

5G LTE-A Cognitive Multiclass Scheduling Scheme for Internet of Things¹Radhia Khdir , ²Aymen Belghith¹JOUF University,

Computer Science Department,

Saudi Arabia

rkhdir@gmail.com

²Saudi Electronic University, Computer Science Department, Saudi Arabia

aymen.belghith@gmail.com

ABSTRACT

Due to the exponential growth of many devices and objects creating a fully interconnected world, Internet of Things (IoT) have important impact on our social life. For example, in computer network and communication area, IoT systems are considered as a promising technology aiming the improvement of the fourth generation (4G) presented by Long Term Evolution Advanced (LTE-A) and even the support of the fifth generation (5G) systems. The Call Admission Control (CAC) and scheduling mechanisms are still considered as important challenges because they greatly contribute to achieve our goals. In this paper, we consider the uplink traffics in 5G systems and we present a new approach based on Quality of Service (QoS) requirements of the uplink IoT demands. The proposed algorithm prioritizes human to human users (H2H) without neglecting Machine to machine (M2M) users. We show, through simulation results, that our proposed algorithm positively influences both the served users and throughput performance. Furthermore, our proposal also ensures an important accepted service flows.

Key words : QoS; Scheduling; LTE-A; MURAP; Throughput

1. INTRODUCTION

Internet of Things (IoT) is the idea to integrate many machines and devices with Internet. In our society, these devices are expanded in all areas of our life (industrial, transportation, health, electricity, industrial, etc. ...) [1]. Face to network technologies growth, five generation networks (5G) is the appropriate solution to bear the IoT services. Indeed, 5G Long Term Evolution-Advanced (LTE-A) is the most promising technology for IoT applications. 5G LTE-A operators are considering Machine-to-Machine (M2M) communication as a big market potential. Indeed, 5G LTE-A technology plays an essential role in M2M communication deployment, as well its applications [2]. In fact, many improvements are defined and developed such as high flexibility, high data rate, seamless connectivity, low cost, low latency, etc. However, the improvements in M2M

communications must not disrupt the Human-to-Human (H2H) communications. Indeed, 5G LTE-A initially provides and optimizes the flexibility and high data rates for the H2H communication including the classical mobile services. M2M communication differentiates to that H2H communication in many characteristics such as the enormous number of autonomous devices, rare regular-basis transmission of small packets, and the high uplink traffic. In this context, various key issues and challenges will be studied to satisfy Quality of service (QoS) provisioning. Allocation and scheduling resources in 5G LTE-A are two examples of these key issues. Therefore, several researchers have been fixated on incorporating M2M applications using radio resource management (RRM) for IoT communications [3-5]. Most of these researches focus to study channel resource capacity without considering the QoS of M2M applications. Therefore, we are motivated to propose a scheduling algorithm for multiclass 5G LTE-A services by assigning different priorities to the M2M and communications. The proposed scheme uses a different M2M queues services to improve the global channel capacity without degrading the H2H services. Indeed, the M2M flows are classified into many queues (video, web, Voice over IP (VoIP), etc.). So, H2H communication is given the highest priority. However, the flows of each M2M service class can have different priority levels.

The remain of this paper is organized as follows: next Section describes the existing scheduling schemes. In section 3, we design the system model we present the proposed work. Performance analysis is provided in section 4. Numerical results are given in Section 5 and finally Section 6 concludes this work.

2. RELATED WORK AND MOTIVATION

In recent years, challenging and extensive researches have been focused in H2H and M2M communications. The most challenging issue as we mentioned is the optimal scheduling and resources allocation which handling diverse M2M devices and H2H users in 5G LTE-A technology.

In this section, we present the main review of the existing 5G LTE-A scheduling algorithms. In [6], an adaptive congestion control scheme is proposed. This scheme applies an Access Class Barring (ACB) factor to delay transmission access requests from numerous communicated M2M devices. Based on this ACB factor, the control scheme aims to reduce flows overload.

The authors presented in [7] a new prioritised random access algorithm. This algorithm, allocates the random access slots to various M2M classes flows and also uses a class dependent backoff procedure. Also, a prioritised random access is designed for M2M communications using OFDMA-based systems [8].

In [9], Zhang et al. proposed a tree structure which presents the multiplexing conditions of H2H users with various periods into a single channel. Basing on this tree structure, authors proposed a recursive resource allocation algorithm serving the different H2H users in the system. Khdir et al. [10] proposed a novel algorithm named Multi-Class Scheduling Algorithm (MuCSA) to serve flows having different classes rather than maximizing the fairness and throughput in LTE network.

Author in [11] proposed a scheduler with cognitive-based access strategy. This scheduler used a priority-based queuing model that is presented for LTE with M2M/H2H communication. The M2M communications are divided into two traffics types: M2M real time and M2M non-real time traffics. The proposed scheduler attributes the highest priority to the H2H communications while the M2M non-real time traffic has the lowest priority. Furthermore, to assess the performance of the system in terms of blocking and forced termination probabilities, service completion rate, and M2M traffic mean queuing delay, a Markov chain model with continuous-time is developed. Finally, the proposed model can be applied to ameliorate the M2M communication networks system performance. The main drawback of this work is to consider only two flows classes while our contribution has the ability to consider several flows classes.

A M2M scheduling algorithm, in 5G LTE-A technology including H2H and M2M traffics, was presented in [12]. The authors used bipartite graph and proposed formula for sharing the various resources between M2M and H2H communications. In addition, they proposed a power control algorithm to minimize the degradation of the H2H performance due to concurrent M2M traffics.

A new scheme named contention-based LTE transmission was proposed in [13]. This scheme can handle a huge number of traffics that transfer small and infrequent data in the LTE network. To avoid the simultaneous machine-type

communications which may cause threats to the LTE system, the author proposed an overload control scheme [14]. This proposed scheme considers a scenario for two types of devices depending on their priority (low and high). In [13] and [14], authors did not consider the classical H2H communication. Moreover, due to the high number of devices and so the huge volume of traffics, available resources may be scarce and therefore H2H communication may suffer from service denial.

In order to maximize bandwidth utilisation and minimize the excessive overall call blocking probability, a dynamic channel borrowing model is proposed in [15]. The resources borrowing mechanism is applied in dynamic manner. The available as well as the required resources are borrowed from the adjacent cells by the reference cell.

In [16], a hybrid preemptive / non-preemptive priority algorithm was designed in LTE. It considers the delay requirement of the M2M services. The M2M traffics are classified depending on the delay sensitivity. The algorithm priorities the delay-sensitive of M2M class but the priorities can be changed if services are interrupted.

In [17], Alqahtani et al. proposed a new scheme for H2H coexistence with M2M network. The authors present an analytical model to evaluate the performance of the H2H throughput when M2M communications exist in the system. In this model the asynchronous M2M transmission and the synchronous H2H transmission is considered at a time-slotted MAC structure of the H2H protocol.

The research works in [18] was based on clustering mechanism to design a robust algorithm to ensure several levels of QoS in the network.

Authors in [19] proposed a new scheduler named QoS-aware scheduling algorithm (QuAS) which increases the QoS fairness and overall throughput of the edge users without causing a significant degradation in overall system throughput.

The previous cited works classify services into two main types: H2H and M2M. The most of these works prioritizes the H2H users without considering the QoS requirements of M2M communications. Therefore, in the case of insufficient the resource, the H2H users will be served and it may cause starvation of the M2M communications. Furthermore, M2M traffics have not the same QoS requirements and therefore should be also classified. Therefore, we are motivated to design a cognitive scheme for 5G LTE-A networks using priority for multi-classes flows named Cognitive Multiclass Scheduling Scheme (CMuSS). Mainly, CMuSS scheduler handles several H2H and M2M traffics classes while improving the channel capacity and fairness.

3. SYSTEM MODEL

In this work, we consider one cell (Figure 1). All types of users are connected to one 5G LTE-A eNodeB. In this paper, two types of users are handled, i.e. H2H users and M2M devices. To serve all these users by our proposed scheme, we classify all user's requests in queues; one queue for H2H requests and K queues for M2M devices. Queue K has the highest priority. Each queue represents one service class (VoIP, video streaming, etc.). We suppose that the H2H requests queue is always prioritized than M2M queues. Also, we assume that the arrival rate of all H2H and M2M services follows the Poisson probability distribution. $\lambda_1, \lambda_2 \dots \lambda_K$ are the arrival rate respectively for H2H, M2M1 (i.e.M2M class C1) ... M2MK (i.e.M2M class CK) classes (services flows). An exponential probability distribution is followed by all request flows times.

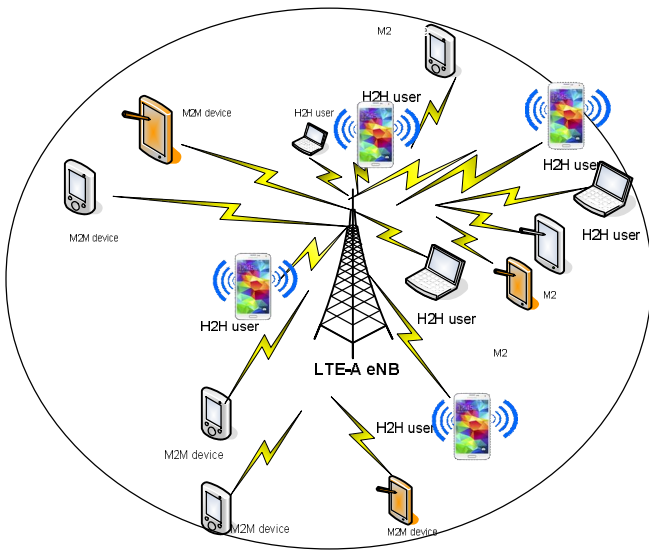


Figure 1: The system model of 5G LTE-A networks

4. PROPOSED COGNITIVE MULTICLASS SCHEDULING SCHEME (CMUSS)

Our optimization problem is to maximize the global system throughput. This optimization problem can be formulated as follows:

Our proposed scheduler handles two types of users. As it is mentioned in section 3, these users have several requests which are classified then filled to the adequate queue; one queue for H2H requests (this queue is always prioritized) and K queues for different M2M requests classes (see Fig.2).

We present several requests in matrix M (see table1). Indeed, the matrix receives $(N+Q)$ requests and m RB. Recall that N and Q represent the number of H2H and M2M requests, respectively. Therefore, the matrix dimension is $[(N + Q) \times m]$. Requests are ordered on the matrix from top to down according to the priority. Therefore, the H2H requests are

$$O^* = \sum_{i \in H, k \in N} R_{i,k}^H + \beta \sum_{j \in M, k \in Q} R_{j,k}^M \quad (1)$$

Where $R_{i,k}^H$ and $R_{j,k}^M$ are the achievable data rate of the k^{th} request of the i^{th} H2H UE and the k^{th} request of the j^{th} M2M UE, respectively. N is the total number of H2H requests and Q is the total number of M2M requests. β is the factor who guarantees that any M2M request must be served only after serving all H2H requests if the number of available RBs is insufficient. β is calculated as follows:

$$\beta = \begin{cases} 1 & \text{if } N_{RH} = N_{SH} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Where N_{RH} is the total number of requested RBs by H2H and N_{SH} is the total number of H2H requests served.

If the number of available RBs is sufficient, a distribution of resource is applicable for serving H2H and M2M requests. A highest percentage of resources is reserved to H2H requests. For example, if we propose that the total number of resources is 100 RB. We reserve for example 60% of RBs for H2H requests and 40% of RBs for M2M requests.

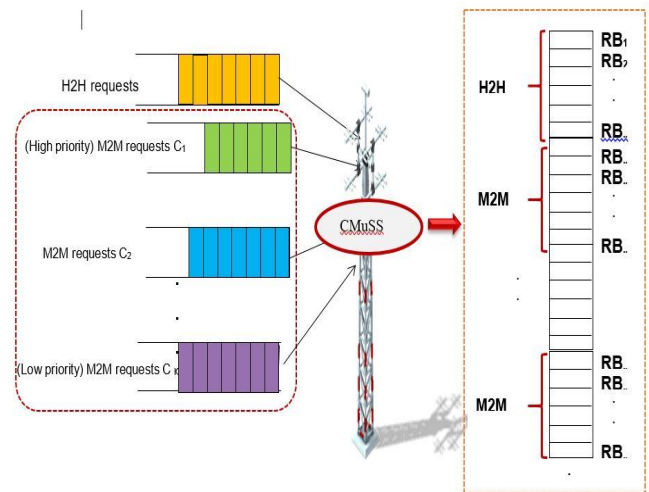


Figure 2: Priority requests based CMUSS scheduler

presented at the top of the matrix, which are followed by the M2M requests. Indeed, M is divided into two submatrix $M1$ and $M2$. $M1$ and $M2$ represent H2H requests and M2M requests, respectively. $M1$ submatrix values and $M2$ submatrix values are the achievable data rate if we assigned an RB to any request of any UE. In addition, $M1$ is divided in others submatrix following the type of any request. So, $M1$ contains some requests of different classes flows: $F_1, F_2 \dots, F_K$. In this work, we consider that there are three classes of service flows for M2M request; F_1, F_2 and F_3 which are Voice over Internet Protocol (VoIP), Video streaming, and File Transfer Protocol (FTP), respectively.

Table 1: Scheduling matrix

| | RB ₁ | RB ₂ | RB ₃ | RB ₄ | RB ₅ | RB ₆ | RB ₇ | RB ₈ | RB ₉ | RB ₁₀ | RB ₁₁ | RB ₁₂ | RB ₁₃ | RB ₁₄ | RB ₁₅ | RB ₁₆ | RB ₁₇ | RB ₁₈ | ... | RB _m |
|--------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----|-----------------|
| $k_{1,1}^H$ | 9 | 6 | 5 | 3 | 8 | 5 | 8 | 5 | 9 | 7 | 4 | 5 | 5 | 6 | 4 | 1 | 2 | 3 | ... | ... |
| $k_{2,1}^H$ | 5 | 5 | 4 | 2 | 5 | 4 | 5 | 2 | 2 | 5 | 1 | 2 | 7 | 5 | 2 | 4 | 7 | 4 | | |
| $k_{2,3}^H$ | 2 | 5 | 4 | 5 | 2 | 1 | 2 | 3 | 2 | 2 | 10 | 3 | 3 | 8 | 7 | 4 | 8 | 8 | | |
| $k_{3,4}^H$ | 1 | 2 | 4 | 2 | 5 | 3 | 8 | 7 | 8 | 4 | 2 | 4 | 5 | 8 | 1 | 4 | 3 | 7 | | |
| $k_{nb,N}^H$ | 5 | 5 | 1 | 2 | 8 | 3 | 10 | 2 | 2 | 3 | 7 | 2 | 5 | 4 | 1 | 5 | 5 | 1 | | |
| $k_{1,1}^M$ | 4 | 5 | 4 | 5 | 5 | 2 | 5 | 7 | 7 | 5 | 6 | 8 | 9 | 1 | 5 | 4 | 7 | 1 | | |
| F_1 | | ... | ... | | | | | | | | | | | | | | | | | |
| F_2 | | | | | | | | | | | | | | | | | | | | ... |
| F_K | | | | ... | | | | ... | | | | | | | | | | | | |
| $k_{n,Q}^M$ | 4 | 5 | 4 | 5 | 2 | 2 | 5 | 7 | 7 | 5 | 5 | 8 | 9 | 1 | 8 | 4 | 7 | 7 | ... | ... |

M1 Sub matrix1: H2H requests

M2 Sub matrix2: M2M requests

To achieve our optimization objective (see eq.1), our proposed scheduler serves H2H and M2M requests beginning with $M1$ submatrix values basing on the achievable data rate values and serves the request of UE corresponding to this highest value. Then, it “expands” the scheduling mechanism on both sides of the matrix and repeat the same reasoning. To conclude, our proposal handles $M1$ and $M2$ independently from the other. Also, the different classes flows are handled independently beginning with F_1 to F_K . The steps of our proposed scheduler are summary in table 2.

We explain our proposed algorithm though an example. we reserved for example 60% of RBs for H2H requested and 40% of RBs for M2M requests. As we mentioned, we consider also that there are three services classes for M2M type. Then 40% of RBs are divided between $F1$ (VoIP) $F2$ (Video streaming) and $F3$ (FTP). 20% of RBs will be assigned for $F1$, 15% RBs will be assigned for $F2$ and 5% for $F3$. Therefore, our scheduler allocates the RBs first of all to the H2H requests by selecting the highest value in the sub matrix $M1$. In this example, the highest value is equal to 10 (in green). After that, it “expands” the allocation on both sides of the sub matrix $M1$, the user obtains finally PRB_{10} , PRB_{11} and PRB_{12} (in blue). The second maximum matrix value is 8. We apply the same principle of scheduler to serve

all H2H requests of $M1$. After serving all H2H requests, we begin to serve the first service class in M2M requests (VoIP) using the mentioned distribution of resources (20%) and applying the same principle for H2H requests (searching the highest value, etc.). If the 20% from RB are allocated to $F1$, we then serve the Video streaming and finally the FTP service classes. The algorithm is presented as follows.

1. Present the H2H queue
2. Present the M2M queues with K classes flows (F_1, F_2, \dots, F_K)
3. Divide all requests (of different queues) into two submatrices ($M1, M2$) according to their types (H2H and M2M).
4. Calculate the matrix values; calculate the achievable data rate of each request of any UE if we assign to it a RB.
5. Handel each matrix $M1$ and $M2$
6. Start with H2H requests ($M1$)
 - 6.1 Search the highest value (achievable data rate) of $M1$ matrix values
 - 6.2 Serve the UE corresponding to this data rate value
 - 6.3 “Expand” the scheduling mechanism on both sides of the $M1$ matrix

- 6.4 Repeat 6.1 to 6.2 steps until all requests of all H2H users are served
- 7. Serve M2M requests (M2)
 - 7.1 For (($f=F_1$ to $f=F_K$), $f=f+1$, 1 until Q request)
 - 7.2 Search the highest value of F_f matrix values
 - 7.3 Serve the UE corresponding to this data rate value
 - 7.4 "Expand" the scheduling mechanism on both sides of the F_f matrix
 - 7.5 End For
 - 7.6 End

Table 2: The steps of proposed scheduler

5. PERFORMANCE EVALUATION

Our principal scheduler objective is to maximize the channel utilization and global system throughput. On the one hand, our scheduler always prioritizes users having the highest estimated throughput. On the other hand, our algorithm can distinguish between users according to their requested resources and according to their priorities. Therefore, our algorithm serves the users who will increase the overall throughput capacity (see fig.3). The same reason may explain the increase number of served H2H users. Indeed, as we mentioned in section 4, our proposed algorithm always gives priority to H2H users since they are classified in a separate queue of other service classes. Figure 4 confirms the results shown in Figure 3.

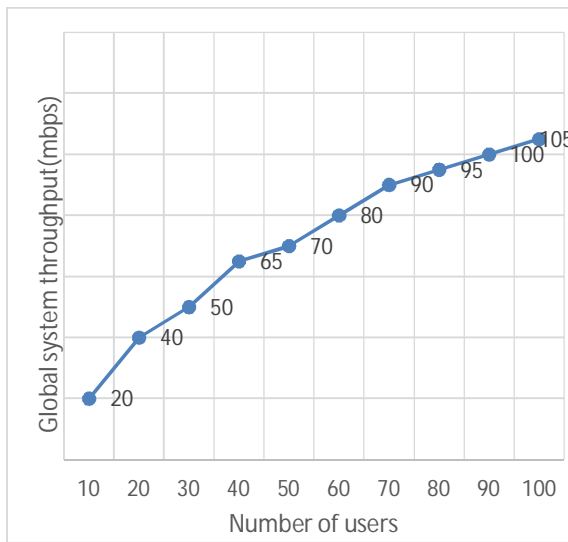


Figure 3: Global system throughput

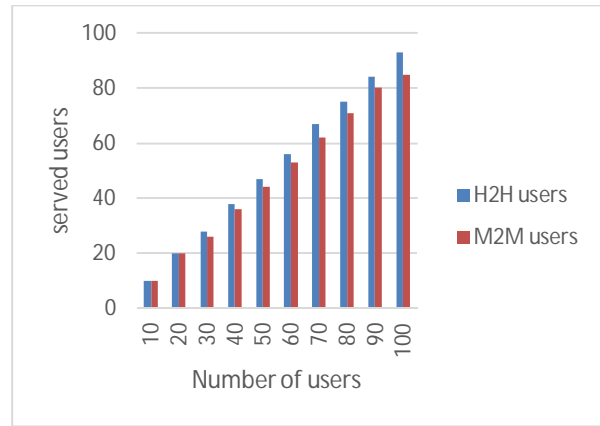


Figure 4: Number of served users

Figure 5 shows the rejected rate of H2H and M2M flows. It is noted that the blocking rate of M2M flows is higher than H2H flows. This result is expected since H2H flows are always prioritised than other types. In additional, recall that CMuSS prioritizes H2H flows because, from the user’s point of view, disruptions during mobility are considered more objectionable than M2M flows blocking.

Note that the end-to-end delay is composed of waiting, propagation, and transmission delays (see fig.6). In this case, we assume that the propagation and the transmission delays are the same for all users. However, the waiting time is related to the strategy of each scheduler. CMuSS serves a high number of H2H users. Therefore, the waiting time is always lower than that for M2M flows. This can explain the end-to-end delay for H2H flows. The same reasoning for M2M flows of $F1$ type are more prioritized than the second type $F2$. Then, the end-to-end delay will be lower for VoIP ($F1$) than for Video streaming ($F2$).

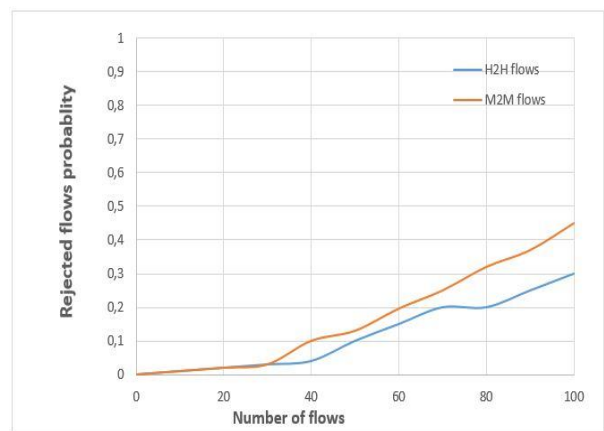


Figure.5: Rejected flows probability

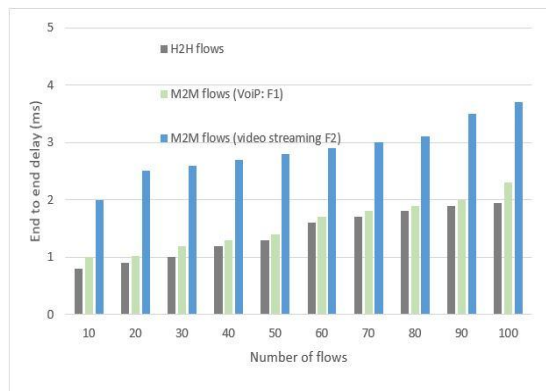


Figure.6: End to end delay

6. CONCLUSION

In this paper, we have proposed a new scheduler that manages two types of services flows; H2H and M2M flows. Therefore, two queues are managed using our proposed scheduler. H2H flows are always prioritized by CMuSS without neglecting the M2M flows. In addition, the principal objective of our proposal is to maximize the global system throughput. Indeed, CMuSS assigns resources and manages the sub matrices M_1 for H2H requests and M_2 for M2M requests respecting the order of each service flows. The performance of our scheme is proved by investigating several performance metrics. Indeed, CMuSS scheduler is studied in terms of end to end delay performance and rejected probability for each service flows as well as in terms of global system throughput. Our proposal has an important positive impact on global system throughput, served users, and end to end delay.

As future work, we plan to extend the proposed CMuSS when entering the Remote Radio Head (RRH) in the 5G system. In that case, the 5G with several RRHs may use an efficient scheduler to avoid flow accumulation in queues.

REFERENCES

[1] Hussain, S., Prasad, A., Kunz, A., Papageorgiou, A., Song, J.: **Recent trends in IoT/M2M related standards**. J. Inf. Commun. Converg. Eng. 4, S6n (2014)

[2] Ali, A.; Hamouda, W.; Uysal, M. “**Next generation M2M cellular networks: challenges and practical considerations**”, IEEE Commun. Mag. 53, 18–24 (2015) <https://doi.org/10.1109/MCOM.2015.7263368>

[3] Potsch, T., Khan Marwat, S.N., Zaki, Y., et al.: ‘**Influence of future M2M communication on the LTE system**’. 6th Joint IFIP Wireless and Mobile Networking Conf. (WMNC), 2013, pp. 1–4 <https://doi.org/10.1109/WMNC.2013.6549000>

[4] Sadeghi, S., Yaghmaee Moghddam, M.H., Bahekmatt, M., et al.: ‘**Modeling of smart grid traffics using non-preemptive priority queues**’. 2nd Iranian Conf. on Smart Grids (ICSG), 2012, pp. 1–4

[5] Giluka, M.K., Rajoria, N., Kulkarni, A.C., et al.: ‘**Class based dynamic priority scheduling for uplink to support M2M communications in LTE**’. IEEE World Forum on Internet of Things (WF-IoT), 2014, pp. 313–317 <https://doi.org/10.1109/WF-IoT.2014.6803179>

[6] Duan, S., Shah-Mansouri, V., Wang, Z., et al.: ‘**D-ACB: adaptive congestion control algorithm for bursty M2M traffic in LTE networks**’, IEEE Trans. Veh. Technol., 2016, 99, pp. 1–15 <https://doi.org/10.1109/TVT.2016.2527601>

[7] Jen-Po, C., Chia-han, L., Tzu-Ming, L.: ‘**Prioritized random access with dynamic access barring for RAN overload in 3GPP LTE-A networks**’. IEEE GLOBECOM Workshops (GC Wrkshps), 2011, pp. 368–372.

[8] Taehoon, K., Kab Seok, K., Dan Keun, S.: ‘**Prioritized random access for machine-to-machine communications in OFDMA based systems**’. IEEE Int. Conf. on Communications (ICC), 2015, pp. 2967–2972

[9] Zhang, Y.: ‘**Tree-based resource allocation for periodic cellular M2M communications**’. IEEE Wirel. Commun. Lett. 3(6), 621–624 (2014) <https://doi.org/10.1109/LWC.2014.2366769>

[10] R. Khdhir, A. Belghith, K. Mnif, and L. Kamoun, “**Scheduling algorithm for multiclass services in LTE network**,” in Proc. Int. Wireless Commun. Mobile Comput. Conf., 2016, pp. 982–987. <https://doi.org/10.1109/IWCMC.2016.7577192>

[11] AlQahtani, S.: ‘**Performance analysis of cognitive-based radio access strategy in multi-channel LTE—a networks with M2M/H2H coexistence**’. IET Commun. 11, 655–663 (2017) <https://doi.org/10.1049/iet-com.2016.0469>

[12] Hamdoun, S., Rachedi, A., & Ghamri-Doudane, Y. (2016). **A flexible M2M radio resource sharing scheme in LTE networks within an H2H/M2M coexistence scenario**. In 2016 IEEE international conference on communications (ICC) <https://doi.org/10.1109/ICC.2016.7511237>

[13] Andreev, S., Larmo, A., Gerasimenko, M., et al.: ‘**Efficient small data access for machine-type communications in LTE**’. IEEE Int. Conf. on Communications (ICC), 2013, pp. 3569–3574. <https://doi.org/10.1109/ICC.2013.6655105>

[14] Dementev, O., Galinina, O., Gerasimenko, M., et al.: ‘**Analyzing the overload of 3GPP LTE system by diverse classes of connected-mode MTC devices**’. IEEE World Forum on Internet of Things (WF-IoT), 2014, pp. 309–3012 <https://doi.org/10.1109/WF-IoT.2014.6803178>

[15] Chowdhury M.Z, Hossain M.A, Ahmed S, Jang Y.M, **Radio resource management based on reused frequency allocation for dynamic channel borrowing scheme in wireless networks**, Wireless Networks, (2015), 21(8), 2593-2607. <https://doi.org/10.1007/s11276-015-0937-9>

[16] Salman A. AlQahtani; **Analysis of a Hybrid Priority Scheme for Multiclass M2M Communications Over LTE-A Networks**, Wireless Networks, (2017), 6981–6993.

<https://doi.org/10.1007/s13369-017-2936-x>

[17] Salman A. AlQahtani; **Modeling and performance analysis of unlicensed bands MAC strategy in multi-channel LTE-A networks with M2M/H2H coexistence** *Wireless Networks*, (2018).

<https://doi.org/10.1007/s11276-017-1449-6>

[18] Salman A. AlQahtani; **A route stability-based multipath QoS routing protocol in cognitive radio ad hoc networks**, *Wireless Networks*, (2019).

<https://doi.org/10.1007/s11276-019-02014-6>

[19] Osman Gokhan Uyan, Vehbi Cagri Gungor **QoS-aware LTEA downlink scheduling algorithm: A case study on edge users**, *Wireless Networks*, (2019).