Maximizing Downlink User Connection Density in NOMA-aided NB-IoT Networks Through a Graph Matching Approach

Shashwat Mishra, Lou Salaun, Jean-Marie Gorce, Chung Shue Chen

Maximizing Downlink User Connection Density in NOMA-aided NB-IoT Networks Through a Graph Matching Approach

Shashwat Mishra*†, Lou Salaün*, Jean-Marie Gorce†, Chung Shue Chen*
*Nokia Bell Labs, Paris-Saclay Center, 91620 Nozay, France
†Institut National des Sciences Appliquées (INSA) de Lyon, 6 Av. des Arts, 69621 Villeurbanne, France
Email: shashwat.mishra@nokia.com, lou.salaun@nokia-bell-labs.com, jean-marie.gorce@insa-lyon.fr,
chung_shue.chen@nokia-bell-labs.com

***Abstract*—We develop a framework for maximizing the number of transmitted packets for devices in a Narrowband Internet of Things (NB-IoT) network using non-orthogonal multiple access (NOMA) in the downlink. The base station (BS) chooses one of the multiple available physical resource blocks (PRBs) that are well separated in frequency for a device, giving them the advantage of exploiting frequency diversity. The scheduling strategy focuses on the two-fold problem involving efficient device clustering and optimum power allocation. This problem is a mixed-integer non-convex problem. We propose a bipartite graph matching approach, termed minimum weight full matching with pruning (MWFMP), to address the problem over multiple PRBs and solve it under the quality-of-service (QoS), allowable PRB, power budget, and interference constraints. Additionally, we provide a comparison with a greedy heuristic, the multi-PRB stratified device allocation (MPSDA), where we extend our previous work for a single PRB connectivity problem. Furthermore, we compare our algorithms to orthogonal multiple access (OMA) scheduling, which is prevalent in legacy LTE networks. We show that our algorithms steadily outperform the connectivity performance offered by OMA.**

***Index Terms*—Industrial IoT, Massive Connectivity, B5G, NOMA, Bipartite Graph Matching, Machine-to-Machine Communication.**

I. Introduction

With the development of new use cases like smart vehicles, digital twins, and smart cities, it is expected that billions of communicating entities will be connected to the Internet soon. The number of devices connected to IP networks will be more than three times the global population by 2023. The number of networked devices will increase to 29.3 billion, up from 18.4 billion in 2018 [1]. Additionally, machine-to-machine (M2M) connections will comprise half of the globally connected devices by 2023. M2M communication is instrumental to realize the Internet of things (IoT), where machines communicate with each other autonomously with little to no human interactions.

The integration of M2M communication with cellular networks is a viable solution to the realization of high device density use cases since cellular networks already have mature infrastructure and provide wide-area coverage. Narrow-band IoT (NB-IoT) is proposed by the 3rd generation partnership project (3GPP) as a cellular-supported M2M communication network using the long-term evolution (LTE) standard. For such use cases, the key goal is having a huge number of devices intermittently connected to the network. Therefore, maximizing the metrics popular for traditional mobile systems such as sum-rate is not best suited to optimize cellular-based M2M networks.

Massive machine type communication (mMTC) is an instrumental technology for enabling beyond 5G networks providing ubiquitous connectivity. This use case is often characterized by sporadic short packet traffic originating from limited capability devices densely packed in a service area [2]. It is known that the traffic for mMTC is usually uplink dominated, however there is still a need to develop downlink solutions to facilitate seamless bi-directional massive connectivity, for example industrial controllers needing to send downlink commands to actuators. Taking the above points into consideration, we see that there is an immediate need for new resource allocation methods that take into account the connectivity requirements specific to such M2M networks. In this paper, we study the downlink resource allocation problem in cellular-based M2M networks supported by NB-IoT.

In contrast to previous works in the literature, we do not study the sum-rate related or outage related objectives. Rather, we consider the connection density objective that is better positioned for the massive access problem in such networks. We aim at maximizing the number of devices satisfying their QoS requirement. We employ the non-orthogonal multiple access (NOMA) technique, comprising of superposition coding at the transmitter, which is the base station (BS) in our scenario considering the downlink. Successive interference cancellation (SIC) is used at the receiver [3], which is the user equipment (UE). There has been some studies regarding the uplink connectivity, most notably the authors in [4] develop a convex optimization based heuristic framework to show the connectivity gains achieved by NOMA in an NB-IoT system. For the same objective, authors in [5] use a reinforcement learning based framework to facilitate joint user and power allocation on sub-carriers. Additionally, the authors in [6] show that the the ergodic sum-rate for NOMA can be multi-fold even in single antenna case and it grows linearly with the number of antennas in a multi-antenna system. It is evident

therefore that the combination of NB-IoT supported by NOMA offers an efficient solution to the massive connectivity demand within the limited radio resources available to current cellular systems. However, the downlink for an NB-IoT network is very different from the uplink in terms of the allowed number of sub-carriers for each user and more crucially having a per PRB power budget as opposed to a per user power budget in the uplink.

In this work, we maximize the number of connected UEs by grouping them into the available resource blocks, scheduling their transmission according to the per PRB downlink power constraint. First, we propose a novel bipartite graph matching framework that solves this problem using a minimum weight assignment strategy. To the best of our knowledge, no work has used a graph matching approach to address the downlink connectivity metric so far, which is trickier to solve compared to the uplink due to a PRB based sum-power constraint. Furthermore, there are no findings so far on efficient parallel multi-PRB device assignment where massive gains can be obtained through careful matching. Additionally, we propose an extension of the stratified device allocation algorithm (SDA) presented in [7] to solve the connectivity problem over multiple PRBs.

II. System Model

We consider a single-cell homogeneous NB-IoT system with one BS at the center of the cell and D randomly located UEs denoted by $\mathcal{D} \triangleq \{1, \ldots, D\}$. We aim to connect as many devices as possible in the downlink such that all devices are guaranteed a minimum data rate while enforcing the system power budget. We study this problem over one transmit time interval (TTI). Let $\mathcal{S} \triangleq \{1, \ldots, S\}$ be the set of all sub-carriers. Each sub-carrier has a bandwidth of B_s Hz, and the total system bandwidth is of $B = B_s \times S$ Hz. Each sub-carrier can have at most L devices superimposed on it due to SIC implementation constraint [8], which we refer to as the L SIC levels of a sub-carrier. It is assumed that perfect SIC decoding is possible at the receiver.

We assume that there are M distinct PRBs, represented by the index set $\mathcal{M} \triangleq \{1, \ldots, M\}$. For any $m \in \mathcal{M}$, we denote by $PRB_m \subset \mathcal{S}$ the set of sub-carriers belonging to the m-th PRB. These sets form a partition of $\mathcal{S}$. That is, there are no common sub-carriers between any two PRBs and the union of PRB_m is equal to $\mathcal{S}$, for all $m \in \mathcal{M}$. The system frequency resources (PRBs, sub-carriers, SIC levels, and frequency allocation slots) and their notations are illustrated in Figure 1.

Each device may be allocated at most one sub-carrier in a PRB in the downlink following the 3GPP NB-IoT standards [9]. The PRBs are well separated in frequency such that the device has a different fading state on each PRB. Physically, this implies that instead of having an expensive wide-band receiver that works over the entire available spectrum, each device can function well instead with just a narrow-band receiver tuned to the frequency range of its allowable PRBs. However, this implies that a device can be assigned one PRB out of a smaller set of contiguous PRBs, for example, say, PRBs $\{2, 3, 4\}$ for device d, rather than all the available PRBs. We shall refer to the allowable set of PRBs for device d as $\mathcal{M}_d \subseteq \mathcal{M}$. Throughout the rest of the paper, we shall refer to the smallest unit of frequency allocation as a frequency slot, which is one of the SIC levels on one sub-carrier in a given PRB. Thus, we have a total of $L \times S$ frequency slots. Due to the PRB selection constraint, all devices cannot access all the frequency slots.

We represent the composite channel gain of device $d \in \mathcal{D}$ on sub-carrier $s \in \mathcal{S}$ by $h_{s,d}$. The channel is assumed to be flat-fading over one PRB because of the low bandwidth. That is, each device has the same channel gain on all sub-carriers $s \in PRB_m$, but different channel gains on different PRBs. The channel gain consists of path-loss and small-scale Rayleigh fading such that $h_{s,d} = b_{s,d} PL_{s,d}^{-1/2}$, where $b_{s,d} \sim \mathcal{CN}(0, 1)$ and $PL_{s,d}$ is the distance-dependent path-loss. The amplitude of the channel gain for device d on sub-carrier s is $g_{s,d} \triangleq |h_{s,d}|^2$. We assume that perfect CSI is available at the BS for scheduling which is realistic for industrial IoT deployments with limited mobility. Efficient CSI acquisition methods for NOMA based NB-IoT system have been well investigated by studies such as [10]. For example, the CSI may be acquired through pilot signaling when the devices make the initial access request and subsequently estimated by a time-predictive model.

The additive white Gaussian noise on each sub-carrier s is given by $N \triangleq N_0 \times B_s \times 10^{\frac{F}{10}}$, where N_0 is the noise spectral density expressed in W/Hz and F is the noise figure expressed in dB. For the sake of simplicity, we will assume that all devices have the same noise power, although our solution approach remains valid even when different devices have different noise powers.

A. NOMA Framework

We use power domain NOMA for enhancing user connection density. The index of the device receiving the l-th encoded packet on sub-carrier s is denoted by $x_s(l)$. In other words, the first decoded packet is $x_s(1)$, followed by $x_s(2)$, etc. We define $\boldsymbol{X}_s \triangleq \{x_s(1), \ldots, x_s(l)\}$ as the set of devices allocated to sub-carrier s. The cardinality of this set is denoted by $|\boldsymbol{X}_s|$ and must satisfy $|\boldsymbol{X}_s| \leq L$, due to the SIC constraint. In a practical implementation, $\boldsymbol{X}_s$ is represented by a list sorted in the SIC decoding order, so that accessing any element $x_s(l)$ from its decoding order l can be done in constant time.

In downlink NOMA, the optimal SIC decoding order on sub-carrier s is in the increasing order of its channel gains, since a strong device can decode both its own signal and the signal for the interferer while a weaker device treats the other devices' signals as noise and can decode only its own signal [8]. In this paper, we only consider allocations satisfying the optimal downlink decoding order, thus we have:

$$\forall\, s \in \mathcal{S}, \quad g_{s,x_s(1)} \leq g_{s,x_s(2)} \leq \cdots \leq g_{s,x_s(|\boldsymbol{X}_s|)}.$$

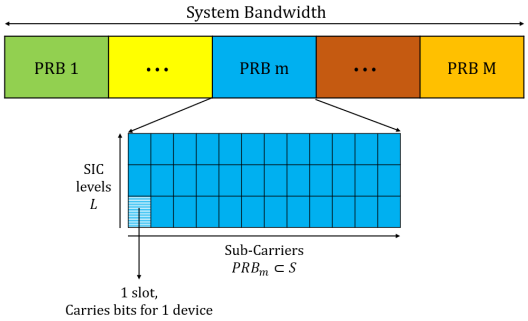

Fig. 1. Representation of system resources in terms of frequency slots

The signal-to-interference-plus-noise ratio (SINR) for the l-th decoded device on sub-carrier s can be expressed as:

$$\gamma_{x_s(l)} = \frac{g_{s,x_s(l)} p_{x_s(l)}}{I_{s,l} + N},$$

where $p_{x_s(l)}$ is the transmit power for device $x_s(l)$, and $I_{s,l}$ is the interference caused by other devices on the same sub-carrier, which is expressible as:

$$I_{s,l} \triangleq g_{s,x_s(l)} \sum_{i=l+1}^{|\boldsymbol{X}_s|} p_{x_s(i)}. \tag{1}$$

B. Constraints and Data Rate Requirements

We consider the following constraints in our system. First, we have the power budget:

$$\forall\, m \in \mathcal{M}, \quad \sum_{s \in PRB_m} \sum_{l=1}^{|\boldsymbol{X}_s|} p_{x_s(l)} \leq P_{\max}. \tag{2}$$

Here, $P_{\max}$ is the maximum total transmit power on any given PRB, as specified by the 3GPP cellular IoT standard [11]. Additionally, each device must achieve a minimum critical data rate to remain viable in the network. The achievable data rate for device $x_s(l)$ can be expressed as:

$$r_{x_s(l)} = B_s \log_2 \left(1 + \frac{g_{s,x_s(l)} p_{x_s(l)}}{I_{s,l} + N}\right). \tag{3}$$

We denote the aforementioned minimum service rate by R. The device is said to be connected if:

$$r_{x_s(l)} \geq R. \tag{4}$$

The power required to achieve this minimum rate R can be obtained from (3) and (1) as:

$$p_{x_s(l)} = \frac{\zeta}{g_{s,x_s(l)}} (I_{s,l} + N), \tag{5}$$

where $\zeta = 2^{\frac{R}{B}} - 1$ is the target SINR to achieve the desired rate of R kbps. The total transmit power for all devices in a given PRB is capped at $P_{\max}$, which is called the PRB power budget.

III. Problem Formulation

In this paper, we assume that the channel state information $g_{s,d}$ is available perfectly at the BS, for any device $d \in \mathcal{D}$ and sub-carrier $s \in \mathcal{S}$. Recall that p_d is the transmit power for device d, we define the vector of all transmit powers as $\boldsymbol{p} \triangleq (p_1, \ldots, p_D)$. We denote by $\boldsymbol{p}_{x_s} \triangleq (p_{x_s(1)}, \ldots, p_{x_s(|\boldsymbol{X}_s|)})$ the powers of all devices on sub-carrier s. We define $Z(\boldsymbol{p}_{x_s}, \boldsymbol{X}_s)$ as the number of connected devices on sub-carrier $s \in \mathcal{S}$:

$$Z(\boldsymbol{p}_{x_s}, \boldsymbol{X}_s) \triangleq \sum_{l=1}^{|\boldsymbol{X}_s|} \mathbb{1}\left(r_{x_s(l)} \geq R\right). \tag{6}$$

Here, $\mathbb{1}\left(r_{x_s(l)} \geq R\right)$ is the indicator function that takes value 1 if the rate for device $x_s(l)$ is greater or equal to the required service rate R, and its value is 0 otherwise. Thus, $Z(\boldsymbol{p}_{x_s}, \boldsymbol{X}_s)$ has the QoS constraint (4) implicitly enforced in its definition. Using (2), (4) and (6), the downlink connectivity maximization problem given perfect CSI can be stated as follows:

$$\begin{aligned}
\text{maximize} \quad & \sum_{s=1}^{S} Z(\boldsymbol{p}_{x_s}, \boldsymbol{X}_s) && (\mathcal{P}) \\
\text{subject to} \quad & C1: \sum_{s \in PRB_m} \sum_{l=1}^{|\boldsymbol{X}_s|} p_{x_s(l)} \leq P_{\max}, \forall\, m \in \mathcal{M}, \\
& C2: \sum_{m \in \mathcal{M}_d} \sum_{s \in PRB_m} |\{d\} \cap \boldsymbol{X}_s| \leq 1, \forall\, d \in \mathcal{D}, \\
& C3: \; d \notin \boldsymbol{X}_s, \forall\, d \in \mathcal{D}, m \in \mathcal{M} \setminus \mathcal{M}_d, s \in PRB_m, \\
& C4: \; |\boldsymbol{X}_s| \leq L, \forall\, s \in \mathcal{S}.
\end{aligned}$$

The objective function in $\mathcal{P}$ aims to maximize the number of transmitted packets for devices over all sub-carriers in $\mathcal{S}$ that satisfy their QoS requirement. Constraint C1 refers to the maximum total transmit power for each PRB as defined in (2). Constraint C2 implies that each device can be allocated at most one sub-carrier according to 3GPP NB-IoT [9]. The above formulation is a generalization of the problem in [7] on resource allocation and user pairing joint optimization over multiple PRBs, more specifically to a restricted set of PRBs $\mathcal{M}_d \subset \mathcal{M}$ for each device d, as specified by the constraint C3. We propose this multi-PRB connection paradigm keeping in mind the massive densification requirements of the network that would be much easily served using multiple resource blocks. Note that this difference however make it much more challenging and requires novel solution strategies. Finally, constraint C4 states the SIC limit of the system of supporting at most L devices superimposed per sub-carrier.

A. Bipartite Graph Matching Framework

We tackle the following problem $\mathcal{P}$ through a graph matching based heuristic termed minimum weight full bipartite matching and pruning (MWFMP). Given D devices, S sub-carriers, and a system SIC limit of L, in general we can connect at most $\max\{D, S \times L\}$ devices.

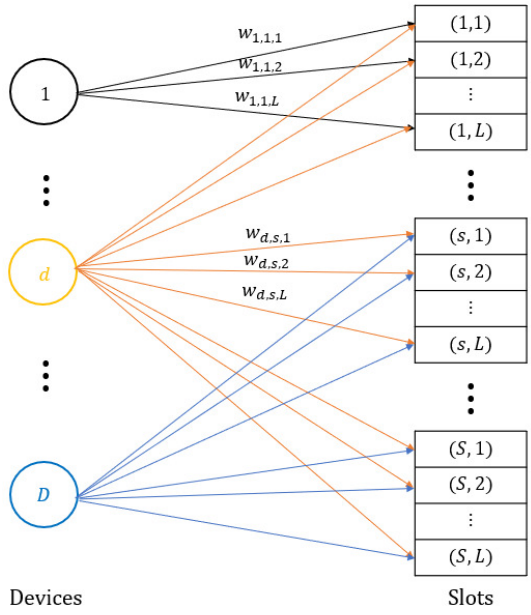

Fig. 2. Construction of the bipartite graph for device-slot pairing

The sum-power of all devices on sub-carrier s can be written using (1) and (5) as:

$$\sum_{l=1}^{L} p_{x_s(l)} = \sum_{l=1}^{L} \left(\frac{\zeta N}{g_{s,x_s(l)}} + \sum_{i=l+1}^{L} \zeta p_{x_s(i)} \right), \\ = \sum_{l=1}^{L} \frac{N\zeta(1+\zeta)^{(L-l)}}{g_{s,x_s(l)}}, \tag{7}$$

where (7) is obtained by re-arranging the iterative equation (26) for interference and accumulating all the terms corresponding to $g_{s,x_s(l)}$ in [12]. From (7) we can infer that the power contribution of a device on a sub-carrier relies only on its own channel gain and the SIC level to which it is assigned. Using this, we can calculate the power contribution of a device on any sub-carrier without the apriori knowledge of other devices superimposed on that sub-carrier. Note that however the power contribution of a device is not the same as the transmit power of the device. It is just a notional re-arrangement of total transmit power on a sub-carrier for ease of calculation.

We now construct a bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ as follows. $\mathcal{V}$ is the set of vertices, which is divided in two parts, by definition of bipartite graph. The first part contains the set of devices $\mathcal{D}$, and the second part has the set of frequency slots (s,l) for all $s \in \mathcal{S}$ and $l \in \{1,\ldots,L\}$. The set of edges is denoted by $\mathcal{E}$. We put an edge between device d and frequency slot (s,l) if there exists a PRB $m \in \mathcal{M}_d$ such that $s \in PRB_m$. Due to the PRB selection constraint C3, there does not necessarily exist an edge between all devices and all frequency slots. In other words, the graph $\mathcal{G}$ is not a complete bipartite graph. The set of weights is denoted by $\mathcal{W}$. The weight of an edge, $w_{d,s,l}$, is the power contribution of device d when connected to frequency slot (s,l), obtained using (7) as:

$$w_{d,s,l} = \frac{N\zeta(1+\zeta)^{(L-l)}}{g_{s,d}}. \tag{8}$$

When assigning devices to a sub-carrier, the first device gets assigned to the frequency slot corresponding to SIC level L, such that it faces no interference. When a second device gets added to frequency slot $L-1$, it faces interference only from the device in frequency slot L. An example of this construction is shown in Figure 2.

We now elaborate Algorithm 1 for maximizing user connection density across multiple PRBs. On line 1, we construct the bipartite graph $\mathcal{G}$ as described in the last paragraph. Then we obtain the minimum weight full matching $\mathcal{F}$ of $\mathcal{G}$ using an optimal linear sum assignment algorithm such as [13] as shown on line 2. More precisely, $\mathcal{F}$ is a sub-graph of $\mathcal{G}$ such that the sum of all edge weights is the least among all full matchings, where each device gets connected to one frequency slot.

However, this allocation doesn't necessarily satisfy the per PRB power budget as given by constraint C1. This necessitates a pruning step. We iterate over each PRB (see line 4) and calculate the sum power of all the allocated devices (see line 5). If the sum power of the devices in the PRB exceeds the power budget as evaluated on line 6, we remove the device with the weakest channel gain amplitude (see line 7) from the current device allocation (see line 15). Additionally, we modify the power of all the other devices on the same sub-carrier as the removed device by moving them up by one SIC level, as shown on lines 10–14. This step reduces the size of the matching and the resultant matching is no longer a full matching. However, since we only remove devices from the initial global minimum sum power assignment, the resulting configuration still uses the lowest possible power for connecting the remaining devices. The complexity of the algorithm is governed by the graph matching step which is $O\left(\max(S,D)^3\right)$.

B. Extending Stratified Device Allocation to Multiple PRBs

We augment the stratified device allocation (SDA) algorithm developed in [7] to serve multiple PRBs. SDA is optimal in terms of downlink connectivity when there is only a single PRB to be served. However, when each device has multiple available PRB with different channel gains, as is the case in our system, SDA is not optimal. Nevertheless, it is a simple and reasonably efficient algorithm. Therefore, we extend it as the multi-PRB SDA (MPSDA), presented in Algorithm 2 as a baseline for MWFMP and for comparison.

In essence, we run SDA for each PRB by sorting all contesting devices by their channel gains in that PRB. The device with the greatest channel gain in PRB m (see line 6) is allocated the first sub-carrier, say s_1 at the first level of SIC, indexed as $x_{s_1}(1)$. This device is then removed from the device contention pool $\mathcal{D}'$ as shown on line 12. Subsequently, we fill the first SIC level of all sub-carriers of PRB m with

Algorithm 1 Minimum weight full matching with pruning (MWFMP)

Input: $\mathcal{D}$, $\mathcal{S}$, $\mathcal{M}$, B, L, $P_{\max}$, PRB_m for all $m \in \mathcal{M}$, and $g_{s,d}$ for all $d \in \mathcal{D}, s \in \mathcal{S}$

1: **Initialization:** Form bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ as described in Section III-A
2: Compute a minimum weight full matching $\mathcal{F}$ by running LAPJV [13] on $\mathcal{G}$
3: For all $s \in \mathcal{S}$, $\boldsymbol{X}_s \leftarrow \{d : \exists\, l \le L,\ (d,s,l) \in \mathcal{F}\}$
4: **for** $m = 1$ to M **do**
5: $\quad P \leftarrow \sum_{(d,s,l)\in\mathcal{F}:\, s\in PRB_m} w_{d,s,l}$
6: $\quad$ **while** $P > P_{max}$ **do**
7: $\quad\quad (d', s', l') \leftarrow \operatorname{argmin}_{(d,s,l)\in\mathcal{F}:\, s\in PRB_m} g_{d,s}$
8: $\quad\quad \mathcal{F} \leftarrow \mathcal{F} \setminus \{(d', s', l')\}$
9: $\quad\quad P \leftarrow P - w_{d',s',l'}$
10: $\quad\quad$ **for** $l = 1$ to $L - 1$ **do**
11: $\quad\quad\quad d \leftarrow x_{s'}(l)$
12: $\quad\quad\quad \mathcal{F} \leftarrow (\mathcal{F} \cup \{(d, s', l+1)\}) \setminus \{(d, s', l)\}$
13: $\quad\quad\quad P \leftarrow P + w_{d,s',l+1} - w_{d,s',l}$
14: $\quad\quad$ **end for**
15: $\quad\quad \boldsymbol{X}_{s'} \leftarrow \boldsymbol{X}_{s'} \setminus d'$
16: $\quad$ **end while**
17: **end for**
18: Derive the power vector $\boldsymbol{p}$ using Eqn. (1) and (5)

Output: $\boldsymbol{p}$ and $\boldsymbol{X}_s$ for all $s \in \mathcal{S}$

Algorithm 2 Multi-PRB stratified device allocation (MPSDA)

Input: B, M, N, $P_{\max}$ and $g_{s,d}, \forall d \in \mathcal{D}, \forall s \in \mathcal{S}$

Initialization: $\forall s \in \mathcal{S}$, $\boldsymbol{X}_s \leftarrow \emptyset$, $\mathcal{D}' \leftarrow \mathcal{D}$, $\boldsymbol{p} \leftarrow \boldsymbol{0}$
for $m = 1$ to M **do**
$\quad P \leftarrow 0$
$\quad$ **for** $l = 1$ to L **do**
$\quad\quad$ **for** s in PRB_m **do**
$\quad\quad\quad d \leftarrow \operatorname{argmax}_{d'\in\mathcal{D}'} g_{d',s}$
$\quad\quad\quad p \leftarrow$ the required power of device d calculated using (5)
$\quad\quad\quad$ **if** $P + p \le P_{\max}$ **and** $\mathcal{D}' \neq \emptyset$ **then**
$\quad\quad\quad\quad \boldsymbol{X}_s(l) \leftarrow d$
$\quad\quad\quad\quad p_d \leftarrow p$
$\quad\quad\quad\quad P \leftarrow P + p_d$
$\quad\quad\quad\quad \mathcal{D}' \leftarrow \mathcal{D}' \setminus \{d\}$
$\quad\quad\quad$ **else**
$\quad\quad\quad\quad$ **break**
$\quad\quad\quad$ **end if**
$\quad\quad$ **end for**
$\quad$ **end for**
end for

Output: $\boldsymbol{p}$ and $\boldsymbol{X}_s$, for all $s \in \mathcal{S}$

successively weaker devices. Then we roll over to the first sub-carrier and allocate the next strongest device to the second level of SIC on the first sub-carrier, indexed as $x_{s_1}(2)$. This strategy is implemented by lines 4–5.

The maximum sum power of all the allocated devices in a PRB must not be greater than the system power budget $P_{\max}$. If either of these limits is reached on line 8, the device allocation for a PRB is terminated. Furthermore, the last condition on line 8 is triggered if the total number of allocated devices exceeds the number of devices contending for connection, which stops any further allocation. Note that the time complexity for this algorithm is $O(\min(D, LS))$ [7].

IV. Simulation Results

We analyze the performance of MWFMP and also MPSDA through computer simulations under an NB-IoT system setting. Key system parameters are given in Table I, which are taken from 3GPP standards for NB-IoT [11] and are for the industrial wireless IoT use case [14]. We evaluate the performance in terms of number of devices achieving their data rate requirement, which we call "connected" devices.

We assume frequency flat Rayleigh fading over the system bandwidth as the frequency bandwidth is small. We consider that perfect SIC for NOMA can be carried out by the receiver. We consider that all UEs have a single receive antenna. Devices are deployed randomly following a uniform distribution in a hexagonal cell. All devices have the target data rate requirement of R kbps. Each device can transmit on one out of the available PRBs. All simulation results are averaged over 1000 independent trials.

TABLE I
Key System Simulation Parameters

Carrier Frequency	900 MHz
PRB Bandwidth	180 kHz
Sub-Carrier Bandwidth	15 kHz
Path-Loss (dB)	$120.9 + 37.6 \log \frac{D}{1000} + G + L_P$
UE Antenna Gain (G)	-4 dB
Indoor Penetration Loss (L_P)	10 dB
$P_{\max}$	23 dBm / PRB
No. of PRBs	10
Percentage of Indoor Users	80%
AWGN Power	-174 dBm/Hz
Noise Figure	5 dB

We present the performance of MWFMP and MPSDA with $M = 2$ and $M = 3$, respectively. Each PRB of 180 kHz bandwidth has 12 sub-carriers which can accommodate up to 24 devices with $M = 2$ and 36 devices with $M = 3$. All devices have a target data rate $R = 20$ kbps. In Figure 3, we show the performance of the algorithms with different cell radii. At a cell radius of 2000 m, MWFMP with $M = 3$ connects about 233 devices on average, while MPSDA connects 218 and OMA connects only 119. Thus, MWFMP connects 94% more devices than OMA and 7% more devices than MPSDA with $M = 3$. MWFMP maintains superior performance with $M = 2$ as well, connecting 46% more devices than OMA and 9% more devices than MPSDA. As expected, the connectivity of all the algorithms decreases with the increasing cell radius. Meanwhile, MWFMP outperforms OMA and MPSDA in all cases.

In Figure 4, we compare the performance with varying

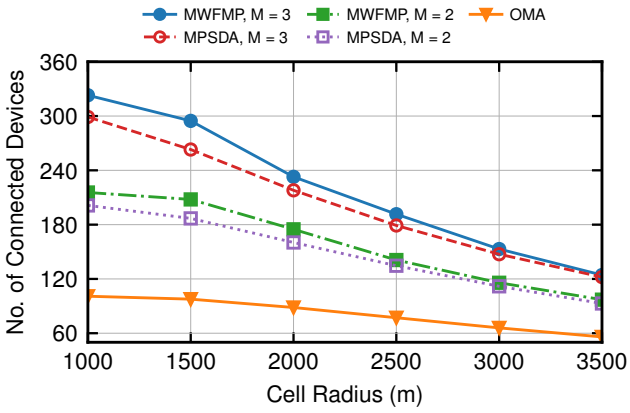

Fig. 3. Connectivity performance with varying cell radius, R = 20 kbps

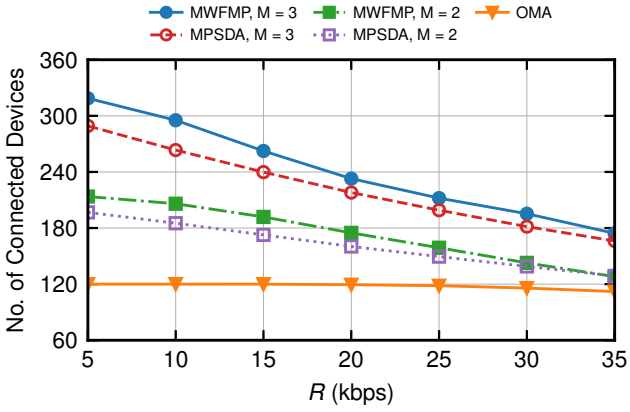

Fig. 4. Impact of the target data rate R on the connectivity of the system, cell radius = 2000 m

target data rate R. The cell radius for all cases is kept as 2000 m. As expected, there is a downward trend of connectivity with the increase in the service rate. At R = 15 kbps, MWFMP with M = 3 connects about 262 devices on average, while MPSDA connects 239 and OMA connects 120, respectively. Thus, MWFMP connects 118% more devices than OMA and 9% more devices than MPSDA with M = 3. When M = 2, MWFMP connects 60% more devices than OMA and 11% more devices than MPSDA. Yet again, MWFMP outperforms OMA and MPSDA over all the required service rates.

In figure 5, we show the connectivity gains of the different techniques with M = 2 under various power budgets. Note that the system has in total 10 PRBs, i.e., 120 frequency sub-carriers. We see that MWFMP steadily provides greater connectivity than OMA and MPSDA over all available power. The connectivity gains are greater when more power is available as with NOMA we can connect more devices per time-slot through superposition compared to OMA where we can connect only as many devices as the number of sub-carriers. The efficient slot-device matching in MWFMP allows for

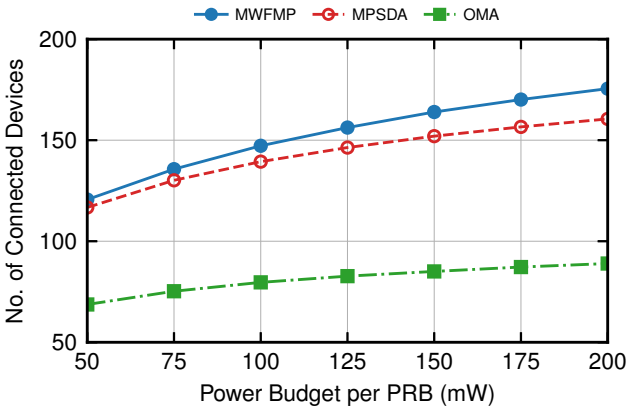

Fig. 5. Connectivity performance with different power budgets, cell radius = 2000 m, R = 20 kbps

TABLE II
POWER EFFICIENCY (mW/bit) WITH DIFFERENT SYSTEM LOAD, CELL RADIUS = 2000 METERS, R = 20 KBPS

	OMA	MPSDA	MWFMP
Under-loaded	2.48	2.48	1.62
Critically-loaded	1.47	1.47	1.37
Over-loaded	1.47	1.23	1.14

better power usage and interference management, leading to 98.8% more devices than OMA and 9% more devices than MPSDA.

Table II shows the power efficiency of the different algorithms presented in this study. In the under-loaded regime, we have lesser devices in the system than the number of sub-carriers. Here we see that the power efficiency of MWFMP is better than MPSDA and OMA because MWFMP connects the greatest number of devices using the least power as opposed to OMA where we greedily assign the best channel gain devices to each sub-carrier over each of the PRBs. This greedy assignment usually ends up assigning each device to a sub-optimal PRB which may come up earlier while performing allocation, consuming more power than necessary for connection. The performance of MPSDA and OMA are identical here as both of them use the same device allocation strategy. The same analysis holds for the critically-loaded regime where we have exactly as many devices and the number of sub-carriers. However, in this case the gap between the power efficiency for MFWMP and MPSDA decreases as more power is used up to connect all the devices. In the over-loaded case, we have more devices in the system than the number of sub-carriers. Here OMA can connect a maximum of 120 devices and therefore has the same power efficiency as the critically-loaded case. MWFMP has the best power efficiency, having lesser power per connected device than MPSDA by about 7%.

V. CONCLUSION

In this paper, we study the problem of connection density maximization in the downlink for NB-IoT networks under

multiple PRBs. We formulate the resource allocation optimization problem by the bipartite graph matching framework and propose the minimum weight full bipartite matching and pruning (MWFMP) algorithm for solving it. Besides, we generalize the stratified device allocation strategy in [7] to accommodate multiple PRBs as a baseline solution (MPSDA) for comparison. We conduct computer simulations and study their performance with varying target service rate and cell radius. Results show that in all the scenarios, MWFMP with $M = 2$ and $M = 3$ can steadily outperform in both the connectivity and power efficiency. Meanwhile, it can connect up to 60% and 118% more devices than OMA, respectively.

ACKNOWLEDGMENT

A part of the work has been carried out at the Laboratory for Information, Networking and Communication Sciences (www.lincs.fr).

REFERENCES

[1] Cisco, "Cisco annual Internet report (2018–2023)," Cisco, San Jose, CA, USA, Tech. Rep., March 2020.

[2] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 615–637, 2021.

[3] S. Mishra, L. Salaün, C. S. Chen, and K. Giridhar, "Analysis of downlink connectivity in NB-IoT networks employing NOMA with imperfect SIC," in *Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 2021, pp. 520–525.

[4] A. E. Mostafa, Y. Zhou, and V. W. S. Wong, "Connection density maximization of narrowband IoT systems with NOMA," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4708–4722, 2019.

[5] Z. Mlika and S. Cherkaoui, "Massive IoT access with NOMA in 5G networks and beyond using online competitiveness and learning," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 624–13 639, 2021.

[6] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of NOMA over OMA in uplink communication systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 536–568, 2020.

[7] S. Mishra, L. Salaün, C. W. Sung, and C. S. Chen, "Downlink connection density maximization for NB-IoT networks using NOMA with perfect and partial CSI," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 305–11 319, 2021.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. USA: Cambridge University Press, 2005.

[9] Y. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A primer on 3GPP narrowband Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 117–123, 2017.

[10] L. Chetot, M. Egan, and J.-M. Gorce, "Joint identification and channel estimation for fault detection in industrial IoT with correlated sensors," *IEEE Access*, vol. 9, pp. 116 692–116 701, 2021.

[11] 3GPP, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 45.820, Dec. 2015, version 13.1.0. [Online]. Available: http://www.3gpp.org/DynaReport/45820.htm

[12] J.-M. Gorce, P. Mary, D. Anade, and J.-M. Kélif, "Fundamental limits of non orthogonal multiple access (NOMA) for the massive Gaussian broadcast channel in finite block-length," *Sensors*, vol. 21, no. 3, Jan. 2021. [Online]. Available: https://hal.inria.fr/hal-03095234

[13] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, pp. 325–340, 2005.

[14] W. Wang, S. L. Capitaneanu, D. Marinca, and E.-S. Lohan, "Comparative analysis of channel models for industrial IoT wireless communication," *IEEE Access*, vol. 7, pp. 91 627–91 640, 2019.