IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 February 2024, accepted 25 February 2024, date of publication 1 March 2024, date of current version 7 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3372417

RESEARCH ARTICLE

An Energy-Aware Tailored Resource Management for Cellular-Based Zero-Touch Deterministic Industrial M2M Networks

SAHAYA BENI PRATHIBA[1], (Member, IEEE), KATHIROLI RAJA[2], R. V. SAIABIRAMI[2], AND GUNASHREE KANNAN[2]
[1]Centre for Cyber Physical Systems, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India
[2]Department of Computer Technology, Anna University, Chennai 600025, India

Corresponding author: Sahaya Beni Prathiba (sahayabeni@ieee.org)

This work was supported by the Vellore Institute of Technology, Chennai, India.

ABSTRACT Zero-Touch Deterministic Industrial Machine-to-Machine (ZT-DI-M2M) serves as a customized communication solution designed to meet the specific needs of industrial settings. Though 5G is the promising solution for ZT-DI-M2M, optimized scheduling in 5G remains challenging to design over aspects such as network overload and congestion control, especially in the uplink transmission of ZT-DI-M2M. Effective allocation of resources for Machine-Type Communication (MTC) devices stands out as a pivotal hurdle within the domain of 5G networks. This challenge directly influences the longevity of battery-operated devices and the Quality of Service (QoS) experienced by applications. This paper aims to develop a Group-Based and Energy Aware (GBEA) resource allocation algorithm for M2M communication in the 5G networks. The GBEA algorithm solves the resource allocation problem by initially clustering the active nodes concerning their delays. Consequently, inter and intra-cluster resource distribution occurs through the Gaussian Mixture Model Expectation Maximization (GMM-EM) algorithm. The GBEA algorithm optimizes the resource allocation of M2M devices by factoring delay, energy, proximity, and fairness into the allocation process. The simulation outcomes unveil the superiority of the GBEA scheduling algorithm over established state-of-the-art resource allocation methods. It showcases remarkable enhancements in throughput, delay sensitivity, and energy efficiency, boasting nearly 1.5-fold, 1.75-fold, and 6.5-fold respective improvements.

INDEX TERMS Resource allocation, M2M communication, scheduling, 5G networks, delay minimization, energy efficiency.

I. INTRODUCTION

Advancements in self-governing Machine-to-Machine (M2M) communication, alternatively termed as Machine Type Communication (MTC), aligns with the definitions outlined in the 3rd Generation Partnership Project (3GPP) specifications [1]. It pertains to data exchange involving single or multiple entities, independent of human intervention, as defined by the 3GPP standards [2]. The M2M deployment requires a suitable data transfer network to create connectivity among M2M devices, which had previously relied on wired data transfer networks and radio networks [3]. An abundant proliferation of connected devices within the communication network results in a substantial surge in mobile traffic.

Zero-Touch Deterministic Industrial Machine-to-Machine (ZT-DI-M2M) is a specialized communication system tailored for industrial environments [4]. It facilitates seamless and dependable communication among machines and devices, prioritizing automation and precision. This system guarantees not only effortless data exchange, denoted as zero-touch but also exceptional predictability regarding latency and reliability. ZT-DI-M2M finds its utmost significance within industrial domains, where instantaneous data sharing is indispensable for functions like process control,

The associate editor coordinating the review of this manuscript and approving it for publication was Stefano Scanzio.

automation, and real-time monitoring [5]. Its primary role is to ensure that machine-to-machine communication not only operates with high efficiency but also adheres rigorously to demanding standards for minimal latency and unfaltering reliability. 5G is the recent communication network that supports ZT-DI-M2M. However, the 5G network infrastructure must facilitate end device expansion to meet the users' critical and non-elastic service demands [6], [7]. In response to this challenge, 5G networks must adeptly manage the extensive M2M traffic originating from connected devices and machines while ensuring uninterrupted and seamless wireless connectivity.

Many M2M devices necessitate periodic network access to fulfill their data transmissions, often operating at meager data rates. Examples of critical M2M services encompass wireless sensors, environmental monitoring, weather data collection, and vehicular communication [8], [9], [10]. Moreover, wearable sensors designed for assessing user health are gaining prominence. These wearable devices actively monitor patients' well-being and can promptly trigger alerts in the event of a health concern [11]. In 5G networks, such devices must be scheduled to utilize the available network resources for processing with a lower response time [12]. With flexible system protocols that can be modified according to user requirements, 5G networks are expected to accommodate massive critical communication.

In contrast to prior network generations, which predominantly tasked infrastructure with control and processing duties, 5G seeks to rebalance this equation through architectural shifts, transitioning from a cell-centric to a device-centric paradigm [13]. Ultra-Reliable Low Latency Communication (uRLLC), massive Machine Type Communication (mMTC), and Enhanced Mobile Broadband (eMBB) constitute the fundamental use case categories that form the foundation for 5G networks to provide data rates surpassing 10 Gbps with low-latent data transmissions, facilitates the widespread deployment of low-power IoT devices (reaching 1 million devices per square kilometer). However, for M2M communication, most traffic is created and transferred via the uplink [14]. As a result, 5G communication for M2M devices may surface issues like energy efficiency and low latency. Recent studies focus on energy-efficient MTC devices, Quality of Service (QoS)-driven network architectures, and energy efficiency in response to increased network activity, diverse services, and user demands including latency, throughput, and bandwidth. With the implementation of new low data rate schemes and careful design of remote nodes, 5G can be tailored to meet the needs of M2M communications, enabling long-term battery life. The 5G M2M communication supported broad protection, data collection, and data processing and made the communication protocol easy to deploy and maintain. However, resource scheduling of such M2M devices in 5G networks plays a major role in mission-critical applications [15].

Efficient allocation of resources stands out as a pivotal element within the radio resource management system of ZT-DI-M2M, essential for guaranteeing QoS at the gNodeB (gNB) component of the 5G framework. Within the 5G infrastructure, gNBs rely on QoS metrics like energy efficiency, packet latency, and priority rankings to orchestrate packet scheduling. Inefficient scheduling of packet transmission can swiftly lead to access congestion, culminating in resource shortages, inefficiencies, or superfluous allocations [16]. As a result, the essential task in ZT-DI-M2M is to develop an effective scheduling algorithm that meets the different QoS requirements. If resource allocation is solely guided by the objective of efficiency, it may result in an inequitable distribution of resources to M2M end users. Those in more favorable conditions, such as proximity to the base station or superior channel quality, could receive a disproportionately larger share of services, potentially leading to an unfair allocation. On the other hand, if the main goal was to allocate resources fairly to users, this could result in an inefficient outcome due to a lack of exploitation of individuals' circumstances. Hence, harmonizing these two objectives is the greatest challenge; however, it is a task of significant importance and complexity [9], [12], [16], [17], [18], [19], [20], [21], [22]. Therefore, this work aims to develop a resource allocation scheme that is fair and efficient while complying with the M2M devices' requirements in ZT-DI-M2M environments. The following are the major contributions of this research work.

1) A Group-Based Energy Aware (GBEA) algorithm is proposed for effective resource scheduling for M2M communication in the 5G environment.
2) The GBEA algorithm considers the processing time of the M2M devices requesting resources and groups the M2M devices for inter and intra-resource distribution.
3) GBEA optimizes the resource allocation through the proposed Hybrid Hungarian Algorithm (HHA) by considering parameters such as factoring delay, energy consumption, proximity, and fairness.
4) Extensive analysis substantiates the GBEA algorithm's performance regarding throughput, responsiveness to delays, and energy utilization.

The subsequent sections of this paper are organized as follows: In Section II, we investigated recent developments in the field. Section III outlines the envisioned operational framework. Moving on to Section IV, we presented the details of the GBEA algorithm and the optimization approach. Section V is dedicated to showcasing simulation results and delving into the achieved performance metrics. Finally, in Section VI, we draw our conclusions.

II. RELATED WORK

Various approaches have been taken to structure the scheduling strategy for resource allocation of M2M devices. The authors in [23] utilize the block length coding analysis of definite size to resolve the errors in transmitting information and effectively maintain the network's bandwidth and capacity. To allocate the resources effectively, computationally matching algorithms are formulated. Similarly, in [17] the resources are allocated for the 5G MTC environment in uplink communication through the introduced Dynamic Priority-Based Resource Allocation (DPBRA). This scheme employs a two-staged dynamic priority system that acts as a prioritized contention mechanism, considering intra and inter-frame wait times and transmission-awaiting M2M devices. In [12], a Sleep-Scheduling Scheme is introduced based on Deep Reinforcement Learning, denoted as DSS. Also, a joint resource allocation algorithm is introduced referred to as DSS-JRAA. This approach involves activating dormant devices in sleep mode to save energy, all the while ensuring they meet the computational demands. Furthermore, an iterative algorithm is devised for optimizing the allocation of resources for computation and communication in a joint manner.

In [16], a Delay-aware Resource Allocation for Guaranteed Fairness and minimal Loss (DRAGFL) scheme is introduced, with the primary goal of enhancing QoS for real-time services. DRAGFL is designed to maintain consistently low latency for delay-sensitive traffic while simultaneously ensuring service fairness and minimizing data loss during periods of network congestion. This is achieved through a comprehensive optimization of the MAC layer scheduler, executed in three distinct phases. Initially, DRAGFL constructs a weighted matrix based on delay metrics to facilitate the allocation of Resource Blocks (RBs) to radio bearers. Subsequently, the scheme employs a greedy-based mechanism to prioritize the assignment of RBs to individual flows. Lastly, it accentuates data rates by applying a channel-aware principle that considers available resources and scheduled data, thus optimizing overall network performance. The authors in [18] employ degrees of freedom for optimal resource allocation to the MTC. Violations over the delay threshold and the QoS guaranteed technique effectively exploit the network's available bandwidth and capacity. The authors in [19] employ a mixed-integer nonlinear programming approach, tackling NP-hardness, and propose competitive algorithms integrated into a reinforcement learning framework for efficient power allocation. In the context of enhancing the determinacy of latency in wireless networks, [21] presents a stochastic game formulation. To optimize this complex problem effectively, a Random Graph-based Sparse Long Short-Term Memory (LSTM) network (RGSL) is formulated. This formulation takes into account critical factors including the control of transmission power, allocation of frequency spectrum, and the selection of base stations.

Additionally, alternative strategies have emerged. In [20], the authors introduce a Non-Orthogonal Multiple Access (NOMA) scheme within the power domain, coupled with user clustering, for Non-Binary Internet of Things (NB-IoT) systems. This approach involves formulating an optimization problem aimed at maximizing the overall network throughput by optimizing the allocation of resources for MTC devices and the organization of NOMA clusters while ensuring compliance with transmission power limits and QoS prerequisites. In [25], the authors direct their attention to the allocation of wireless resources for multicast communication within a software-defined hybrid satellite-terrestrial communication network. They introduce a Multicast solution based on Service Rate Constraints (MSRC). Their proposal centers on a hybrid communication network architecture, incorporating satellite and terrestrial components, enhanced with Software-Defined Networking (SDN) capabilities. The multicast solution is designed around the constraints related to service rates. Reference [22] presents an innovative approach to expedite uplink grant scheduling, drawing inspiration from the Multi-Armed Bandits (MABs) theory. The authors devise a customized mixed QoS metric, which incorporates factors such as data packet value, maximum acceptable access delay, and data rate. This unique metric serves as the basis for a reward function within a MAB framework, which is employed to identify the optimal Multi-Armed Bandit for scheduling at each time instance.

In [26], the authors put forward a scheme for sleep-scheduling and combined allocation of computation and communication resources. The primary objective is to reduce overall system energy consumption while adhering to task delay constraints. This scheme is designed for application in M2M-assisted and NOMA-based Mobile Edge Computing (MEC) networks tailored for 5G IoT scenarios. Subsequently, a multi-DQN (Deep Q-Network) approach to sleep scheduling is introduced, coupled with an algorithm for jointly allocating computation and communication resources. The overarching aim is to minimize the overall energy consumption of the system. In [27], the authors have presented a 5G framework. This framework is adept at reserving RBs for emergency data while formulating the resource allocation strategy as a task focused on maximizing energy efficiency. Subsequently, an iterative algorithm, leveraging the Lagrange dual method, is meticulously crafted. This algorithm is employed to explore and identify the global optimal allocation scheme aimed at resolving the problem at hand. Also, in [28], the authors introduce a cross-layer approach. This approach is designed to streamline the resource allocation process for M2M devices. It takes into account the diverse QoS requirements of the devices while concurrently working towards the minimization of their energy consumption. It can be derived that QoS metrics must be managed parallelly without prioritizing and compromising one another.

The above studies show that energy and delay play important roles in the devices' ability to attain resources. Therefore, this paper proposes a delay-sensitive and energy-aware scheduling algorithm for providing optimized resource allocation with minimized latency to the M2M devices in a 5G environment.

III. BACKGROUND AND PROBLEM FORMULATION

A. OVERVIEW OF 5G

In 5G networks, advanced and intelligent architectural designs come into play. Within these designs, Radio Access

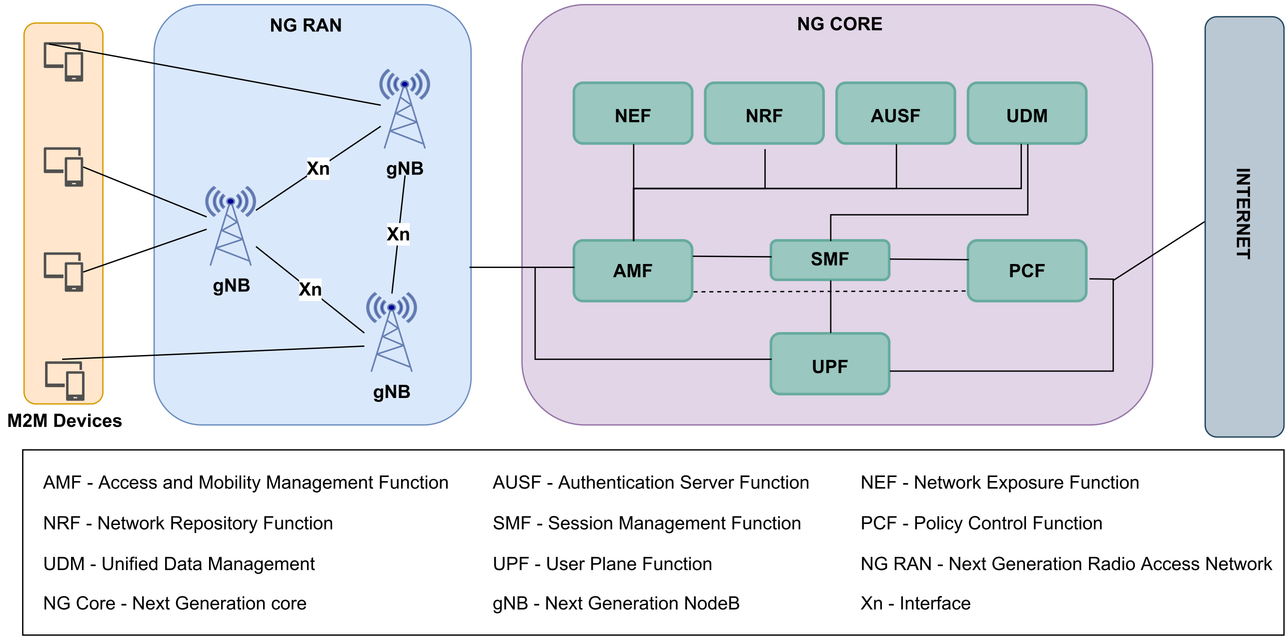

FIGURE 1. Typical 5G core architecture [24].

Networks (RANs) are liberated from the limitations of proximity to base stations or intricate infrastructure requirements. 5G paves the path for the development of dis-aggregated, adaptable, and virtual Radio Access Networks (RANs). This progress is facilitated through the introduction of fresh interfaces that generate additional data access points. Thus, 5G serves as the pivotal catalyst for ZT-DI-M2M. The 3GPP introduces a pair of fundamental architectural alternatives for 5G deployment, stemming from LTE: Non-Standalone (NSA) and Standalone (SA). SA architecture encompasses a solitary Radio Access Technology (RAT), permitting the full range of 5G enhancements tailored exclusively for the 5G New Radio (NR) SA framework. In contrast, NSA architecture raises more significant considerations, as it cannot accommodate all the 5G-specific services like mMTC, URLLC, and eMBB. Standalone 5G integrates the 5G Next Generation – Radio Access Network (NG-RAN) with the 5G Core (5GC). A multitude of new network functions is included in 5GC, each of which performs quite specific roles whereas 4G, a given network node will have numerous roles to fulfill.

The network architecture of 5G Core (5GC), illustrated in Fig. 1, is characterized by its exceptional flexibility, modularity, and scalability [24]. It encompasses a multitude of functions, including the capability for network slicing, and catering to diverse and specific customer demands. Additionally, it incorporates distributed cloud computing, Network Functions Virtualization (NFV), and Software-Defined Networking (SDN).

B. RESOURCE BLOCKS IN 5G

In 5G networks, the allocation of wireless resources occurs in both the time and frequency domains, mirroring the approach employed in LTE networks. Fig. 2 represents the constituents of a frame in the 5G networks. The time slot represents the most diminutive unit for scheduling data in the time domain, comprising a total of 14 Orthogonal Frequency Division Multiplexing (OFDM) symbols [29]. The system bandwidth is divided into numerous subchannels in the frequency domain, each of which has 12 consecutive subcarriers. Every RB corresponds to one subchannel and occupies a single time slot, establishing the RB as the fundamental unit for transmission scheduling. To efficiently harness the available network resources, 5G networks furnish M2M devices with a precisely defined quantity of resources essential for data transmission within each millisecond (ms).

As an outcome, conflict springs up in which devices of the same network compete for the necessary number of RBs. When the quantity of accessible RBs falls short of the count of active nodes seeking services, the scenario deteriorates, leading to contention-related issues. Unlike LTE, 5G NR allows different subcarrier spacings and a wide range of transmission bandwidths.

C. SCHEDULING AND CHALLENGES IN M2M

Scheduling is a mechanism in which eNB determines the entities that are to be provided with resources and the number of resources to be provided. In 5G, scheduling is performed based on subframes for every 1 ms and this is governed by the scheduler. Scheduling is the main function of the MAC layer of the protocol stack. It considers system configuration, QoS information, and channel quality information before scheduling. As the number of devices connected to the eNB increases, the M2M communication over 5G networks faces significant control signaling overhead [30]. Another notable challenge lies in orchestrating the devices to optimize

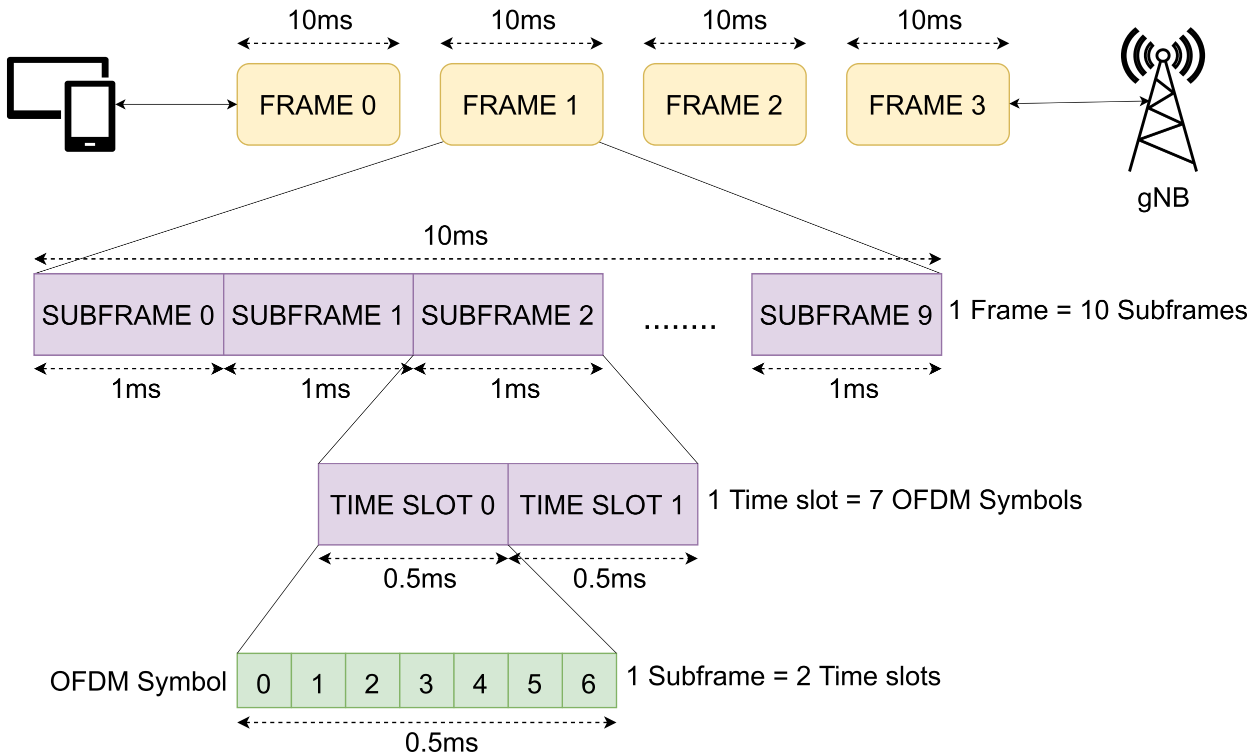

FIGURE 2. Frame structure of 5G networks [29].

overall system resource utilization, all the while taking into account the distinctive QoS prerequisites of each M2M device. Generally, M2M systems have low cost and energy consumption requirements as well as diverse QoS requirements. Certain M2M devices related to fields such as emergency notifications, road safety, and medical care, operate on very low latency necessitating applications that demand effective and interminable connectivity.

Huge control signaling traffic, inefficient resource usage, and difficulty in QoS-aware scheduling are some of the challenges that M2M communication encounters. Contention typically arises when the count of active nodes requiring scheduling surpasses the available quantity of RBs. Further, delay-sensitive applications require quicker allocation systems, and precise fairness bounds/measures are difficult to uphold. Since the 5G networks are used for both M2M and traditional communication, network congestion arises, resulting in queuing delays, packet loss, interference, and other constraints that can decrease the overall throughput. As a result, frequent network overloads occur leading to a degradation in the network's performance. These problems with service quality deterioration and unjustified resource exploitation prove the problem of resource scheduling as a challenge to the designers. Consequently, the need for a better methodology for scheduling M2M communication arises.

IV. PROPOSED WORK

If different devices in ZT-DI-M2M have different delay budgets, the traditional scheduling algorithms should be tweaked to consider the delay budgets for better coordination among the devices. For example, when device A has a lower delay budget of 10 ms as opposed to device B with a delay budget of 20 ms, it necessitates the scheduling process to prioritize device A over device B for efficient resource allocation. Thus, the usage of delay budget as a basis for grouping devices fits as a suitable parameter in clustering as it would serve the delay-sensitive M2M applications better while also eliminating the need for the devices to be physically close. However, when the number of available RBs is lesser than the devices requesting service, contention arises. To handle this problem of scheduling, the proposed GBEA algorithm optimizes the resource allocation of M2M devices by considering its delay, energy, proximity, and fairness aspects in the process of resource allocation thereby being both group-based and energy-aware. The resultant set of M2M devices will be allocated with the RBs leaving behind the sub-optimal set and this scenario of RB shortage is considered for the scheduling process.

To overcome the existing drawbacks and optimize the resource allocation of M2M devices the following methodologies are employed in the proposed GBEA scheme:

1) Initially, clustering is performed using the Gaussian Mixture Model Expectation Maximization (GMM-EM) algorithm to divide the M2M devices requesting services into "virtual groups".
2) Following the aggregation of M2M devices into clusters, the per-cluster RB distribution is achieved using a game theoretic approach inspired by the Shapley value theory.

3) The intra-cluster RB allocation is then carried out using a combination of probability-based selection and HHA algorithm where the devices are chosen based on their cost functions and the RBs are assigned to those chosen individual M2M devices respectively.

V. GBEA ALGORITHM

The scheduling algorithm assumes a pivotal role in dictating the distribution of RBs to the MTC devices within each Transmission Time Interval (TTI), which spans a duration of 1 ms. In MTC, there are an enormous amount of devices with data to be transmitted over a network. MTC devices by virtue are characterized to send more uplink data. This results in multiple MTC devices requesting access to transmit data over the same network simultaneously, thereby inundating the eNB scheduler which leads to network congestion. To address this issue and mitigate network congestion caused by numerous MTC devices simultaneously requesting access to transmit data over the same network, the GBEA resource allocation algorithm is introduced.

In [28], it becomes evident that a fundamental memetic algorithm demonstrates efficiency primarily when the count of active devices remains below 100. Moreover, the general usage of the K-means algorithm for M2M communications demands the devices be physically close to each other. To thwart this, the devices are grouped according to their delay budget thereby forming virtual clusters. The MTC devices are grouped based on their delay tolerance constraints and are not necessarily physically close to each other because the clusters that are formed using the delay constraint tend to be a virtual collection i.e., the MTC devices within a group may be physically sparse. This approach reflects the divide-and-conquer strategy while also respecting the device's delay sensitivity. After clustering, the per-cluster RB distribution is executed using a game theoretic approach inspired by the Shapley value theory. This is used to quantify the number of RBs to be allocated to each cluster.

Finally, the intra-cluster RB distribution is executed in two stages. Stage one comprises a probability-based selection followed by stage two where the HHA algorithm for the RB allocation is performed. Stage one utilizes a cost function and selects the set of candidate nodes from the pool of nodes present in the cluster while stage two avails those nodes and allocates the RBs to them to minimize the delay constraint violation. The architecture of the proposed GBEA approach is given in Fig. 3.

A. CLUSTERING OF ACTIVE M2M DEVICES

In GBEA, the aggregation of active M2M devices is performed using the GMM-EM algorithm [31]. The user's delay budget contributes to the key attribute in the process of device aggregation, especially when the number of M2M devices is high as it eliminates the need for the MTC devices to be within proximity. A Gaussian Mixture Model (GMM) employs a parametric probability density function, which is expressed as a weighted combination of Gaussian component densities. In GMM, the assumption is that all data points are generated from a mixture of a finite number of Gaussian distributions, each characterized by specific parameters such as the mean (μ) defining the peak of the distribution, the covariance (Σ) denoting the width of the distribution and a constant (π) labeling the fraction of data points belonging to a class k.

The Gaussian Mixture can be described as a function consisting of multiple Gaussians, with each Gaussian denoted by $k \in \{1, 2, \ldots, K\}$, where K represents the total number of clusters under consideration. Dataset X is defined as the set of delay budgets of all the active M2M devices. The probability density function of Gaussian distribution [18] for the datapoint n in dataset X is defined as follows in equation (1),

$$G\left(X_n|\mu_k, \Sigma_k\right) = \frac{1}{\Sigma\sqrt{2\pi}} e^{\frac{-(X-\mu)^2}{2\Sigma^2}} \tag{1}$$

The standard and powerful tool used in GMM called the Expectation Maximization (EM) algorithm works by fitting the K Gaussian components to the dataset X by parameterizing the μ, Σ, and π of each cluster, where K is the total number of clusters. If there are K clusters in the dataset, GMM-EM finds and fits these K clusters to the dataset X by optimizing its parameters (μ_k, Σ_k, and π_k). The probability of observing data is given by the below formula.

$$P\left(X|\pi, \mu, \Sigma\right) = \prod_{n=1}^{N} \left[\sum_{k=1}^{K} \pi_k G\left(X_n|\mu_k, \Sigma_k\right)\right] \tag{2}$$

The aim is to maximize equation (2) thereby maximizing the probability of observing and fitting the data point to its corresponding cluster. This probability function P aids in the determination of the best fit for any datapoint X to cluster K. The EM process follows an iterative methodology that comprises two stages: the Expectation (E step) and the Maximization (M step). Commencing with an initial, randomly chosen, and sensible parameter value, the EM procedure unfolds. In the E step, the latent variable $\gamma(Z_{nk})$ is computed, which denotes the probability of seeing the observation X_n in class K where Z_{nk} is the indicator denoting if X_n belongs to a class K or not. Also, N_k is calculated, which is the sum of $\gamma(Z_{nk})$ over all the data points. Equations (3), (4), and (5) denote the above-discussed variables respectively.

$$Z_{nk} = \{1, if\ X_n\ in\ class\ K; 0, if\ not\} \tag{3}$$

$$\gamma\left(Z_{nk}\right) = P\left(Z_{nk} = 1|X_n\right) = \frac{\pi_k G\left(X_n|\mu_k, \Sigma_k\right)}{\sum_{j=1}^{K} \pi_j G\left(X_n|\mu_j, \Sigma_j\right)} \tag{4}$$

$$N_k = \sum_{n=1}^{N} \gamma\left(Z_{nk}\right) \tag{5}$$

During the M step, the three parameters are updated based on the previously calculated variables as given by the equation (6), (7), and (8).

$$\mu_k = \frac{1}{N_k\left[\sum_{n=1}^{N} \gamma\left(Z_{nk}\right) X_n\right]} \tag{6}$$

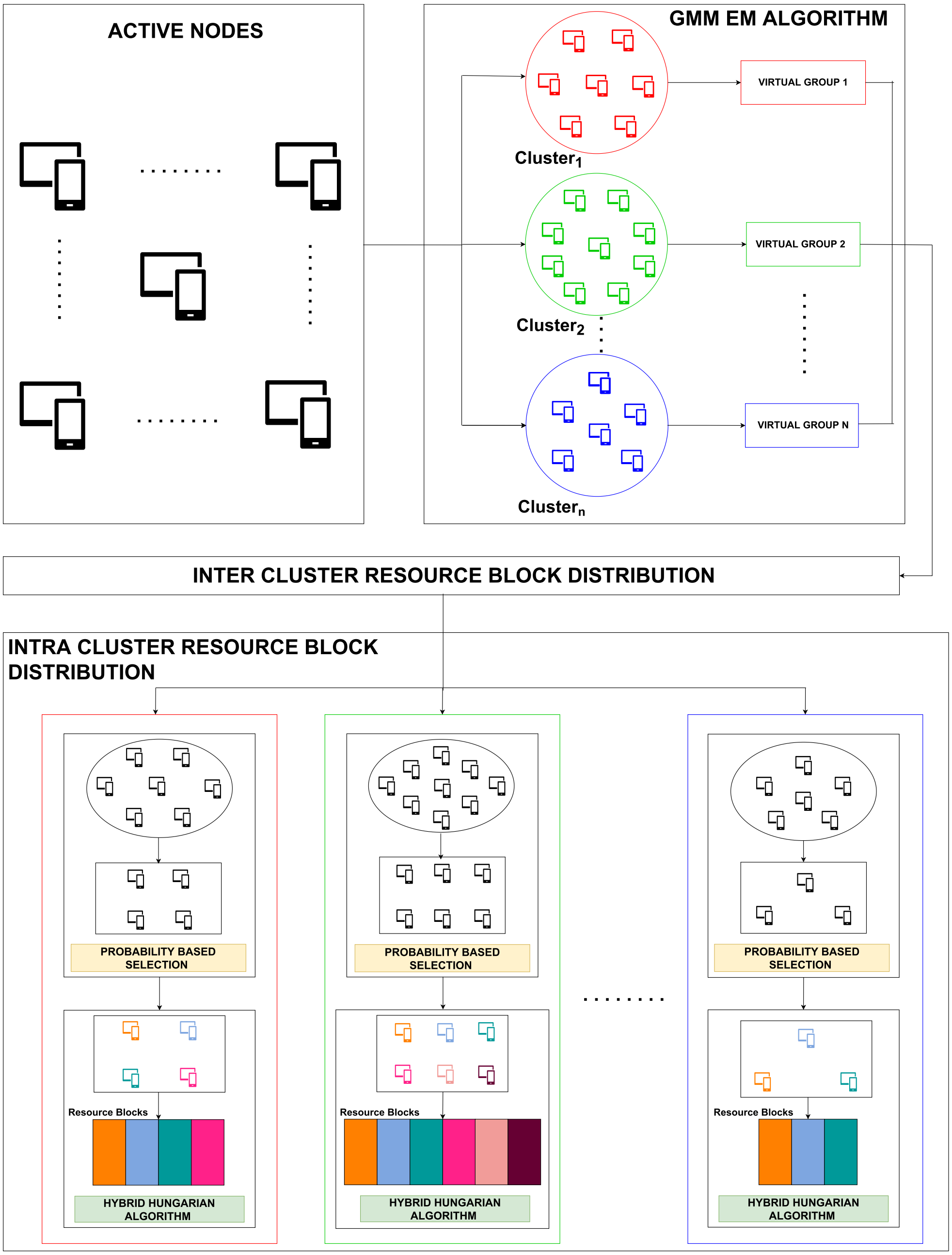

FIGURE 3. Group-Based and Energy Aware (GBEA) architecture.

$$\Sigma_k = \frac{1}{N_k\left[\sum_{n=1}^{N} \gamma\left(Z_{nk}\right)\left(X_n - \mu_k\right)\left(X_n - \mu_k\right)^T\right]} \tag{7}$$

$$\pi_k = \frac{N_k}{N} \tag{8}$$

As a result, the EM algorithm repeats the E and M steps until it identifies the optimal parameter values, facilitating the convergence of these parameters. This convergence aids in the optimal fitting of M2M devices into their respective clusters.

B. INTER CLUSTER RESOURCE BLOCK DISTRIBUTION

This section presents the inter-cluster RB distribution using a game theoretic approach. After the active M2M devices have been clustered, the distribution of the number of RBs for each cluster is quantified. The resource distribution game is a solution concept in cooperative game theory. This allows the cluster classes to play a resource distribution game among themselves at the end of which RBs are distributed among the cluster classes. The inter-cluster RB distribution game is characterized by the utilization of a pair (N, ν), where N represents the count of players, and ν serves as the characteristic function signifying the coalition value among these players. This game is a competition among the virtual clusters to distribute the RBs. Within a set of players, a coalitional game establishes the fitness or utility associated with each player participating in the game. The game theory's outcome is a set of RBs granted to each virtual group. In the context of the coalitional game (N, ν), the allocation of payoffs to the players is determined by the equation (9):

$$\phi_i = \frac{1}{N!} \sum_{S \subseteq Ni}^{\infty} |S|! \, (|N| - |S| - 1)! \, [\nu (S \cup \{i\}) - \nu (S)] \quad (9)$$

where for each player i, the marginal contribution is equal to the term $\nu (S \cup \{i\}) - \nu (S)$. Here S is the coalition of the player cluster set where players who cooperate in making a decision are referred to as coalitions and $S \subset C$, where C is the set of all players. This value is influenced by the number of ways $|S|!$ that the set S could have been constructed before the inclusion of i, as well as the $(|N| - |S| - 1)!$ ways in which the remaining players could have been added. The characteristic function for each coalition is calculated by summing up the respective cluster sizes. Ultimately, the sum for all the feasible sets S is computed, and the average is determined by dividing it by $N!$ representing the total number of possible orderings of all players. Thus, ϕ_i represents the number of RBs distributed to each cluster i and the resultant is a vector $\phi = \{\phi_1, \phi_2, \phi_3, \ldots.., \phi_k\}$, where k is the number of players.

Algorithm 1 Per Cluster Resource Block distribution

Input: Player Cluster Set: $C = \{C_1, C_2, \ldots, C_K\}$, Cluster Sizes: $N = \{N_1, N_2, \ldots, N_K\}$
Output: RBs allocated to each cluster $\phi = \{\phi_1, \phi_2, \ldots, \phi_K\}$

```
for i=1,2,3,...,K do
    φ_i = 0
end for
Let S be the power set of C
for S_i ∈ S do
    v(S_i) = 0
end for
for each S_i ∈ S do
    for each ele ∈ S_i do
        v(S_i) = v(S_i) + N_ele
    end for
end for
for each C_i in C do
    for each S ⊂ C do
        if C_i ∈ S then
            φ_i = φ_i + [(v(S) − v(S − {i})( |S|!(|C|−|S|−1)! / |C|! )]
        end if
    end for
end for
return φ = {φ_1, φ_2, ..., φ_K}
```

Thus, the clustering algorithm and inter-cluster resource block allocation mechanism in the proposed system contribute to achieving a global solution by leveraging game theoretic principles and cooperative strategies among clusters. The clustering algorithm partitions the active M2M devices into distinct clusters based on relevant criteria such as delay requirements, energy consumption, and proximity to base stations. This clustering process facilitates the formation of coalitions or virtual clusters, each representing a group of M2M devices with common resource allocation needs. Similarly, the inter-cluster resource block allocation involves a cooperative game among the virtual clusters to distribute resource blocks optimally. The characteristic function (ν) quantifies the value of each coalition (virtual cluster) based on the number of resource blocks allocated to them. Through strategic decision-making and negotiation, clusters aim to maximize their utility or payoff while considering the overall performance of the network.

C. INTRA CLUSTER RESOURCE BLOCK DISTRIBUTION

In this section, to allocate RBs to individual M2M devices, the intra-cluster RB distribution phase has been divided into two subsequent stages. The first stage comprises a probability-based selection and the second stage performs the HHA algorithm. The virtual clusters, after being allocated with the RBs by the inter-cluster RB distribution are fed into the first stage of the intra-cluster RB distribution process. After stage 1, only the selected candidate nodes are passed onto stage two where the HHA algorithm is performed. The undesirable 'Node-RB' pair assignments of stage 2 are allocated to the unselected nodes of stage 1 to avoid RB wastage.

1) STAGE 1: PROBABILITY BASED SELECTION

The probability-based selection aids in the elimination of unfit nodes thereby selecting the candidate nodes for the next stage. In probability-based selection, the cost function is computed for every node of the respective clusters. A cost function determines the competency of the machine node. It plays a crucial role in the node's selection probability (i.e.) the probability that a machine node will be selected for stage 2 is based on its cost function. The proposed cost function's parameters include the delay tolerance value (D_m) of the m^{th} M2M device and the data periodicity (DP_m) factor which is the data transmission interval of the m^{th} device.

The distance between the eNB and the m^{th} M2M device called the Proximity (P_m), is also factored into the calculation of the cost function. The cost function is further dependent on the fairness for each m^{th} M2M device (F_m), denoted by equation (10), and is calculated by dividing the total number of RBs allocated so far to the m^{th} device (R_m) by the total number of RBs allocated to the cluster (R_t) to which it belongs.

$$F_m = \frac{R_m}{R_t} \quad (10)$$

This equation quantifies the fairness for each m^{th} M2M device by taking the ratio of the total number of RBs allocated to that device (R_m) to the total number of RBs allocated to the entire cluster (R_t). This ratio reflects the proportion of resources assigned to a specific device relative to the overall allocation in its cluster. Consequently, higher F_m values signify that a machine node has received a more substantial share of the total resources within its cluster, indicating a fairer distribution compared to other devices in the same cluster. The cost function takes the energy consumption factor

(E_m) into account as well where the energy consumed for each m^{th} device is calculated by,

$$E_m = 1 - \frac{E_i}{E_t} \tag{11}$$

where, E_t denotes the total energy capacity of the m^{th} device and E_i denotes the energy consumed so far by the m^{th} device. Another factor called the scope factor of the m^{th} M2M device (S_m) is obtained as the product of the m^{th} device's proximity and energy factor. The scope factor is calculated as demonstrated by,

$$S_m = P_m \, * \, E_m \tag{12}$$

The scope factor (S_m) in Equation (12), defined as the product of the m^{th} M2M device's proximity (P_m) and energy factor (E_m), is indeed a term coined within the context of the given system. The scope factor is a composite metric that combines spatial proximity and energy considerations for a specific M2M device. In the decision-making process, devices with a higher scope factor are given precedence over their counterparts with lower scope factors. The rationale behind this lies in the fact that a higher scope factor indicates a more favorable combination of remaining energy and proximity, making the device a more suitable candidate for selection when compared to other competitive devices. The proposed cost function depends on factors such as data periodicity, delay tolerance, proximity, and scope, which are the product of energy and delay tolerance factors. Thus, the cost function (C_m) for each M2M device can be computed as follows:

$$C_m = \frac{DP_m}{D_m * S_m * F_m} \tag{13}$$

Equation (13) helps in finding the cost of each M2M device. The formulation of the cost function in Equation (13) is designed to capture and balance multiple factors that are crucial in assessing the competency of M2M devices for the subsequent stage of the GBEA system. The choice of this specific form arises from a consideration of the key parameters influencing the GBEA system's performance. By incorporating data periodicity (DP_m), delay tolerance (D_m), scope factor (S_m), and fairness (F_m), the cost function attempts to evaluate the suitability of each M2M device comprehensively. The multiplication of these factors in the denominator of the equation reflects the interdependence and combined influence of delay tolerance, spatial scope, and fairness on the overall competency assessment. This formulation allows for a weighted consideration of factors, ensuring that devices with lower delay tolerance, limited spatial scope, and unfair resource allocations receive higher cost function values, signifying lower competency.

Following this, a constant called C_{max} is calculated, which is the sum of all the cost functions of a particular cluster. Each device's cost function is then normalized by dividing it by C_{max} to get the net cost function value which will be used for the probability-based selection. A random number between (0,1) is generated and each node is selected using range matching.

The probability of a particular node getting selected is proportional to its cost function value. The number of candidate nodes selected from a cluster is equal to the number of RBs allocated to cluster K (ϕ_K). This is to avoid the situation of unbalanced assignment in the HHA algorithm that is performed in the next stage.

Algorithm 2 Probability Based Selection

Input: Cluster Nodes: $I = \{I_1, I_2, \ldots, I_N\}$
Output: Selected Optimal Nodes: $O = \{O_1, O_2, \ldots, O_{\phi_K}\}$
for $i = 1, 2, 3, \ldots, N$ **do**
    $Costfunction_i = \frac{DP_i}{D_i * S_i * F_i}$
end for
Let $C_{max} = 0$
for $i = 1, 2, 3, \ldots, N$ **do**
    $C_{max} = C_{max} + Costfunction_i$
end for
for $i = 1, 2, 3, \ldots, N$ **do**
    $Costfunction_i = \frac{Costfunction_i}{C_{max}} * 100$
end for
for $i = 1, 2, 3, \ldots, \phi$ **do**
    $Prob \leftarrow rand(0, 1)$
    for $k = 1, 2, 3 \ldots N$ **do**
        $j = k - 1$
        if $Costfunction_j < Prob \leq Costfunction_k$ **then**
            $O_i = k$
        end if
    end for
end for
return $O = \{O_1, O_2 \ldots O_{\phi_K}\}$

In Algorithm 2, the cost function ($Costfunction_i$) is calculated for each cluster node (I_i) using the formula $Costfunction_i = \frac{DP_i}{D_i * S_i * F_i}$ where DP_i is the delay penalty, D_i is the delay, S_i is the scaling factor, and F_i is a factor related to the cluster's performance. The cost function is inversely proportional to the delay, implying that lower delays result in higher values of the cost function. The delay tolerance or budget (D_m) is not explicitly used in the cost function calculation, but the inverse relationship implies that nodes with delays closer to or within the delay budget will have higher cost function values.

The Probability based selection algorithm also uses a random probability ($Prob$) in the range [0, 1] for probabilistic node selection. However, the cost function values are normalized. The purpose of this normalization is to create a relative scale for comparison such that a node with a higher cost function value is more likely to be selected. Thus, the term probability implies the likelihood of selection based on the comparison of cost function values ($Costfunction_j < Prob \leq Costfunction_k$).

While the delay budget (D_m) is not directly integrated into the cost function calculation, the inverse relationship suggests that nodes with delays within the budget will have higher cost function values. This, in turn, influences the probabilistic selection of optimal nodes, indirectly incorporating the delay budget constraint into the algorithm. The probabilistic nature of the selection allows for flexibility within the specified delay budget range [5 ms - 300 ms].

2) STAGE 2: HYBRID HUNGARIAN ALGORITHM

The HHA algorithm is a combinatorial optimization technique that efficiently addresses the assignment problem within a polynomial time frame. This algorithm eliminates the need for the brute force method which evaluates all the $n!$ ways of assignment in an attempt to solve the problem with a minimum delay constraint violation. However, it might end up with two cases namely the balanced assignment and the unbalanced assignment. In a balanced assignment, the number of candidates is the same as the number of resources. While in the unbalanced assignment, they are not equal. The HHA algorithm uses a matrix representation and for a balanced assignment, a square matrix is obtained.

In the case of an unbalanced assignment, it is converted to a square matrix (balanced assignment) by adding dummy rows or columns. However, in stage 1, the number of candidate nodes is chosen to be equal to the number of RBs allocated to that particular cluster, and the balanced assignment is calculated. In the proposed HHA algorithm, the Cost Matrix [CM] is calculated first where each row depicts a time interval of 1 ms and each column represents an individual M2M device. The value '1' is assigned to a cell if it satisfies the delay constraint of the node else, the cell is assigned a value of '100' indicating that it has violated the delay constraint of the node.

The 'Node-RB' assignments are represented as an Assignment Matrix [AM] and since we try to minimize the assignment cost, any node that has been allocated with the value '100' is considered an undesirable assignment. These less desirable assignments are not discarded; instead, they are allocated to the nodes that were not selected in stage 1, taking into account their delay constraints. The HHA algorithm operates in two distinct parts: Part 1 involves row and column reduction, while Part 2 focuses on optimization. In Part 1, the algorithm conducts row and column reduction by subtracting the minimum value of each row and column from all the entries within that respective row and column.

Consecutively in part 2, the aim is to try and figure out if all the nodes have been allocated with RBs. This is done by row/column scanning followed by row/column deletion. Once unique 'Node-RB' pairs for all the nodes are procured, an optimized solution is reached. Finally, the undesirable assignments (if any) are mapped to an unselected node of stage 1 to its delay constraint. In summary, each M2M device that was fed into stage two is allocated with an RB to minimize the delay constraint violation and if any undesirable assignment is to arise, the RB wastage is also counteracted.

Algorithm 3 Hybrid Hungarian Algorithm

Input: Selected Nodes: $I = \{I_1, I_2, \ldots, I_\phi\}$
Number of RBs Allocated: ϕ
Output: Optimal Set of M2M Nodes

Let $Assignment = \{\emptyset\}$
for $i = 1, 2, 3, \ldots, \phi$ **do**
  for $j = 1, 2, 3, \ldots, \phi$ **do**
    if $Delay_j \geq i$ **then**
      $Costmatrix_{ij} = 1$
    else
      $Costmatrix_{ij} = 100$
    end if
  end for
end for
$Assignment \leftarrow HybridHungarian(Costmatrix)$
for $i = 1, 2, 3, \ldots, \phi$ **do**
  if $Assignment_i = 100$ **then**
    $Assignment_i \leftarrow Unselected\ Node \ni delay\ budget \geq i$
  end if
end for
return $Assignment$

TABLE 1. Simulation parameters.

Parameter	Value
Bandwidth	1.4MHz
Number of Antennas	1
Type of Antenna	Omni Antenna
Propagation Model	Two Ray Ground
Propagation Loss Model	Free Space
No of Base Stations	1
Node Mobility	Stationary
Area	800 m x 600 m
Maximum number of Nodes	250
Node Placement	Random
Delay Budget	5 ms - 300 ms

VI. SIMULATION RESULTS AND ANALYSIS

The performance of the proposed GBEA algorithm is simulated in Network Simulator 2 (NS2) and the simulation parameters are summarized in Table 1. We evaluated the GBEA's performance and compared it with Round Robin (RR), Proportional Fair (PF), Memetic, and Memetic + MaxSpace (MS) algorithms [28] considering an uplink scenario. Also, GBEA has been compared to DRAGFL [16], DPBRA [17], DSS-JRAA [12], and RGSL [21] approaches.

In the simulation setup, there is a single gNB serving multiple randomly positioned M2M devices within its coverage area. The simulation focuses on a contention scenario where there are more M2M devices requesting RBs than the available RBs. Thus, the GBEA performs clustering, and inter and intra-RB distribution to select the optimal set of M2M devices. The algorithm's performance is assessed across various metrics as the number of active devices is systematically increased.

A. PERFORMANCE ANALYSIS OF GBEA WITH POPULAR AND CONVENTIONAL ALGORITHMS

For comparing the performance of GBEA, we consider six different metrics such as the percentage of packet loss, mean throughput, percentage of satisfied nodes, percentage of unsatisfied nodes, fairness, and energy consumption. The PF, RR, Memetic, and Memetic + MS [28] and GBEA algorithms are judged based on the above-mentioned metrics.

Figure 4 shows the percentage of packet loss among the different algorithms compared where the packet loss rate is the measure of packets that have been lost in the way, in essence, the packets that have not reached the destination. The packet loss ratio is calculated as the ratio of packets lost to the total packets generated. In the figure, it is observed that the GBEA algorithm's packet loss is low until 200 devices and slowly rises after. Quantitatively, the percentage reduction in packet loss by GBEA over the PF algorithm is more than 5% at the 250 devices' mark. It indicates that the GBEA algorithm can handle packet losses better than the PF algorithm when the number of devices gets higher.

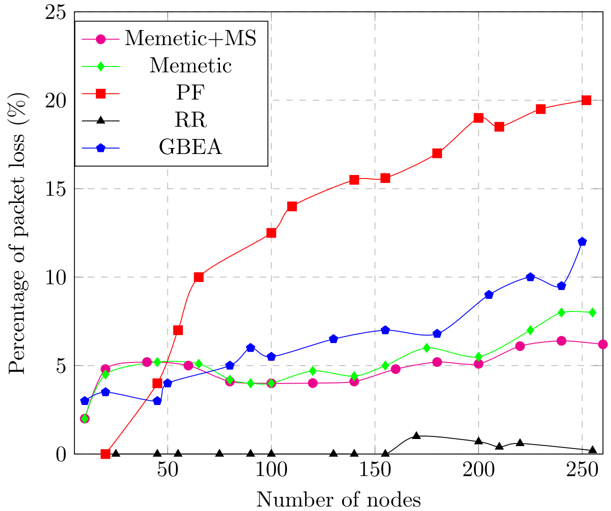

FIGURE 4. **Percentage of packet loss analysis of GBEA with popular and conventional algorithms.**

Fig. 5 depicts the mean overall energy consumption for the considered scheduling algorithms. Energy consumption is one of the crucial factors for M2M communication as it directly influences their lifetimes. It can be observed from the figure that with the increase in the total number of devices, the overall energy consumption across all the scheduling algorithms increases. This is because as the number of devices increases, the eNB is inundated with requests. However, a nearly 6.5-fold decrease in the energy consumption by the GBEA algorithm over the RR and PF algorithms at the 250 devices' mark can be observed, advocating the GBEA algorithm to be the most energy-efficient among the compared algorithms.

Fig. 6 shows the fairness index variations across different scheduling algorithms in terms of the number of M2M devices. Fairness is the measure of how fairly the devices have been bestowed with their turn at data transmission by the equitable allocation of RBs to the devices. The PF algorithm despite its weakness in energy consumption and mean throughput, outwitted the existing algorithms in terms of fairness with its value ranging between 0.4 and 0.8. Furthermore, the GBEA algorithm due to the intra-cluster RB distribution phase performed better with values approximately in the range of 0.4 to 0.6 demonstrating an edge over the two memetic algorithms and the RR.

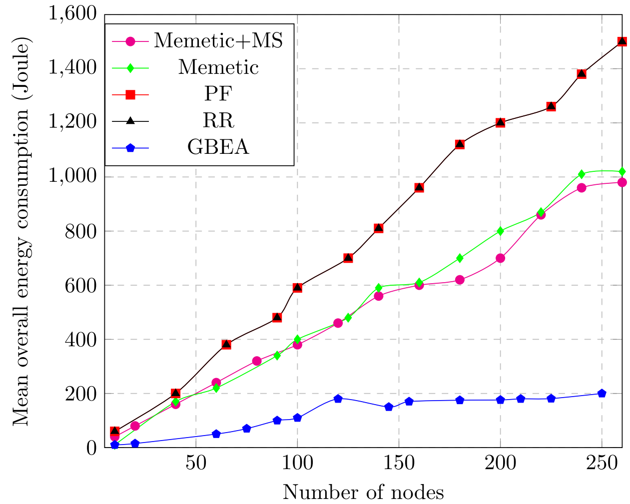

FIGURE 5. **Energy consumptions analysis of GBEA with popular and conventional algorithms.**

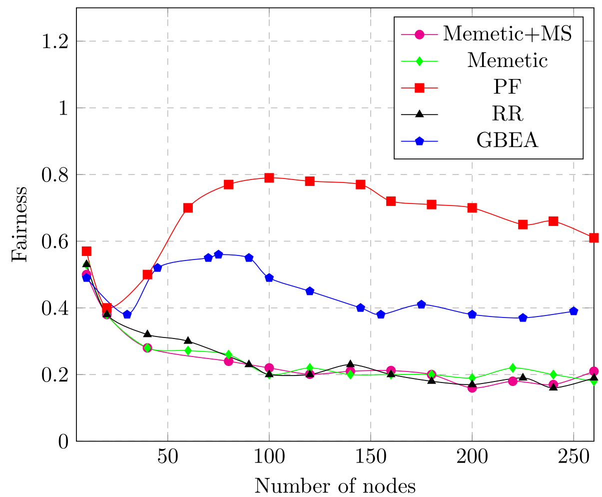

FIGURE 6. **Fairness analysis of GBEA with popular and conventional algorithms.**

Fig. 7 depicts the percentage of unsatisfied nodes to delays, which is the measure of the number of nodes whose delay limit has been violated. This metric reflects each algorithm's trait in considering the delay deadlines of the M2M devices. The PF algorithm performed with the value almost reaching 20% followed by the RR algorithm which ranged between 0 and 5%. Whereas, the GBEA algorithm exhibited a progressive performance with its value almost remaining near zero constantly.

Fig. 8 compares the mean throughput variation where the throughput is the amount of data transmitted from a source to a destination in a specified amount of time. As the number of devices increases, it can be observed that the throughput decreases across all the algorithms yet the GBEA algorithm shows a significant throughput improvement when compared with the existing algorithms with its value reaching a maximum of 25kbps even when the number of devices is

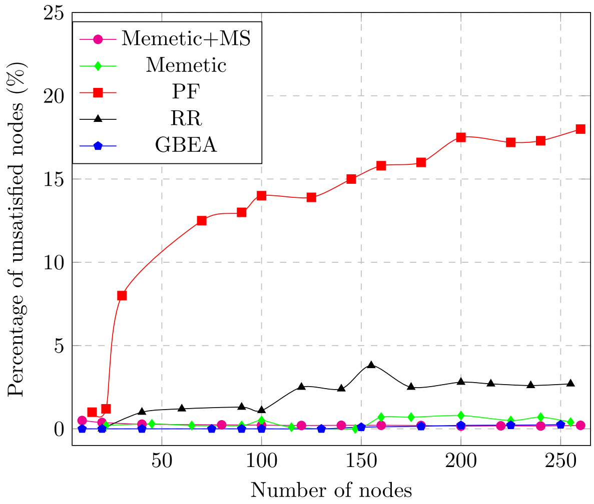

FIGURE 7. **Percentage of unsatisfied nodes analysis of GBEA with popular and conventional algorithms.**

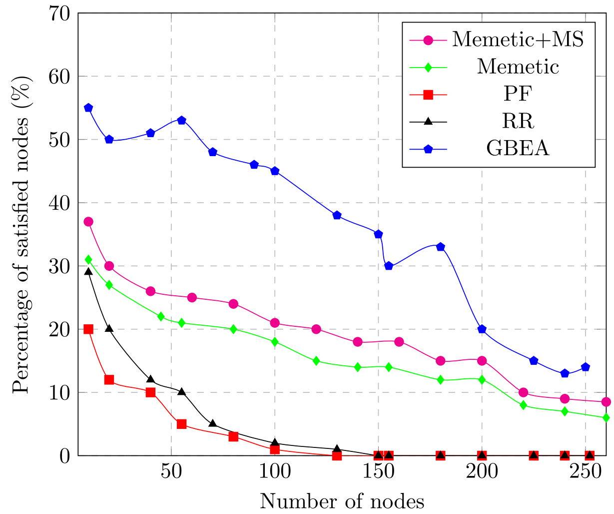

FIGURE 9. **Percentage of satisfied nodes analysis of GBEA with popular and conventional algorithms.**

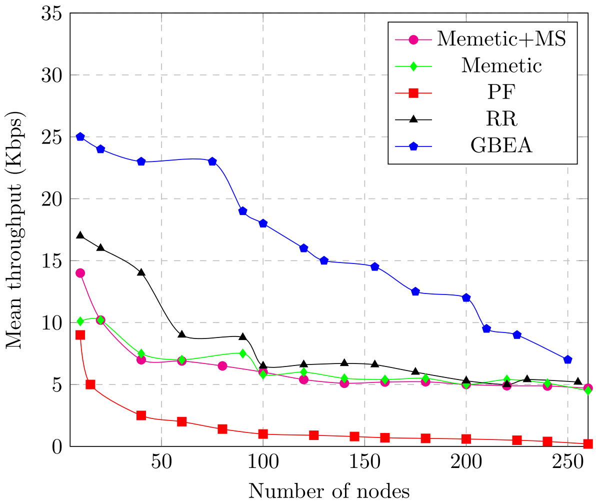

FIGURE 8. **Throughput analysis of GBEA with popular and conventional algorithms.**

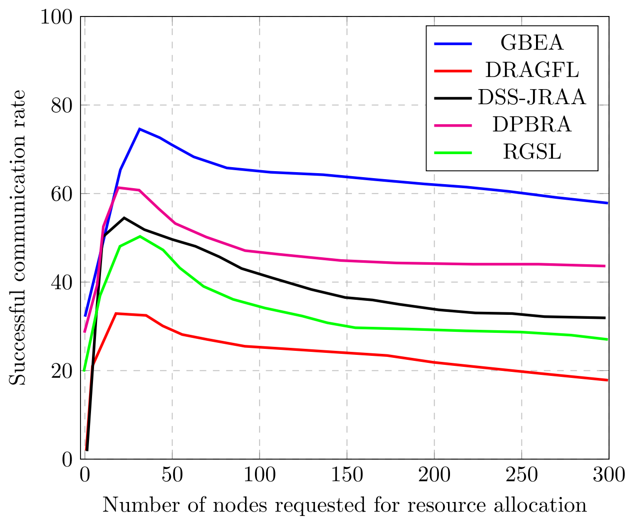

FIGURE 10. **Successful communication analysis of GBEA with recent algorithms in the literature.**

low. The graph proves GBEA to be more efficient than the compared algorithms.

The performance in terms of the percentage of satisfied nodes across all the scheduling algorithms is illustrated in Fig. 8 and the metric evaluates the fraction of M2M nodes that were served with the RBs. As it can be perceived from the figure, even in dense situations, the performance of the GBEA was acceptable with its value approximately ranging from 10% to 55%, outplaying the other alternatives. It can also be seen that the GBEA algorithm performed a cut above the others even when the number of M2M devices was large such as when |M2M| = 100.

B. PERFORMANCE ANALYSIS OF GBEA WITH RECENT ALGORITHMS IN THE LITERATURE

To evaluate GBEA against recent algorithms, we assess its performance using metrics including the number of successful communications as the number of requested MTC devices increases, resource partition percentage over time (ms), energy efficiency relative to the number of MTC devices, fairness index among MTC devices, and throughput concerning the number of MTC devices. This analysis is conducted for comparison with DRAGFL, DPBRA algorithm, DSS-JRAA algorithm, and RGSL algorithm.

Fig. 10 analyses the number of successful communications of the nodes in contrast with the number of resource allocation requests that have been received. As the number of requests increases, it can be observed that the number of successful communications decreases gradually but the GBEA algorithm shows the best performance compared to the DPBRA algorithm followed by DSS-JRAA, RGSL, and DRAGFL with GBEA's success rate (nodes succeeding communication) being the highest at 50 or above and DRAGFL below 20.

Fig. 11 analyses the resource partition rate to time. Resource partition rate can be defined as the efficiency of

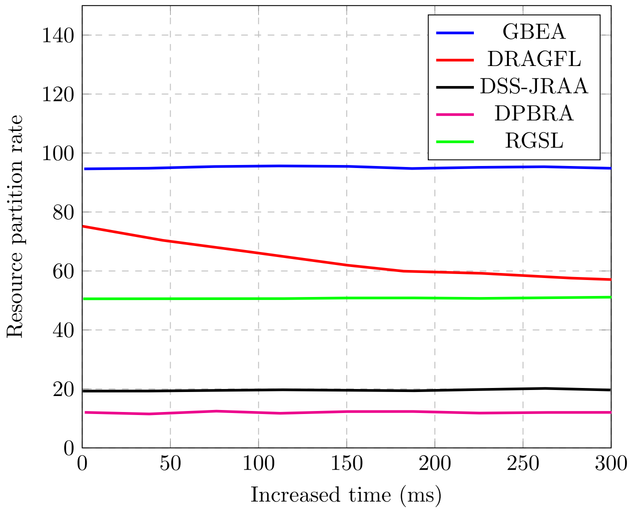

FIGURE 11. Resource partition rate analysis of GBEA with recent algorithms in the literature.

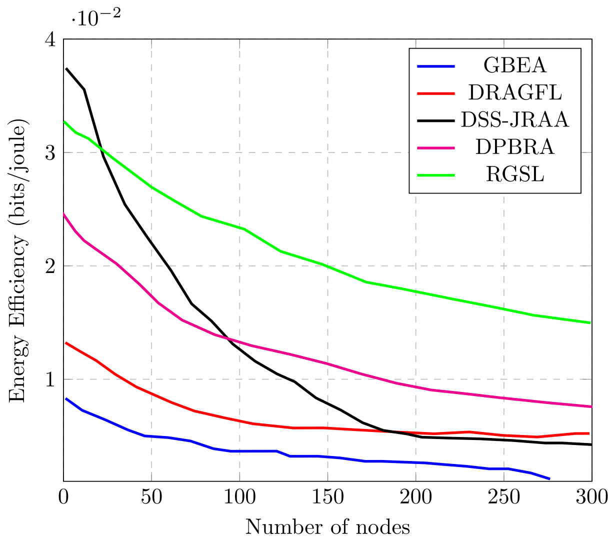

FIGURE 12. Energy Efficiency analysis of GBEA with recent algorithms in the literature.

an algorithm to allocate a sustainable number of resources for the requested resource allocations fairly over time. As the time increases it can be observed that DSS-JRAA, RGSL, and DPBRA algorithms maintain a stable rate except for DRAGFL which has a significant rate decline with a rate below 60 ms. The graph proves GBEA to outperform the existing algorithms with the highest resource partition rate being 90 ms and above throughout.

Fig. 12 analyses the energy consumption concerning the number of nodes. It can be observed that there is an overall gradual decline in the energy consumption for each algorithm. As the number of nodes increases, it can be observed that RGSL has the highest energy consumption followed by DSS-JRAA, DPBRA, and DRAGFL. It is noticeable that GBEA outperforms the existing algorithms with the lowest energy consumption.

Fig. 13 analyses the energy consumption concerning the number of nodes. It can be observed that there is an

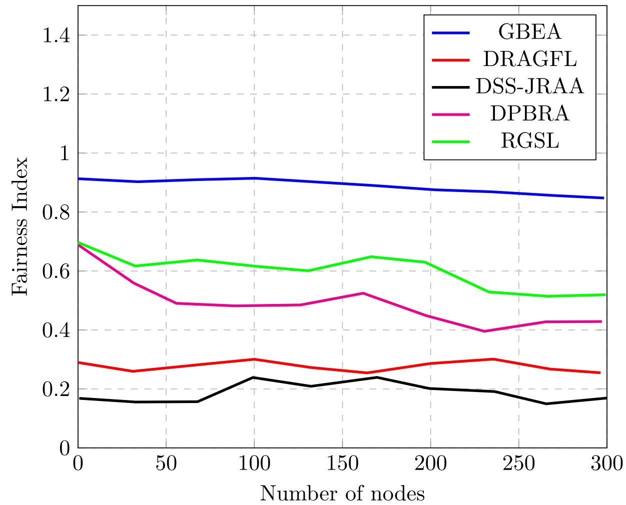

FIGURE 13. Fairness index analysis of GBEA with recent algorithms in the literature.

overall gradual decline in the energy consumption for each algorithm. As the number of nodes increases the RGSL has the highest energy consumption followed by DSS-JRAA, DPBRA, and DRAGFL. Thus, it is envisioned that GBEA outperforms the existing algorithms with the lowest energy consumption.

Fig. 14 analyses the overall throughput and the service portability of the algorithms. Service portability refers to the ability of a node to retain existing communications without impairment of quality, reliability, or convenience. It can be observed that DSS-JRAA and DPBRA have the lowest rate correspondingly. GBEA outperforms the existing algorithms for both throughput and service portability at a rate of 95% and above followed by DRAGFL and RGSL.

Overall, the comparative analysis of the GBEA algorithm with the traditional and recent resource scheduling algorithms proves the improved performance of GBEA, resulting in an optimal resource scheduling algorithm for M2M devices in 5G-supported ZT-DI-M2M environments.

C. ANALYSIS OF CONVERGENCE AND OPTIMALITY OF THE PROPOSED ALGORITHMS

1) SHAPLEY VALUE ALGORITHM

The Shapley value algorithm converges when it reaches a stable allocation of resource blocks (RBs) among cluster nodes. The stability is achieved when no cluster has an incentive to change its RB allocation, meaning the Shapley values have been computed in a way that satisfies fairness and cooperation among clusters.

The algorithm follows the Shapley value concept, which considers all possible permutations of clusters and calculates the marginal contribution of each cluster to the overall coalition. Convergence is reached when these marginal contributions settle into a stable distribution.

The optimality of the Shapley value algorithm is based on its properties, including fairness and efficiency. The Shapley value provides a unique and fair distribution of RBs

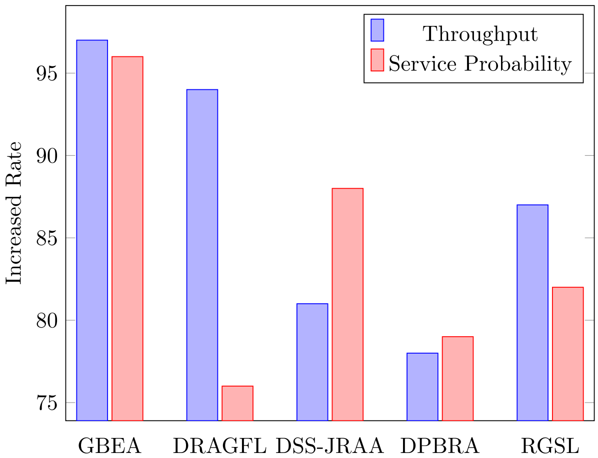

FIGURE 14. Throughput and service probability of GBEA and existing algorithms.

among clusters, ensuring that each cluster receives a share proportional to its contribution to the overall system. This allocation is considered optimal in terms of fairness.

2) PROBABILITY-BASED SELECTION

The probability-based selection algorithm converges when a specific number of nodes from each cluster are chosen for the next stage of the Hungarian algorithm. The convergence condition is set based on the initial RB allocation determined by the Shapley value. Once the required number of nodes is selected, the algorithm proceeds to the next stage.

Optimality in this context is linked to the cost function used for node selection. The cost function considers factors such as delay, size, and frequency, aiming to optimize the selection of nodes that collectively contribute to the overall system's efficiency. The normalization step ensures that the probabilities sum to 1, providing a valid probability distribution.

3) HYBRID HUNGARIAN ALGORITHM

The convergence of the Hybrid Hungarian Algorithm occurs when all nodes are assigned an RB. The algorithm is executed in stages, with the first stage involving the selected nodes from the probability-based selection. The subsequent stages use the Hungarian algorithm to assign RBs optimally, considering delay constraints. The convergence is achieved when RBs are assigned to all selected nodes.

Optimality in the Hybrid Hungarian Algorithm is measured by its ability to assign RBs optimally, considering the cost matrix that incorporates delay constraints. The algorithm ensures that RBs are allocated efficiently while adhering to the delay budgets of the selected nodes. The final assignment, considering both selected and unselected nodes to avoid RB wastage, contributes to the overall optimality of RB distribution.

4) OVERALL OPTIMALITY

The overall optimality of the entire system is achieved through the combination of these algorithms. The Shapley value ensures fairness in the initial distribution of RBs among clusters. The probability-based selection optimizes the node selection process within clusters based on the cost function. Finally, the Hybrid Hungarian Algorithm refines the RB assignments, considering delay constraints and achieving an optimal overall allocation.

VII. CONCLUSION

This paper proposes the GBEA scheme for ZT-DI-M2M communications, which is an efficient resource allocation algorithm that is both group-based and energy-conscious. The GBEA algorithm comprises three consecutive steps namely the virtual clustering using the GMM-EM algorithm, inter-cluster RB distribution inspired by Shapley value theory, and intra-cluster RB distribution with probability-based selection and the HHA algorithm. The proposed scheme has been incorporated into the NS2 platform, and its performance has been evaluated in a practical scenario. The simulation findings show that the proposed GBEA scheduling algorithm outperforms the other existing resource allocation algorithms in terms of throughput, delay sensitivity, and energy consumption, with 1.5, 1.75, and 6.5 fold, improvements respectively. After analyzing the results, the proposed resource allocation scheme demonstrates progressive performance when compared to the state-of-the-art techniques. In future work, an extension of this research could involve implementing GBEA in real-time environments with strict time intervals (TTIs) for resource allocation decisions and optimizing the algorithm's efficiency, particularly in large-scale networks, for faster execution and reduced computational load.

REFERENCES

[1] S. N. K. Veedu, M. Mozaffari, A. Höglund, E. A. Yavuz, T. Tirronen, J. Bergman, and Y.-P. E. Wang, "Toward smaller and lower-cost 5G devices with longer battery life: An overview of 3GPP release 17 RedCap," *IEEE Commun. Standards Mag.*, vol. 6, no. 3, pp. 84–90, Sep. 2022.

[2] M. S. Khalaf, A. A. A. Wahab, H. A. Al-Wzwazy, A. S. Al-Khaleefa, M. A. Alazzawi, and M. Almohamadi, "A look at the 3GPP standard's current positioning of 5G networks," in *Proc. Al-Sadiq Int. Conf. Commun. Inf. Technol. (AICCIT)*, Jul. 2023, pp. 28–33.

[3] U. Singh, A. Dua, N. Kumar, and M. Guizani, "QoS aware uplink scheduling for M2M communication in LTE/LTE—A network: A game theoretic approach," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4156–4170, Apr. 2022.

[4] E. Coronado, R. Behravesh, T. Subramanya, A. Fernández-Fernández, M. S. Siddiqui, X. Costa-Pérez, and R. Riggio, "Zero touch management: A survey of network automation solutions for 5G and 6G networks," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2535–2578, 4th Quart., 2022.

[5] R. Muzaffar, M. Ahmed, E. Sisinni, T. Sauter, and H.-P. Bernhard, "5G deployment models and configuration choices for industrial cyber-physical systems—A state of art overview," *IEEE Trans. Ind. Cyber-Phys. Syst.*, vol. 1, pp. 236–256, 2023.

[6] D. Moltchanov, E. Sopin, V. Begishev, A. Samuylov, Y. Koucheryavy, and K. Samouylov, "A tutorial on mathematical modeling of 5G/6G millimeter wave and terahertz cellular systems," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1072–1116, 2nd Quart., 2022.

[7] X.-R. Xu, Y.-H. Xu, W. Zhou, and A. Nallanathan, "Energy efficient resource allocation for UAV-served energy harvesting-supported cognitive industrial M2M networks," *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1454–1458, Aug. 2023.

[8] C. Zhuansun, K. Yan, G. Zhang, C. Huang, and S. Xiao, "Hypergraph based joint channel and power resource allocation for cross-cell M2M communication in IIoT," *IEEE Internet Things J.*, vol. 10, no. 17, pp. 15350–15361, Sep. 2023.

[9] S. B. Prathiba, G. Raja, S. Anbalagan, K. Dev, S. Gurumoorthy, and A. P. Sankaran, "Federated learning empowered computation offloading and resource management in 6G-V2X," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3234–3243, Sep. 2022.

[10] 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.

[11] K. S. Arikumar, S. B. Prathiba, R. S. Moorthy, K. Tamilarasi, M. M. V. Chalapathi, and A. Deepak Kumar, "The role of machine learning in IoT: A survey," in *Proc. 3rd Int. Conf. Smart Electron. Commun. (ICOSEC)*, Oct. 2022, pp. 451–457.

[12] N. K. M. Madi, M. M. Nasralla, and Z. M. Hanapi, "Delay-based resource allocation with fairness guarantee and minimal loss for eMBB in 5G heterogeneous networks," *IEEE Access*, vol. 10, pp. 75619–75636, 2022.

[13] Y. Mehmood, N. Haider, M. Imran, A. Timm-Giel, and M. Guizani, "M2M communications in 5G: State-of-the-art architecture, recent advances, and research challenges," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 194–201, Sep. 2017.

[14] D. Kim, M. Ko, S. Kim, S. Moon, K.-Y. Cheon, S. Park, Y. Kim, H. Yoon, and Y.-H. Choi, "Design and implementation of traffic generation model and spectrum requirement calculator for private 5G network," *IEEE Access*, vol. 10, pp. 15978–15993, 2022.

[15] D. Borsatti, C. Grasselli, C. Contoli, L. Micciullo, L. Spinacci, M. Settembre, W. Cerroni, and F. Callegati, "Mission critical communications support with 5G and network slicing," *IEEE Trans. Netw. Service Manage.*, vol. 20, no. 1, pp. 595–607, Mar. 2023.

[16] N. Zhu, X. Xu, S. Han, and S. Lv, "Sleep-scheduling and joint computation-communication resource allocation in MEC networks for 5G IIoT," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–7.

[17] W. U. Rehman, T. Salam, A. Almogren, K. Haseeb, I. Ud Din, and S. H. Bouk, "Improved resource allocation in 5G MTC networks," *IEEE Access*, vol. 8, pp. 49187–49197, 2020.

[18] M. Y. Abdelsadek, Y. Gadallah, and M. H. Ahmed, "Matching-based resource allocation for critical MTC in massive MIMO LTE networks," *IEEE Access*, vol. 7, pp. 127141–127153, 2019.

[19] Z. Mlika and S. Cherkaoui, "Massive IoT access with NOMA in 5G networks and beyond using online competitiveness and learning," *IEEE Internet Things J.*, vol. 8, no. 17, pp. 13624–13639, Sep. 2021.

[20] A. Shahini and N. Ansari, "NOMA aided narrowband IoT for machine type communications with user clustering," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7183–7191, Aug. 2019.

[21] Y.-H. Xu, W. Zhou, Y.-G. Zhang, and G. Yu, "Stochastic game for resource management in cellular zero-touch deterministic industrial M2M networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 12, pp. 2635–2639, Dec. 2022.

[22] S. Ali, A. Ferdowsi, W. Saad, N. Rajatheva, and J. Haapola, "Sleeping multi-armed bandit learning for fast uplink grant allocation in machine type communications," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5072–5086, Aug. 2020.

[23] M. Y. Abdelsadek, Y. Gadallah, and M. H. Ahmed, "A critical MTC resource allocation approach for LTE networks with finite blocklength codes," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5598–5609, May 2020.

[24] M. R. M. Anfar and J. Mwangama, "Machine learning-based service differentiation in the 5G core network," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Apr. 2021, pp. 144–149.

[25] L. Yang, X. Yu, H. Zhu, W. Chen, Y. Cao, Y. Zhao, H. Li, J. Zhang, and L. Yi, "Resource allocation in software defined hybrid satellite-terrestrial communication networks toward 5G," in *Proc. Int. Conf. Netw. Infrastructure Digit. Content (IC-NIDC)*, Aug. 2018, pp. 300–303.

[26] K. Wang, Y. Zhou, Z. Liu, Z. Shao, X. Luo, and Y. Yang, "Online task scheduling and resource allocation for intelligent NOMA-based industrial Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 803–815, May 2020.

[27] Q. Li, H. Tang, Z. Liu, J. Li, X. Xu, and W. Sun, "Optimal resource allocation of 5G machine-type communications for situation awareness in active distribution networks," *IEEE Syst. J.*, vol. 16, no. 3, pp. 4187–4197, Sep. 2022.

[28] S. Dawaliby, A. Bradai, Y. Pousset, and C. Chatellier, "Joint energy and QoS-aware memetic-based scheduling for M2M communications in LTE-M," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 3, pp. 217–229, Jun. 2019.

[29] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, "5G new radio: Waveform, frame structure, multiple access, and initial access," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 64–71, Jun. 2017.

[30] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: Potential, challenges, and solutions," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 178–184, Mar. 2012.

[31] F. Najar, S. Bourouis, N. Bouguila, and S. Belghith, "A comparison between different Gaussian-based mixture models," in *Proc. IEEE/ACS 14th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Oct. 2017, pp. 704–708.

SAHAYA BENI PRATHIBA (Member, IEEE) received the bachelor's, master's, and Ph.D. degrees in computer science and engineering from Anna University, Chennai, in 2022. Currently, she is an Assistant Professor with the Centre for Cyber Physical Systems, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai. Her current research interests include 5G/6G, vehicle-to-everything, software defined networking, autonomous vehicular networks, industry 5.0, and metaverse. She have secured 23rd rank among 2581 candidates in Master of Engineering. She was a recipient of the Anna Centenary Research Fellowship. She is serving as a reviewer for IEEE TRANSACTIONS and reputed Elsevier journals.

KATHIROLI RAJA received the M.E. degree in computer science and engineering and the Ph.D. degree from Anna University, Chennai, India, in 2008 and 2016, respectively. She is currently an Assistant Professor (Sl.Gr.) with the Department of Computer Technology, Anna University. Her current research interests include security in wire and wireless networks, machine learning, and intrusion detection.

R. V. SAIABIRAMI is currently pursuing the Bachelor of Engineering degree in the field of computer science and engineering with the Department of Computer Technology, Anna University, Chennai, India. Her research interests include wireless networking, database management, and machine learning.

GUNASHREE KANNAN is currently pursuing the Bachelor of Engineering degree in the field of computer science and engineering at the Department of Computer Technology, Anna University, Chennai, India. Her previous research contribution including "IoTexpert: Interconnection, Interoperability and Integration of IoT Platforms" got recognized by an international conference. Her research interests include the IoT, wireless networking, big data, and advanced computing.