Flexible Physical Layer based Resource Allocation for Machine Type Communications Towards 6G

Yalcin Sadi *IEEE Member*, Serhat Erkucuk *IEEE Senior Member*, and Erdal Panayirci *IEEE Life Fellow*
Electrical and Electronics Engineering
Kadir Has University, Istanbul, Turkey
{yalcin.sadi, serkucuk, eepanay}@khas.edu.tr

***Abstract*—The exponential growth of Internet of Things applications necessitates the design of next generation cellular systems to provide native support for machine type communications (MTC). While 5G aims at providing this native support under domain of massive MTC (mMTC) as one of the three major domains it focuses; i.e., enhanced mobile broadband, ultra reliable low latency communication, and mMTC, the enabling technologies and communication architectures are still limited and incomplete considering the nearly standardized efforts under 3GPP Releases 15 and 16. Studies towards 6G should elaborate on enabling truly massive MTC flexibly to support fast growing machine-to-machine (M2M) services with massive number of devices and very diverse quality of service (QoS) requirements. In this paper, we study radio resource allocation for mMTC based on the envisioned flexible physical layer architecture for 5G and beyond, possibly including 6G. We first present an overview of the 5G New Radio physical layer aspects particularly focusing on multiple numerologies and discuss the 3GPP features in Releases 15-17 as possible enablers of a flexible radio resource allocation scheme. Then, we propose a polynomial-time persistent resource allocation scheme for M2M communications aiming at meeting diverse QoS requirements of the M2M applications while achieving spectral efficiency. Finally, we present some numerical results and discuss future research directions for access schemes to enable truly massive MTC.**

***Index Terms*—Resource Allocation, Flexible Physical Layer, Machine Type Communications, 5G and Beyond, 6G.**

I. INTRODUCTION

Supporting native machine type communications is of paramount importance for next generation cellular networks, while existing machine-to-machine (M2M) communication architectures are inefficient to support fast growing M2M services with massive number of devices and very diverse quality of service (QoS) requirements.

The initial phase of the 5G system based on Release 15 (Rel-15) developed by 3GPP specifies 5G New Radio (NR) to enhance the flexibility and scalability in supporting three major domains: enhanced mobile broadband (eMBB), ultra reliable low latency communication (URLLC), and massive machine type communications (mMTC) [1]. While 3GPP efforts up to Rel-15 have addressed some mMTC deployment scenarios under Long Term Evolution for Machines (LTE-M) and Narrow Band IoT (NB-IoT), existing technologies and 3GPP features need to be radically improved towards enabling all mMTC use cases, supporting massive connectivity and QoS requirements of diverse Internet of Things applications. Due to its limited flexibility, LTE cannot meet the heterogeneous user and application requirements of possible 5G services [2]. However, 5G NR brings many opportunities to enable truly massive MTC. The flexibility introduced by 5G NR to support heterogeneous requirements of mMTC applications is mainly provided by multiple numerologies in physical (PHY) layer architecture [3].

One of the most challenging tasks to achieve the goal of enabling truly massive MTC towards 6G is to develop grant-free access methods considering the limited radio resources available as identified by the first 6G White Paper [4]. 6G systems will have to flexibly enable different mMTC scenarios, supporting a massive number of low-power and complexity MTC devices while attaining high spectral efficiency [4]. As a progression towards grant-free acces, semi-persistent scheduling is utilized based on the allocation of a sequence of resource units and fixed modulation to a user equipment (UE) for a certain duration of time without additional grants. Semi-persistent scheduling schemes for MTC aims at meeting the diverse QoS requirements of MTC devices [5]–[7] on a basic idea that is clustering MTC devices with common QoS characteristics. These studies allocate an entire Access Grant Time Interval (AGTI) to a cluster without considering spectral efficiency. Moreover, QoS requirements of the low-priority MTC devices are adversely affected by the allocation of all higher priority classes on the same channel due to AGTI based allocation. In [8], [9], we have proposed semi-persistent resource allocation and access grant schemes for MTC in 5G and beyond cellular networks with the objective of minimizing the required bandwidth while satisfying diverse QoS and traffic requirements of MTC applications. While outperforming the previously allocated resource allocations schemes in terms of spectral efficiency and schedulability, they do not exploit the flexible PHY layer architecture in 5G NR. The goal of this paper is to investigate radio resource allocation for mMTC based on the flexible PHY layer architecture envisioned for 5G and beyond cellular systems including 6G, and complementary beyond-Rel-15 3GPP features enabling massive connectivity and grant-free access.

The rest of the paper is organized as follows. In Section II, an overview of the 5G PHY layer aspects are presented along with enhancements in Release 16 (Rel-16) and ex-

---

This work is supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under 3501-Career Development Program (CAREER) Grant #118E920.

978-1-7281-6047-4/20/$31.00

TABLE I
NEW RADIO FLEXIBLE PHYSICAL LAYER

Multiple Numerologies in NR					
Subcarrier Spacing (kHz)	**Cyclic Prefix**	**# Slots per Subframe**	**Slot Duration (ms)**	**# OFDM Symbols per Slot**	**Applicable Band**[a]
15	normal	1	1	14	FR1
30	normal	2	0.5	14	FR1
60	normal	4	0.25	14	FR1,FR2
60	extended	4	0.25	12	FR1,FR2
120	normal	8	0.125	14	FR2
240	normal	16	0.0625	14	FR2

[a] 5G NR supports a wide range of carrier frequencies and channel bandwidths within two possible bands FR1 and FR2 corresponding to $410-7125$ MHz and $24250-52600$ MHz, respectively, in order to have flexibility to meet the needs of various deployment scenarios and applications. While FR1 supports channel bandwidths of 5, 10, 15, 20, 25, 30, 40, 50, 60, 80, 90, and 100 MHz; FR2 supports channels with 50, 100, 200, 400 MHz bandwidths.

pectations from Release 17 (Rel-17). Section III describes an optimization framework for multiple numerology based resource allocation and presents a semi-persistent resource allocation algorithm along with some performance results. Finally, concluding remarks and future directions are discussed in Section IV.

II. 5G NR FLEXIBLE PHYSICAL LAYER AND BEYOND

In this section, we present an overview of the flexible physical layer architecture based on multiple numerologies in 5G NR [1]. Moreover, we discuss 3GPP features specified in Rel-15 as possible enablers of a flexible radio resource allocation scheme for mMTC. In addition, we discuss the enhanncements in Rel-16 and expectations from Rel-17 towards 6G.

A. Multiple Numerologies

NR uses a flexible physical layer architecture supporting *multiple numerologies*, see Table I. A numerology is generally defined by subcarrier spacing and cyclic prefix (CP) overhead. The subcarrier spacing is defined as the separation between the centers of two consecutive subcarriers. Multiple subcarrier spacings can be obtained by scaling a base subcarrier spacing by an integer $N = 2^{i-1}$ for $i \in 1, 2, ...$ such that a subcarrier spacing is an integer divisible by all smaller subcarrier spacing values. The set of subcarrier spacing values defined in NR are 15, 30, 60, 120, and 240 kHz. All possible numerologies, regardless of the CP overhead, align on symbol boundaries every 1 ms. The symbol duration and the CP length are inversely proportional to the subcarrier spacing. Specifically, each symbol length (including CP) of 15 kHz subcarrier spacing equals the sum of the corresponding 2^n symbols of the scaled subcarrier spacing of 15 kHz $\times 2^n$. For instance, for 15 kHz subcarrier spacing, the OFDM symbol duration is approximately 66.6 s and the CP length is approximately 4.7 s. For 30 kHz subcarrier spacing, the OFDM symbol duration and CP length are approximately divided by two compared to the 15kHz subcarrier spacing. The numerology used can be selected based on the frequency band (applicable band) as listed in Table I.

B. Resource Block Model

NR supports a resource block granularity similar to LTE. A resource block consists of 12 contiguous subcarriers and 6 (for extended CP use) or 7 (for normal CP use) consecutive OFDM symbols.

Time is divided into 10 ms radio frames each consisting of 10 subframes of fixed length 1ms. Based on the numerology chosen, each subframe consists of 1, 2, 4, 8, or 16 slots each consisting of 12 (for extended CP use) or 14 (for normal CP use) OFDM symbols. Note that, regardless of the CP overhead, the slot duration is downscaled similar to the symbol duration by a factor of 2^{-n} for a subcarrier spacing of 15 kHz $\times 2^n$, see Table I. While one slot consists of two RBs, generally transmissions are carried over one slot.

Based on the foregoing discussion, we can generalize the resource block model as follows: An RB consists of α subcarriers and β symbols. Let f_{rb}^k and τ_{rb}^k denote the bandwidth and the length of an RB based on the selected numerology k with subcarrier spacing Δf^k and symbol duration T_s^k. Then, $f_{rb}^k = \alpha \Delta f^k$ and $\tau_{rb}^k = \beta T_s^k$. Each RB has $\alpha \times \beta$ resource elements defined by one subcarrier and one symbol regardless of the numerology selected. However, as discussed in Sections II-C4 and II-C5, alternative RB models can be envisioned towards 6G.

C. 3GPP NR Features for Flexibility

In the following, we briefly present the features supported by NR, specified in Rel-15 [3], as possible enablers of a flexible radio resource allocation scheme for mMTC.

1) Non-orthogonal Multiple Access: For uplink transmissions, in addition to the scheduling based orthogonal approach, NR also supports non-orthogonal multiple access targeting mMTC applications. This feature allowing the use of common resources by MTC devices may enable massive connectivity characteristic of mMTC applications.

2) Bandwidth Parts: NR introduces a new term called bandwidth part (BWP). A BWP is a partitioned band within a channel over which the communication is maintained with the same numerology. The bandwidth and the numerology of the BWP can be controlled at the base station based on the application needs and network requirements. UEs can save power since they do not need to monitor the entire bandwidth as in LTE but only process the BWPs assigned to them. This is very useful for mMTC applications with low-power battery-operated devices. Moreover, multiple BWPs with different numerologies can be multiplexed within a single NR carrier to support different types of mMTC applications with heterogeneous service requirements. A resource allocation scheme creating BWPs for different MTC applications is proposed in Section III.

3) Uplink Semi-Persistent Scheduling: For uplink transmissions, semi-persistent scheduling based on the repetition of resource allocations for periodic transmissions is supported. This allows grant-free transmissions particularly for time-controlled mMTC applications where MTC devices generate data packets periodically. Grant-free transmission schemes are crucial to reduce the control signaling burden on the network considering the limited radio resources.

4) Uplink Sub-PRB Allocation: Sub-resource allocations are introduced to improve the uplink spectral efficiency. Allocation sizes correspond to $1/2$ PRB (6 subcarriers) and $1/4$ PRB (3 subcarriers). While this may be beneficial for power saving and uplink coverage, it may also facilitate allocation of multiple MTC devices on a PRB. While the latter is not supported yet, one can envision multiple resource block architectures for use based on different UE and application needs.

5) Mini-slot Allocation: As stated in Section II-B, while transmissions are generally carried over one slot, they can also be carried out over a fraction of a slot, with the minimum set to only two symbols. Such very short transmissions mainly target usage cases requiring very low latency, such as some URLLC applications as well as some time-critical mMTC applications.

D. Towards 6G

Elaborating on the flexible NR design specified in Rel-15, Rel-16 [10], [11] introduces the following enhanced features that may possibly be used for mMTC applications:

- Support for multiplexing and pre-empting different traffic types in uplink transmission allowing prioritization among different applications. This is beneficial to support time controlled semi-persistent traffic flow and event-triggered applications in the same band while satisfying their heterogeneous QoS requirements.
- Support for multiple active grant-free uplink transmission configurations to accommodate different application data flows.

Rel-17 [12] is expected to introduce NR-Light to address use cases that cannot be met by current mMTC supported via LTE-M and NB-IoT. NR-Light is planned to meet lower latency and higher reliability requirements than LTE-M and NB-IoT. NR-Light is expected to use BWP concept introduced in Rel-15. We can also envision that Rel-17 will increase the number of numerologies to satisfy a wider range of service requirements.

III. MULTIPLE NUMEROLOGY BASED RESOURCE ALLOCATION FOR MTC

In this section, we first describe and mathematically formulate the multiple numerology based resource allocation in an optimization framework. Upon illustrating the NP-hardness of the problem, we present an efficient semi-persistent polynomial-time algorithm.

A. Optimization Framework

Multiple numerology based resource allocation problem, denoted by $\mathcal{P}$, is described as follows: Given a set of MTC devices $i \in \mathcal{I}$ with packet generation period p_i and maximum allowable jitter j_i for each device i, and a set of PHY layer numerologies $k \in \mathcal{K}$ represented by RB bandwidth f^k and duration τ^k, what is the minimum bandwidth resource allocation such that each MTC device is allocated one RB periodically without violating maximum allowable jitter requirements?

$\mathcal{P}$ can be mathematically formulated as follows:

$$\textbf{minimize} \quad \sum_{n=1}^{n_{max}} z_n f_n \tag{1a}$$

$$\textbf{subject to} \quad \sum_{n=1}^{n_{max}} x_{in} = 1, \quad \forall i \in \mathcal{I} \tag{1b}$$

$$\sum_{i \in \mathcal{I}} x_{in} \leq |\mathcal{I}| z_n, \quad \forall n \in [1, n_{max}] \tag{1c}$$

$$\{x_{in}, p_i, j_i, f_n\}_{i \in \mathcal{I}} \in S^f, \quad \forall n \in [1, n_{max}] \tag{1d}$$

$$\textbf{variables} \quad x_{in} \in \{0,1\}, \quad \forall i \in \mathcal{I}, \forall n \in [1, n_{max}] \tag{1e}$$

$$z_n \in \{0,1\}, \quad \forall n \in [1, n_{max}] \tag{1f}$$

$$f_n \in \{f^1, f^2, ..., f^{|\mathcal{K}|}\}, \quad \forall n \in [1, n_{max}] \tag{1g}$$

The variables of the optimization problem (1) are x_{in}, binary variable with value 1 if MTC device i is allocated to band $n \in \{1, ..., n_{max}\}$ where n_{max} denotes the maximum number of bands to be used in a channel and 0 otherwise; z_n, binary variable with value 1 if at least one MTC device is allocated to band n; i.e., band n is used, and 0 otherwise; and f_n, discrete variable from set $\{f^1, f^2, ..., f^{|\mathcal{K}|}\}$ corresponding to the RB bandwidth of the numerology used in band n. The objective given by Eq. (1a) is minimization of the total bandwidth used by the allocation of the set $\mathcal{I}$ of MTC devices. Eq. (1b) states that each MTC device is allocated to one band. Note that this is important to have a persistent scheme. Eq. (1c) specifies the use of a band if at least one MTC device is allocated on that band. Finally, Eq. (1d) represents the QoS requirements of the MTC devices; i.e., whether the allocation of a set of devices on a particular band is feasible or not, where S^f denotes the set of all possible feasible allocations on a band.

While Eq. (1d) can be replaced by alternative sufficient or necessary and sufficient conditions satisfying the QoS requirements, we formulate it as follows [9]:

$$\tau_n + \sum_{l \in H_i} \lceil \frac{p_i}{p_l} \rceil \tau_n \leq j_i, \quad \forall i \in \mathcal{I}, \tag{2}$$

where H_i denotes the set of MTC devices with higher priority than device i allocated in the same band n, and τ_n is the RB duration corresponding to the numerology used on band

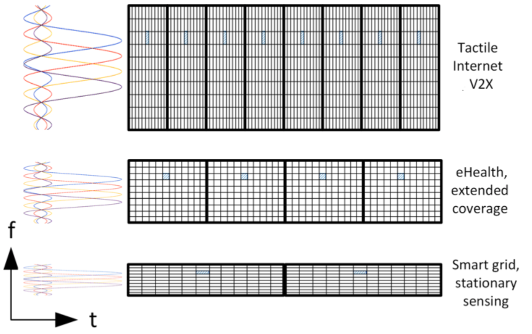

Fig. 1. Multiple numerologies for heterogeneous service requirements.

n. Note that Eq. (2) is a sufficient QoS satisfying condition for a feasible allocation of a set of MTC devices in the same band. Necessary and sufficient conditions mostly require exact non-polynomial tests as in the scheduling of real-time processing jobs [13] which are intractable considering massive connectivity of MTC devices.

Single numerology version of problem $\mathcal{P}$; i.e., for $|\mathcal{K}| = 1$, is proven to be NP-hard in [8]. Hence, $\mathcal{P}$ is also NP-hard meaning that it cannot be solved optimally in polynomial-time. Moreover, a fully-dynamic optimal resource allocation eliminates the possibility of creating a persistent or grant-free allocation scheme. On the other hand, massive connectivity expected for 5G and beyond networks requires computationally simple and at least semi-persistent schemes due to the huge control channel signaling overhead. In the following, we propose a fast and efficient semi-persistent resource allocation algorithm.

B. Multiple Numerology Resource Allocation Algorithm

Before presenting the algorithm, we first discuss the numerology design for different MTC applications with heterogeneous service requirements. A significant part of MTC traffic comes from a large number of stationary sensors (e.g., smart homes, metering) deployed over wide areas and produces sporadic and small amount of data. For such kind of applications, the Doppler effect is negligible, hence using narrow subcarrier spacing is more convenient especially when the application is delay tolerant. In addition, instead of increasing power spectral density to extend coverage area, using smaller subcarrier spacing stretches transmission over time allowing usage of cheaper and less energy-consuming battery-operated devices. There are other MTC applications with high-mobility characteristics (e.g., V2X) and using narrow subcarrier spacing will lead to high Doppler spread causing an increase in intercarrier interference (ICI). Therefore, subcarrier spacing value should be wide enough to alleviate Doppler spread while keeping the CP overhead aceptable. There may be further MTC applications with stringent latency requirements which cannot be served by the currently used transmission time interval (TTI) length. For such applications (e.g. e-Health), TTI length can be shortened by increasing subcarrier spacing. Fig. 1 depicts the idea of using multiple numerologies to meet various MTC application requirements.

Algorithm 1 Multiple Numerology Resource Allocation

Input: $\{f_n, \tau_n, p_i, j_i\}$ for C_{mn}, $\forall\{m, n\}$
Output: B^k_{mn}, $\forall\{m, n\}$

```
1:  for $n = 1 : N$ do
2:      $k = 0$;
3:      while $\exists m \in \{1, ..., M\}$ s.t. $|C_{mn}| \neq 0$ do
4:          $k = k + 1$;
5:          for $m = 1 : M$ do
6:              $j^*_m = 0$;
7:              for $l = 1 : m - 1$ do
8:                  $j^*_m = j^*_m + B^k_{ln} * \left\lceil \frac{p_m}{p_l} \right\rceil * \tau_n$;
9:              end for
10:             $j_{rem} = j_m - j^*_m$;
11:             if $j_{rem} > 0$ then
12:                 $B^k_{mn} = \min\{|C_{mn}|, \left\lfloor \frac{j_{rem}}{\tau_n} \right\rfloor\}$;
13:             end if
14:             $|C_{mn}| = |C_{mn}| - B^k_{mn}$;
15:         end for
16:     end while
17: end for
```

Next, we propose Multiple Numerology Resource Allocation (MNRA) Algorithm, given in Algorithm III-B, as described next. MTC devices are first grouped into M QoS classed based on the packet generation period and the maximum allowable jitter; i.e., devices with common packet generation period and maximum allowable jitter are grouped together. Then, each QoS class $m \in \{1, ..., M\}$ is further partitioned into numerology classes based on the foregoing numerology design discussion, basically considering different application characteristics. Each MTC cluster C_{mn} consists of MTC devices with packet generation period p_m and maximum allowable jitter j_m, and belonging to numerology class n having RB bandwidth and duration of f^n and τ^n, respectively. MNRA algorithm prioritizes numerology over QoS requirements and starts allocation of MTC clusters belonging to first numerology class $n = 1$ (Line 1). k denotes the current subchannel of bandwidth f^n used for the allocation of MTC devices with numerology class n. For each band k used for MTC devices in numerology class n, MNRA performs resource allocation for the MTC clusters in increasing order of packet generation period p_m; i.e., $p_1 \leq p_2 \leq ... \leq p_M$ (Line 5). For each cluster C_{mn} belonging to QoS class m, MNRA first determines the jitter j^*_m experienced due to the allocation of MTC devices with higher priority QoS classes on the same band k; i.e., $\{C_{ln} | 1 \leq l < m\}$ (Lines 6−9). Then, the difference between the maximum allowable jitter for C_{mn} and the experienced jitter j^*_m is calculated to determine the number of MTC devices in C_{mn} that can be allocated in band k (Line 10). Since each MTC device in C_{mn} should be allocated one RB of duration τ^n (due to numerology), the number of devices B^k_{mn} from C_{mn} to be allocated in band k is calculated

TABLE II
MNRA PERFORMANCE OVER AGTI-BASED SCHEME

Scenario		Performance		
Cluster Index	$[p_i, \delta_i]$ (ms)	Scheduled Clusters	BW Ratio (%) OMA (N=1)	BW Ratio (%) NOMA (N=10)
1	[10,2]	[1]	30.7735	5.6834
2	[20,4]	$[1:2]$	45.0418	8.4366
3	[20,6]	$[1:3]$	46.3511	8.5391
4	[40,12]	$[1:4]$	48.1223	8.8787
5	[100,50]	$[1:5]$	48.1730	8.8347
6	[100,60]	$[1:6]$	47.3685	8.7881
7	[200,80]	$[1:7]$	48.6371	8.9561
8	[250,100]	$[1:8]$	48.1238	8.8685
9	[500,150]	$[1:9]$	48.2183	8.9566
10	[500,200]	$[1:10]$	47.9935	8.8909
11	[10^3,500]	$[1:11]$	47.9083	8.9047
12	[10^5,10^4]	$[1:12]$	48.6099	9.6321

such that the maximum allowable jitter requirements are not violated based on Eq. (2)(Lines $11-13$). The remaining number of unallocated devices in C_{mn} is updated consequently (Line 14). The algorithm maximizes the utilization of each band k by allocating the maximum number of devices from each MTC cluster and then moves to the next band (Line 4) to allocate all devices belonging to numerology class n (Line 3). MNRA terminates after completing allocation of all numerology classes $n = \{1, ..., N\}$ (Line 1). The set of bands allocated with MTC devices belonging to same numerology class consitutes a BWP with common numerology as discussed in Section II-C2.

C. Results

In Table II, we illustrate the performance of the proposed resource allocation algorithm MNRA over a conventional AGTI-based resource allocation scheme designed for LTE systems in [6]. We consider a 20 MHz channel on $FR1$ band on which 3 numerologies with subcarrier spacing values of 15, 30, and 60 kHz are supported as given in Table I. 12 different QoS classes are considered with different p_i and j_i values as listed in Table II where each QoS class has a number of MTC devices uniformly distributed within the range $[20, 100]$. MTC devices in each QoS class are further assigned to one of 3 numerology classes randomly. Results are averaged over 1000 independent runs for each scenario.

Simulations are performed for different number of QoS classes from 1 to 12 and for orthogonol multiple access (OMA) and non-orthogonal multiple access (NOMA) scenarios. In the OMA scenario, each RB is allocated to a single MTC device; whereas in NOMA, multiple devices can use the same RB. In the NOMA scenario, each MTC device independently transmits its packet with maximum transmit power and the base station applies a successive interference cancellation (SIC) mechanism to decode multiple signals using the distinctness in the uplink channel gains of each MTC device in a NOMA cluster. MTC devices belonging to the same QoS and numerology class are further clustered into NOMA clusters based on the uplink clustering scheme proposed in [14] considering a maximum cluster size of $N = 10$ MTC devices. For OMA, the bandwidth reduction is more than 50% for all scenarios with a maximum of 70%. For NOMA based allocation, the spectral efficiency is improved by more than 90% with an average NOMA cluster size of around 5.

IV. CONCLUDING REMARKS AND PATH TOWARDS 6G

In this paper, we investigate radio resource allocation for mMTC based on 5G NR and beyond-Rel-15 3GPP features. We first overview 5G NR PHY layer aspects with a particular focus on multiple numerologies and 3GPP features in Releases 15-17 useful for grant-free acces paradigm. Then we characterize an optimization framework for multiple numerology based resource allocation and present a semi-persistent resource allocation scheme for mMTC applications with the objective of maximizing spectral efficiency while satisfying diverse QoS and traffic requirements of MTC devices.

While designing truly grant-free access schemes is a challenging task based on the current technologies, 3GPP Rel-15 provides a solid path towards 6G complemented by Rel-16 enhancements. Rel-17 is expected to provide different technologies such as NR-Light and increase flexibility of the PHY layer through greater number of numerologies and alternative resource block architectures targeting different user and application needs. Besides, existing powerful techniques such as NOMA and artificial intelligence algorithms may prove their use in enabling truly massive MTC towards 6G.

REFERENCES

[1] *Release 15, document TR 21.915 V15.0.0, 3GPP.* Sep 2019.

[2] A. Yazar, B. Pekz, and H. Arslan, "Fundamentals of multi-numerology 5g new radio," 2018.

[3] *5G; Study on New Radio (NR) access technology, document TR 38.912 version 15.0.0, 3GPP.* Sep 2018.

[4] M. Latva-aho and K. Leppanen, *Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence (White Paper).* Sep 2019.

[5] S. Lien and K. Chen, "Massive access management for qos guarantees in 3gpp machine-to-machine communications," *IEEE Communications Letters*, vol. 15, pp. 311–313, March 2011.

[6] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3gpp machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, 2011.

[7] A. G. Gotsis, A. S. Lioumpas, and A. Alexiou, "Evolution of packet scheduling for machine-type communications over lte: Algorithmic design and performance analysis," in *2012 IEEE Globecom Workshops*, pp. 1620–1625, Dec 2012.

[8] Z. H. Hussien and Y. Sadi, "Flexible radio resource allocation for machine type communications in 5g cellular networks," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, May 2018.

[9] G. Karadag, R. Gul, Y. Sadi, and S. Coleri Ergen, "Qos-constrained semi-persistent scheduling of machine-type communications in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 2737–2750, May 2019.

[10] *System Architecture for the 5G System, document TS 23.501 V16.1.0, 3GPP.* June 2019.

[11] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5g evolution: A view on 5g cellular technology beyond 3gpp release 15," *IEEE Access*, vol. 7, pp. 127639–127651, 2019.

[12] *Key Directions for Release 17, document RP-190831, 3GPP.* Jun 2019.

[13] M. Joseph and P. Pandya, "Finding Response Times in a Real-Time System," *The Computer Journal*, vol. 29, pp. 390–395, 01 1986.

[14] M. 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.