JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 37, 901-915 (2021)
DOI: 10.6688/JISE.202107_37(4).0010

A Priority-based Predictive Resource Scheduling Algorithm for MTC in LTE Networks

CHUNG-NAN LEE[1], CHUN-YEN TSENG[2] AND MING-FENG LEE[3]
Department of Computer Science and Engineering
National Sun Yat-sen University
Kaohsiung, 804 Taiwan
E-mail: cnlee@mail.cse.nsysu.edu.tw[1]; sb82115@gmail.com[2]; mflee@mail.nsysu.edu.tw[3]

Current communication networks are mainly designed for Human-to-Human (H2H) communication requirements. However, future Internet of Things (IOT) and fifth-generation mobile communications system (5G) will connect a lot of mechanical equipment. Therefore, this paper proposes a priority-based predictive resource scheduling algorithm that integrates the predictive method and priority scheduling for the uplink of Massive Machine Type Communication (mMTC). The predictive method is used to reduce the delay of a device sent scheduling request to the eNodeB to obtain the approved resources for data transmission. This is combined with a priority ranking sequence to develop the optimal resource allocation strategy to reduce the total delay time, and improve the transmission success rate and resource utilization. Simulation results show that the proposed algorithm can effectively reduce delay and increase the transmission success rate, compared with the existing methods.

Keywords: MTC, machine type communication, uplink scheduling, resource allocation, scheduling request, priority

1. INTRODUCTION

The third-generation mobile communications system (3G) and fourth-generation mobile communications system (4G) communication networks generally focus on Human-to-Human (H2H) applications. However, the incoming 5G has a comprehensive service for the machine-to-machine (M2M) oriented needs. At present, there are three major aspects of 5G according to the situation of use: (a) Enhanced Mobile Broadband (eMBB): this aspect serves mainly to achieve the ultra-high throughput required for 5G applications, such as ultra-high quality images, Ultra High Definition (UHD), Virtual Reality (VR) and ultra-real-time applications; (b) Massive Machine Type Communications (mMTC): the Internet of Things is the most important service in 5G. With various user devices connected to the Internet, there are smart homes, smart meters, and automated factories and so on forming a smart city; (c) Ultra-reliable and Low Latency Communications (URLLC): the Internet of Vehicles (IoV) must have high reliability and extremely low latency due to its high mobility, as well as other security and disaster alert systems.

In the foreseeable future, machine communication protocols will be one of the mainstream 5G applications, which can be used for environmental monitoring, video surveillance and disaster notification, and such applications require little human intervention. However, current LTE network architecture is not suitable for MTC, and 5G must adapt its resource allocation algorithm and the overall network architecture to facilitate the development and application of MTC. The 5G network architecture is based on the specific

Received October 15, 2019; revised February 8, 2020; accepted March 7, 2020.
Communicated by Ren-Hung Hwang.

design principles proposed in the NGMN 5G white paper [1]. It has an SDN (Software Defined Network) structure and NFV (Network Function Virtualization). The design principles are to make full use of the spectrum, achieve cost-effective deployment, coordinate and eliminate interference, support dynamic wireless topologies, offer flexible capabilities, support new value creation, enhance security and privacy, and streamline operations and management.

This paper mainly focuses on forward-looking 5G MTC research under the existing LTE platform. More specifically, this study constructs a network environment simulating LTE, and applies the proposed resource allocation algorithm to MTC to carry out simulation tests. Current LTE research focuses mainly on H2H communications. The data traffic travels mainly from the base station (eNodeB) to the device, that is, the downlink. In the case of MTC, the data traffic moves in the opposite direction. The device is directed to the eNodeB, which is the uplink. Under the LTE network architecture, the downlink is mainly for high-speed data transmission, and is effective against frequency selective attenuation, while the uplink average power consumption is relatively low, which can improve the power transmission efficiency of the device and can reduce the power consumption of the device battery, prolonging the service life cycle. In terms of uplink scheduling, the research on MTC applications can be divided into two categories. The first aims to optimize the power consumption of the user device, and reduce the frequency of transmission or the rate of transmission in order to improve the operation time. The second aims to reduce transmission delay time.

The main contributions of this paper are as follows. First, for the application of MTC, an algorithm that can effectively use the limited radio resources of eNodeB is proposed, and the priority is assigned based on the device state according to the different states in which the device is located. Secondly, a predictive approach is used to effectively schedule resources to reduce possible latency. Since the applications of MTC include sensing of various environments, emergency situations, and notification of disasters, if the data is not transmitted to the eNodeB within the time limit, it will be out of date. Therefore, it is very important that data reach the eNodeB within the time limit. In order to meet this requirement, the overall delay must be reduced.

The rest of this paper is organized as follows: In Section 2, we review the related works. Section 3 describes the proposed method. Section 4 gives simulation results. Conclusions are drawn in Section 5.

2. RELATED WORK

Massive MTC is an important application scenario in 5G. MTC has the opposite characteristics to Human-to-Human (H2H) communications over LTE/LTE-A Networks. In terms of radio resource management and allocation, most of the H2H scheduling algorithm focuses on downlink, how to effectively allocate radio resources to achieve the best fairness, efficiency and resource utilization. In terms of H2H uplink, it is also a resource designed to maximize performance and maintain continuity allocated to specific devices, however these characteristics are not applicable to MTC [13, 14]. We then further discuss the correlation of various algorithms or methods in the uplink aspect with MTC.

The application of uplink scheduling in MTC can be divided into two directions. One is to optimize the power consumption of the user device. By reducing the transmission frequency or the transmission rate, the power is reduced and the battery operation time can be longer. The other is to reduce the delay time of transmission. Since all the data transmitted by MTC has its time limit, if the data is received late due to the long delay time, the availability of this data will be greatly reduced, especially in emergency and disaster alert information. Kalil *et al.* [15] proposed an energy-efficient LTE uplink scheduling, taking into accounts the QoS (Quality of Service) requirements and channel fading. The main goal is to ensure the user's QoS while minimizing the transmission power.

Essassi [16] proposed a centralized and decentralized resource allocation algorithm based on genetic inheritance. They use decentralized alternate resource allocation optimization and reduces the size of the cluster. With centralized, the resources in the entire cluster are determined by the central cell. This method has three steps. First, the central cell collects the perceived interference in the resource blocks of neighboring cells, then the central cell executes the genetic optimization algorithm, and finally sends a new resource block allocation decision to the neighboring cells in the next cycle. The research in [17] considers the efficient allocation of limited resources and the reduction of inter-cell interference in the case of small cell networks with high density. The method divides and organizes cells, and allocates available resources according to the number to avoid the problem of resource snatching. Ragaleux *et al.* [18] proposed an adaptive and potential aware scheduling scheme (APASS) in the LTE-A environment, which operates by combining three algorithms. The first algorithm is to solve the frequency-domain packet scheduling (FDPS) problem without the maximum number of users. Then the second algorithm is used to schedule resource allocation. Finally, the third algorithm is used to modify the second resource allocation block. APASS overcomes the limitations and characteristics of LTE to find a better solution

Although most studies have used LTE as an environment to simulate MTC, there are many difficulties in deploying MTC in LTE / LTE-A. Ghavimi and Chen [19] discussed some requirements for implementing M2M applications under LTE / LTE-A, such as low mobility, time control, delay tolerance, and small data transmission. After all, LTE / LTE-A is mainly designed for H2H. Generally, H2H uplink traffic is lower than downlink, but M2M is the opposite situation.

Azair and Miao [2] examined M2M uplink scheduling and transmission power control under LTE networks to maximize device lifetime in cellular networks. In order to achieve simplified scheduling and save energy, a distributed clustering scheme was proposed to explore how to optimize the cluster size. Su and Hwang [3] proposed an efficient resource allocation with grey relational analysis (ERAGRA), which is a channel-aware resource allocation algorithm under 3GPP LTE. Its main purpose is to maximize the system throughput and resource utilization fairness. In the simulated environment, ERAGRA has better results in terms of performance or fairness compared to the round robin and the best-CQI methods.

Abu-Ali *et al.* [4] noted that current LTE/LTE-A studies focus on how to modify the downlink scheduling to improve the overall performance or load balance. They also reviewed many LTE/LTE-A uplink scheduling related papers and analyzed them. They specifically discussed the integration of technologies such as single-carrier frequency-division multiple access, coordinated multipoint Tx/Rx, and carrier aggregation with MTC, and then discussed which types of algorithms should be matched in different situations to achieve better results. The research in [5] targeted several algorithms commonly used in LTE uplink scheduling to perform simulation and performance evaluation, such as first maximum expansion (FME), recursive maximum expansion (RME), and carrier-by-carrier in turn, with the channel dependent matrix and the proportional fairness matrix, simulated under different factors, and finally summarized the challenges of the uplink and future research directions. Ali *et al.* [6] proposed an uplink resource allocation algorithm that allows the devices to make decision of resource allocation blocks according to their battery status and the QoS metric that relates their application's power profile.

The method proposed by Mostafa and Gadallsh [7] is based on different data type setting properties for statistics to prioritize and save resources and energy by reducing unnecessary transmissions. In terms of environmental monitoring data, a threshold value is set for comparison. For video data, the importance of this data is also judged according to the similarity of each frame, so as to avoid repeated waste of energy and resources on the frame with low difference. The statistical priority is calculated by evaluating the specific statistical properties of the data. The importance of the data value is determined by testing the upper and lower thresholds to determine whether the data needs to be transmitted. Elhamy and Gadallah [8] proposed a balanced alternating uplink scheduling technique, Balanced Alternating Technique (BAT). The main design concept of BAT is to achieve a trade-off between throughput maximization and meeting system deadlines. BAT is adaptive, and its scheduling metrics can be adjusted to be based on delay or based on channel state, or a combination of both, depending on network operating conditions and priorities. Mehaseb *et al.* [9] analyzed most LTE uplink scheduling technologies and classified them for M2M applications. Based on the different applications in the M2M environment, they discussed some existing algorithms for these applications.

Giluka *et al.* [20] proposed a Class-Based Dynamic Priority (CBDF) scheme. This scheduling technology is most effective for hybrid schedulers with H2H and M2M. Their algorithm mainly supports M2M and has a small impact on H2H. The CBDP scheme not only meets the QoS requirements of H2H, but also meets the QoS requirements of M2M.Chao and Wang [21] studied the small data uplink transmission of the cellular network, which is mainly used in the scenario of massive machine type communications (mMTC). They explored that contention-based transmission (CBT) is guaranteed for delay time and signal consumption in comparison with scheduling-based transmission (SBT). However, CBT faces challenges in the random access collision and resource use conflict. Chao and Wang also made changes to the MAC layer to improve the problem of random access collisions. The simulation also clearly shows that if there is no proper MAC mechanism, CBT will perform worse than SBT.

Han *et al.* [22] studied the power-domain non-orthogonal multiple access (NOMA) technology to support energy-efficient massive MTC networks. In order to maximize the effective energy efficiency of large MTC networks based on uplink NOMA, they considered the characteristics of short packet-communication and studied the subchannel allocation and power control policy. Their method obtains the optimal transmission power policy by approximating the achievable effective rate of uplink NOMA-based short packet communications. Miuccio *et al.* [23] studied the problem of uplink wireless resource allocation in the mMTC scenario, which contains a large number of battery constrained MTC devices generating small-sized traffic. By adopting Sparse Code Multiple Access (SCMA) technology in the physical uplink shared channel (PUSCH), they proposed a dynamic load-aware physical random access channel (PRACH) and PUSCH resource allocation. Simulation results show that the Miuccio *et al.* allocation scheme can increase the number of successful communications while ensuring low power consumption. Lin *et al.* [24] pointed out that in MTC services, improving energy efficiency and quality of service (QoS) are both important issues, but many studies have not been able to improve both. Therefore, Lin *et al.* Proposed an uplink scheduling algorithm assisted by Multi-access Edge Computing (MEC) to simultaneously reduce power consumption and guarantee QoS. Finally, Lin *et al.* used OpenAirInterface as a simulation platform to evaluate the performance of their uplink scheduling algorithm. Compared with Round-Robin (RR) and Proportional fair (PF), their uplink scheduling algorithm has better energy consumption and the system performance.

3. PROPOSED ALGORITHM

The main concept of the proposed algorithm is to improve the immediacy of the overall system, so the delay time of each message can be reduced. In the case of MTC, the states of the devices (or MTC devices) can be classified into the following three modes:

1) Regular mode: In this mode, the data to be transmitted by the device is generated by a specific sensor to constantly detect various phenomena in the environment, including temperature, humidity, pressure and light intensity; and the data can be collected for recording and archiving.
2) Alarm mode: In this mode, the data generated by an abnormal situation is mainly caused by an event trigger. This mode is used for various emergencies such as fire alarms or unauthorized entrance.
3) Idle mode: In this mode, the device does not sense data or is in an idle state that is not triggered by an event.

The main purpose of dividing the device states into three types is to observe whether the resources not used in idle mode can be effectively utilized under different numbers of devices and different packet arrival rates. The proposed algorithm can be divided into two parts: the construction of the traffic model and the scheduling algorithm, as described below.

3.1 Traffic Model

The traffic model is established according to the method proposed in [10]. The original concept of the research in [10] is that the device has only two states and does not have an idle mode. In reality, however¸ a device will not always be in a state of transmission, and will not always be receiving data transmitted from base station. Therefore, idle mode is included when establishing the traffic model. As shown in Fig. 1, the blue dotted line indicates that the device is transmitting data in regular mode; the red circles represent locations where emergencies are occurring, and devices near the circles will be affected, and enter alarm mode. The red lines represent the transmission of emergency data.

In the traffic model, a circle centered on a base station and with R as the radius is the range covered by the base station, and N devices are spread evenly in this range. Traffic is simulated according to the state and application of the device. Fig. 2 is a Markov chain model with three states that are regular (R), alarm (A) and idle (I). Events are randomly generated and appeared within the coverage of the base station. Assume that when an event occurs, the device within a radius of 50 meters centered on the location will be triggered into alarm mode with an arbitrary probability of 0.002 to simulate the scarce event.

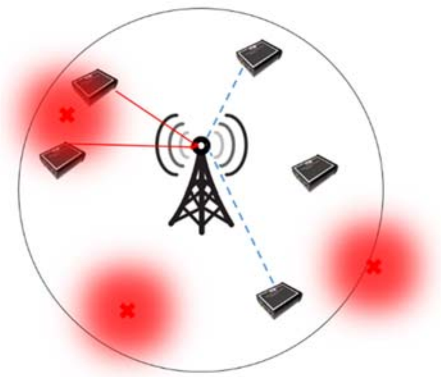
Fig. 1. Traffic model diagram.

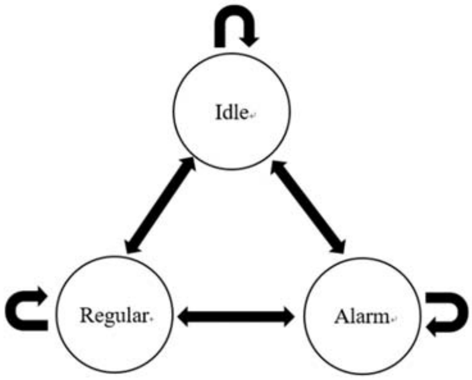

Fig. 2. Device state transition diagram.

3.2 Scheduling Algorithm

Since the latency tolerance of MTC applications is quite strict, if the data is not delivered within the time limit, it may be useless. Therefore, based on this fact, Brown and Khan [11] proposed a reduced delay prediction allocation algorithm. However, their algorithm only discusses basic problems such as the feasibility of prediction, and does not measure the priority of the device in various situations. Therefore, this study will focus on the priority issue. Fig. 3 shows a device waiting on the uplink in the RRC_CONNECTED state of standard LTE which is the waiting time that must be consumed before a request is sent. The horizontal axis in Fig. 3 represents the time intervals. Where (I) (II) (II) indicate the delay. (I) is variable delay depending on the SR periodicity and offset, (II) is a variable delay depending on an eNodeB scheduler policy, (III) is a 4ms fixed delay. Each of x, y, and z represents a period of time. The superscripts (1), (2), (3) and (4) are the sequence of actions between UE and eNodeB. In (1) a data enters UE transmit buffer; (2) the device first sends an SR (Scheduling Request) notification to the eNodeB; (3) Then the eNodeB returns the authentication to the device when sufficient resources are available; (4) When the device receives the authentication, it starts transmitting data. In this case, it will have a lot of unnecessary waiting time.

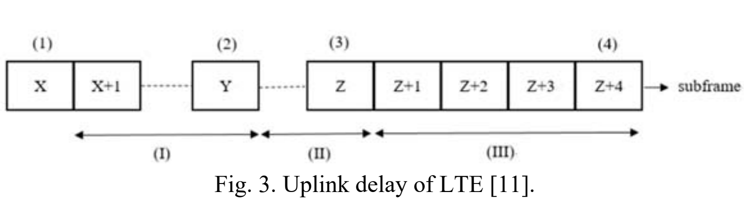

Fig. 3. Uplink delay of LTE [11].

Table 1. SR period and subframe offset configuration [12].

SR configuration Index I_{SR}	SR periodicity(ms) $SR_{PERIODICITY}$	SR subframe offset $N_{OFFSET,SR}$
0-4	5	I_{SR}
5-14	10	$I_{SR} - 5$
15-34	20	$I_{SR} - 15$
35-74	40	$I_{SR} - 35$
75-154	80	$I_{SR} - 75$
155-156	2	$I_{SR} - 155$
157	1	$I_{SR} - 157$

Table 1 lists the SR period and subframe offset configuration. Let n_f be the number of the frames, and $\lfloor n_s/2 \rfloor$ be the number of the subframes, where $\lfloor n_s/2 \rfloor$ indicates the largest integer not exceeding $n_s/2$. Then a device transmits SR that satisfies Eq. (1) to yield the required resources for sending SRs.

$$(10 \times n_f + \lfloor n_s/2 \rfloor - N_{OFFSET,SR}) \bmod SR_{PERIODICITY} = 0 \tag{1}$$

From Eq. (1) one can calculate how many subframes are needed for transmitting SR on PUCCH. For example, given SR configuration index, it gives the corresponding system frame number and subframe number.

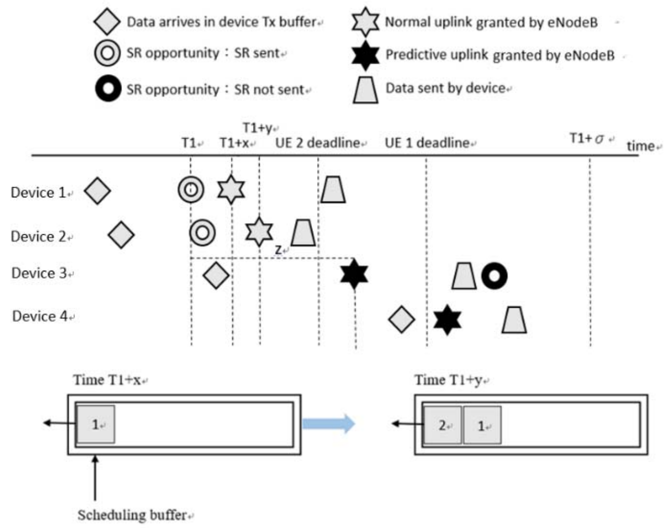

Fig. 4. Predictive uplink allocation concept.

The general idea of the proposed scheduling algorithm that is similar to the one in [11] can be seen from Fig. 4. This shows the waiting time that consumed before a request is sent by applying the proposed algorithm. The major difference between the concept of the proposed scheduling algorithm and that of the Brown-Khan's scheduling algorithm is that the proposed algorithm sorts priorities according to the transmission deadline, which is more suitable for MTC applications. In Fig. 4, both device 1 and device 2 transmit the data after the SR request is normally sent and authorized. Then device 3 starts the prediction process after device 1 has sent out the SR. Because there are still some resources available, the eNodeB sends a predicted uplink authorization and allocates resources to device 3 for data transmission. This can reduce the time taken normally by device 3 to send the SR, and for the eNodeB to perform the normal uplink authorization. Device 4 does not obtain the predicted uplink authorization at the same time because the authorization and required resources have been allocated to device 3.

Let P(SR) be the probability that device $m+1$ sends an SR, and P(Pred) is the probability that the eNodeB schedules a predictive allocation toward the device $m+1$ after it has received an SR from device m. It is also assumed that P(U) is the probability that the device $m+1$ is unsuccessfully predicted and allocated, and P(S) represents the probability of successful allocation toward the device $m+1$. Note that the transmission probability of different devices is different. It can also be seen that the device will only transmit the SR without prediction and prediction failure, and its probability can be expressed as Eq. (2):

$$P(SR) = 1 - P(Pred) + P(U) = 1 - P(S). \tag{2}$$

The difference between device 1 and device 2 is their data transmission deadlines that determine the priority. The device with the more urgent deadline has higher priority in the scheduling buffer. According to different data types, there are different SR periodicity, which can be obtained by substituting them into Eq. (1) by SR configuration index and SR subframe offset, according to Table 1.

The flowchart of the proposed scheduling algorithm is shown in Fig. 5, where the predictive algorithm is detailed in Fig. 6. Before the predictive allocation, the data type

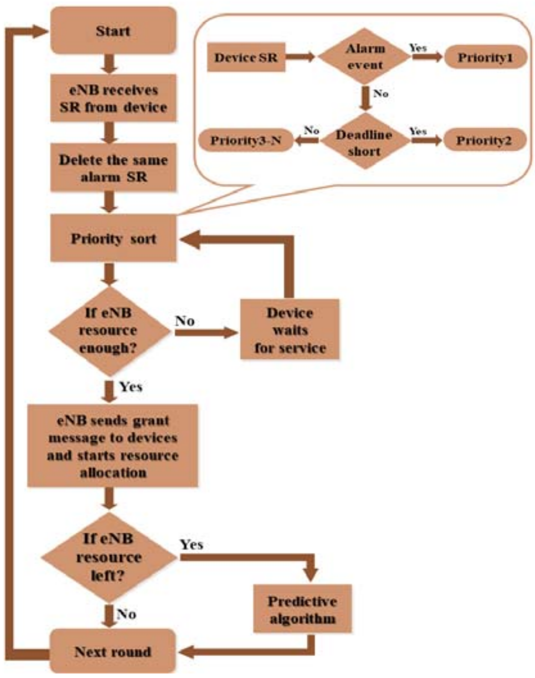

Fig. 5. Flowchart of the proposed scheduling algorithm.

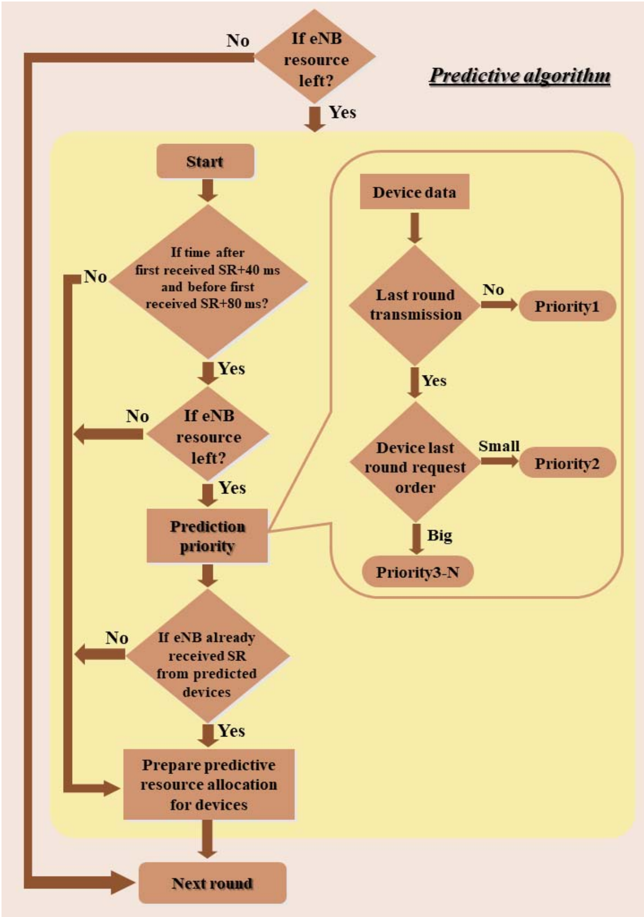

Fig. 6. Flowchart of the proposed predictive algorithm.

and the data deadline are used as the reference for sorting and allocating. After the priority is scheduled, the redundant resources will be predicted and allocated. After the second round, the predicted priority allocation will also be considered. The main advantage of this predictive algorithm is to reduce the delay time by using prediction. At the same time, in combination with the priority order, the best resource allocation strategy is formulated to reduce the total delay time and improve the transmission success rate and resource utilization. However, it is possible that low priority service request might not be served for a long time, when the resource blocks is limited.

Note that the case in Fig. 4 only occurs if all devices belong to the same application and do not have any priority differences. It is assumed that device 4 is in the case of alarm mode, and the SR requirement has been transmitted to the eNodeB. If the required resources are not immediately allocated to device 4 for emergency return, it may cause significant harm after a delay. Therefore, in order to address this problem, some resource blocks (RBs) should be reserved in the prediction allocation behavior for devices in alarm mode that is used to transmit emergency messages.

4. SIMULATION

This study first used the simulation to build the LTE network architecture, then wrote the MTC traffic model in C++, and finally imported the proposed scheduling algorithm into the model to observe its performance. Table 2 lists the parameters configured for the simulation.

Table 2. Simulation parameters.

Parameter	Value
Duplex technique	TDD
Bandwidth	20 MHz
Number of eNodeBs	1
Number of devices	100-500
Distance dependent pathloss (*R* in km)	$128.1 + 37.6 \times \log_{10}(R)$dB
SINR	Signal / (Interference+Noise)
eNodeB transmitting power	43 dBm
device transmitting power	23 dBm
Cell radius	1000 m
Prediction time	80 ms
Prediction reserved RB	10%
Minimum and Maximum delay	6ms/279ms

OFDMA is used in the LTE uplink, with ms as the time unit. Each slot has 100 RBs to be used, two slots are equal to 1 ms and 10 ms is a frame. The number of RBs required for each device is calculated according to the number of required resources and the amount of bits that can be carried by each RE (Resource Element) corresponding to the CQI. The CQI value of the device is converted using SINR judgment, mainly referring to the FPGA implementation of the LTE system channel estimation in the time-varying channel environment. Note that the proposed algorithm uses 90% of RBs to perform predictive allocation of general conditions, and reserves 10% of RBs without predictive allocation. The reservation of 10% of resources without predictive allocation is to allow emergency equipment and safety devices to transmit data in the event of an emergency.

Fig. 7 shows the delay time for starting prediction at different time points under the First In First Out (FIFO) algorithm and the proposed algorithm, respectively, where *z* rep-

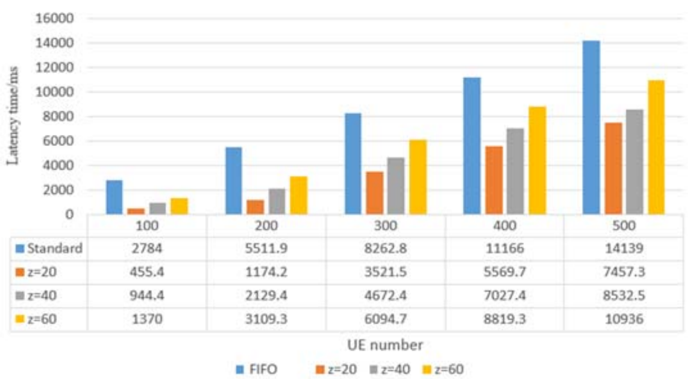

Fig. 7. Delay time for starting prediction at different time points.

resents the time point at which to start the prediction. It can be seen that when the number of devices is only about 100 and 200, the proposed algorithm has obvious latency reduction. In addition, the difference of the proposed algorithm under different *z* values can be seen. The smaller the *z* value, the earlier the prediction starts, thus the more latency can be reduced. However, when the number of devices reaches 300, the effect of the *z* value on the delay time is less significant, so the *z* value has a better impact on the delay time only when the number of devices is small, or there are sufficient resources. In general, the proposed algorithm can effectively reduce the delay time of SR transmission and grant message backhaul with different numbers of devices.

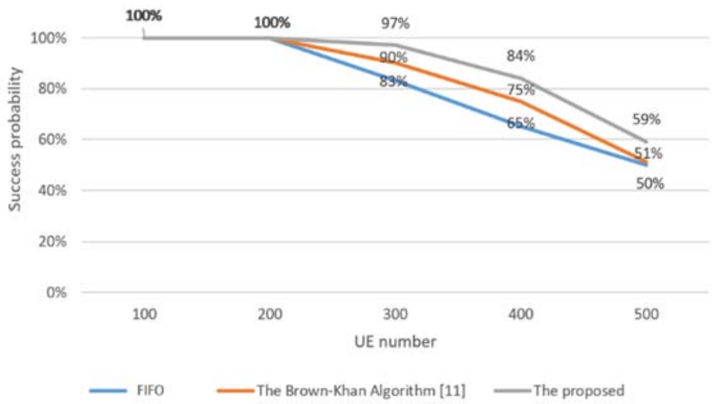

Fig. 8. Transmission success rate.

Fig. 8 shows the success rate of the device data transmitting to the base station, and the data arrival rate is 80% and data arrival rate indicates the probability that the device sending data periodically may send data next time. Non-prediction means that the resources are allocated according to the timing of the SR transmission, and the first requester obtains the resources first. At the beginning of the simulation, the resources are sufficient to meet the scheduling request needs of 200 MTC devices, so all methods can reach 100% transmission success rate. After more than 200 MTC devices, the resources started to be insufficient, so the advantages and disadvantages of each algorithm begin to show up. It can be seen that the transmission success rate of the transmission drops very fast when there are more than 200 devices. When the number of devices reaches 300, the success rate begins to decline, and it is about 50% when the number of devices reaches 500. The success rate of the Brown-Khan predictive algorithm begins to decrease at 300 devices, and the decline is particularly noticeable in the range between 400 to 500 devices. This is because the prediction is made in the case of idle resources, but there is no priority classification for the deadline of the device data, resulting in some device data exceeding the deadline, to cause the transmission failure. When the number of devices is 500, the performance of the Brown-Khan predictive algorithm is similar to that of non-prediction. The proposed predictive algorithm predicts the priority according to the data deadline, and its success rate only begins to decline after 400 devices. At 500 devices, it can still maintain a success rate of about 59%, which is obviously better than the other two.

Fig. 9 shows the resource utilization simulation results for cases involving 300 and 400 devices. The data arrival rate is 70%. It can be seen that the proposed algorithm performs better than the FIFO or the Brown-Khan predictive algorithm in improving resource utilization. Note that when nearly 400 devices are involved, the resource utilization is close to saturation, so there is not much difference between 300~400 devices and more than 400 devices. In the case of many small cells for massive IOT situation, if the devices in each small cell are mostly within 400, our method can have good results.

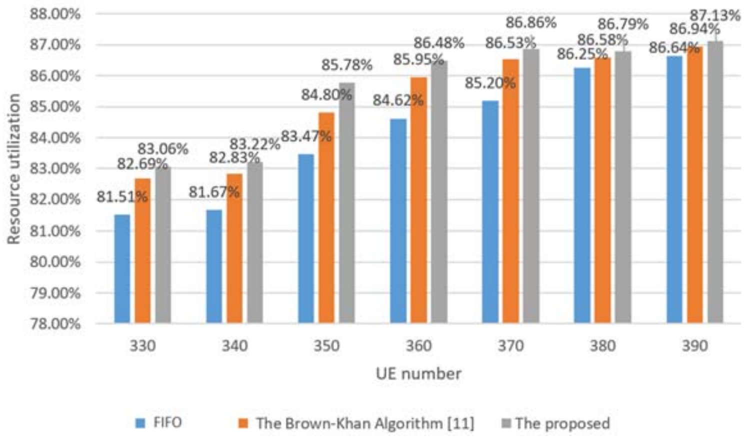

Fig. 9. Resource utilization.

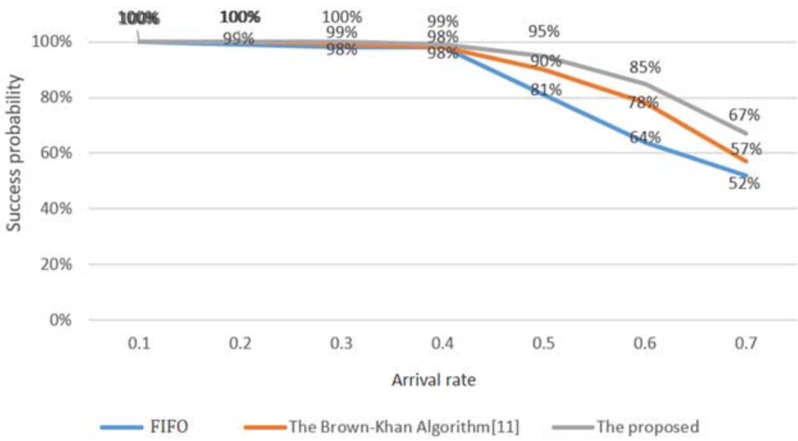

Fig. 10. Success rate under different data arrival rates.

Fig. 10 is the simulation result of the case where the number of devices is involving 500 devices under different data arrival rates. It can be seen that the performance of the unpredicted algorithm only has a 52% success rate when it reaches 0.7. In contrast, the success rates of the Brown-Khan predictive algorithm and the proposed algorithm are 57% and 67%, respectively.

5. CONCLUSION

This paper proposes a priority-based predictive resource scheduling algorithm that can reduce device scheduling request delay to improve MTC transmission immediacy and success rate. In the proposed algorithm, the base station reduces the delay caused by the general scheduling or pre-scheduled request. In addition, data transmission deadlines are used to prioritize device requests, and resources are allocated according to the priority to avoid data expiration, which improves the overall data transmission success rate. Simulation results show that in terms of delay time reduction, if the number of devices and the scheduling request are within a reasonable amount, the proposed algorithm achieves better performance, and the transmission success rate is better than those of the competing methods.

REFERENCES

1. R. E. Hattachi and J. Erfanian, "5G white paper," Technical report of NGMN Alliance, Ver. 1.0, 2015, pp. 1-125.
2. A. Azair and G. Miao, "Lifetime-aware scheduling and power control for M2M communications in LTE networks," in *Proceedings of IEEE 81st Vehicular Technology Conference*, 2015, pp. 1-5.
3. R. Su and I. Hwang, "Efficient resource allocation scheme with grey relational analysis for the uplink scheduling of 3GPP LTE networks," in *Proceedings of IEEE International Conference on Industrial Technology*, 2016, pp. 599-603.
4. N. Abu-Ali, A. M. Taha, M. Salah, and H. Hassanein, "Uplink scheduling in LTE and LTE-advanced: Tutorial, survey and evaluation framework," *IEEE Communications Surveys and Tutorials*, Vol. 16, 2013, pp. 1239-1625.
5. H. Safa and K. Tohme, "LTE uplink scheduling algorithms: Performance and challenges," in *Proceedings of the 19th International Conference on Telecommunications*, 2012, pp. 1-6.
6. A. Ali, G. A. Shah, and J. Arshad, "Energy efficient resource allocation for M2M devices in 5G," *Sensors*, Vol. 19, 2019, pp. 1-16.
7. E. Mostafa and Y. Gadallah, "A statistical priority-based scheduling metric for M2M communication in LTE networks," *IEEE Access*, Vol. 5, 2017, pp. 8106-8117.
8. A. Elhamy and Y. Gadallah, "BAT: A balanced alternating technique for M2M uplink Scheduling over LTE," in *Proceedings of IEEE 81st Vehicular Technology Conference*, 2015, pp. 11-14.
9. M. A. Mehaseb, Y. Gadallah, A. Elhamy, and H. Elhennawy, "Classification of LTE uplink scheduling techniques: An M2M perspective," *IEEE Communications Surveys and Tutorials* Vol. 18, 2016, pp. 1310-1335.
10. H. Thomsen, C. N. Manchon, and B. H. Fleury, "A traffic model for machine-type communications using spatial point processes," in *Proceedings of IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications*, 2017, pp. 1-6.
11. J. Brown and J. Y. Khan, "A predictive resource allocation algorithm in the LTE uplink for event based M2M applications," *IEEE Transactions on Mobile Computing*, Vol. 14, 2015, pp. 2433-2466.

12. 3GPP TS 36.213, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 12), V12.5.0, Apr. 2015.
13. A. Chakrapani, "Efficient resource scheduling for eMTC/NB-IoT communications in LTE Rel. 13," in *Proceedings of IEEE Conference on Standards for Communications and Networking*, 2017, pp. 66-71.
14. T. P. C. Andrade, C. A. A. Trujillo, and N. L. S. Fonseca, "Allocation of control resources for machine-to-machine and human-to-human communications over LTE/LTE-A Networks," *IEEE Internet of Things Journal*, Vol. 3, 2016, pp. 366-377.
15. M. Kalil, A. Shami, and A. Al-Dweik, "QoS-aware power-efficient scheduler for LTE uplink," *IEEE Transactions on Mobile Computing*, Vol. 14, 2015, pp. 1672-1685.
16. S. Essassi, M. Siala, S. Cherif, R. Hamila, and M. O. Hasna, "Centralized-decentralized RB allocation based on genetic algorithm and coordination over X2 interface in LTE uplink," in *Proceedings of International Symposium on Wireless Communication Systems*, 2015, pp. 121-125.
17. S. Essassi, M. Siala, R. Hamila, M. O. Hasnsa, and S. Cherif, "Power control and RB allocation for LTE uplink," in *Proceedings of International Wireless Communications and Mobile Computing Conference*, 2016, pp. 417-421.
18. A. Ragaleux, S. Baey, and M. Karaca, "Standard-compliant LTE-A uplink scheduling Scheme with Quality of Service," *IEEE Transactions on Vehicular Technology*, Vol. 66, 2017, pp. 7207-7222.
19. F. Ghavimi and H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: architectures, service requirements, challenges, and applications," *IEEE Communications Surveys and Tutorials*, Vol. 17, 2015, pp. 525-549.
20. M. K. Giluka, N. Rajoria, A. C. Kulkarni, V. Sathya, and B. R. Tamma, "Class based dynamic priority scheduling for uplink to support M2M communications in LTE," in *Proceedings of IEEE World Forum on Internet of Things*, 2014, pp. 313-317.
21. H. Chao and Y. Wang, "Contention-based uplink small data transmission for massive MTC in 5G," in *Proceedings of IEEE 28th Annual International Symposium on Personal*, *Indoor*, *and Mobile Radio Communications*, 2017, pp. 1-6.
22. S. Han, X. Xu, Z. Liu, P. Xiao, K. Moessner, X. Tao, and P. Zhang, "Energy-efficient short packet communications for uplink NOMA-based massive MTC networks," *IEEE Transactions on Vehicular Technology*, Vol. 68, 2019, pp. 12066-12078.
23. L. Miuccio, D. Panno, and S. Riolo, "Dynamic uplink resource dimensioning for massive MTC in 5G networks based on SCMA," in *Proceedings of the 25th European Wireless Conference*, 2019, pp. 108-113.
24. Z. N. Lin, S. R. Yang, and P. Lin, "Edge computing-enhanced uplink scheduling for energy-constrained cellular Internet of Things," in *Proceedings of the 15th International Wireless Communications and Mobile Computing Conference*, 2019, pp. 1391-1396.

Chung-Nan Lee (李宗南) received the B.S. and the M.S. degrees, both in Electrical Engineering, from National Cheng Kung University, Tainan, Taiwan, in 1980 and 1982, respectively, and the Ph.D. degree in Electrical Engineering from the University of Washington, Seattle, in 1992. Since 1992, he has been with National Sun Yat-sen University, Kaohsiung, Taiwan, where he was an Associate Professor in the Department of Computer Science and Engineering from 1992 to 1999; he was the Chairman of the Department of Computer Science and Engineering from August 1999 to July, 2001; Director of Wireless Broadband Communication Protocol Cross-Campus Research Center from 2013 to 2017; President of Taiwan Association of Cloud Computing from 2015 to 2017; currently, he is a Distinguished Professor and Director of Cloud Computing Research Center; Board of Governor of APSIPA from 2017. In 2016 he received the outstanding engineering professor from Chinese Institute of Engineers, Taiwan. His current research interests include multimedia over wireless networks, cloud computing, and evolutionary computing.

Chun-Yen Tseng (曾俊彥) received the M.Sc. degree in Computer Engineering from National Sun Yat-sen University in 2018. His research interests include the 5G wireless networks, machine-to-machine communications, and radio resource allocation. He is currently working in Industrial Technology Research Institute.

Ming-Feng Lee (李名峰) received the Ph.D. degree in Computer Science from the University of Bristol, United Kingdom, in 2014. He is currently a Postdoctoral Fellow in Cloud Computing Research Center, National Sun Yat-sen University, Taiwan. His current research interests include privacy, communication security and 5G security.