Performance Analysis of URLL Random-Access NOMA-Enabled IoT Networks with Short Packet and Diversity Transmissions

Mohammad Reza Amini[1], and Mohammed W. Baidas[2]
[1]Department of Electrical Engineering, Islamic Azad University, Borujerd, Iran
[2]Department of Electrical Engineering, College of Engineering and Petroleum, Kuwait University, Kuwait
(Email: mr.amini@iaub.ac.ir, and m.baidas@ku.edu.kw)

***Abstract*—Recently, non-orthogonal multiple-access (NOMA) has been proposed to improve spectrum-efficiency and throughput of 5G cellular networks, and is also considered a key-enabler for ultra-reliable and low-latency (URLL) communications. Moreover, the Internet-of-Things (IoT) paradigm has emerged to provide massive-connectivity for intelligent devices and systems, which entail spectrum-efficient transmission schemes. Hence, exploiting NOMA for URLL transmissions in IoT networks is inevitable. Furthermore, random-access (RA) techniques are also considered essential to enable massive URLL IoT networks, since they reduce signaling overhead and packet latency, especially when massive numbers of clustered IoT devices with sporadic traffic behaviour are considered. In this paper, the performance of uplink RA-NOMA in URLL IoT networks with short packet and diversity transmissions is analyzed. Specifically, network metrics—such as average packet latency, reliability, and GoodPut—are mathematically derived. Additionally, the effect of transmission diversity, and number of data bits per blocklength on the different network metrics has been extensively evaluated, illustrating several tradeoffs between the different network metrics as well as highlighting the importance of carefully selecting the network parameters to satisfy the URLL requirements.**

***Index Terms*—Internet-of-Things, low-latency, NOMA, packet latency, ultra-reliability.**

I. Introduction

The explosive growth in the number of smart mobile devices, applications, and services in the current 5G cellular and emerging Internet-of-Things (IoT) networks entails intelligent spectrum- and energy-efficient transmission techniques to meet the demands for massive connectivity, high data rates, and low-latency. Also, the advent of non-orthogonal multiple-access (NOMA) has improved spectrum utilization by allowing multiple users to share the same resource blocks (RBs) simultaneously, while employing successive interference cancellation (SIC) for multi-user detection. Specifically, NOMA has been shown to outperform the conventional OMA schemes in terms of transmission latency and capacity [1]. Therefore, incorporating NOMA into IoT networks is expected to cater for the high spectral-efficiency and massive connectivity requirements [2]. On the other hand, ultra-reliable and low-latency (URLL) transmissions are considered to be the main features that distinguish 5G from previous generations, and serve as foundations for many 5G and IoT applications [3]. 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 packet. To realize low-latency, short packet transmissions via finite blocklength (FBL) codes can be utilized [4]. The combination of the aforementioned transmission techniques is extremely attractive to realize URLL transmissions in NOMA-based IoT networks. In general, IoT networks incorporate massive numbers of smart devices and systems that are characterized by sporadic traffic behavior and spectrum access, which motivates the use of random-access (RA) techniques. Hence, analyzing RA-NOMA IoT networks with URLL requirements is of paramount importance.

Recently, several research works have considered NOMA communications. Despite the importance of URLL communications, only few studies have considered NOMA-enabled URLL communications in grant-free scenarios, which are essential for providing massive connectivity, minimizing signaling overheads, and reducing latency, especially in URLL IoT networks [5]. For instance, the authors in [6] 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 proposed scheme has been shown to be primarily dominated by the collision probability. RA with multichannel ALOHA in uplink (UL) NOMA systems is considered in [7]. 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. Based on NOMA, a non-orthogonal random-access (NORA) scheme has been proposed in [8] to alleviate the access congestion problem. Particularly, NORA is based on the difference of times 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 access success probability, preamble collision probability, and throughput. In [9], the authors proposed UL NORA techniques for 5G cellular networks. Particularly, the channel inversion technique is utilized, whereby 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 the maximum achievable throughput has been shown to exceed 0.7, in comparison to 0.368 of the conventional S-ALOHA technique. A throughput-oriented NORA scheme is proposed in [10]

978-1-7281-6535-6/20/$31.00 ©2020 IEEE

for massive machine-type communications (mMTC) networks. Specifically, tagged preambles are employed by multiple MTC devices (MTCDs), forming power-domain NOMA groups. An optimization problem is formulated for throughput maximization, and a particle swarm optimization (PSO) algorithm is devised. The proposed scheme has been shown to outperform existing schemes in terms network throughput performance.

Due to the lack of studies in URLL grant-free NOMA in IoT networks, this paper considers an URLL-IoT network with UL RA-NOMA transmissions and clustered IoT nodes, where 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 the IoT nodes' sporadic traffic behavior. In deriving the 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 the transmission diversity, and number of transmitted data bits per blocklength on the RA-NOMA network performance metrics is investigated. *To the best of the authors' knowledge, no prior work has analyzed UL RA-NOMA IoT networks with URLL requirements, and provided analytical expressions and numerical comparisons for various IoT network metrics.*[1]

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. Finally, conclusions are drawn in Section V.

II. System Model

The system model is based on UL IoT network consisting of a BS, and N transmitting nodes with URLL transmission requirements[2]. The IoT nodes transmit their sporadic data packets to the BS in the FBL regime via RA-NOMA. The IoT nodes are paired in clusters of two nodes according to some predefined clustering strategic [12], and hence, there are $M = N/2$ clusters in the IoT network. 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 m^{th} cluster, for $m \in \{1,\dots,M\}$. 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. All the N links from the nodes to the BS experience independent but not necessarily identically distributed Rayleigh block fading, which are also assumed to be constant during each transmission frame due to the short packet transmission. The channel coefficient between $U_{i,m}$ and the BS is denoted $h_{i,m}$. In turn, the corresponding channel gain $|h_{i,m}|^2$ follows an exponential distribution with mean $d_{i,m}^{-\nu}$, where $d_{i,m}$ is the corresponding distance, while ν is the path-loss exponent. Moreover, 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 convenience, let $U_{1,m}$ and $U_{2,m}$ be the nodes with the higher and lower channel gains over the selected RB, respectively (i.e. $|h_{1,m}|^2 > |h_{2,m}|^2$), and hence $P_{1,m} \geq P_{2,m}$ [13].

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 [13]. 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

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

1) Frame Structure and Channel Access: The IoT nodes transmit their data packets in a frame-based structure, where the IoT nodes in the same cluster access the same RB by exploiting power-domain NOMA, while different clusters access the RBs randomly[3]. 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 sequence on the randomly selected RB r (for $r \in \{1,\dots,R\}$) to the BS[4]. 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. When the BS detects only one preamble sequence on a RB, it starts decoding the only received signal without applying SIC.

2) Link Specifications: To achieve UR communications, transmission diversity is adopted, in which multiple replicas of each frame, say K, are transmitted successively. For LL transmissions, the FBL codes are adopted [4], [15]. 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 $m \in \{1,\dots,M\}$) at the BS can be approximated as [16]

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

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.

[1] A longer version of this paper can be found in [11].

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

[3] From a practical perspective, this can be performed by feeding the random number generators with the same seed [14].

[4] The preamble phase enables the BS to detect the stronger/weaker nodes within each cluster in order and 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.

Remark 4: The packet arrival process of IoT node $U_{i,m}$ is assumed to be sporadic and independent of the other nodes. Particularly, it is modeled as a Poisson arrival process with rate λ_d, where each IoT node has a buffer in which the packets are stored and queued to be sent. It is further assumed that each data packet can be sent over one RB within T_d seconds.

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 lasts $K \times T_f$ seconds.

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 (in terms of non-redundant data bits per time unit) received successfully at the BS.

B. Average Packet Latency

The average packet latency for a typical IoT nodes' packet is the sum of packet transmission cycle and average waiting time in the buffer. The former lasts for $K \times T_f$ seconds, while the latter is derived according to queuing theory as follows. Let $\mathcal{D}^n_{i,m}$ be the number of data packets in the buffer of node $U_{i,m}$ just after that the n^{th} packet leaves the node's buffer. One can easily conclude that $\mathcal{D}^n_{i,m}$ constitutes a discrete-time Markov chain, and the underlying packet queue follows the $M/D/1$ queuing model, since the data arrival process is Poisson according to **Remark 4**, and the average packet latency of $K \times T_f$ seconds represents the service time. Then, the average number of packets in the buffer of a typical IoT node's is $\overline{\mathcal{D}} = \frac{\rho}{1-\rho}\left(1 - \frac{\rho}{2}\right)$ [17], in which $\rho = \lambda_d K T_f$ is the link utilization of the IoT node. Thus, according to Little's formulae [17], the average packet latency equals $\mathcal{L} = KT_f\left(\overline{\mathcal{D}}+1\right)$. Furthermore, the probability that a typical IoT node does not transmit data in a frame (due to not having any data packet in its buffer) is $\Pi_0 \triangleq 1 - \rho = 1 - \lambda_d K T_f$.

C. Reliability

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; RA collisions due to inter-cluster transmissions, and the decoding error at the BS. **Lemma 1** provides the reliability for an IoT node[5].

Lemma 1: The reliability of IoT node $U_{i,m}$ in a NOMA cluster m is derived as

$$\mathcal{R}_{i,m} = 1 - \left(1 - \Pr\left(E^{NIC}_m\right) \Pr\left(E^{NDE}_{i,m}\right)\right)^K. \tag{4}$$

In particular, $\Pr\left(E^{NIC}_m\right)$ is the probability of no inter-cluster collisions (NIC) experienced by cluster m over the selected RB, as given by (5), where ω_l is given as

$$\omega_l = \begin{cases} 1, & l = 0, \\ \left(\frac{R-1}{R}\right)^l, & \text{otherwise.} \end{cases} \tag{6}$$

Furthermore, $\Pr\left(E^{NDE}_{i,m}\right)$ is the probability of no decoding error (NDE), which is as given in (7). Also, in (7), $\overline{\Upsilon}_{i|i,m}(n_b, n_d)$ is given as (8), while $\overline{\Upsilon}_{i|j,m}(n_b, n_d)$ is given by

$$\overline{\Upsilon}_{i|j,m}(n_b, n_d) = \frac{d^\nu_{i,m}\sigma^2}{P_{i,m}} \int_0^\infty \Upsilon(\gamma, n_b, n_d) e^{-\frac{d^\nu_{i,m}\sigma^2}{P_{i,m}}\gamma} d\gamma. \tag{9}$$

Proof Sketch: By definition, the 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. Thus, deriving $U_{i,m}$'s reliability relates to formulating the probability of receiving one replica of a typical packet successfully at the BS. Now, when $U_{i,m}$ transmits a data

[5]Proof of **Lemma 1** is not given in this paper due to space limitation. Instead, a proof sketch is given, while the full proof can be found in [18].

$$\Pr\left(E^{NIC}_m\right) = \sum_{l=0}^{M-1} \omega_l \binom{M-1}{l} \left(1 - \Pi_0^2\right)^l \left(\Pi_0^2\right)^{M-1-l} \tag{5}$$

$$\Pr\left(E^{NDE}_{i,m}\right) = \frac{P_{i,m} d^\nu_{j,m}\left(1 - \overline{\Upsilon}_{i|i,m}(n_b, n_d)\right) + P_{j,m} d^\nu_{i,m}\left(1 - \overline{\Upsilon}_{i|j,m}(n_b, n_d)\right)\left(1 - \overline{\Upsilon}_{j|j,m}(n_b, n_d)\right)}{P_{j,m} d^\nu_{i,m} + P_{i,m} d^\nu_{j,m}}\left(1 - \Pi_0\right) + \left(1 - \overline{\Upsilon}_{i|j,m}(n_b, n_d)\right)\Pi_0 \tag{7}$$

$$\overline{\Upsilon}_{i|i,m}(n_b, n_d) = \int_0^\infty \Upsilon(\gamma, n_b, n_d) \frac{d^\nu_{i,m} d^\nu_{j,m} \sigma^2 \left(P_{i,m} d^\nu_{j,m} + P_{j,m} d^\nu_{i,m}\gamma\right) + \left(P_{i,m} P_{j,m} d^\nu_{i,m} d^\nu_{j,m}\right)}{\left(P_{i,m} d^\nu_{j,m} + P_{j,m} d^\nu_{i,m}\gamma\right)^2} e^{-\frac{d^\nu_{i,m}\sigma^2}{P_i}\gamma} d\gamma \tag{8}$$

packet, the packet is delivered successfully if no IoT node from another cluster transmits over the same RB (i.e. no inter-cluster collision happens), and no decoding error occurs at the BS. To calculate the probability that no inter-cluster collision happens, one must determine 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, except the underlying cluster m). Hence, by applying conditional expectation on such a random event, and incorporating the probability that none of the transmitting clusters select the same RB as cluster m, the probability of no inter-cluster collision is derived. To calculate the probability of no decoding error occurring at the BS, the order of SIC decoding at BS must be employed. Since the SIC order is determined by the strength of the preamble signals received at the BS, the stronger/weaker nodes are determined based on received preamble at the BS. Hence, by conditioning on the event that $U_{i,m}$'s signal is stronger than $U_{j,m}$'s ($\forall i,j \in \{1,2\}, i \neq j$), and exploiting the decoding error probability in the FBL regime (i.e. via (3)), the probability of no decoding error at the BS can be derived. Finally, the reliability can be obtained by combining the two derived probability metrics (i.e. no inter-cluster collisions and no decoding error at the BS).

■

D. GoodPut

As per the definition of GoodPut provided in subsection III-A, the GoodPut $\mathcal{G}_{i,m}$ of IoT node $U_{i,m}$ is defined as the average effective rate received successfully at the BS from $U_{i,m}$'s transmission. In turn, the successfully transferred "non-redundant" data bits per time unit from $U_{i,m}$ to the BS must be considered. By noting that during a transmission cycle, the number of effective successfully delivered bits is $n_d\mathcal{R}_{i,m}$, then the GoodPut of node $U_{i,m}$ is determined as

$$\mathcal{G}_{i,m} = \frac{n_d \mathcal{R}_{i,m}}{K T_f}. \tag{10}$$

IV. Numerical Results

This section numerically evaluates the effect of the number of replicas K of an IoT node's packets, and number of data bits n_d in each packet per blocklength on the network packet latency, reliability, and GoodPut. Table I summarizes the simulated network parameters [19][6]. In the simulations, and for simplicity, the $N = 50$ nodes are assumed to be split into two sets of nodes; near and far [21], while 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, $\forall m \in \{1,\ldots,25\}$[7].

TABLE I: Simulation Parameters

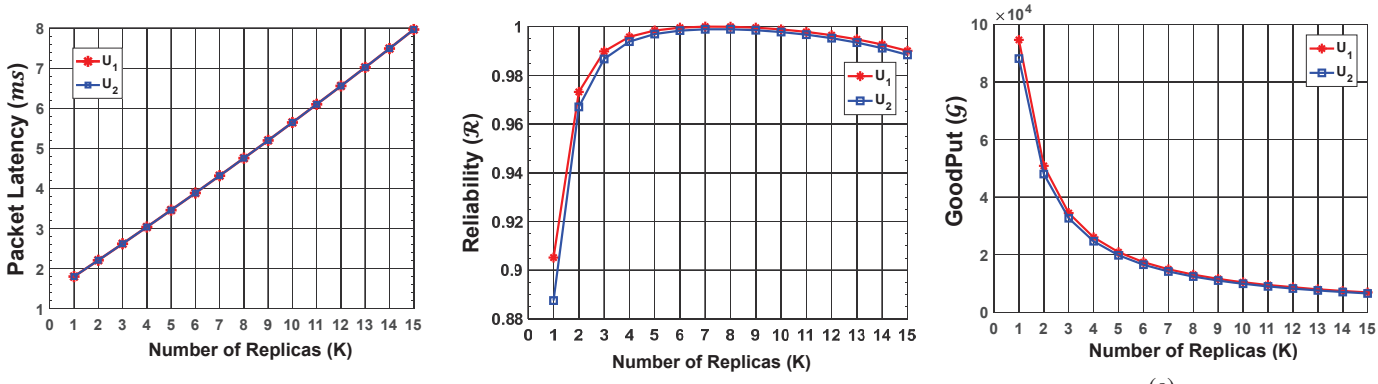
Parameter	Value	Parameter	Value
N	50	R	5 RBs
ν	3	λ_d	10 Packets/s
n_b	200	P_1, P_2	0.2 W
T_d, T_p	0.2 ms	d_1, d_2	0.7, 1 Km
W	1 MHz	N_0	-174 dBm/Hz

A. Effect of Number of Replicas K

The average packet latency as a function of K for both the stronger and weaker nodes is depicted in Fig. 1a. Clearly, the packet latency increases with the increase in K (as would be expected), which is due to the increase in the transmission cycle. Fig. 1b illustrates the nodes' reliability as a function of K for the stronger and weaker nodes. As can be seen, the reliability for both nodes increases sharply as K increases. This is expected since by transmitting multiple replicas of a typical packet, the number of successful transferred information bits increases. However, the reliability starts to decrease when the number of re-transmitted packets increases beyond $K = 8$. This is due to the effect of 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. Hence, the RA collisions due to inter-cluster interference increase. In turn, the optimal value of K must be carefully chosen according to the URLL requirements, the

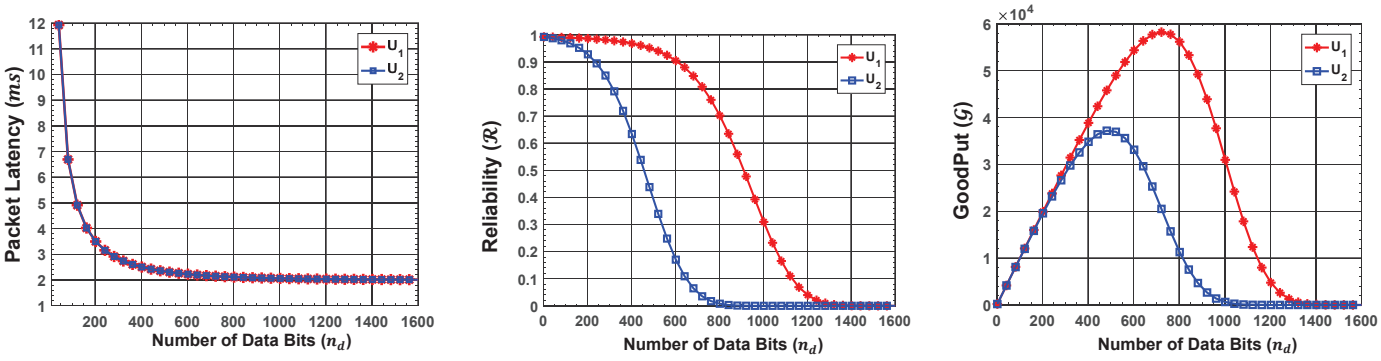
[6]Note that the preamble time T_p is that adopted by RA-LTE as PRACH preamble format #4 for UL-TX [20].

[7]From this point onward, the subscript referring to the cluster index is dropped. Hence, the stronger and weaker nodes in each cluster will be referred to as U_1 and U_2, respectively.



(a) (b) (c)

Fig. 1: Network Performance Metrics vs. Number of Replicas K - $n_d = 200$



(a) (b) (c)

Fig. 2: Network Performance Metrics vs. Number of Data Bits n_d - $K = 5$

number of nodes N in the network, and their data traffic pattern (i.e. packet arrival rate at each frame), which is due to the impact of N and λ_d on the inter-cluster interference. Moreover, the reliability of the stronger nodes is higher than the weaker nodes, as would be expected. The effect of transmit diversity on GoodPut for the RA-NOMA nodes is plotted in Fig. 1c. 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.

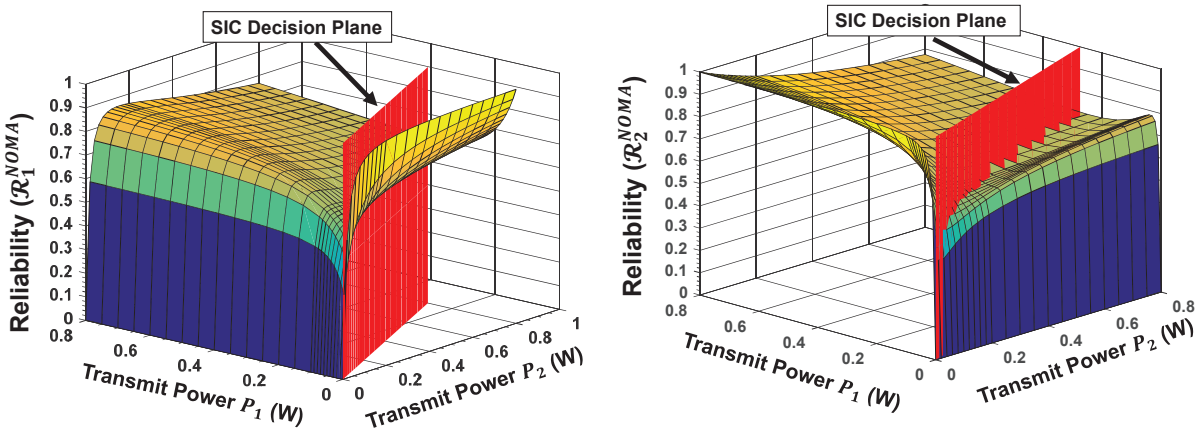
B. Effect of Number of Data Bits n_d

To explore the effect of the 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. Fig. 2a shows the average packet latency, where it can be seen that the higher the number of data bits per 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. Fig. 2b illustrates the reliability, where one can see that the reliability decreases when n_d increases. This is because the higher coding rate in fixed blocklength (or equivalently the higher n_d) results in higher decoding error probability when short packet lengths are adopted. Furthermore, the stronger node's reliability can be verified to be superior to that of the weaker node, since the stronger node has higher SINR at BS than the SNR of the weaker one. The GoodPut of nodes U_1 and U_2 is plotted in Fig. 2c as a function of n_d. For both RA-NOMA nodes, the Goodput increases as n_d increases, peaks at some values, and then starts to decrease. Particularly, for U_1, the GoodPut peaks at $n_d = 722$ while for U_2, it peaks at $n_d = 482$. The reason for this phenomenon is that the higher the number of data bits is, the higher GoodPut, in agreement with (10). However, as explained earlier, increasing n_d also increases the decoding error probability, which lowers the reliability. Thus, the GoodPut starts to decrease simultaneously with the decrease in reliability, since the excessive increase in n_d outweighs the improvement in the GoodPut. Moreover, and in general, $\mathcal{G}_1$ is higher than $\mathcal{G}_2$, which is due to the higher U_1's SINR at BS than U_2's SNR. It is worth mentioning that for the same transmit power of nodes U_1 and U_2, the peak value in $\mathcal{G}_1$ is greater than $\mathcal{G}_2$, indicating that the stronger node can convey higher number of data bits (or equivalently has higher GoodPut) per time-slot than the weaker node in FBL regime.

Finally, the reliability of the RA-NOMA nodes U_1 and U_2 along with the SIC decision plane (SDP) is illustrated in Figs. 3a and 3b, respectively. Plotted in transparent red, SDP is the plane in where the received power[8] of the stronger node U_1 at BS equals that of the weaker node U_2. This is a threshold plane for the BS to determine the SIC order. As can be seen from Fig. 3a, on the left side of the SDP where the U_1's received power is higher than U_2's (i.e. $P_1 > P_2$), $\mathcal{R}_1$ 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. 3a, 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[9]. Thus, U_1 experiences interference-free decoding. Now, going back to the right side of the SDP in Fig. 3a, it is observed that for small values of P_1, $\mathcal{R}_1$ 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, $\mathcal{R}_1$ improves, since U_1's SNR is increased. However, since $\mathcal{R}_1$ 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 $\mathcal{R}_1$. Such decrease in $\mathcal{R}_1$ 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), $\mathcal{R}_1$ increases, since the BS changes the SIC decoding order. The same observations for $\mathcal{R}_2$ can be made in Fig. 3b 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.

[8]Recall that the preamble sequences are orthogonal and hence, in the preamble phase, the BS is able to measure the received SNR for the signal 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 [22].

[9]Note that in right side of the SDP in Fig. 3a, U_1 experiences interference-free signal decoding, since $P_1 < P_2$.



(a) (b)

Fig. 3: Reliability vs. Transmit Powers for RA-NOMA Nodes: (a) Stronger node (b) Weaker node - $K = 2$

V. Conclusions

In this paper, an UL RA-NOMA IoT network with URLL requirements has been considered. Particularly, the IoT nodes have been clustered and SIC has been employed through preamble transmissions. To realize URLL, short packets with transmission diversity have been adopted. Additionally, network metrics such as average packet latency, reliability and GoodPut have been analytically derived. Furthermore, the effect of number of packet replicas and number of transmitted data bits per time-slot on the network metrics for the RA-NOMA IoT clusters have been explored numerically, illustrating several tradeoffs between the different network metrics. It is concluded that the number of packet replicas and transmitted data bits per time-slot must be carefully selected to meet the URLL requirements.

Acknowledgement

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

References

[1] 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.

[2] 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.

[3] 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.

[4] M. Shirvanimoghaddam, M. 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 Comms. Mag.*, vol. 57, no. 2, pp. 130–137, 2019.

[5] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Comms. Surveys & Tutorials*, vol. 22, no. 3, pp. 1805–1838, May 2020.

[6] 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.

[7] J. Choi, "NOMA-based random access with multichannel ALOHA," *IEEE JSAC*, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.

[8] 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.

[9] 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.

[10] 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.

[11] M. R. Amini and M. W. Baidas, "Random-access NOMA in URLL energy-harvesting IoT networks with short packet and diversity transmissions," *IEEE Access*, vol. 8, pp. 220 734–220 754, Dec. 2020.

[12] 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 IoT Journal*, vol. 1, no. 3, 2014.

[13] 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 IoT Journal*, vol. 6, no. 3, pp. 5119–5131, Jun. 2019.

[14] 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.

[15] 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 Comms. Surveys & Tutorials*, vol. 21, no. 3, pp. 2488–2524, 2019.

[16] 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.

[17] U. N. Bhat, *An Introduction to Queueing Theory, Modeling and Analysis in Applications*. Birkhäuser Basel, 2015.

[18] M. R. Amini and M. W. Baidas, "Performance analysis of URLL random-access NOMA-enabled IoT networks with short packet and diversity transmissions," *IEEE ICCSPA*, Mar. 2021, Online: https://www.dropbox.com/s/1ewufvf485ir09x/Proof.pdf?dl=0.

[19] 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

[20] 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

[21] 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.

[22] 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 Comms. Letters*, vol. 21, no. 10, pp. 2246–2249, 2017.