UNIVERSITY of York

This is a repository copy of *Priority-based learning automata in Q-learning random access scheme for cellular M2M communications*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/176817/

Version: Published Version

Article:
Shinkafi, N, Mohammed Bello, Lawal, Shu-aibu, D et al. (1 more author) (2021) Priority-based learning automata in Q-learning random access scheme for cellular M2M communications. ETRI Journal. ISSN 1225-6463

https://doi.org/10.4218/etrij.2020-0091

Reuse
Other licence.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Received: 12 March 2020 | Revised: 18 July 2020 | Accepted: 15 October 2020

DOI: 10.4218/etrij.2020-0091

ORIGINAL ARTICLE

ETRI Journal WILEY

Priority-based learning automata in Q-learning random access scheme for cellular M2M communications

Nasir A. Shinkafi[1] | **Lawal M. Bello**[1] | **Dahiru S. Shu'aibu**[1] | **Paul D. Mitchell**[2]

[1]Department of Electrical Engineering, Bayero University, Kano, Kano, Nigeria

[2]Department of Electronic Engineering, University of York, York, UK

Correspondence
Paul D. Mitchell, Department of Electronic Engineering, University of York, Heslington, York, UK.
Email: paul.mitchell@york.ac.uk

Abstract

This paper applies learning automata to improve the performance of a Q-learning based random access channel (QL-RACH) scheme in a cellular machine-to-machine (M2M) communication system. A prioritized learning automata QL-RACH (PLA-QL-RACH) access scheme is proposed. The scheme employs a prioritized learning automata technique to improve the throughput performance by minimizing the level of interaction and collision of M2M devices with human-to-human devices sharing the RACH of a cellular system. In addition, this scheme eliminates the excessive punishment suffered by the M2M devices by controlling the administration of a penalty. Simulation results show that the proposed PLA-QL-RACH scheme improves the RACH throughput by approximately 82% and reduces access delay by 79% with faster learning convergence when compared with QL-RACH.

KEYWORDS

Learning automata, LTE network, machine to machine, Q-learning, RACH congestion

1 | INTRODUCTION

Machine-to-Machine (M2M) communication, also called machine-type communication (MTC), is predicted to be one of the major applications of current and future cellular communications [1,2]. M2M communication, as defined in [3], enables communication between various devices without or with limited human intervention. Different devices such as sensors, actuators, meters, and radio frequency tags are used as M2M devices to read the status of machines and share information, either on a wireless network, wired network, or a hybrid of both to a target destination [4,5]. M2M devices are an important part of the emerging "Internet of Things" and "Smart City" paradigms [6,7], which are expected to provide solutions to current and future socioeconomic demands. In addition, M2M devices engender new applications in areas such as building and industrial automation, remote and mobile healthcare, and many more, as described in [8]. According to [9], the number of M2M devices is expected to significantly outnumber the world population [10]. This creates a significant gap and makes it practically impossible for humans to control them. Therefore, there is a need for these devices to autonomously interact among themselves.

The envisaged growth of M2M applications has led to many research studies on protocols and products oriented to support M2M services. The 6LowPAN protocol suite is a popular technology for low-power devices [11], the IEEE 802.15.4 standard is used for low-bit rate short-range transmission [12], and Zigbee (which utilizes the 802.15.14 standard) is for M2M device interconnection in short-range

This is an Open Access article distributed under the term of Korea Open Government License (KOGL) Type 4: Source Indication + Commercial Use Prohibition + Change Prohibition (http://www.kogl.or.kr/info/licenseTypeEn.do).
1225-6463/$ © 2021 ETRI

wireless sensor networks [13]. Additional studies that could add value to M2M applications with communication protocols include the European Installation Bus/Konnex, Local Operating Network, and Building Automation and Control Network for home automation [14]. Most of the existing communication access protocols or techniques are incapable of fulfilling the demand for ubiquitous access. Although short-range network solutions, such as Zigbee/6LoWPAN or the IEEE 802.11ah extension for M2M communications support the interconnection of M2M devices in the same local area, there is a need for a long-range network to provide end-to-end communications [15]. A ubiquitous radio technology that can provide wide coverage with energy efficiency, minimal cost per bit, and low latency is what M2M communication needs. Cellular networks with their existing infrastructure, capacity, and ubiquity have all the necessary requirements to enable M2M long-range communications [16]. Current cellular network technologies will not be able to accommodate the projected growth of M2M traffic, as they have been primarily designed to support human-to-human (H2H) traffic.

A cellular system has been designed primarily for H2H devices that have significant data transfer requirements, whereas M2M communication typically corresponds to a large number of devices that require sporadic transmission of short packets. Heavy M2M traffic occurs when many such devices are activated simultaneously, generating many instantaneous attempts to access the cellular network through the initial signaling random access channel (RACH), which leads to its overload and congestion [17]. As a result of the heavy traffic generated by M2M devices, the current RACH access approach is not sufficient [16,17]. This is recognized as a major challenge for wireless cellular systems, and it needs to be addressed to support significant M2M traffic without impacting H2H communication services. There is a need to design an effective RACH access technique to overcome these challenges to accommodate additional M2M traffic in cellular M2M communication. Numerous RACH access protocols have been proposed to address these challenges. Prominent among them are the reinforcement learning-based techniques such as priority-based learning automata (PLA) [18] and Q-learning RACH (QL-RACH) [19], with associated modification schemes presented in [20–22].

A PLA scheme is proposed in this paper to improve the performance of the QL-RACH scheme [19]. The scheme is called PLA-QL-RACH and uses a learning automata (LA) technique to improve RACH throughput performance [19] by minimizing interaction and collision among M2M devices or with H2H devices sharing the RACH resources. This scheme also eliminates the excessive punishment suffered by M2M devices by controlling the administration of a penalty factor applied in [19]. Simulations were undertaken to assess the performance of PLA-QL-RACH compared with the existing schemes. The results show that the PLA-QL-RACH scheme significantly improves the overall RACH throughput and reduces the access delay through faster learning convergence.

The remainder of this paper is structured as follows. Section 2 summarizes the related research. The system model is introduced in Section 3, and the proposed PLA-QL-RACH scheme is described in Section 4. Section 5 provides a detailed performance evaluation, and the paper is concluded in Section 6.

2 | RELATED WORK

A number of techniques have been proposed to deal with the RACH overload challenges when M2M devices coexist with H2H devices in cellular networks. These are either reinforcement learning (RL) based or non-RL based, which includes separating M2M and H2H users in the RACH contest by allocating separate RACH resources, access class barring (ACB), MTC-specific back-off, and pull-based techniques. This section provides a summary of RACH access scheme research to support the co-existence of M2M and H2H traffic over cellular networks using RL and non-RL-based approaches.

A self-optimizing overload control (SOOC) scheme is outlined in [23] to handle the physical RACH (PRACH) overload using resource separation. The scheme uses a mechanism that collects and monitors information on RACH overload at each random access (RA) cycle. Accordingly, long-term evolution (LTE) is structured in such a way that an evolved node B (eNB) adapts the number of RA slots within the RA cycles. The M2M device enters an overload control mode when it does not manage to secure an RA slot during the first attempt. To regulate RA retries following collision, a classic p-persistent mechanism is applied in this mode. The scheme also adds high- and low-priority access classes for time-tolerant and time-sensitive M2M devices. It sets different p values depending on the device access class. The SOOC protocol monitors the congestion level, making the eNB react dynamically by adjusting the number of PRACH RA slots within successive cycles, thereby maintaining a target maximum collision probability for the system. The SOOC scheme handles high-traffic load situations for two-time dependent priority classes but suffers RACH congestion/overload when a large number of M2M devices with diverse priority classes are considered. To handle the overhead generated from a large number of M2M devices, Taleb and Kunz [24] proposed a bulk M2M signaling scheme as a resolution mechanism for congestion/overload. The proposal worked on the assumption that M2M signaling messages are moderately delay tolerant. This makes it feasible to minimize overheads at the eNB by exploiting bulk processing and aggregating signaling data from M2M devices before forwarding them to the core network. The scheme efficiently

handles the traffic generated by a large number of channel access requests. However, it is restricted to only M2M traffic without considering that of H2H in sharing the RACH. In an attempt to improve [24], a slotted access scheme was proposed in [25] to provide RA cycle requests for M2M. The scheme employs RA slots for dedicated access, reserved for each M2M device and accessed in a collision-free manner. The reserved slots for every M2M device are generated from the International Mobile Subscriber Identity, which uniquely identifies each device, while the cycle parameter is broadcasted by the eNB within each cycle. The scheme protects H2H devices from the impact of M2M access, but may result in delays due to the dedicated RA slots used for each M2M device. This further creates access collisions, and the scheme is not efficient for heavily delay-constrained M2M applications. To minimize the probability of collision and average access delay for a large number of fixed M2M devices, an RA scheme for fixed-location M2M communication was proposed in [26]. This RA scheme exploits the resource allocation procedure in terms of fixed uplink timing alignment (TA) between the devices and the eNB according to five RA steps. The process is similar to the traditional LTE RA procedure except in step 3, where TA information is used to lower the probability of collision during transmission of Message 3. The TA value of a static M2M device is assumed to remain constant over time. In addition, once the TA received from the eNB in Message 2 varies from that of the M2M device, there is a high probability that Message 2 is meant for a distinct M2M device that is transmitted on the same PRACH. The M2M device evades transmission of Message 3 in step 3, which minimizes the probability of collision at step 4 and, in turn, the access delay. The scheme minimizes the likelihood of collision and access delay; however, it is partial in resource allocation owing to the failure to transmit Message 3, which can bring about a rise in access delay and poor quality of service (QoS) performance.

A reinforcement learning-based eNB selection algorithm (Q-learning) to reduce access delay was proposed in [27]. The RL-based algorithm allows M2M devices to select an eNB in a self-organized fashion. The algorithm yields a lower access delay when compared to random eNB selection. However, the algorithm does not take throughput into account when determining the QoS performance. In [28], a game theoretic scheme was proposed to enhance system throughput in an RACH overload scenario. RA resources are organized into three groups: for H2H, M2M, and hybrid usage. Different RACH preamble pools are earmarked as RH, RM, and RB where RH is the preamble reserved for H2H usage, RM is for M2M usage, and RB is for both H2H and M2M usage. The M2M devices pull the preamble either in the M2M-dedicated pool, in the shared one, or remain silent with a probability distribution that is determined based on the outcome of a game. The scheme attains an improved system throughput for both M2M and H2H devices, but at the expense of a high-access delay. To minimize congestion and high-access delay, the Fast Adaptive Slotted ALOHA (FASA) scheme was developed in [29] as an appropriate option for RA control of event-driven M2M communications. Slots are considered to have various states: idle, successful, or collided. The scheme employs these states to accelerate the process of tracking the network status by adjusting the transmission probability of a p-persistent Slotted ALOHA (s-ALOHA) system with the aim of estimating the number of active devices in a slot. The FASA scheme is shown to be an effective and stable s-ALOHA scheme suitable for event-driven M2M communications and other systems characterized by bursty traffic. However, the scheme also suffers from high-access delay and congestion when different classes of M2M devices communicate with different probabilities, thereby lowering system throughput because it is limited to event-driven M2M communications.

In [19], a Q-learning-based RACH (QL-RACH) access scheme was introduced to lower collisions among M2M devices. The QL-RACH scheme uses an intelligent slot assignment mechanism to avoid collisions between M2M devices. It allows M2M and H2H devices to share RACH resources. The devices are categorized into two groups: learning M2M and non-learning H2H. The learning M2M devices used the QL-RACH access scheme while the non-learning H2H devices maintained the conventional s-ALOHA RACH (SA-RACH) scheme. The QL-RACH scheme significantly reduces collisions between the M2M devices, but the RACH throughput ultimately collapses owing to collisions resulting from the disturbance coming from the uncontrolled H2H traffic at high-load levels. In addition, slots may be wasted when the mean RACH request rate is higher than M2M frame time. Furthermore, since every M2M device maintains a Q-value for each slot in the M2M frame to record transmission history in consecutive frames, the mechanism is energy inefficient for battery-limited M2M devices.

A frame-based back-off QL-RACH (FB-QL-RACH) scheme was proposed as a modification of QL-RACH in [20]. The scheme lowers the probability of collision between H2H and M2M devices when sharing the same frame for both the initial access and the back-off. The scheme also minimizes the slot wastage introduced by the M2M back-off in the QL-RACH scheme. The scheme enhances the RACH throughput performance of QL-RACH because the effect of M2M back-off is eliminated. The challenges presented by RL-based and non-RL-based schemes were resolved by [18], where a priority-based adaptive access barring scheme for M2M communications in LTE networks was developed using LA to support different M2M priority classes during the resource allocation procedure. The scheme dynamically assigns RA resources to different M2M device classes based on specific priorities and demands. In addition, the scheme

fine-tunes the barring factor for each class to control the possible overload. This scheme minimizes access delay and resource wastage, but causes poor QoS when considering both H2H and M2M devices. To resolve the challenges brought about by the effect of the penalty factor in QL-RACH, LA was used to classify M2M according to QoS classes, thereby producing an LA-based QL-RACH (LA-QL-RACH) access scheme for cellular M2M communications, as discussed in [22]. The scheme includes a mechanism to remove the excessive punishment experienced by M2M devices by regulating the use of a penalty factor in the QL-RACH scheme. It classifies M2M devices according to three QoS classes and assigns RACH resources on demand. The classification minimizes the level of interaction and collision between the M2M and H2H devices without forcing the M2M into another Q-learning process, thereby resolving the problem of the disturbance from the non-learning H2H devices. Although the scheme enhanced RACH throughput and reduced the end-to-end access delay, it was restricted to the control of Q-learning penalty administration without prioritizing the cellular M2M traffic.

None of the papers mentioned have employed a combination of PLA and Q-learning to improve the QoS performance of cellular-based M2M. Therefore, in this study, PLA and Q-learning are used together to develop a PLA approach to enhance the Q-learning random access scheme (PLA-QL-RACH) for cellular M2M communications. This is possible by employing a PLA technique to improve the RACH throughput performance of QL-RACH and eliminate the excessive punishment suffered by M2M devices.

3 | SYSTEM MODEL

3.1 | RA procedure in LTE

In LTE, the first step of the RA procedure is for the user equipment (UE) to connect to the network through the RACH in an uplink transmission mode. The RA procedure is performed either in a contention-free or contention-based manner. In a contention-free scenario, the eNB assigns a unique preamble to a particular user, guaranteeing its access to the network, as in the case of handover. In contrast, in a contention-based approach, the individual UE initiates the access request. The contention-based mode of the RA procedure is the most appropriate for cellular M2M communication. The preambles are randomly selected by the users through the RA slots according to the four RA steps, as presented in [16,17].

The structure of a cell in LTE consists of up to 64 assigned preambles, some of which are reserved for contention-free access while the remaining are made available for contention-based RA [17]. The LTE frame structures and modes of operation are described in detail in [16]. RACH collision occurs if one preamble is selected simultaneously by more than one M2M device in the same RA slot [16].

3.2 | Q-learning and LA

Q-learning is an off-policy or model-free RL algorithm that seeks to acquire a policy that maximizes the whole reward [21]. The algorithm searches for the optimal action to take at any given instant. It is considered to be off policy because the Q-learning function learns from actions that are outside the present policy. It is viewed as model free because it does not require a model of the environment, and it can handle problems with stochastic transitions and rewards, without requiring adaptation [21]. LA is equally a RL model that is employed in many applications that involve adaptive cognitive processes. It is seen as a self-operating learning model with the power to work in an environment with unknown characteristics. LA is analogous to an automaton that enhances its functionality by obtaining knowledge of the behavior of the random environment [30]. It employs the knowledge gained previously for future cognitive processes. The response of the environment to the chosen LA action comes as feedback, which is either a reward or penalty. With the help of the feedback, the choice of probability of subsequent actions is updated. P-Model LA is employed in this work which includes a set of environmental responses that take only the binary values of 1 and 0, respectively, for penalty and reward [18,22].

3.3 | LA-QL system model

We assume that one RA slot occurs in a cycle and 50 preambles are earmarked in each RA slot for use by the three priority classes. Additionally, the M2M devices are presumed to be spread within an eNB coverage area in a unit cell of an LTE network, each having applications with different priorities and QoS requirements, as shown in Figure 1. Each M2M device is initiated within the interval $[0, \tau_s]$ with probability that of beta distribution, as presented in [31].

As presented in [30], LA is proven to be effective in guaranteeing adaptation to systems operating in environments with changing or unknown characteristics. The adaptation feature is used in our simulation, as indicated in Figure 1, where a number of M2M devices are made to contend in an RA cycle.

The quantity of contending M2M devices in each cycle is unknown and depends on the stochastic arrival process of RA requests of the UE. However, in this work, the UEs that represent M2M in this work attempt to access the network based on priorities and demands for uplink resources [32,33]. PLA is the proposed scheme developed in this work, which classifies

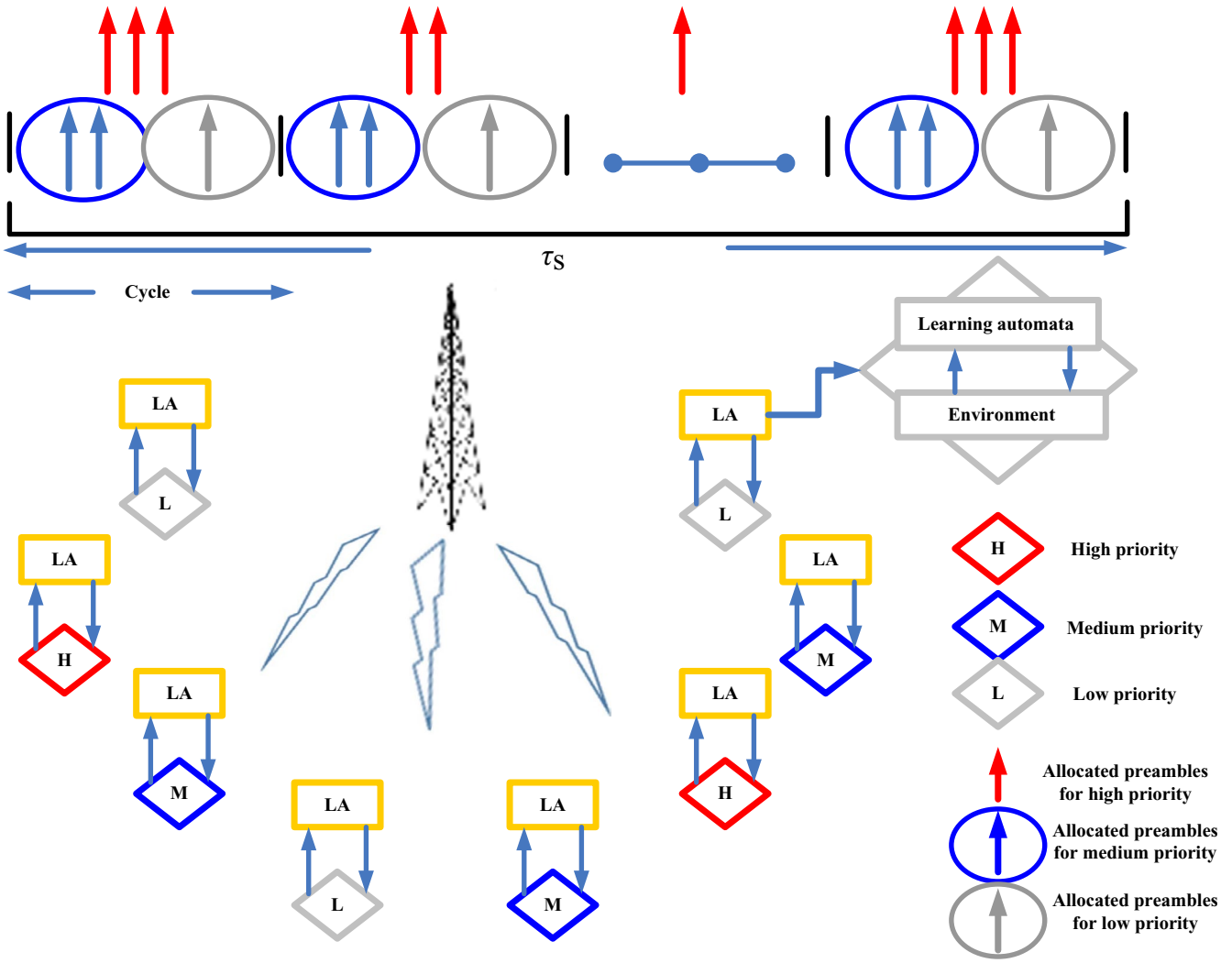

FIGURE 1 Machine-to-machine devices with different priorities in LTE networks (adopted from [18])

M2M devices in line with three priority classes: *High (H)*, *Medium (M)*, and *Low (L)*, where $x \in \{H, M, L\}$. The scheme fine-tunes the RACH resource allocation probability ($\eta_x(t)$), LA feedback ($c_x(t)$), and ACB parameter ($\alpha_x(t)$) for each priority class x and within the tth LA cycle. The parameters are adjusted to control any possible overload or collision for a particular priority class when the number of M2M devices contending for access from that class is higher or lower than the resources allocated. The number of contending M2M devices in each priority class is tracked, and RACH resources are allocated according to prioritization rules.
The rules are as follows:

- Each priority class uses a certain quantity of available resources, which is determined based on its priority class and average requirement.
- The unused resources that have already been allocated to a particular priority class are proportionally allocated to other priority classes demanding resources.

According to the rules, the steady-state performance of the technique is achieved as follows:

1. When the quantity of M2M devices demanding access from a priority class is below the maximum RACH resources available for that class, then the number of allocated preambles to class x, ($M_x(t)$), is obtained as follows:

$$M_x(t) = k_x(t), \tag{1}$$

where $k_x(t)$ is the number of M2M devices for priority class x.

2. When the quantity of M2M devices contending for access from a priority class is above the maximum RACH resources available for that class, then all the resources are used by the devices belonging to this class and the ACB parameter is adjusted as follows:

$$\alpha_x(t) = \frac{M_x(t)}{k_x(t)}. \quad (2)$$

Each M2M device from priority class x participates in the RA procedure according to (2) and randomly selects a preamble with probability computed as

$$P = \frac{1}{M_x(t)}. \quad (3)$$

The probability that a definite preamble is picked by an M2M device from priority class x is given by

$$P_x^m(t) = \frac{\alpha_{x(t)}}{M_x(t)}. \quad (4)$$

The operation of the proposed PLA-QL-RACH scheme, which is based on the model presented above, is provided in detail in Section 4.

4 | PLA-QL-RACH ACCESS SCHEME

In this section, a modification of the QL-RACH scheme, called PLA-QL-RACH, is described. To provide an appropriate context, the shortcomings of the QL-RACH scheme are initially presented. QL-RACH uses Q-learning to regulate M2M devices while coexisting with H2H devices in sharing the RACH channel of a cellular network. These devices are classified into two groups: learning M2M and non-learning H2H. Coexisting together in a combined RACH access scheme, the learning M2M devices go through the QL-RACH access scheme while the non-learning H2H devices retain the conventional s-ALOHA RACH (SA-RACH) access scheme. The learning was realized by designing a virtual M2M frame with a size equal to the number of M2M devices. The random effect of H2H traffic as it approaches the s-ALOHA capacity hinders the performance of the QL-RACH scheme. As the H2H traffic load approaches the s-ALOHA capacity, the probability of collisions between the H2H and M2M devices increases, leading to RACH throughput collapse. The collisions are caused by the failure to prioritize M2M traffic when coexisting with H2H traffic, and excessive reward and punishment.

To address the aforementioned problem, a PLA-QL-RACH access scheme is proposed. First, the technique considers the probability that a preamble remains idle, as given by:

$$P_x^{\text{idle}}(t) = \left(1 - P_x^m(t)\right)^{k_x(t)}, \quad (5)$$

and the probability that the preamble is successfully used by a device is

$$P_x^{\text{succ}}(t) = \binom{k_x(t)}{1} P_x^m(t) \left(1 - P_x^m(t)\right)^{k_x(t)-1}, \quad (6)$$

while the probability that the preamble suffers collision is given by

$$P_x^{\text{coll}}(t) = 1 - k_x(t) P_x^m(t) \left(1 - P_x^m(t)\right)^{k_x(t)-1} - \left(1 - P_x^m(t)\right)^{k_x(t)}. \quad (7)$$

Information on the number of idle, successful, and collided preambles at the end of each LA cycle is provided to the eNB by the PLA technique. Although the eNB is not aware of the number of M2M devices demanding access from priority classes in each cycle, it is conscious of the access attempt from an M2M device based on the probability that an attempted preamble converges per state. The convergence of a preamble in the idle, successful, and collision states is compared with the maximum throughput achieved through s-ALOHA of e^{-1}, $2e^{-1}$, and $1 - 2e^{-1}$, respectively. This is achieved through the adjustment of $\eta_x(t)$ and $\alpha_x(t)$ [18,29], where feedback is also produced. The feedback, which is collision dependent, avoids the conventional RA attempt retrials, which leads to RACH overload and throughput collapse. Instead, it triggers a resource allocation procedure that guarantees a collision-free RA procedure and better throughput with lower delay. This behavior is further explained when the feedback is received by the LAs of all the activated M2M devices for each class and takes a binary value as reward or punishment. It can be presented in the form of an array as

$$c(t) = \left(c_{\text{H}}(t), c_{\text{M}}(t), c_{\text{L}}(t)\right), \quad (8)$$

where $c_{\text{H}}(t)$, $c_{\text{M}}(t)$, and $c_{\text{L}}(t)$ represent feedback for the high-, medium-, and low-priority classes, respectively. Within a $[0\ \tau_s]$ interval, each active M2M device transmits a small data packet to the eNB during the RA procedure. The activation interval τ_s is distributed into Z_s cycles having two identical parts: the first part is used for transmitting the preambles, and the second part for transmitting Message 3 of the RA procedure. At the end of each cycle, the eNB monitors P_x^{coll} for class x and is generated $c_x(t)$ by comparing it with the expected value of $g = 1 - 2e^{-1}$ [18,22], computed as

$$c_x(t) = \begin{cases} 0 & \text{if } P_x^{\text{coll}}(t) < g \\ 1 & \text{if } P_x^{\text{coll}}(t) \geq g \end{cases}. \quad (9)$$

The eNB communicates the generated feedback $c_x(t)$ at the end of each cycle through the downlink broadcast channel. Whenever $P_x^{\text{coll}}(t) \geq g$, a unit feedback is produced to raise $\eta_x(t)$ and lower $\alpha_x(t)$ as new input to the PLA-QL-RACH

scheme. Furthermore, when $P_x^{\text{coll}}(t) < g$, a null feedback is generated, which results in a decrease in $\eta_x(t)$ and an increase in $\alpha_x(t)$. In steady-state conditions, when $P_x^{\text{coll}}(t) = g$, two scenarios that are both determined by the feedback occur:

1. The PLA-QL-RACH scheme stabilizes when $\eta_x(t)$ is updated as follows:

$$\eta_x(t+1) = \begin{cases} \eta_x(t) + \gamma\varepsilon_1 & \text{if } c_x(t) = 1 \\ \eta_x(t) - \gamma\varepsilon_2 & \text{if } c_x(t) = 0 \text{ and } \alpha_x(t) = 1 \end{cases}, \quad (10)$$

where γ is the Q-learning rate; $0 < \varepsilon_1 < \Omega_x - \eta_x(t)$; $0 < \varepsilon_2 < \eta_x(t) - \theta_1$; and θ_1 is a very small value that guarantees a positive non-zero percentage of resources allocated per class even when that class has no access request. Moreover, Ω_x is the maximum value of $\eta_x(t)$, which is statically assigned by the eNB. The eNB communicates this value to the M2M devices at the beginning of the activation interval through the system information blocks (SIBs). Furthermore, $\alpha_x(t)$ is updated as follows:

$$\alpha_x(t+1) = \begin{cases} \alpha_x(t) + \gamma\varepsilon_1 & \text{if } c_x(t) = 0 \\ \alpha_x(t) - \gamma\varepsilon_2 & \text{if } c_x(t) = 1 \text{ and } \eta_x(t) = \Omega_x \end{cases}, \quad (11)$$

where θ_2 is a small value and ε_1 and ε_2 are the LA learning variables that are chosen in such a way that $\eta_x(t)$ and $\alpha_x(t)$ converge to the optimal value asymptotically. The values of $\gamma\varepsilon_1$ and $\gamma\varepsilon_2$ respectively determine the estimation accuracy and the convergence speed of the automaton, and hence the stability of the PLA-QL-RACH scheme.

2. The outcome of the LA feedback determines the penalty factor ($R(t)$) in the QL-RACH scheme to regulate the Q-learning punishment technique at time t, as follows:

$$R(t) = \begin{cases} +1 & \text{if } c_x(t) = 0 \\ -1 & \text{if } c_x(t) = 1 \end{cases}. \quad (12)$$

The second scenario is necessary to eliminate the chances of pushing the M2M devices into another Q-learning process using the updated Q-value from QL-RACH, as follows:

$$Q' = (1-\gamma)\,Q + \gamma c_x'(t). \quad (13)$$

with

$$c_x'(t) = R(t), \quad (14)$$

where $c_x'(t)$ is the steady-state LA feedback.

The prioritization of the M2M traffic by the PLA technique eliminates the RACH collisions and controls the QL-RACH reward and punishment technique using Algorithm 1.

Algorithm 1 PLA-QL-RACH algorithm implementation on collided M2M devices while coexisting with H2H devices during a RACH contest. M2M, machine-to-machine; H2H, human-to-human; RACH, random access channel; ACB, access class barring; QL, Q-learning

for every device RACH contest **do**
  Route H2H via SA-RACH
  Route M2M via QL-RACH
end for
if M2M collision occurs in QL-RACH **then**
    Classify the M2M devices **else**
    Route the collided devices via PLA-QL-RACH according to their classes
end if
for every M2M device using PLA-QL-RACH to contest RACH resources **do**
Calculate probability of collision ($P_x^{coll}(t)$) and compare it with the expected value g
  Calculate steady state LA feedback ($c_x'(t)$) value of 0 / 1
end for
If probability of collision is less than the expected value and LA feedback is 0, **then**
Calculate ACB parameter ($\alpha_x(t)$)
Decrease RACH resource allocation probability ($\eta_x(t)$) using (11) when ACB parameter is 1.
  Update ACB parameter using (11)
  Reward QL penalty factor (R(t)) by 1 using (13)
else if LA feedback is 1, **then**
    Calculate ACB parameter ($\alpha_x(t)$)
Update RACH resource allocation probability ($\eta_x(t)$) using (10)
Decease ACB parameter using (11) when RACH resource allocation probability reaches maximum value (Ω_x)
    Penalize QL penalty factor (R(t)) by −1 using (13)
end else if
end if

A flow chart of the algorithm is presented in Figure 2.

5 | PERFORMANCE EVALUATION

5.1 | Simulation scenario

Simulation was used to assess the performance of the PLA-QL-RACH scheme using MATLAB. The scheme, which is a modification of the QL-RACH scheme, was evaluated using the same simulation parameters as in [19].

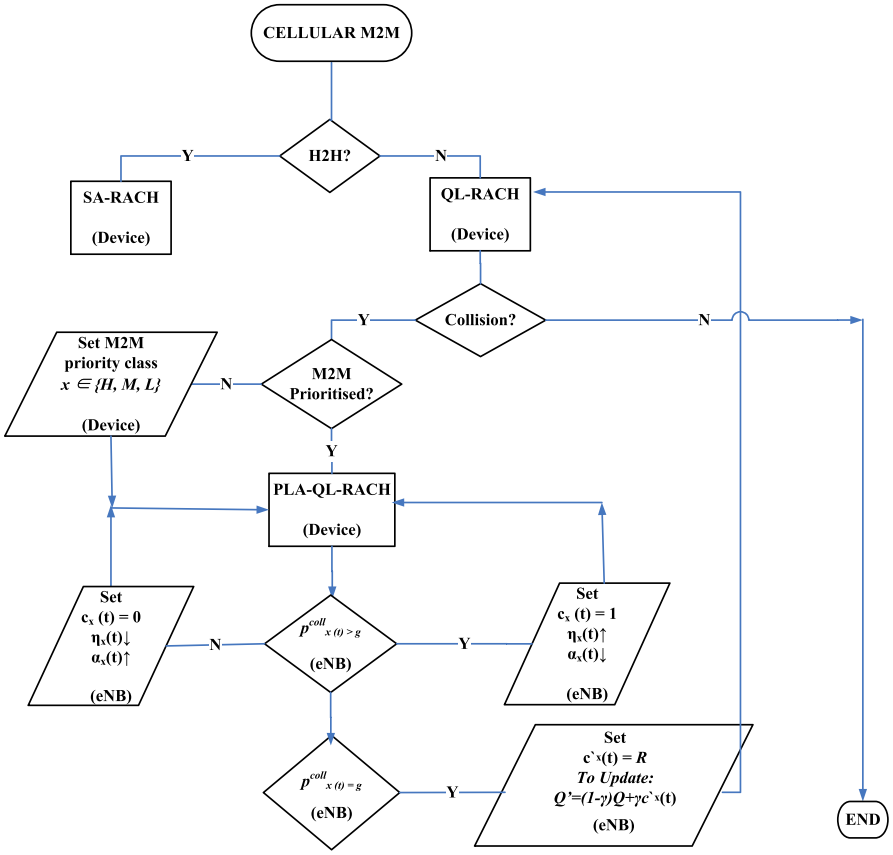

FIGURE 2 Proposed PLA-QL-RACH scheme flow chart

A fixed allocation of resources was implemented at this stage such that the values of Ω_H, Ω_M, and Ω_L were set to 0.5, 0.3, and 0.2, as 50%, 30%, and 20% ratios, respectively. In the fixed allocation approach, a fixed number of preambles are pre-allocated to each class statically by the eNB according to the priority, and the average number of M2M devices attempting to access RACH resources in that class is within a τ_s interval. The choice of preamble allocation ratios for the three traffic classes is restricted by the fact that they should sum to 1 and provide effective prioritization of resources among the three classes.

TABLE 1 Simulation parameters

Parameter	Value
PRACH configuration index	12
RA slot period	1 ms, 1 cycle
1 RA slot	50 preambles
Preamble format duration	1 ms
Back-off period/AC-Barring Time	28 ms
Number of allowed retransmissions	7
RACH allocation probability (Ω_H, Ω_M, Ω_L)	0.5, 0.3, 0.2
Learning rate	0.01

5.2 | Simulation parameters

Table 1 present details of the parameters used in this simulation, based on the LTE standard.

In Table 1, the PRACH configuration index of 12 was selected to determine the PRACH preamble type and PRACH preamble timing. The index also shows which frame and sub-frame M2M devices are permitted to transmit a PRACH preamble. In each frame of 10 ms, there are 10 sub-fames of 1 ms each, and each sub-frame has two slots of 0.5 ms. In addition, a learning rate of 0.01, which determines the speed of the convergence of the QL-RACH was set to ensure that it is within the same low value as the penalty factor. Additionally, an ACB time (ac-Barring time) of 28 ms was used as the back-off period, which indicates when retransmission will occur after collision has occurred. The values of the RACH allocation probability (η_x) were selected as the ratio of the pre-allocated preambles per class x for use by all M2M devices.

5.3 | Simulation results and discussion

In this section, the performance of the proposed PLA-QL-RACH scheme is evaluated along with five other RACH access schemes: SA-RACH [29], QL-RACH [19], FB-QL-RACH [20], Framed-ALOHA for QL-RACH (FA-QL-RACH) [21], and LA-QL-RACH [22]. The schemes are evaluated in terms of throughput and average access delay by means of simulation. The reporting procedure used by [19]

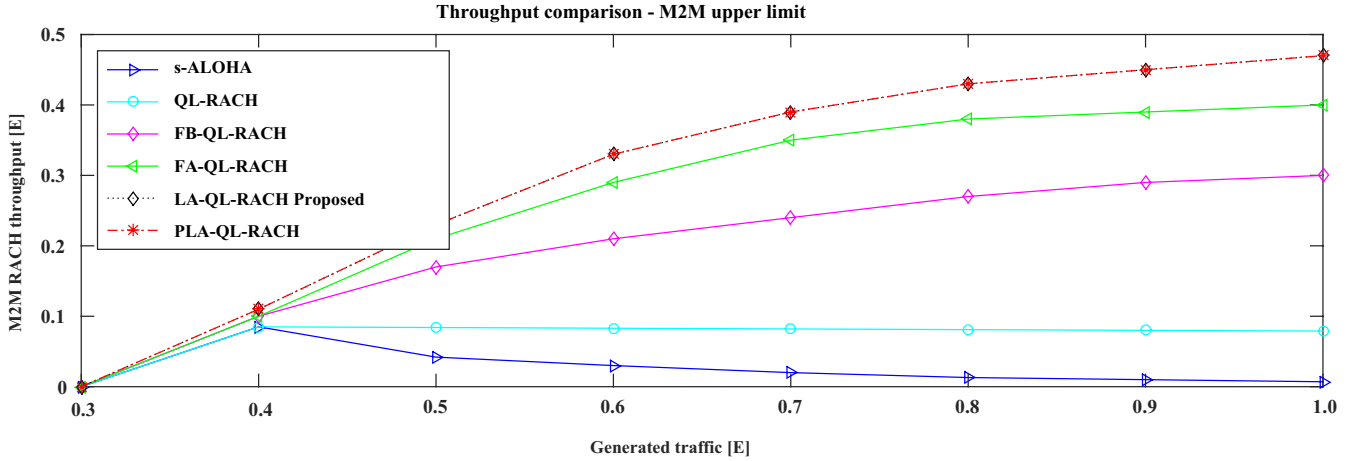

FIGURE 3 Proposed PLA-QL-RACH throughput comparison—M2M lower limit (H2H = 0.1E)

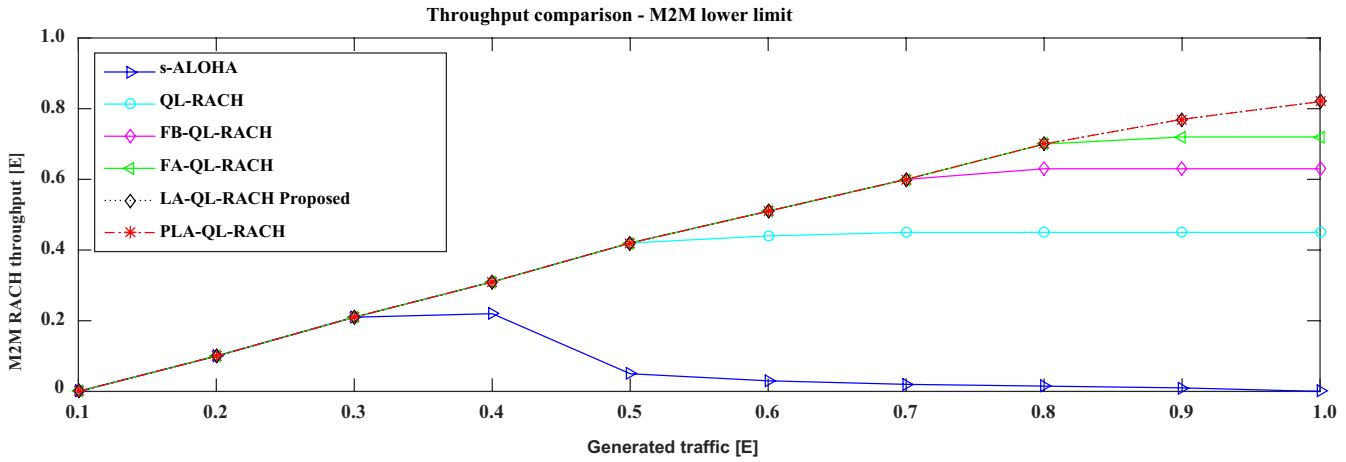

FIGURE 4 Proposed PLA-QL-RACH throughput comparison—M2M upper limit (H2H = 0.3E)

was adopted in presenting our result. The procedure considers the s-ALOHA throughput capacity (e^{-1}) in Erlangs (E) at both the upper and lower limits as a threshold for traffic prediction. The Erlang is a unit of traffic density in a telecommunication system. One Erlang is the equivalent of one call (including call attempts and holding time) in a specific channel for 3600 s (in an hour). An upper limit of 0.3 E is selected because it is closer to the load limit, whereas a lower limit of 0.1 E is chosen as it is away from the load limit. These limits are assumed to be the average peak-hour load generated by H2H devices during their interaction with M2M devices and are used as a measure of RACH stability. The effect of the proposed PLA-QL-RACH scheme is shown by carefully assessing its delay performance and comparing its RACH throughput with the existing schemes. Figure 3 shows the RACH throughput performance of the proposed PLA-QL-RACH scheme at the M2M upper limit with H2H traffic fixed closer to the load limit (0.3 E). When the generated traffic is above the s-ALOHA capacity (0.368 E), it indicates that the overall total traffic comprises the fixed H2H traffic and a variable but additional M2M traffic load.

Figure 3 demonstrates that the schemes exhibit identical performance from 0.3 E to 0.4 E because the generated traffic is below the s-ALOHA capacity. As the generated traffic rises from 0.4 E to 1.0 E, the s-ALOHA RACH scheme performance starts to decline, whereas the QL-RACH scheme maintains its channel stability as it approaches the load limit. However, above the load limit, the proposed PLA-QL-RACH is on par with the LA-QL-RACH scheme but performs better than the other compared schemes in terms of the RACH throughput. Furthermore, Figure 3 demonstrates that the proposed scheme is stable at 1.0 E with 47% RACH throughput at steady state. Consequently, above the s-ALOHA capacity, the RACH throughput remains at 0.47 E for the proposed PLA-QL-RACH scheme, which is 17% higher than that of the FB-QL-RACH scheme and 7% better than the FA-QL-RACH scheme.

Figure 4 illustrates the throughput comparison of the schemes at the M2M lower limit. At the lower limit, the H2H traffic is set away from the load limit (0.1 E), which is well below the s-ALOHA capacity. Figure 4 shows that all six schemes exhibit similar behavior from 0.1 E to approximately 0.3 E because the generated traffic within this range is below the s-ALOHA capacity. As the generated traffic approaches 0.368 E (s-ALOHA capacity), the s-ALOHA RACH scheme starts to decline as it is unable to support additional traffic, unlike the other schemes that maintain identical behavior up to 0.5 E. The figure also shows that all the other schemes exhibit similar behavior from 0.5 E to 0.7 E, except the QL-RACH scheme, which falls at 0.6 E due to the random effect of the H2H traffic. Additionally, the figure demonstrates that the behavior of the FB-QL-RACH, the LA-QL-RACH, and the proposed PLA-QL-RACH schemes are similar from 0.7 E to 0.8 E. However, from 0.8 E to 0.9 E, the proposed scheme remains on par with the LA-QL-RACH but outperforms the FB-QL-RACH scheme owing to the prioritization of M2M traffic. Hence, below the s-ALOHA capacity limit, the RACH throughput sits at 1.0 E, for PLA-QL-RACH and is 10% higher than that of FB-QL-RACH and 19% better than the FA-QL-RACH scheme. The recorded enhancement in the throughput performance of the PLA-QL-RACH scheme results from the influence of the prioritization (PLA) technique on QL-RACH.

Figure 5 shows the RACH throughput comparison per priority class against the M2M upper limit with the H2H traffic set closer to the load limit. The analysis of the RACH throughput per priority class illustrates the performance of each class, how it responds to the fixed allocation of resources, and how fast the scheme converges to steady state. In Figure 5, it can be seen that the RACH throughput performance at 1.0 E for the proposed scheme is 19.7% higher for the H priority class, 11.9% higher for the M priority class, and 8.1% higher for the L priority class when compared with the LA-QL-RACH

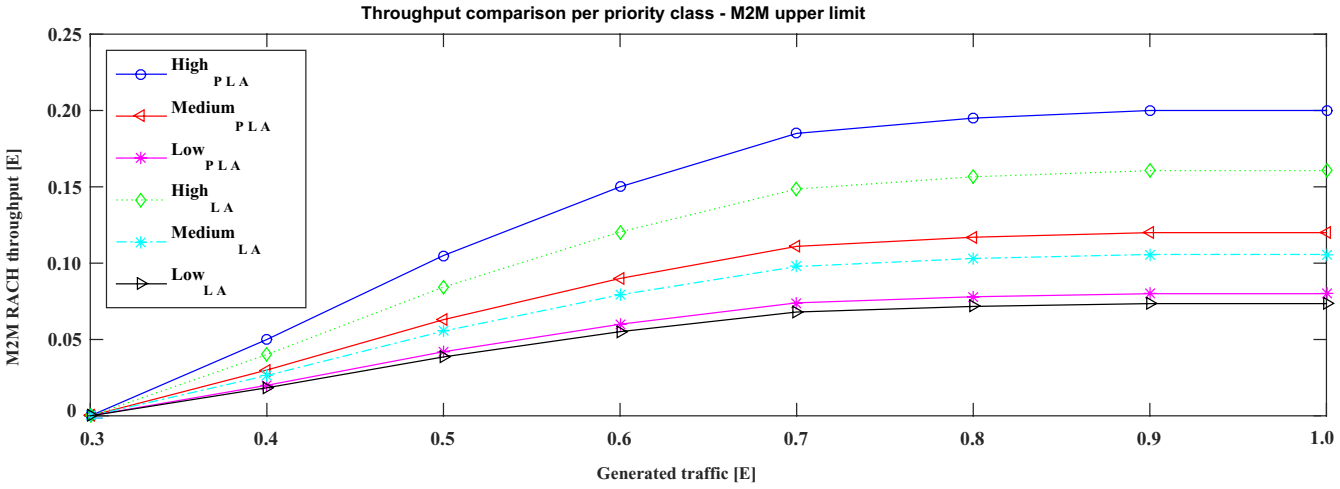

FIGURE 5 Proposed PLA-QL-RACH throughput comparison per priority class—M2M upper limit

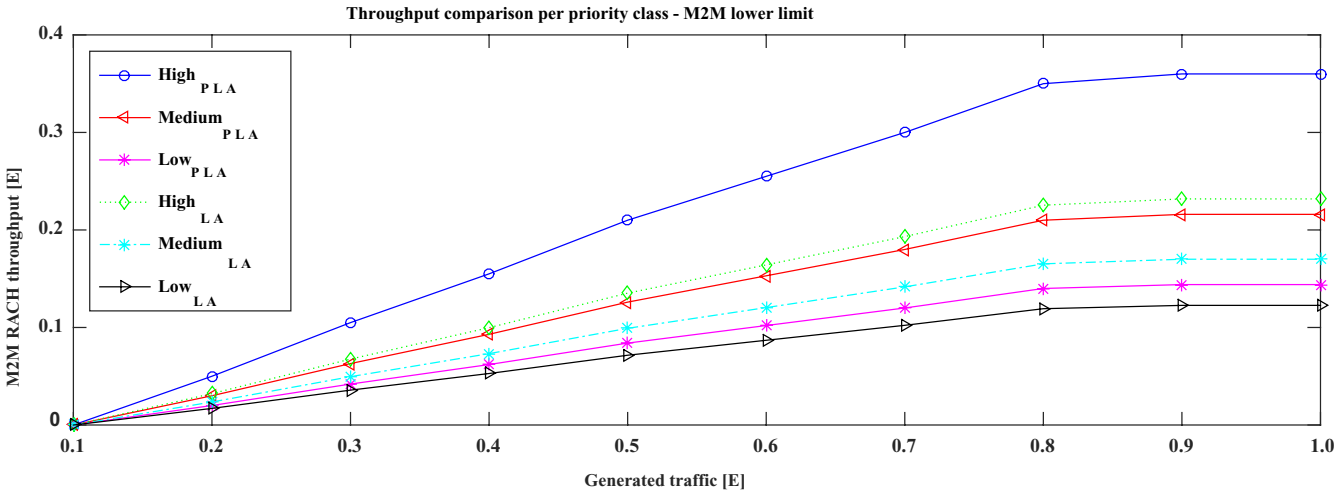

FIGURE 6 Proposed PLA-QL-RACH throughput comparison per priority class—M2M lower limit

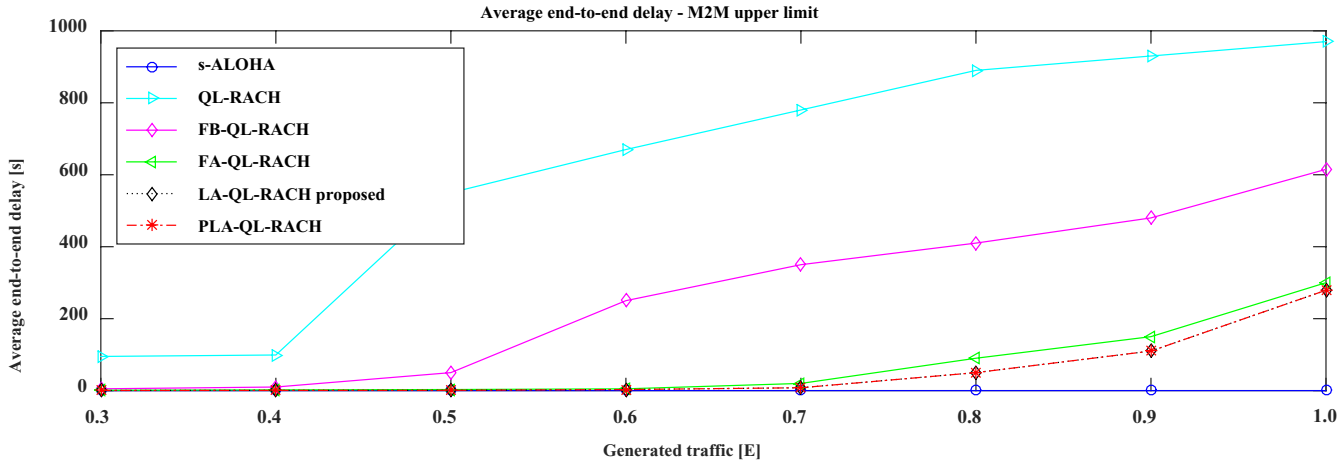

FIGURE 7 Proposed PLA-QL-RACH average end-to-end delay comparison—M2M upper limit

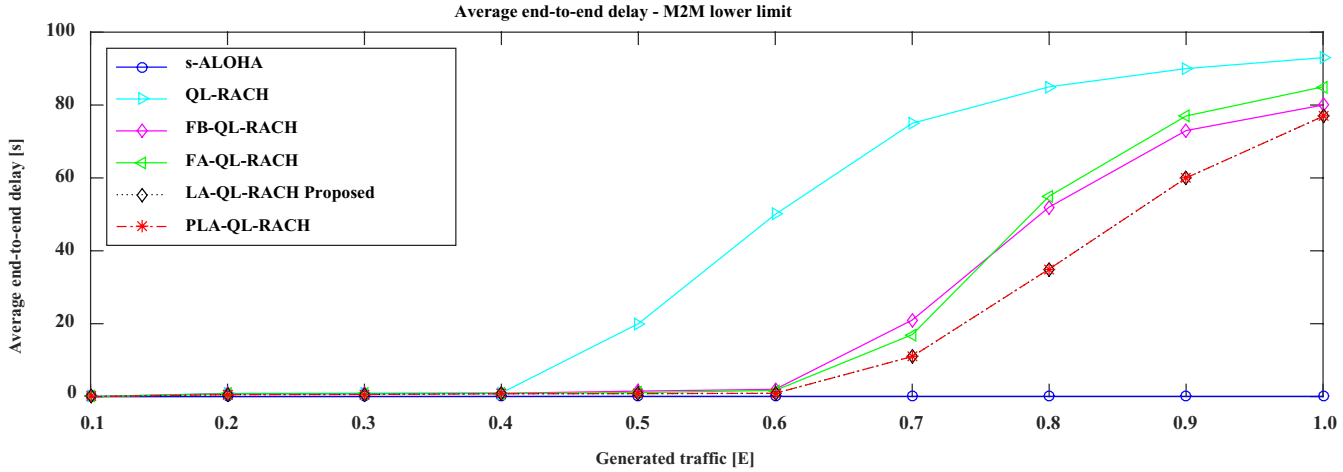

FIGURE 8 Proposed PLA-QL-RACH average end-to-end delay comparison—M2M lower limit (H2H = 0.1E)

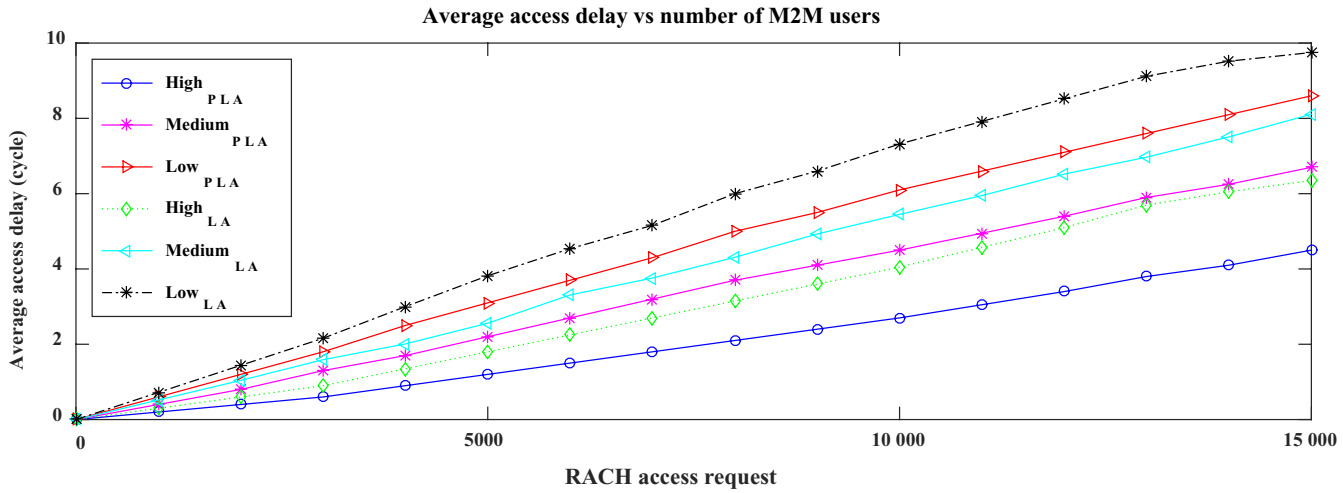

FIGURE 9 Proposed PLA-QL-RACH average end-to-end delay comparison per device RACH access request

scheme. This improvement was recorded because of the impact of the PLA technique employed by the proposed scheme, which enhances the speed of learning convergence over LA-QL-RACH, which was limited to categorizing M2M devices without prioritizing them. A similar trend is observed in Figure 6, which illustrates the RACH throughput comparison per priority class for the M2M lower limit when the H2H traffic is fixed away from the load limit.

Figure 6 indicates that when the generated traffic is below the s-ALOHA capacity, the RACH throughput performance of each priority class is proportional to the fixed allocated resources. It is shown that at 1.0 E, the throughput performance of the PLA-QL-RACH scheme is 35.6% higher for the H priority class, 21.3% higher for the M priority class, and 14.8% higher for the L priority class when compared with the LA-QL-RACH scheme. This specifies that the proportionate distribution and allocation of resources to the classes agrees with their respective priorities and QoS demand, which is possible because of the effect of collision elimination and penalty factor regulation on QL-RACH.

Figure 7 presents the average end-to-end delay comparison of the schemes at the M2M upper limit when the H2H traffic is set closer to the load limit (0.3 E). It is observed that the average end-to-end delay experienced by the proposed PLA-QL-RACH scheme is on par with the LA-QL-RACH scheme but 79% lower than that of FB-QL-RACH and 21% lower than the FA-QL-RACH scheme. Up to 0.6 E of generated traffic, the effect of the continuous regulation of the QL-RACH penalty factor is insignificant. However, as the generated traffic rises from 0.6 E to 1.0 E, the enhancement in the end-to-end delay of the proposed PLA-QL-RACH scheme compared to the previous schemes is witnessed. The enhancement is achieved by the continuous regulation of the QL-RACH penalty factor, which reduces the repetitive collisions experienced in the dedicated M2M slots.

Additionally, Figure 8 shows the average end-to-end delay comparison of the schemes at the M2M lower limit when the H2H traffic is fixed away from the load limit (0.1 E). It is observed that the proposed PLA-QL-RACH has a minimal impact on the LA-QL-RACH scheme but offers 20% lower delay than FB-QL-RACH and 22% lower delay than the FA-QL-RACH scheme. The average end-to-end delay performance of the schemes remains identical from 0.1 E to 0.5 E of the generated traffic when the H2H traffic is set below the s-ALOHA capacity. This is possible because the probability of collision is normally very low below the s-ALOHA capacity. As the generated traffic increases from 0.5 E to 1.0 E, the average end-to-end delay of the proposed PLA-QL-RACH scheme appears lower than all the previous schemes. This behavior is attained because of the absence of collisions resulting from the low level of contention at the lower levels of generated traffic, in which 8 out of 10 M2M transmission attempts are successful. Furthermore, the average end-to-end delay comparison per device RACH access request for the proposed PLA-QL-RACH and LA-QL-RACH schemes is illustrated in Figure 9.

In Figure 9, it is observed that at 15 000 RACH access requests, the average delay experienced by the PLA-QL-RACH scheme is 29% lower than that of LA-QL-RACH for the H priority class. Similarly, for the M priority class, the proposed scheme attains an average delay of 6.7 cycles, which is 17% lower than that of the LA-QL-RACH scheme. In contrast, for the L priority class, the proposed scheme reached a steady-state average delay of 8.6 cycles, which is 12% lower than the 9.75 cycles recorded by the LA-QL-RACH at 15 000 RACH requests. The lower delay cycles recorded by the PLA-QL-RACH scheme with respect to the LA-QL-RACH scheme represent how quickly the scheme reached its stability based on training completion and learning convergence. Furthermore, it is shown that at steady state, the higher the priority, the lower the access delay, which is due to the effect of prioritization, proportionate distribution, and allocation of resources. Consequently, the result illustrates that the average access delay for each priority class depends on the percentage of resources available to that class, as each class takes resources from the maximum RACH resources assigned to it. Additionally, when the access requests from a priority class are treated, the remaining free RACH resources for this class are assigned proportionally to the other classes in a collision-free manner. Therefore, the delay performance of the proposed PLA-QL-RACH scheme is improved because at steady state, there are no collisions at the lower levels of generated traffic owing to the minimal contention. Figure 9 further shows that variations around the preamble allocation ratios do not have any impact on the relative performance characteristics of the different schemes. However, they only subtly change the absolute values, and these values effectively prioritize the traffic and provide distinct performance levels for the high-, medium-, and low-priority classes.

6 | CONCLUSION

In this paper, the PLA-QL-RACH scheme was proposed to improve the performance of the QL-RACH access scheme. The novel technique classifies M2M devices according to three QoS priority classes and assigns RACH resources based on their respective demands. The classification minimizes the level of interaction and collision between the M2M and H2H devices without pushing the M2M into another Q-learning process. In this scheme, the s-ALOHA capacity (e^{-1}) has been used as an indicator of the preamble utilization status, which is either idle, successful, or collision. The collision state determines the response of the LA through feedback, which regulates the resource allocation process and the use of a penalty factor in QL-RACH. The simulation results show

that at 1.0 E, the proposed PLA-QL-RACH converges faster than the LA-QL-RACH scheme and achieves an RACH throughput performance that is 10% higher than that of FA-QL-RACH and 19% higher than that of the FB-QL-RACH scheme. Overall, the proposed scheme improves the RACH throughput to 82% and access delay to 79% with a speed of convergence that is faster than that of existing schemes.

ACKNOWLEDGEMENTS

The authors acknowledge the support and facilities available at their respective institutions which has allowed this research to be fulfilled.

AUTHOR CONTRIBUTION

All authors have contributed to the ideas reported in the paper and preparation of the manuscript for publication. The order of the author list reflects the magnitude of contribution, with the first author undertaking most of the work.

ORCID

Nasir A. Shinkafi https://orcid.org/0000-0003-2334-1441
Paul D. Mitchell https://orcid.org/0000-0003-0714-2581

REFERENCES

1. Cisco, *Cisco visual networking index: Global mobile data traffic forecast update 2010–2015*, Cisco, San Jose, CA, USA, 2011, mobile-white-paper-c11-520862.
2. 3GPP, *Study on RAN improvements for machine-type communications*, 3GPP TR 37.868 V11.0.0, Tech. Rep. 2011.
3. D. S. Watson et al., *Machine to machine (M2M) technology in demand responsive commercial buildings*, in Proc. ACEEE Summer Study Energy Effic. Build. (Washington D.C, USA), Aug. 2004, pp. 1–14.
4. B. Emerson, *M2M: The internet of 50 billion devices*, 2010, available at http://www.huawei.com/ [last accessed June 2017].
5. S. K. Tan, M. Sooriyabandara, and Z. Fan, *M2M communications in the smart grid: Applications, standards, enabling technologies, and research challenges*, Int. J. Digital Multimed. Broadcast. **2011** (2011), article no. 289015, https://doi.org/10.1155/2011/289015
6. A. Zanella et al., *Internet of things for smart cities*, IEEE Internet Things J. **1** (2014), no. 1, 22–32.
7. L. Atzori, A. Iera, and G. Morabito, *The internet of things: A survey*, Comput. Netw. **54** (2010), 2787–2805.
8. P. Bellavista et al., *Convergence of MANET and WSN in IoT urban scenarios*, IEEE Sens. J. **13** (2013), 3558–3567.
9. Standards on Machine to Machine Communications, ETSI Mob. World Congr. (2011), ETSI Newsletter, pp. 1–12.
10. PopulationPyramid, *Population Pyramids of the World from 1950 to 2100*, available at https://populationpyramid.net/world/2020/ [last accessed January 2017].
11. G. Montenegro et al., *Transmission of IPv6 packets over IEEE802.15.4 networks*, Tech. Rep. 4944, 2007, available at http://tools.ietf.org/pdf/rfc4944.pdf
12. J. Zheng and M. J. Lee, *A comprehensive performance study of IEEE 802.15.4*, in Sensor Network Operations, IEEE, 2006, pp. 218–237.
13. P. Kinney et al., *Zigbee technology: Wireless control that simply works*, in Proc. Commun. Des. Conf. Oct. 2003, pp. 1–20.
14. H. Merz et al., Building automation: Communication systems with EIB/KNX, LON and BACnet, Springer, Berlin, Germany, 2009.
15. A. Biral et al., *The challenges of M2M massive access in wireless cellular networks*, Digital Commun. Netw. **11** (2015), no. 1, 1–19.
16. A. Laya, L. Alonso, and J. Alonso-Zarate, *Is the random access channel of LTE and LTE-A suitable for M2M communications? a survey of alternatives*, IEEE Commun. Surv. Tutor. **16** (2014), 4–16.
17. 3GPP, *Medium Access Control (MAC) Protocol Specification*, 3GPP TS 36.321 V11.0.0, 2012.
18. F. Morvari and A. Ghasemi, *Priority-based adaptive access barring for M2M communications in LTE networks using learning automata*, Int. J. Commun. Syst. **30** (2017), no. 16, https://doi.org/10.1002/dac.3325.
19. L. M. Bello, P. Mitchell, and D. Grace, *Application of Q-learning for RACH access to support M2M traffic over a cellular network*, in Proc. European Wirel. Conf. (Barcelona, Spain), May 2014, pp. 1–6.
20. L. M. Bello, P. Mitchell, and D. Grace, *Frame based back-off for Q-learning RACH access in LTE networks*, in Proc. Telecommun. Netw. Appl. Conf. (ATNAC), (Southbank, VIC, Australia), Nov. 2014, pp. 176–181.
21. L. M. Bello et al., *Q-learning based random access with collision free RACH interactions for cellular M2M*, in Proc. Int. Conf. Next Generation Mob. Appl., Serv. Technol. (Cambridge, UK), Sept. 2015, pp. 78–83.
22. N. A. Shinkafi et al., *Learning automata based Q-learning RACH access scheme for cellular M2M communications*, in Proc. IEEE Glob. Conf. Internet of Things (GCIoT), (Dubai, United Arab Emirates), Dec. 2019, pp. 1–6.
23. A. Lo et al., *Enhanced LTE-advanced random-access mechanism for massive Machine-to-Machine (M2M) communications*, in Proc. World Wirel. Res. Forum Meeting, (Dusseldorf, Germany), Oct. 2011.
24. T. Taleb and A. Kunz, *Machine type communications in 3GPP networks: Potential, challenges and solutions*, IEEE Commun. Mag. **50** (2012), 178–184.
25. M. Beale, *Future challenges in efficiently supporting M2M in the LTE standards*, in Proc. IEEE Wirel. Commun. Netw. Conf. Work. (WCNCW), (Paris, France), Apr. 2012, pp. 186–190.
26. K. S. Ko et al., *A novel random access for fixed-location machine-to-machine communications in OFDMA based systems*, IEEE Commun. Lett. **16** (2012), 1428–1431.
27. M. Hasan, E. Hossain, and D. Niyato, *Random access for machine-to-machine communication in LTE-advanced networks: Issues and approaches*, IEEE Commun. Mag. **51** (2013), 86–93.
28. Y. C. Pang et al., *Network access for M2M/H2H hybrid systems: A game theoretic approach*, IEEE Commun. Lett. **18** (2014), 845–848.
29. H. Wu and C. Zhu, *FASA: Accelerated S-ALOHA using access history for event-driven M2M communications*, IEEE/ACM Trans. Netw. **21** (2013), 1904–1917.
30. MTC LTE Simulations. 3rd Generation Partnership Project (3GPP). TSG RAN WG2 v11.0.0, 2011.
31. J. Sarker and S. Halme, *An optimum retransmission cut-off scheme for slotted ALOHA*, Wirel. Personal Commun. **13** (2000), 185–202.
32. P. Nicopolitidis, G. I. Papadimitriou, and A. S. Pomportsis, *Distributed protocols for ad-hoc wireless LANs: A learning-automata-based approach*, Ad Hoc Netw. **2** (2004), 419–431.
33. F. Hussain, A. Anpalagan, and R. Vannithamby, *Medium access control techniques in M2M communication: Survey and critical review*, Trans. Emerg. Telecommun. Technol. **28** (2017), e2869.

AUTHOR BIOGRAPHIES

Nasir A. Shinkafi received his BEng, Meng, and PhD degrees from Bayero University, Kano, Nigeria, in 1998, 2006, and 2020 respectively. He has extensive experience working with giant telecommunications companies and mobile network operators in Nigeria and abroad for over 20 years. He is currently the chief technology officer at Backbone Connectivity Networks (BCN), Abuja, Nigeria.

Lawal M. Bello received his BEng degree in electrical engineering from Bayero University, Kano, Nigeria, in 2004, MSc degree in RF and wireless communication from the University of Leeds, Leeds, UK, in 2009, and PhD from the University of York, York, UK, in 2015. He is currently a senior lecturer at Bayero University, Kano, Nigeria. His research interests include intelligent access strategies, machine-to-machine communication, cellular networks, and the Internet of Things.

Dahiru S. Shu'aibu received his BEng and MEng degrees from Bayero University, Kano, Nigeria, in 1999 and 2005 respectively. He obtained his PhD from Universiti Teknologi Malaysia, Johor Bahru, Malaysia. He has been working with Bayero University, Kano, Nigeria, since 2001 and is currently a professor in the Department of Electrical Engineering. His research interests include radio resource management for OFDMA based systems.

Paul D. Mitchell received his MEng. and PhD degrees from the University of York, York, UK in 1999 and 2003, respectively. He has over 20 years research experience in wireless communications, and industrial experience gained at BT and DERA (now QinetiQ). He has been a member of academic staff at the Department of Electronic Engineering since 2005 and is now full professor. His primary research interests lie in underwater acoustic communication networks, terrestrial wireless sensor networks, and communication protocols.