Artificial Intelligence Aided Low Complexity RRM Algorithms for 5G-MBS

Ernesto Fontes Pupo, Claudia Carballo González, Jon Montalban, Pablo Angueira, Maurizio Murroni, Eneko Iradier

***Abstract*—For the upcoming 5G-Advanced, the multicast/broadcast services (5G-MBS) capability is one of the most appealing use cases. The effective integration of point-to-multipoint communication will address the ever-growing traffic demands, disruptive multimedia services, massive connectivity, and low-latency applications. This paper proposes novel approaches for the dynamic access technique selection and resource allocation for multicast groups (MGs) subject to the 5G-MBS paradigm. Our proposal is oriented to address and contextualize the complexity associated with multicast radio resource management (RRM) and the implications of fast variations in the reception conditions of the MG members. We propose a solution structured by a multicast-oriented trigger to avoid overrunning the algorithm, a K-means clustering for group-oriented detection and splitting, a classifier for selecting the most suitable multicast access technique, and a final resource allocation algorithm. To choose the multicast access technique that better fits the specific reception conditions of the users, we evaluate heuristic strategies and machine learning (ML) multiclass classification solutions. We consider the conventional multicast scheme (MCS) and sub-grouping based on orthogonal/non-orthogonal multiplex access (OMA/NOMA) as access techniques. We assess the effectiveness of our solution in terms of the quality of service (QoS) parameters and complexity. The proposed technical solution is validated through extensive simulation for a single-cell 5G-MBS use case in the microwave (μWave) and millimeter wave (mmWave) band with different mobility behaviors.**

***Index Terms*—5G-MBS, Computational Complexity, Machine Learning, mmWave, Multicast Access Techniques, NOMA**

I. Introduction

The 5G New Radio (NR) development stage has a bottleneck in fulfilling future enhanced mobile broadband (eMBB) applications, such as augmented/virtual/extended reality, massive connectivity, and low latency. For this reason, academia and industry are extending their attention to the 5G-Advanced systems and integrating the best available and under-development wireless communication technologies to fulfill future demands and forecast performance [1]. Nevertheless, this upcoming evolution can significantly increase the network complexity, adding new computational levels, constraints, and hardware needs. Consequently, radio resource management (RRM) solutions must effectively cope with the trade-off between optimal network performance and computational complexity (CC) [2]. Moreover, this upcoming evolution will be tied to a three-dimensional ultra-dense heterogeneous network (HetNet) with differentiated services, tight requirements, and the always best-connected (ABC) vision, making managing and exploiting network resources even more complex [3].

Over the years, point-to-multipoint (PTM) communications have proved high efficiency in delivering high-quality multimedia services over wireless networks. It can enable a considerable capacity gain into the B5G ecosystem through cost-effective and high-quality delivery mechanisms optimized to face the requirements of the future massive Internet of Things (IoT) deployments, huge software updates, and massive multicast multimedia content delivery. For this reason, the 3rd Generation Partnership Project (3GPP) is working on the new phases of the 5G standard in the novel multicast/broadcast services (5G-MBS) paradigm that provides an effective convergence between unicasting and the multicast/broadcast capability [4].

Multicasting to users with different channel conditions without a tailored resource allocation strategy can degrade the quality of service (QoS) of the whole multicast group (MG) and produce an unfair resource allocation [5]. In recent years, works such as [5]–[9] have addressed this challenging trade-off in the 5G context, with solutions mainly based on grouping the MG members according to their specific reception conditions and using a multi-rate modulation coding scheme (MCS) to deliver the service taking advantage of orthogonal/non-orthogonal multiplex access (OMA/NOMA) techniques, the spatial diversity of the users, and single/multi-antenna approaches. However, none of the previous research delves into the implications of fast variations in the reception conditions of the MG member in the RRM strategies CC. Significant variations in the channel quality conditions of the MG members imply the recalculation of the delivery solution. A non-optimized multicast RRM could exponentially increase CC and associated delay with a constant recalculation toward an optimal solution, which could not be tolerated in 5G-MBS scenarios. Thus, novel multicast RRM solutions should consider the CC a critical element in network performance [10].

Considering the previous analysis, this paper proposes novel low-complex multicast RRM strategies for dynamic access technique selection and resource allocation subject to the 5G-MBS paradigm. Our proposal is oriented to address and contextualize the complexity associated with the multicast resource allocation process and the implications of fast variations in the reception conditions of the MG members due to the users' mobility behaviors and the impact of mmWave propagation. The research's main contributions can be summarized as follows:

(i) We address the CC associated with the dynamic multicast RRM strategies in 5G-MBS use cases and highlight the impli-

cations of fast variations in the MG members' reception conditions; (ii) We evaluate the advantage of the dynamic selection of the multicast access technique based on [5] and propose two alternative solutions aided multiclass classification machine learning (ML) algorithms with multi-layer perceptron (MLP) and extra tree classifier (ETC). The proposed solutions allow for tailoring the radio resource allocation regarding users' distributions, multimedia service constraints, and network parameters; (iii) We propose the use of a K-means clustering unsupervised ML approach for detecting and splitting group-oriented MGs based on the collected channel quality indicator (CQI) values at the base station (BS); (iv) We propose a multicast oriented trigger to avoid overrunning the entire algorithm subject to the temporal variations of the MG CQI distribution, reducing the induced latency over the timeslots lattice; (v) Our proposed approaches allow addressing the trade-off between optimal network performance and CC by maximizing specific QoS parameters through non-optimal-solutions considerably reducing the CC of conventional exhaustive mechanism for specific 5G-MBS use cases.

The remainder of the paper is structured as follows. Section II presents the related works, including some theoretical fundamentals and CC analysis oriented to the RRM. Section III details the system model with the mathematical formulation. In Section IV, we formulate the problem by highlighting the CC. Section V gives the proposed solutions. Next, section V discusses the conducted simulations and their results. Conclusions are drawn in section VI.

II. Basic Concepts and Related Works

This section summarizes related works and defines several concepts to understand better and frame our proposal.

A. 5G-MBS and Radio Resource Managment

For the upcoming 5G-Advanced, the RRM is progressively increasing its critical role. As defined in the survey [10], the RRM schemes must implement flexible resource allocation that dynamically allocates available resources with different constraints, such as system throughput and energy awareness.

Over the years, PTM communication has proved its efficiency in delivering high-quality QoS multimedia services to multiple users through cost-effective mechanisms. This is an appealing capability for the upcoming B5G standards, as an essential element of the RRM toolbox, helping to address the ever-growing traffic demands, massive connections, and stringent service requirements [1]. In [11], the authors analyzed the integration of 6G NTNs and the M/B capability to enable scalable services through efficient broadcasting strategies, streaming content to large areas, and offloading popular content to the network edge caching. In [12], Zhou *et al.* delved into the importance of the 5G-MBS capability for 6G massive vehicular IoT in emergency information delivery as a fundamental component of modern transportation systems. As defined in [13], the 5G-MBS standardization is being conducted for the overall 5G system architecture from next-generation radio access and core networks perspectives. The future Release 18 will embrace artificial intelligence technologies providing data-driven RRM solutions [14]. These new potentialities will be the base for the upcoming MBS over B5G.

B. Multicast Access Techniques

The multicasting techniques are divided into single-rate and multi-rate [15]. The single-rate solutions are based on CMS, where the MG is served with the same MCS selected based on the user with the worst channel quality conditions. This traditional solution does not handle the multiuser diversity and tends to suffer from a low capacity efficiency and unfair resource allocation [16].

In the multi-rate schemes, the MG members are served with different MCSs according to the particularities in the channel quality conditions defined by the spatial users' distribution and the heterogeneity of the wireless channels. This approach overcomes the limitations of the single-rate techniques by exploiting the multiuser diversity throughout a group-oriented resource allocation. In [17], the authors proposed the subgrouping techniques as a practical scheme to serve all the users requesting the same content by allocating group-oriented resources to each MG subgroup. This solution allocated the resources to the different users applying subgrouping based on OMA (SOM) as the baseline multi-rate technology. Nevertheless, in [6], [18], the authors proposed using subgrouping based on NOMA (SNOM) in the envisaged 5G environments, where different quality video services were delivered to a group of users interested in the same content. Moreover, in the previous stage of this research [5], we proposed a dynamic multicast access technique selection algorithm (DMATS) based on CMS, SOM, and SNOM for 5G-MBS use cases. The proposal shows the advantages of dynamically selecting among the available multicast access techniques. In [19], we defined theoretical thresholds of outperformance among CMS, OMA, and NOMA, where the most effective access technique is selected based on multiple factors (i.e., CQI variations of the MG members, the percent of users grouped as low channel quality users, and the service constraints). In the last years, multiple works have been specifically oriented to compare the performance gain of NOMA over OMA, such as [20]–[22].

C. Computational Complexity of the Multicast RRM

As defined in [10], recent research about RRM has been mostly oriented on optimizing resource management, spectrum utilization, and interference mitigation, but the associated complexities have been given minor attention. Nevertheless, upcoming RRM solutions and use cases like 5G-MBS significantly increase the network complexity, adding new computational levels, constraints, and hardware needs to provide seamless connectivity and real-time response [10]. In [2], the authors presented a survey on ML-based RRM solutions in B5G networks. The authors explained how ML-based solutions can relax the RMM-associated computational burden and face the trade-off between optimal network goals and CC. In [23] was highlighted how, for the delay aware services, high

induced latency cannot be tolerated during the RRM, making the CC critical during the design of the solutions.

In such a context, one of the challenges of multicast RRM is the CC of the subgrouping techniques for an effective subgroup creation and resource allocation that maximizes a specific cost function. To simplify such complexity, in [24], [25], the authors proved that the subgrouping techniques achieve the best results when creating just two subgroups. The proposed solution achieved optimal subgrouping with reduced complexity considering each user's SINR and corresponding CQI. In [6], the author considered the CC as a KPI in the RRM algorithm. They defined the CC assisted to the implemented solution for multicast subgrouping based on OMA and NOMA and analyzed the implication of the resource blocks and the injection level. Regarding the use of NOMA, in [26], the authors proposed several methods to reduce the implementation complexity and delay of NOMA-based transmission.

In [9], the authors proposed using ML techniques for optimal multicasting, evidencing a trade-off between accuracy and CC for the multicast grouping task. The results offer a CC-efficient solution for optimal multicast grouping in mmWave with directional multi-beam antennas. Moreover, in [27] was considered a resource allocation framework for a 5G mmWave MBS system with beamforming techniques to identify a low-complexity and resource-efficient mechanism that allocates the resources to different UE groups, considering users' priority.

To the authors' knowledge, none of the previous research delves into the implications of fast variations in the reception conditions of the MG member in the dynamic multicast RRM strategies over 5G-MBS use cases. Our proposal aims to fill this gap by analyzing and recreating specific conditions that increase CC due to the users' mobility behaviors or the implications of mmWave propagation and evaluating its implications. Moreover, we propose multicast access techniques solutions to cope with the specific network condition, effectively handling the trade-off between optimal network performance and CC. The proposal is inserted into an HetNet scenario with multiple 5G next-generation BSs (gNBs), such as urban macro/micro stations (UMa/UMi) or unmanned aerial vehicles (UAVs) acting as BSs, supporting differentiated traffic with tight requirements and multiple concurrent users with different mobility behaviors. Assuming a 5G-MBS use case, the BSs can deliver a multicast multimedia service to numerous users requesting the same multicast-oriented content, as shown in Fig. 1.

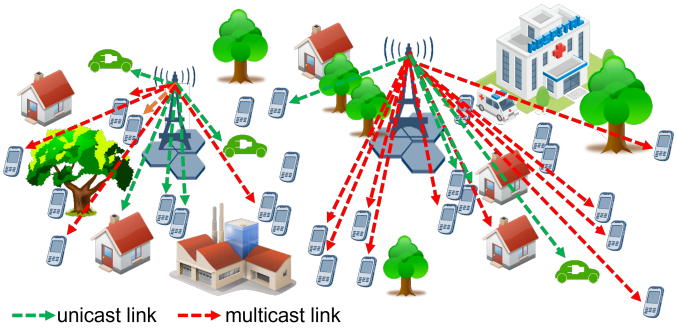


Fig. 1: 5G-MBS capability embedded in an HetNet scenario.

TABLE I: Mathematical Notations.

Notation	Definition
R	Total number of RBs of the BS
R_M	Number of RBs assigned to the multicast session
R_M^*	Effective number of RBs used for resource allocation
B_0	Bandwidth of the RB
Th_{min}, Th_{max}	Minimum and maximum throughput of the service
P_{users}	Percent of users to be served
$\mathcal{K}$, K	Set of users in the MB and total number of users
u_k	User k belonging to $\mathcal{K}$
CQI_{MG}	Set of CQI values reported by the K users in MG
CQI_{min}	Minimum CQI in CQI_{MG}
$M \leq 15$	Number of different CQI values in CQI_{MG}
$SINR_{\text{min}}$	Minimum SINR in the MG
S	Number of subgroups, $S = 2$
$G1$, $G2$	Group 1 and Group 2
$\mathcal{K}_{\text{G1}}$, K_{G1}	Set of users belonging to $G1$, number of users of $G1$
$\mathcal{K}_{\text{G2}}$, K_{G2}	Set of users belonging to $G2$, number of users of $G2$
CQI_{min}^{G1}, CQI_{min}^{G2}	Minimum CQI in $G1$ and $G2$
C_k	Capacity of user k
C_{CMS}, C_{SOM}, C_{SNOM}	Capacity of CMS, SOM, and SNOM
eff	efficiency value of the MCS
$eff_{\text{min}} = eff_{\text{min}}^{G1}$	Minimum eff in the MG and $G1$
$SINR_k^{\text{UL,G1}}$	SINR of the UL after SINR adaptation
$SINR_k^{\text{LL,G2}}$	SINR of the LL after SINR adaptation
$eff_{\text{min}}^{\text{UL,G1}}$	Minimum eff in $G1$ after SINR adaptation
$eff_{\text{min}}^{\text{LL,G2}}$	Minimum eff in $G2$ after SINR adaptation

III. System Model

We consider a gNB with an effective channel bandwidth equal to B and a total number of scheduling resource blocks (RBs) equal to R. One RB is the smallest frequency resource that the gNB can allocate. An RB corresponds to 12 consecutive and equally spaced subcarriers [28]. The 5G NR defines multiples numerologies (μ) for different Subcarrier Spacing (SCS) values, according to $\Delta f = 15 \times 2^{\mu}$, expressed in kHz. B_0 is the bandwidth of an RB, being $B_0 = 12 \times \Delta f$ and also equal to B/R [28]. For better understanding, Table I summarizes the main mathematical notations used throughout the paper.

We assume R_M RBs dedicated to the enabled multicast session, with minimum and maximum throughput (Th_{min}, Th_{max}) as multimedia service constraints. The number of allocated resources to accomplish the multimedia service requirement is R_M^*, with $R_M^* \leq R_M \leq R$. As a constraint, the resource allocation must guarantee to deliver correctly at least the Th_{min} to a predefined percent of the MG users (P_{users}, e.g., 95 %, 98 %, 100 %). We assume K users requesting the same multicast multimedia service, where the sub-index k identifies each user u_k in the MG, with $k \in \{1, 2, 3, \ldots, K\}$. Let $\mathcal{K}$ be the set of MG users, with $K = |\mathcal{K}|$.

Each u_k reports the experienced channel quality through the uplink channel state information (CSI) throughout the CQI

[28]. The array of CQI values reported by MG is $CQI_{\text{MG}} = \{CQI_{k=1}, CQI_{k=2}, \ldots, CQI_{k=K}\}$. We define M as the number of different CQI values reported by the MG (in the service area of the gNB) with $M \leq 15$. Let us define the users' CQI distribution vector $U_{\text{MG}} = \{u_{CQI=1}, u_{CQI=2}, \ldots, u_{CQI=15}\}$ from the CQI_{MG} array, where $u_{CQI=x}$ is the number of users in $\mathcal{K}$ that report a CQI equal to x. Each CQI has associated a specific code rate and efficiency (eff) value of the corresponding MCS according to Table 5.2.2.1-3 in [28].

The reported CQI_k is directly related to the SINR experienced by the u_k. This SINR is higher than or equal to the minimum SINR ($SINR_{\text{min}}$) required to decode the MCS associated with the reported CQI correctly. In our proposal, we assume perfect CSI estimation. The impact of imperfect CSI estimation in the presented results is out of the scope of this research. Nevertheless, in [24], the authors concluded that the imperfect CSI estimation minimally impacts the results when adopting a subgrouping approach.

During the subgrouping creation for SOM and SNOM, we assume that the MG is split into S subgroups. In our proposal, we assume $S = 2$, according to the results presented in [24], [29]. We use the terms $G1$ and $G2$ to identify subgroups 1 and 2, respectively. $G1$ is confirmed by the low channel quality users, identified as the set $\mathcal{K}_{\text{G1}}$ with K_{G1} users (u_{k1}, with $k1 \in \{1, 2, 3, \ldots, K_{\text{G1}}\}$). The set $\mathcal{K}_{\text{G2}}$ identifies the high channel quality users grouped into $G2$, with K_{G2} users (u_{k2}, with $k2 \in \{1, 2, 3, \ldots, K_{\text{G2}}\}$). $K = K_{\text{G1}} + K_{\text{G2}}$, $\mathcal{K} = \mathcal{K}_{\text{G1}} \cup \mathcal{K}_{\text{G1}}$, and $\mathcal{K}_{\text{G1}} \cap \mathcal{K}_{\text{G2}} = 0$. Moreover, we define $CQI^{G1}_{\text{MG}} = \{CQI_{k1=1}, CQI_{k1=2}, \ldots, CQI_{k1=K_{G1}}\}$ and $CQI^{G2}_{\text{MG}} = \{CQI_{k2=1}, CQI_{k2=2}, \ldots, CQI_{k2=K_{G2}}\}$, as the CQI values of the MG, grouped in $G1$ and $G2$, respectively, with $CQI_{\text{MG}} = CQI^{G1}_{\text{MG}} \cup CQI^{G2}_{\text{MG}}$.

The capacity (C) assigned to each u_k, expressed in bit/s, is calculated as $r_k \times B_0 \times eff_k$, where r_k is the number of RBs assigned to the u_k and eff_k, represented in bits/s/Hz, is the efficiency value associated to the CQI feedback of u_k. From the multicast perspective, we define the C for the considered multicast technique as

$$C_{\text{CMS}} = R^*_M \times B_0 \times eff_{\text{min}}, \tag{1}$$

$$C_{\text{SOM}} = r_{G1} \times B_0 \times eff^{G1}_{\text{min}} + r_{G2} \times B_0 \times eff^{G2}_{\text{min}}, \tag{2}$$

$$C_{\text{SNOM}} = r_{G1} \times B_0 \times eff^{\text{UL,G1}}_{\text{min}} + r_{G2} \times B_0 \times eff^{\text{LL,G2}}_{\text{min}}. \tag{3}$$

In the case of CMS, the eff_{min} is the minimum eff corresponding to the MCS assigned to the u_k that reports the lowest CQI_k (CQI_{min}) in the MG. Therefore, the CMS technique assigns the same capacity C_{CMS} to the entire MG, based on the user u_k with the lowest channel quality.

For SOM, the eff^{G1}_{min} is the eff corresponding to the MCS assigned to the u_{k1} belonging to G1 that reports the lowest CQI (CQI^{G1}_{min}, after the subgrouping creation). Therefore, the SOM technique assigns the same capacity $C1_{\text{SOM}}$ to all members of the multicast subgroup G1.

The eff^{G2}_{min} is the eff corresponding to the MCS assigned to the user u_{k2} belonging to the G2 that reports the lowest CQI (CQI^{G2}_{min}). Therefore, the capacity $C2_{\text{SOM}}$ is assigned to the entire multicast subgroup G2 members. The CQI^{G2}_{min} is always higher than or equal to CQI^{G1}_{min}.

In the SOM technique, the r_{G1} and r_{G2} ($r_{G2} = R^*_M - r_{G1}$) represent how the R^*_M are split into G1 and G2 to accomplish with the throughput constraints of the multicast multimedia service. The resulting capacity of G1 (C_{G1}) and G2 (C_{G2}) are the first and second part of (2), respectively. The final resource allocation is s.t. $Th_{min} \leq C_{\text{G1}} < C_{\text{G2}} \leq Th_{max}$.

In the case of the multicast technique SNOM, the $eff^{\text{UL,G1}}_{\text{min}}$ and $eff^{\text{LL,G2}}_{\text{min}}$ are the result of the SNR adaptation [30] in the layer division multiplexing (LDM) approach. We assume that the lowest channel quality users grouped into G1 are powered multiplexed into the upper layer (UL) and the highest channel quality users G2 into the lower layer (LL). The SINR adaptation allows calculating from the real SINR (in dB) of each u_k ($SINR_k$) the corresponding SINR (in dB) for the UL ($SINR^{\text{UL,G1}}_k$) or for the LL ($SINR^{\text{LL,G2}}_k$), as defined in [30]. We consider the IL values between -25 and -5 at steps of 1 and between -5 and 0 at steps of 0.5, with l equal to 31 [6].

A. QoS Performance Criteria

To evaluate the performance of the proposal from the QoS perspective, we consider the following criteria: the system ADR, proportional fairness (PF), and minimum dissatisfaction index (MDI) as defined in [18], [24].

The ADR is the sum of the C_k delivered to each u_k in the MG. For the subgrouping approaches, it will be the sum of the total capacity delivery to G1 and G2, respectively. It enables us to compare how efficiently the proposed multicast access strategies allocate their resources, maximizing the sum of the total experienced throughput of the MG. The drawback of this metric is the lack of information about how fair the available multicast RBs were distributed among users regarding their CSI diversity [5].

The PF and MDI are two well-known metrics to assess how fair the available resources were allocated among the MG members [5]. In [31], the authors demonstrated that one unique fair allocation exists, and it is obtained by maximizing the sum of the logarithm of the data rate of each u_k. The MDI is computed as the sum of the ratio between the data rate achieved by each u_k and the maximum possible data rate value achieved if all RBs (R^*_M) are assigned to the u_k [18].

IV. Problem Formulation: Complexity associated with the dynamic multicasting strategies

An effective 5G-MBS RRM strategy aims to maximize the QoS of the MG members by efficiently allocating the available resources of the multicast sessions. The metrics to evaluate how efficiently the resources were allocated could be the above-mentioned or a weighted combination of them and additional complex metrics. Nevertheless, the envisaged ultra-dense HetNet (Fig. 1) with differentiated services and tight QoS requirements is a dynamic scenario, making managing and exploiting network resources even more complex [3].

In such a context, the necessity of a constant recalculation and the intrinsic convergence time of the best multicast

access technique and resource allocation (to maximize the CF) strategies could become a critical factor from the computational complexity point of view and for delay-sensitive traffic, increasing the control plane delay, the total end two end delay (e2eDelay) and inducing a critical extra latency in the communication. Therefore, CC is also considered a KPI for our proposed dynamic access technique selection and resource allocation solution for MGs subject to the 5G-MBS paradigm.

A. Dynamic Multicasting Complexity

The problem space of the dynamic selection of the best multicast access technique and resource allocation, from the CC perspective, can be divided into three main phases: (a) the selection of the best multicast access technique (i.e., MCS, SOM, or NSOM for our approach); (b) the MG subgrouping for the multi-rate solutions; (c) the selection of the optimal resource allocation that maximizes a specific CF.

Selecting the best multicast access technique through an exhaustive search strategy (ESS) implies evaluating the N available multicast access techniques (i.e., $N = 3$ in our analysis). This ESS multicast access technique selection (ESS-MAT) has a CC equal to the sum of the intrinsic CC of the N techniques, as presented below.

The CC of CMS (CC_{CMS}), when the Th_{min}, Th_{max} and P_{users} are considered as the constraints of the service, can be defined as $\mathcal{O}(M * R_M)$. M is the number of different CQI values reported by the MG in CQI_{MG} ($M \leq 15$), and R_M the available RBs to iterate over subject to the throughput constraints. If the P_{users} is assumed as 100 %, the CMS algorithm only considers the CQI value corresponding to the low channel quality user (i.e., the CQI_{min} of the MG). Therefore, the M term has no impact on the CC. Moreover, if Th_{max} is assumed as infinity, R^*_M is equal to R_M, and R_M have not impact in the complexity. For these specific conditions, the CC of CMS can be assumed as insignificant. In the following, we will assume P_{users} as 100 % and Th_{max} as ∞ to simplify the analysis.

The subgrouping process for the multi-rate solutions has an essential contribution to the CC of these techniques. In [24], the authors defined that the subgrouping process solved through an ESS has a CC equal to $\mathcal{O}(M^S)$, where S is the number of considered subgroups, assumed as equal two (i.e., $\mathcal{O}(M^2)$). Following the assumption of P_{users} as 100 %, such CC can be redefined as $\mathcal{O}(M)$ because the CQI^{G1}_{min} is fixed to the CQI_{min} of the MG, then the algorithm only has to iterate over the remaining $M - 1$ combinations.

The multi-rate approaches for each subgrouping alternative must evaluate the best resource allocation that maximizes the specific CC, returning the best combination overall. For SOM, the CC (CC_{SOM}), including the subgrouping process (for ESS), can be defined as $\mathcal{O}(M^S * R_M)$ and redefined after the assumptions as $\mathcal{O}(M * R_M)$. In the case of SNOM, the CC (CC_{SNOM}), including the subgrouping process (for ESS), can be defined as $\mathcal{O}(M^S * R_M * l)$, and redefined after the assumptions as $\mathcal{O}(M * l)$. Regarding the assumptions, SNOM uses all the available R_M; therefore, this factor is not considered in the CC_{SNOM}.

We can summarize the analysis by defining the CC of the dynamic selection of the best multicast access technique and resource allocation based on an ESS, ($CC_{ESS-MAT}$) as equal to $CC_{CMS} + CC_{SOM} + CC_{SNOM}$, defined as:

$$CC_{ESS-MAT} = \mathcal{O}(M * R_M) + \mathcal{O}(M * l). \quad (4)$$

According to the above-mentioned assumptions, the CC_{CMS} is considered insignificant. The inefficiency of the ESS-MATS approach in terms of CC reaches its maximum expression each time CMS is selected as the best multicast access technique. The major drawback of ESS-MATS is that $CC_{ESS-MAT}$ is constant independently of the MG CQI distribution and increases linearly with the number of M, R_M, and l.

To assess the effectiveness of our proposals and the benchmark algorithms in terms of CC, we measure the central processing unit (CPU) execution time (Et) in seconds of the evaluated algorithms using a predefined function of the Python library Time [32]. The Et method only measures the CPU's time executing the code without including the time spent waiting for input/output resources or sleeping time.

B. Implications of the Channel Quality Conditions Variation

The complexity analysis can not be just faced from the perspective of the intrinsic CC associated with the dynamic multicast access technique selection and resource allocation process, as defined in the above Subsection. How often the RRM has to recalculate these solutions can directly affect the network performance, increasing the control plane delay and overhead and affecting delay-sensitive traffic.

As the frequency increase, increases as well, the large and small-scale fading, with higher path loss attenuation, higher penetration loss, and higher shadowing and fast fading attenuation [33]. If this frequency dependence is essential at the traditional μWave (below 6G) band, it becomes more critical at mmWave [33]. Another essential factor directly correlated with the dynamics in the channel conditions is the mobility behavior of the users, and the dimensions of its impact are also highly frequency dependent.

Let us define the metric CQI changing ratio (CCR) as the percent of u_k in the MG that change its reported CQI from the instant $t-1$ to the instant t concerning K. Fig. 2 shows how the CCR is affected by the increment in the propagation frequency and the users' velocity. In the Figure, we can see how the CCR increases almost linearly with the speed and the frequency. These simulations were carried out through our homemade numerical link-level simulator [34], for $K = 50$, 60 seconds of simulation with 100 ms of the resolution, and 20 simulation runs. Section VI presents additional details about the simulation setup.

As the worst case, in terms of CC, let us assume a dynamic multicast access technique solution based on ESS-MAT that executes the algorithm at each simulation time-step with a long Et and independently of the user's CQI variation. For these ESS-MAT solutions, the mean Et over 60 s of simulation will remain almost constant (s.t. the intrinsic CC of ESS-MAT) independent of the users' velocity and frequency. This optimal

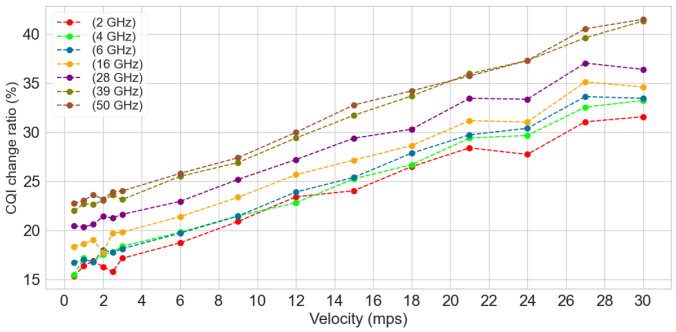


Fig. 2: CQI change ratio regarding the propagation frequency and users velocity.

solution can be considered the superior bound in terms of CC for our specific problem. If we consider a basic trigger that enables to run off the ESS-MAT (btESS-MAT) algorithm if more than 20 % of the MG members change their reported CQI from the time-step $t-1$ to t, it will considerably reduce the mean Et over the 60 s of simulation.

This basic trigger solution based on the 20 % can imply overrunning the algorithm where the CQI changes are not in the low-channel quality users or letting users without service because its CQI change does not represent the 20 %. To address this challenge, we propose an MG-oriented trigger solution that considerably reduces the mean Et and is less affected by users' velocity and propagation at high mmWave.

V. Proposed Solution

Bearing the previous analysis, we aim to propose novel approaches for the dynamic access technique selection and resource allocation for MGs with dynamic reception conditions, specific multimedia service constraints, and network parameters. Our proposal is oriented to highlight and address the complexity associated with the multicast resource allocation process regarding the change ratio in the reception conditions of the MG members.

Our proposal can be divided into two main components: (i) a classifier for the selection of the best multicast access technique that better suits the network conditions, (ii) a K-means clustering for detecting and splitting group-oriented user distributions and a trigger algorithm that allows executing the entire algorithm dynamically.

A. Multicast Access Technique Selection

As presented in [5], selecting the best multicast access technique that better suits the specific network conditions can be found heuristically based on outperformance equations (OEs) among CMS, SOM, and SNOM. Such outperformance equations can be defined as

$$\Delta_{CMS}^{SOM} = 1 - \frac{eff_{min}^{G1}(\kappa\eta)^{-1}}{eff_{min}^{G1} + eff_{min}^{G2}(1+(\kappa\eta)^{-1} - \kappa^{-1} - \eta^{-1})}, \tag{5}$$

$$\Delta_{CMS}^{SNOM} = 1 - \frac{eff_{min}^{G1}(\kappa)^{-1}}{eff_{min}^{UL,G1} + eff_{min}^{LL,G2}(\kappa^{-1}-1)}, \tag{6}$$

Algorithm 1: OE-MAT

Input: $CQI_{min}^{G1}, CQI_{min}^{G2}, B_0, R_M, \mathcal{I}, Th_{min}$
Output: Multicast access technique selection
1: From CQI_{min}^{G1} **and** CQI_{min}^{G2} **determine** eff_{min}^{G1} **and** eff_{min}^{G2} **from Table 5.2.2.1-3 in [28].**
2: Apply SINR adaptation [30] to find: $eff_{min}^{UL,G1}$ and $eff_{min}^{LL,G2}$
3: Compute r_{G1}**:** $r_{G1} = Th_{min}/(B_0 \times eff_{min}^{G1})$
4: Compute κ **and** η**:** $\kappa = K_{G1}/K$; $\eta = r_{G1}/R_M$
5: Compute 5, 6 and 7
if $\Delta_{CMS}^{SOM} \leq 0$ *and* $\Delta_{CMS}^{SNOM} \leq 0$ **then**
  return: CMS
else if $\Delta_{CMS}^{SOM} \geq 0$ *and* $\Delta_{SOM}^{SNOM} \leq 0$ **then**
  return: SOM
else if $\Delta_{CMS}^{SNOM} \geq 0$ *and* $\Delta_{SOM}^{SNOM} \geq 0$ **then**
  return: SNOM
end

$$\Delta_{SOM}^{SNOM} = 1 - \frac{eff_{min}^{G1}\eta + eff_{min}^{G2}(\eta + \kappa^{-1} - \eta\kappa^{-1} - 1)}{eff_{min}^{UL,G1} + eff_{min}^{LL,G2}(\kappa^{-1}-1)}, \tag{7}$$

where $\kappa = K_{G1}/K$ and $\eta = r_{G1}/R_M$. With these equations, we can determine which multicast technique performs best in the QoS criteria system ADR subject to the specific network conditions. Following the methodology presented in [5], we propose a heuristic solution called OE-MAT to select the multicast access technique. The pseudo-code for implementing OE-MAT is presented on Algorithm 1.

To take advantage of the existing correlation among the specific service constraints, the network conditions, and the MG members' CQI distribution by selecting the best multicast access technique, we face the problem as a supervised ML multiclass classification [35]. Supervised ML classification algorithms are oriented to develop a learning model from a labeled dataset [36].

In our specific problem, we have a training dataset ($\mathcal{D}_{train}$) of S_{train} samples of the form (X_f, y), where $f \in \{1, 2, \ldots, F\}$ and $y \in \{1, 2, \ldots, C\}$, with F features and one label y of C classes. The goal of the training process is to obtain a learning model $\mathcal{H}$ such that $\mathcal{H}(X_f) = y$ for unseen samples of a testing dataset ($\mathcal{D}_{test}$) with S_{test} samples [35]. In this multiclass classification problem, we have a C equal to three classes, i.e., CMS, SOM, and SNOM, labeled 1, 2, and 3, respectively.

To obtain a set of features that represent the actual state of the network, s.t. our specific problem, we consider the users' CQI distribution vector U_{MG} divided by the K multicast group members U_{MG}^{K}, such as $\{u_{CQI=1}/K, u_{CQI=2}/K, \ldots, u_{CQI=15}/K\}$. We also included the features K, R_M and Th_{min}, for $F = 18$ features.

In Fig. 3, a diagram is presented with the main components of the dataset creation and ML algorithms training. The complete process was carried out in a Python environment. For dataset creation, at each iteration, we generate a random value

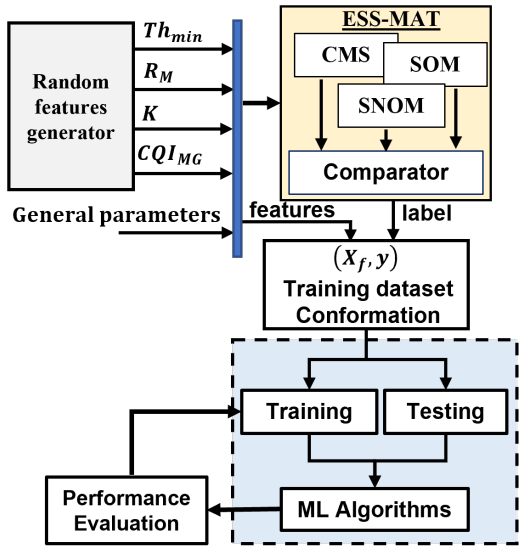


Fig. 3: Diagram of the dataset creation and ML algorithm training process.

of K between 25 and 100 users, a Th_{min} of the multicast multimedia service between 1 Mbps and 60 Mbps, and a R_M value between 100 and 250 RBs (with a resolution of 50) enabled for the multicast session. The Th_{min} range allows to recreate of a wide range of eMBB services, as seen in [37]. For the available RBs, we consider the 5G NR numerology $\mu = 2$ for an SCS equal to 60 kHz and $B_0 = 720$ kHz as recommended in [38] for μWave and mmWave application.

To generate the CQI_{MG}, we define four variables subject to the following bounds: $1 \leq min_{G1} \leq 15$, $min_{G1} \leq max_{G1} \leq 15$, $max_{G1} \leq min_{G2} \leq 15$ and $min_{G2} \leq max_{G2} \leq 15$. We randomly generate these variables as the CQI limits of possible group-oriented CQI distribution. Then, we randomly generate from 30 % to 70 % of the K CQI values, between min_{G1} and max_{G1}, and the remainder between min_{G2} and max_{G2}. This approach enables us to recreate group-oriented CQI distribution with different percentages of the users as low and high-channel quality. Under these constraints, we generate a dataset of 20000 samples.

We opt for a numerically generated dataset because it enables us to obtain extensive data, covering a wide action space and directly oriented to our specific problem. Each sample created by the Random feature generator plus other extra network parameters are used to run the ESS-MAT algorithms for the multicast access technique selection and to add the corresponding label to each feature set X_f in the dataset. The resulting dataset has 9680 samples of SNOM (3), 4545 samples of SOM (2), and 4438 samples of CMS (1). The remaining samples got the label 0 and were deleted from the dataset because it means that for these samples, the randomly generated CQI distribution and R_M values can not satisfy the Th_{min} constraint, with any of the available access techniques. To lead with the unbalanced dataset, we use the Python library SMOTE (Synthetic Minority Oversampling Technique) [39] that oversamples the minority classes in the dataset.

The training process of the ML algorithms includes data normalization, data train/test split, grid-search/cross-validation, and, finally, the evaluation of the algorithms through specific error metrics [40]. We apply the Min-Max scaling method for normalization, transforming all features into the range $[0, 1]$ [40]. The dataset was split into 80 % and 20 % for the training and testing, respectively. We apply Grid-search and k-fold (with $k = 5$) cross-validation to evaluate multiple combinations of the hyperparameters associated with each ML algorithm and use as a tuning criterion the F1 score [40].

To solve the problem, we evaluate in the above-described training process the performance of several scikit-learn native multiclass classifiers [40]. These estimators have multi-learning support as described in [35]. After multiple iterations, the best multiclass classifiers learning models $\mathcal{H}$ were obtained with Multilayer Perceptron (MLP) [41] and Extra-Trees Classifier (ETC) [42] as $\mathcal{H}_{MLP}$ and $\mathcal{H}_{ETC}$, respectively. The ETC algorithm builds an ensemble of unpruned decision or regression trees according to the classical top-down procedure. After the tuning process, the number of estimators to form the ensemble was set to 450, with 100 as the maximum depth of the trees. The number of features to consider for the best split was set as the square root of the total number of features in the dataset. In the case of the artificial neural network (ANN) algorithm MLP, after the tuning process, the selected activation function was ReLU (Rectified linear activation function), Adam (Adaptive Moment Estimation) as the optimization algorithm, a learning rate of 0.03, and one hidden layer of 150 neurons.

To assess the effectiveness of the proposed classifiers, we consider the metrics accuracy and F1 score, including a 3x3 confusion matrix analysis [40]. The accuracy of predicting each one of the classes can be defined as

$$Accuracy = \frac{CP}{T}, \tag{8}$$

where $CP = TP + TN$ is the percent of correct classified predictions, including the true positive (TP) and true negative (TN). $T = TP + TN + FP + FN$ is the total number of predictions, including the false positive (FP) and false negative (TN). The TP, FP, TN, and FN are computed based on the correct and incorrect multicast access technique selection for a specific feature set X_f regarding the actual y label obtained with the optimal ESS-MAT algorithm. The model's accuracy will be the mean of the accuracy of each class.

The F1 score is the harmonic mean of the metrics precision and recall giving a balanced measurement of these metrics [40]. This expression is preferable to evaluate unbalanced datasets with an individual evaluation of multiple classes. The precision, recall, and F1 score metrics are defined in [40].

In Algorithm 2, we present the description of the dynamic multicast access technique selection based on MLP (MLP-MAT), and ETC (ETC-MAT), respectively.

Regarding the CC analysis presented in Section IV, let us define the CC of the proposed dynamic selection algorithms (DSAs) as CC_{DSA}. Then we can define the CC

Algorithm 2: MLP-MAT—ETC-MAT

Input: $\mathcal{H}_{MLP}|\mathcal{H}_{ETC}, CQI_{MG}, R_M, Th_{min}, K$
Output: Multicast access technique selection
1: From CQI_{MG} **and** K **determine:**
$U^K_{\text{MG}} = \{u_{CQI=1}/K, u_{CQI=2}/K, \ldots, u_{CQI=15}/K\}$
2: Conform X_f **as** $\{U^K_{\textbf{MG}}, R_M, Th_{min}, K\}$
3: Predict y **with** $\mathcal{H}_{MLP}|\mathcal{H}_{ETC}$ **and** X_f**:**
$y = \mathcal{H}_{MLP}.predict(X_f)|\mathcal{H}_{ETC}.predict(X_f)$
5: Select the multicast access technique
if $y = 1$ **then**
  return: CMS
else if $y = 2$ **then**
  return: SOM
else if $y = 3$ **then**
  return: SNOM
end

of the proposed solutions, $CC_{OE-MAT}, CC_{MLP-MAT}$ and $CC_{ETC-MAT}$ as

$$CC = \begin{cases} CC_{DSA}, & \text{if } CMS \\ CC_{DSA} + \mathcal{O}(M * R_M), & \text{if } SOM \\ CC_{DSA} + \mathcal{O}(M * l), & \text{if } SNOM \end{cases} \tag{9}$$

where CC_{DSA} equals the CC of the proposed solutions to find the best multicast access technique. In the case of OE-MAT, we can define CC_{DSA} as $\mathcal{O}(l)$ because the solutions have to iterate over the l IL values to compute the outperformance equations in Algorithm 1. Hence, using OE-MAT suppose a CC equal to $\mathcal{O}(l)$ plus the specific CC of the selected technique. As can be seen, the improvement in terms of CC of OE-MAT concerning ESS-MAT in (4) reaches its maximum value when the selected access technique is CMS because ESS-MAT still has to compute the three access techniques.

B. Subgrouping and Trigger based on K-means Clustering

As presented in the above Section IV, the MG subgrouping process significantly contributes to the CC of these techniques.

Aimed to split CQI_{MG}, we based our solution on the ML K-Means [43] algorithm. The K-Means is a simple unsupervised ML algorithm capable of clustering unlabeled datasets in a few iterations [40]. Based on K-Means, our proposed solution applies a partitional clustering of CQI_{MG}, dividing the CQI values into two non-overlapping groups of CQI values labeled as 0 for $G1$ and 1 for $G2$. If all the CQI values are labeled as 0, all the reported CQI in the MG have the same value. Let us assume our K-Means learning model as $\mathcal{H}_{\text{km}}$, with $\mathcal{H}_{\text{km}}(X^{\text{km}}_f) = y_{\text{km}}$, where X^{km}_f will be the $M \leq 15$ unique CQI values in CQI_{MG}, and y_{km} the M labels (0 or 1) for each unique CQI in CQI_{MG}. Algorithm 3 shows a pseudocode that describes the functions of the proposed clustering strategy.

From the K-Means clustering algorithm, we implement an MG-oriented trigger intending only to run the algorithm OE-MAT, MLP-MAT, or ETC-MAT and recompute the resource allocation strategy if either $CQI_{min} = CQI^{G1}_{\text{min}}$ or CQI^{G2}_{min}

Algorithm 3: K-Means clustering

Input: CQI_{MG}
Output: $flag, CQI_{min} = \{CQI^{G1}_{min}, CQI^{G2}_{min}\}, CQI^{G1}_{\text{MG}}, CQI^{G2}_{\text{MG}}$
1: Conform $X^{\textbf{km}}_f$ **as a sorted array with just the unique values in** CQI_{MG}**:**
$X^{\text{km}}_f = sort(unique(CQI_{MG}))$
2: Predict $y_{\textbf{km}}$ **with** $\mathcal{H}_{\textbf{km}}$ **and** $X^{\textbf{km}}_f$**:**
$y_{\text{km}} = \mathcal{H}_{\text{km}}.predict(X^{\text{km}}_f)$
3: If all the $y_{\textbf{km}}$ **are equal 0,** $flag = 0$**:**
if y_{km} **= 0 then**
  return: $flag = 0, CQI_{min} = min(CQI_{MG})$
else if $y_{km} \neq 0$ **then**
  then: $flag = 1$
  4: Find CQI^{G2}_{min} **as the minimum CQI value in** $X^{\textbf{km}}_f$ **labeled as 1:**
  $CQI^{G2}_{min} = min(X^{\text{km}}_f|_{y_{\text{km}}=1})$
  5: Conform $CQI^{G1}_{\textbf{MG}}$ **and** $CQI^{G2}_{\textbf{MG}}$ **from** $CQI_{MG}, CQI^{G1}_{\textbf{min}}, CQI^{G2}_{\textbf{min}}$ **:**
  return: $flag, CQI^{G1}_{min}, CQI^{G2}_{min}, CQI^{G1}_{\text{MG}}, CQI^{G2}_{\text{MG}}$
end

change from the RRM timeslot $t-1$ to t. The trigger can be defined as

$$trigger_t = \begin{cases} 1, & \text{if } flag_t \neq flag_{t-1} \\ 1, & \text{if } CQI_{min,t} \neq CQI_{min,t-1} \\ 1, & \text{if } CQI^{G2}_{min,t} \neq CQI^{G2}_{min,t-1} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $trigger = 1$ means that RRM must recompute the multicast access technique selection and the resource allocation; otherwise, it will not. In Fig. 4, we present a diagram with the interconnection among the proposed K-Means clustering, the trigger, and the dynamic multicast access technique solutions: OE-MAT, MLP-MAT, or ETC-MAT.

VI. Results and Discussion

The evaluation is conducted through two scenarios based on link-level simulation and an extensive simulation set covering a wide range of configurations. Metrics to assess the ML classification task, the QoS, and the CC are the common thread of the results and discussion of this section. In the following, we will refer to the proposed solutions using the flow diagram presented in Fig. 4 as OE-MAT, MLP-MAT, and ETC-MAT. Only to highlight the advantage of the proposed trigger, we include the index "kt" to differentiate the solution with the trigger (ktOE-MAT, ktMLP-MAT, and ktETC-MAT) and without the trigger (OE-MAT, MLP-MAT, and ETC-MAT).

A. Scenarios description

The two recreated validation scenarios are based on an ad-hoc developed Python-based link-level simulator [34] oriented to recreate 5G-MBS use cases. Table II summarizes the main simulation parameters.

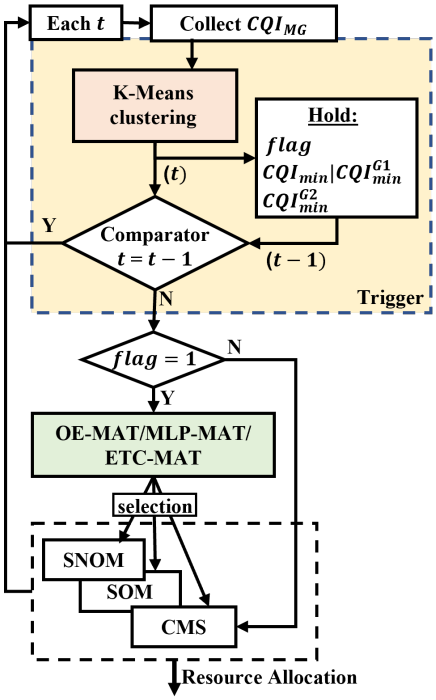


Fig. 4: Flow diagram of the proposed K-Means clustering, the trigger, and the dynamic multicast access technique solutions.

Scenario A) We consider a UMi Street Canyon open area with a single gNB in the middle of a grid of 60x60 meters operating at 28 GHz. We consider the 5G NR numerology $\mu = 2$ for an SCS equal to 60 kHz and $B_0 = 720$ kHz as recommended in [38]. The goal of these scenarios is to generate 1000 samples of uncorrelated network configurations where for each sample, we generate a random number of stationary users K between 25 and 100 users, a Th_{min} of the multicast multimedia service between 1 Mbps and 60 Mbps, and a R_M value between 100 and 250 RBs (with a resolution of 50). The RRM must ensure the service is correctly delivered to 100 % of the users. We consider each sample a network snapshot where the simulator's output will be the instantaneous CQI_{MG} and network parameters. With the outputs of the simulator, we generate a $\mathcal{D}_{test}$ with 1000 samples of unseen data for the proposed dynamic multicast access technique solutions.

Scenario B) We consider a UMi Street Canyon open area with a single gNB in the middle of a grid. We consider the 5G NR numerology $\mu = 2$ for an SCS equal to 60 kHz and $B_0 = 720$ kHz as recommended in [38] compatible with μWave and mmWave applications. The BS provides multicast multimedia service with a Th_{min} of 50 Mbps and 200 available RBs. The RRM has to ensure delivering the service correctly to 100 % of the users. For each simulation, we consider 50 MG members in the service area of the BS, with a random directional mobility model [44], at the same constant speed during 60 s of simulation (with 100 ms of resolution). In this scenario, we test different user's speeds from 0.5 to 3 mps at a resolution of 0.5, and from 3 to 30 mps at a resolution of 3, for 15 different speeds. The selected speed range is oriented to evaluate typical pedestrian and vehicular mobilities. We also evaluate the scenario for different μWave and mmWave frequencies 2, 4, 6, 16, 28, 39, and 50 GHz. We adjust the grid size for each frequency to ensure collecting CQI values from 1 to 15 at each run. For each combination of velocity and frequency 15×7, we generate 20 random simulation runs.

TABLE II: Simulation parameters.

Parameter	*Scenario A*	*Scenario B*
Scenario type	UMi Street Canyon	UMi Street Canyon
Frequency (GHz)	28	2, 4, 6, 16, 28, 39, 50
Numerology μ	2	2
*R*Bs bandwidth (kHz)	720	720
Available RBs, R_M	100-250	200
BS/user height (m)	10/1.5	10/1.5
Transmission power (dBm)	10	10
BS/user antenna gain (dB)	10/0	10/0
Large-scale fading models	[45]	[45]
Small-scale fading model	Jakes	Jakes
Dynamic line of sight	Yes	Yes
Antenna	Sectorial (120°)	Sectorial (120°)
Mobility Models	Stationary	Random Directional

TABLE III: Multiclass classification task evaluation.

OE-MAT / MLP-MAT / ETC-MAT				
Class	precision (%)	recall (%)	F1 (%)	support
CMS	93/95/95	100/97/95	97/96/95	373
SOM	98/99/98	100/96/93	99/97/95	356
SNOM	100/92/94	87/94/94	93/93/91	271
mean F1(%)	96.4/95.7/94.1			
Accuracy(%)	96.0/96.0/94.0			

B. Multiclass Classifier Assessment

To validate the performance of OE-MAT, MLP-MAT, and ETC-MAT to select the best multicast access technique that better suits the specific conditions of the network, we use the dataset $\mathcal{D}_{test}$ from *Scenario A*. We employ the optimal solution found by ESS as a reference result. The validation is based on the accuracy, F1 score, precision, and recall metrics. Table III summarizes the evaluation results.

As shown in Table III, the best multiclass classification performance in terms of mean F1 score and accuracy was achieved with the heuristic solution OE-MAT. The worst result with OE-MAT was achieved for the SNOM samples with an F1 score of 93 %. This degradation is due to the 87 % of recall, meaning SNOM samples were misclassified. In contrast, the precision classifying SNOM was 100 %, which means that all the samples classified as SNOM were SNOM. The best ML-based solution was achieved with MLP-MAT with the same global accuracy of 96 % that OE-MAT and a difference of less than 1 % of the mean F1 score.

For the three solutions, the worst performance was classifying SNOM samples. In the case of the ML solutions, one of the reasons for this behavior could be the unbalanced dataset during the training process, where SOM doubled the number of samples concerning SNOM. In the case of the solution based on MLP, the 92 % of precision classifying SNOM

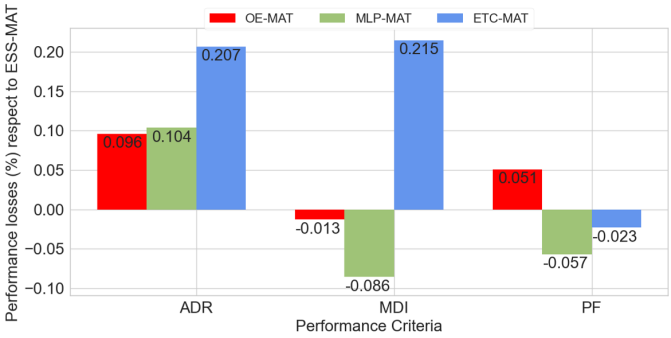


Fig. 5: QoS performance losses for ESS-MAT.

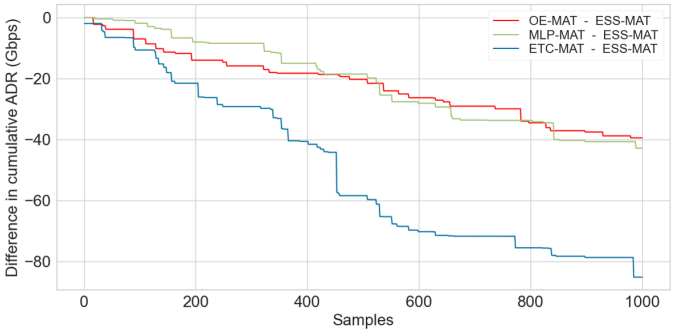


Fig. 6: Difference in cumulative ADR of the proposed solutions and ESS-MAT.

samples is because 1.4 and 0.3 percent of the SNOM samples were wrongly classified as CMS and SOM, respectively. Moreover, 1.1 and 1 percent of the CMS and SOM samples were classified as SNOM, reducing the recall to 94 percent.

The multiclass classification assessment shows that the three proposed solutions are considerably effective, with accuracy and F1 score values higher than 94 percent. In the case of the ML solutions, the results could be improved by increasing the dataset and balancing the classes.

C. QoS Assessment

The performance of the proposed sub-optimal OE-MAT, MLP-MAT, and ETC-MAT was evaluated regarding the metrics system ADR, MDI, and PF. In every case, their performance was compared with the optimal solution ESS-MAT. The validation was based on *Scenario A*, measuring the average performance losses of the proposals concerning the ESS-MAT for the 1000 samples. The results are shown in Fig. 5, evidencing the QoS performance of the proposed solutions and ESS-MAT is almost the same. Specifically, the performance losses of the proposed solutions for ESS-MAT for the three QoS metrics are lower than 0.21 %. The ESS-MAT solution optimizes the system ADR, and OE-MAT and MLP-MAT have a minor ADR performance degradation on the order of 0.1 %. Moreover, OE-MAT and MLP-MAT outperform the ESS-MAT regarding MDI minimally (i.e., a negative value in Fig. 6). It happens because even when the resource allocation of the proposed algorithms is optimized for the ADR, the miss-classification during the selection of the best multicast access technique tends to favor fairer multicast strategies for ESS-MAT. The same happens for the metric PF with MLP-MAT and ETC-MAT.

To help visualize the evolution of these metrics along the 1000 samples, Fig. 7 shows the difference in cumulative ADR between the proposed solutions and ESS-MAT. In this Figure, for the worst case, the cumulative difference over 1000 samples equals 85.1 Gbps. This value, with ETC-MAT, represents a mean system ADR difference of 85.1 Mbps ($85.1\ Gbps/1000$), the 0.207 % respect to the mean ADR of ESS-MAT over the 1000 samples, equal to 41.138 Gbps. An equivalent analysis can be made for OE-MAT and MLP-MAT, where the mean ADR difference over the 1000 samples equals 39.44 Mbps and 42.8 Mbps.

D. CC Assessment

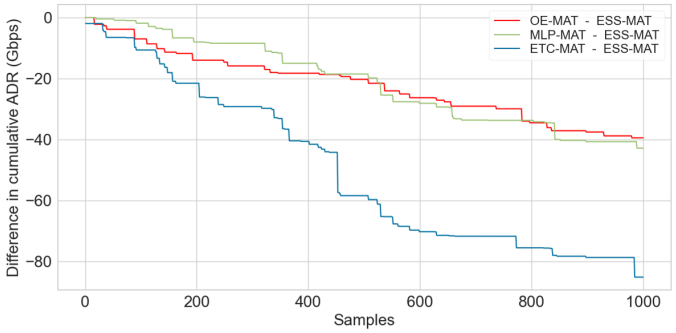
After validating the performance of the proposed solutions in terms of QoS, let us evaluate its CC improvements. As we define above, we evaluate the CC in terms of the Et metric. The simulations were launched in a server with 1 x Barebone Asus ESC4000 G4 1+1 1600W RPSU 4 x GPU 2 x Processor Intel Xeon Gold 6238 2,1GHz 22C 140W 8 x Samsung Memory DDR4 2933MHz 32GB.

Continuing with *Scenario A*, Fig. 8 presents the cumulative Et reduction of the analyzed solutions regarding ESS-MAT. As we can see, the best performance is achieved with MLP-MAT and OE-MAT. The MLP-MAT reduces the mean Et concerning ESS-MAT over 1000 samples by more than 96 %. The solution based on the ETC learning model takes longer to converge, with an improvement of around 15 %. From this result, we can validate the advantage of using a non-optimal solution for the MG splitting and the multicast access technique selection regarding the CC. The proposed algorithms reduce the intrinsic CC associated with this dynamic process in 5G-MBS use cases.

To assess the performance of the proposed MG-oriented trigger, we use *Scenario B* to evaluate the impact of the CQI change ratio increment due to the propagation frequency and user velocity. First, we added the proposed trigger to the ESS-MAT and compared it with the performance of btESS-MAT (ESS-MAT with the 20 % trigger) presented in **??**. In this simulation, we find the mean Et of each solution over 60 seconds and 20 simulation runs for each possible combination of users' speed and frequency.

Fig. 9(a) illustrates the extra Et reduction added by the proposed MG-oriented trigger with ESS-MAT (ktESS-MAT) concerning ESS-MAT and btESS-MAT. For the mmWave cases, the improvement is always bigger than 20 %. This improvement increases with the velocity in all cases. The most important outcome of this simulation is the change in the behavior introduced by the proposed trigger regarding the velocity. We can see how the results for ktESS-MAT have a minimal variation with the velocity increase, which means that the proposed strategy for the trigger is less affected by the increment in the general CQI changing ratio of the MG.

Fig. 9(b) shows the Et reduction of the entire proposal, including the trigger, the K-MEans clustering, and the MLP-based multicast access technique solution. The improvement



(a)

Figs/Difference in cumulative MDI.png

(b)

Figs/Difference in cumulative PF.png

(c)

Fig. 7: Difference in cumulative ADR (a), MDI (b), and PF (c) of the proposed solutions and ESS-MAT.

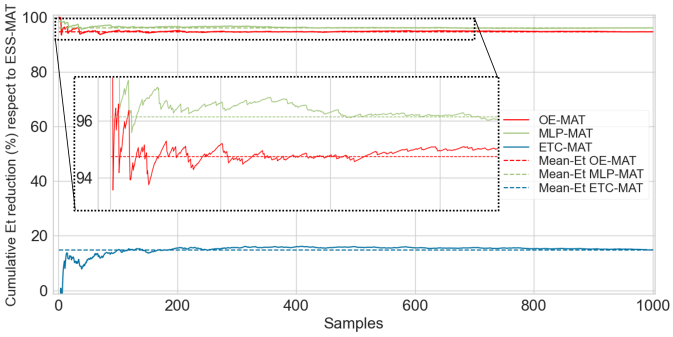
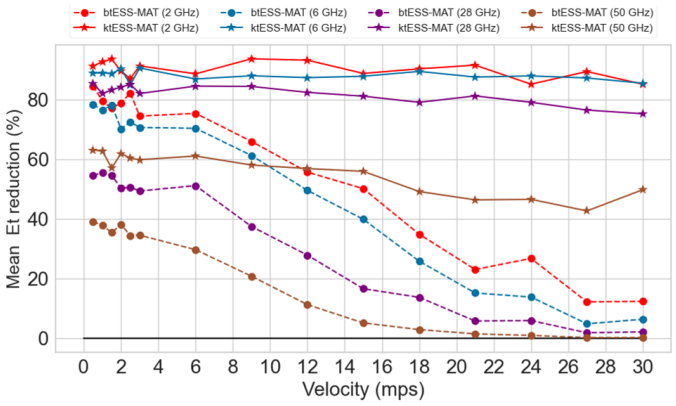
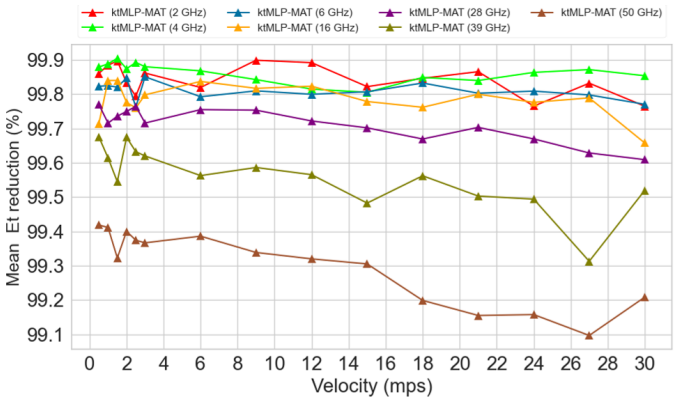


Fig. 8: Cumulative ET reduction regarding ESS-MAT.



(a)



(b)

Fig. 9: Cumulative Et reduction of btESS-MAT, ktESS-MAT (a), and ktMLP-MAT (b) concerning ESS-MAT.

with respect to ESS-MAT is around 99 % for all cases. We can see how the Et reduction increases at around 3 % concerning the result in Fig. 8(b). Such outcomes validate the performance of the proposed solution in terms of CC with an approach that is less sensitive to high CQI changing ratios. This is important in handling high mobility use cases and mmWave scenarios. Moreover, the performance of the solution in terms of QoS is almost the same as the optimal ESS-MAT solution. The validation results show a worthy trade-off between a non-optimal QoS performance of less than 0.25 % with respect to the optimal solution and an improvement of 99 % of CC reduction.

VII. Conclusions

This research was oriented to contextualize and address the complexity associated with the multicast RRM process regarding the fast variations in the reception conditions of the MG members subject to the 5G-MBS paradigm. We propose a dynamic multicast access technique selection and resource allocation strategy based on an MG-oriented trigger, a K-means clustering for detecting and splitting group-oriented user distributions, and a classifier for selecting the best multicast access technique. For the dynamic selection of the multicast access technique that better fits the specific reception conditions of the MG members, we propose a novel approach base on ML-multiclass classification with MLP, ETC algorithms. The proposal results show a worthy trade-off between a non-optimal QoS performance of less than 0.25 % concerning the optimal solution and an improvement of 99 % of CC reduction. The proposed dynamic multicast access technique solution and the following strategies can contribute to the envisaged 5G-MBS use cases. It helps reduce the complexity of RRM and the induced latency in communication. Our approach effectively handled the trade-off between the multicasting gain and the multiuser diversity highlighting the implications of the existing correlation between the users' velocity, the propagation frequency, and the variations in the channel conditions. The proposed solutions allow tailoring the multicast access technique and radio resource allocation regarding users' distributions, multimedia service constraints, and network parameters.

References

[1] C. C. González, S. Pizzi, M. Murroni, and G. Araniti, "Multicasting over 6g non-terrestrial networks: a softwarization-based approach," *IEEE Vehicular Technology Magazine*, 2023.

[2] I. A. Bartsiokas, P. K. Gkonis, D. I. Kaklamani, and I. S. Venieris, "Ml-based radio resource management in 5g and beyond networks: A survey," *IEEE Access*, vol. 10, pp. 83507–83528, 2022.

[3] C. C. González, E. F. Pupo, L. Atzori, and M. Murroni, "Dynamic radio access selection and slice allocation for differentiated traffic management on future mobile networks," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 1965–1981, 2022.

[4] L. Zhang, W. Li, Y. Wu, Y. Xue, E. Sousa, S.-I. Park, J.-Y. Lee, N. Hur, and H.-M. Kim, "Using non-orthogonal multiplexing in 5g-mbms to achieve broadband-broadcast convergence with high spectral efficiency," *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 490–502, 2020.

[5] E. F. Pupo, C. C. González, L. Atzori, and M. Murroni, "Dynamic multicast access technique in sc-ptm 5g networks: Subgrouping with om/nom," in *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–6, IEEE, 2022.

[6] E. Iradier, M. Fadda, M. Murroni, P. Scopelliti, G. Araniti, and J. Montalban, "Nonorthogonal multiple access and subgrouping for improved resource allocation in multicast 5g nr," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 543–556, 2022.

[7] M. N. Dani, D. K. So, J. Tang, and Z. Ding, "Resource allocation for layered multicast video streaming in noma systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 11379–11394, 2022.

[8] A. de la Fuente, G. Interdonato, and G. Araniti, "User subgrouping and power control for multicast massive mimo over spatially correlated channels," *IEEE Transactions on Broadcasting*, vol. 68, no. 4, pp. 834–847, 2022.

[9] N. Chukhno, O. Chukhno, D. Moltchanov, A. Gaydamaka, A. Samuylov, A. Molinaro, Y. Koucheryavy, A. Iera, and G. Araniti, "The use of machine learning techniques for optimal multicasting in 5g nr systems," *IEEE Transactions on Broadcasting*, 2022.

[10] S. Manap, K. Dimyati, M. N. Hindia, M. S. A. Talip, and R. Tafazolli, "Survey of radio resource management in 5g heterogeneous networks," *IEEE Access*, vol. 8, pp. 131202–131223, 2020.

[11] M. M. Azari, S. Solanki, S. Chatzinotas, O. Kodheli, H. Sallouha, A. Colpaert, J. F. M. Montoya, S. Pollin, A. Haqiqatnejad, A. Mostaani, *et al.*, "Evolution of non-terrestrial networks from 5g to 6g: A survey," *IEEE communications surveys & tutorials*, 2022.

[12] F. Zhou, L. Feng, P. Yu, W. Li, X. Que, and L. Meng, "Drl-based low-latency content delivery for 6g massive vehicular iot," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14551–14562, 2021.

[13] V. K. Shrivastava, S. Baek, and Y. Baek, "5g evolution for multicast and broadcast services in 3gpp release 17," *IEEE Communications Standards Magazine*, vol. 6, no. 3, pp. 70–76, 2022.

[14] X. Lin, "An overview of 5g advanced evolution in 3gpp release 18," *IEEE Communications Standards Magazine*, vol. 6, no. 3, pp. 77–83, 2022.

[15] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for ofdma-based systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 240–254, 2012.

[16] Y. Zhang, D. He, Y. Xu, Y. Guan, and W. Zhang, "Mode selection algorithm for multicast service delivery," *IEEE Transactions on Broadcasting*, vol. 67, no. 1, pp. 96–105, 2020.

[17] G. Araniti, M. Condoluci, L. Militano, and A. Iera, "Adaptive resource allocation to multicast services in lte systems," *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 658–664, 2013.

[18] J. Montalban, P. Scopelliti, M. Fadda, E. Iradier, C. Desogus, P. Angueira, M. Murroni, and G. Araniti, "Multimedia multicast services in 5g networks: Subgrouping and non-orthogonal multiple access techniques," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 91–95, 2018.

[19] E. F. Pupo, C. C. González, L. Atzori, and M. Murroni, "Thresholds of outperformance among broadcast/multicast access techniques in 5g networks," in *2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–6, IEEE, 2021.

[20] J. Ghosh, I.-H. Ra, S. Singh, H. Haci, K. A. Al-Utaibi, and S. M. Sait, "On the comparison of optimal noma and oma in a paradigm shift of emerging technologies," *IEEE Access*, vol. 10, pp. 11616–11632, 2022.

[21] J. Ghosh, V. Sharma, H. Haci, S. Singh, and I.-H. Ra, "Performance investigation of noma versus oma techniques for mmwave massive mimo communications," *IEEE Access*, vol. 9, pp. 125300–125308, 2021.

[22] 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 Transactions on Communications*, vol. 68, no. 1, pp. 536–568, 2019.

[23] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, *et al.*, "Enabling technologies for ultra-reliable and low latency communications: From phy and mac layer perspectives," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2488–2524, 2019.

[24] G. Araniti, M. Condoluci, A. Iera, A. Molinaro, J. Cosmas, and M. Behjati, "A low-complexity resource allocation algorithm for multicast service delivery in ofdma networks," *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 358–369, 2014.

[25] G. Araniti, M. Condoluci, M. Cotronei, A. Iera, and A. Molinaro, "A solution to the multicast subgroup formation problem in lte systems," *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 149–152, 2015.

[26] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of noma: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[27] P.-Y. Su, K.-H. Lin, Y.-Y. Li, and H.-Y. Wei, "Priority-aware resource allocation for 5g mmwave multicast broadcast services," *IEEE Transactions on Broadcasting*, 2022.

[28] G. T. . 214, "5g; nr; physical layer procedures for data (3gpp ts 38.214 version 16.2.0 release 16)," 2020.

[29] W. Guo and B. Mouhouche, "A method to tailor broadcasting and multicasting transmission in 5g new radio," in *2019 European Conference on Networks and Communications (EuCNC)*, pp. 364–368, IEEE, 2019.

[30] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H. M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered-division-multiplexing: Theory and practice," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 216–232, 2016.

[31] J.-Y. Le Boudec, "Rate adaptation, congestion control and fairness: A tutorial," *on line*, 2008.

[32] Python, "time-time access and conversions." https://docs.python.org/3/library/time.htmltime.process$_time$, 2023.

[33] M. Shafi, J. Zhang, H. Tataria, A. F. Molisch, S. Sun, T. S. Rappaport, F. Tufvesson, S. Wu, and K. Kitao, "Microwave vs. millimeter-wave propagation channels: Key differences and impact on 5g cellular sys-

tems," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 14–20, 2018.

[34] E. F. Pupo, C. C. González, E. Iradier, J. Montalban, and M. Murroni, "5g link-level simulator for multicast/broadcast services," in *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–6, IEEE, 2023.

[35] M. Aly, "Survey on multiclass classification methods," *Neural Netw*, vol. 19, no. 1-9, p. 2, 2005.

[36] C. C. González, E. F. Pupo, D. P. Ruisanchez, D. Plets, and M. Murroni, "From mfn to sfn: Performance prediction through machine learning," *IEEE Transactions on Broadcasting*, vol. 68, no. 1, pp. 180–190, 2021.

[37] H. Alves, G. D. Jo, J. Shin, C. Yeh, N. H. Mahmood, C. Lima, C. Yoon, N. Rahatheva, O.-S. Park, S. Kim, *et al.*, "Beyond 5g urllc evolution: New service modes and practical considerations," *arXiv preprint arXiv:2106.11825*, vol. 7, 2021.

[38] Y. Kim, Y. Kim, J. Oh, H. Ji, J. Yeo, S. Choi, H. Ryu, H. Noh, T. Kim, F. Sun, *et al.*, "New radio (nr) and its evolution toward 5g-advanced," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 2–7, 2019.

[39] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[40] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[41] B. W. Suter, "The multilayer perceptron as an approximation to a bayes optimal discriminant function," *IEEE transactions on neural networks*, vol. 1, no. 4, p. 291, 1990.

[42] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, pp. 3–42, 2006.

[43] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[44] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 3, pp. 1897–1907, IEEE, 2005.

[45] "Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 14)," tech. rep., 3GPP TR 38.901 V14.1.1, July 2017.