran. He is also an inspector of Iran's Standard Institute and accepted in National Foundation of Elites. He was a recipient of the Outstanding Teaching and Outstanding Researcher Awards in 2012, 2016, 2017 and 2018. His research interests include cognitive radio networks, system implementation and green and energy-harvesting networks.

MOHAMMED W. BAIDAS (M'05–SM'17) received the B.Eng. (Hons.) degree in communication systems engineering from the University of Manchester, Manchester, U.K., in 2005, the M.Sc. degree (with distinction) in wireless communications engineering from the University of Leeds, Leeds, U.K., in 2006, the M.S. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2009, and the Ph.D. degree in electrical engineering from Virginia Tech, Blacksburg, VA, USA, in 2012. Dr. Baidas was a Visiting Researcher with the University of Manchester in the academic years of 2015/2016 and 2018/2019. He is currently an Associate Professor with the Department of Electrical Engineering, Kuwait University, Kuwait, where he has been on the faculty since May 2012. He is also a frequent reviewer for several IEEE journals and international journals and conferences, with over 70 publications. His research interests include resource allocation and management in cognitive radio systems, game theory, cooperative communications and networking, and green and energy-harvesting networks. He also serves as a technical program committee member for various IEEE and international conferences. He was a recipient of the Outstanding Teaching Award of Kuwait University for the academic year of 2017/2018.