https://doi.org/10.31449/inf.v49i12.6573 Informatica **49** (2025) 157–172

Congestion Control of Large-Scale Elevator Terminal Data Access in Large Metro Stations Based on The Internet of Things

Juanjuan Shi*, Mengtian Jiao
College of Information Engineering, Jiaozuo University, Jiaozuo 454000, China
E-mail: jinglove666999@163.com
*Corresponding author

Keywords: hybrid access method, internet of things, congestion control, ACB control parameters

Received: July 6, 2024

Large metro station IoTs used to face congestion while access to terminals was going on a large scale. Due to this, low success rate in access and delay in monitoring critical equipment was observed, which included elevators and escalators. This paper presented a congestion control method for large-scale elevator terminal data access in metro stations using IoT. Business data were categorized based on volume and latency requirements: Slot ALOHA (SA) direct access mode was used for delay-insensitive, small data services, and Access Class Barring (ACB) random access was used for time-sensitive, large data services. ACB control parameters were dynamically adjusted by estimating access requests. Using uniform and Beta distribution models, the method's effectiveness was validated through experiments. With 4000 access requests, the hybrid method achieved a 52.43% success rate and a 76.72 ms average delay under the uniform model, and a 42.07% success rate with an 82.02 ms average delay under the Beta model. These results demonstrated the method's ability to meet Quality of Service (QoS) requirements for high-priority services, ensuring efficient and reliable communication in large-scale IoT environments.

Povzetek: Prispevek predstavlja hibridno metodo za nadzor preobremenjenosti do podatkov naprav IoT, ki uporablja kombinacijo direktnega in naključnega dostopa, s prilagajanjem parametrov glede na obseg zahtev.

1 Introduction

Based on the current urban development and people's travel needs, the number of elevators inside large metro stations is also growing [1]. The stability of elevator operation is closely related to the safety of residents. However, due to the quality, maintenance, supervision, and other influencing factors, elevator accidents often occur. How to conduct unified real-time monitoring of elevator equipment in large and medium-sized Spaces to reduce daily lightweight failures and prevent heavyweight accidents has become a hot topic of scholarly attention [2, 3].

The increasing global population and urbanization have heightened the demand for elevators, necessitating advanced, safe, and efficient systems. China's elevator demand grows by 5%–7% annually due to the need to replace outdated units and comply with new regulations, increasing maintenance workloads and risks. Innovative designs must prioritize safety, including weight capacity, emergency alarms, and secure installation sites. Energy-efficient elevators can reduce operational costs significantly. Traditional monitoring systems, like video surveillance, fail to reflect the elevator's condition and failure rates adequately. With its advantages of low power consumption, significant connection, low delay, and high reliability, the Internet of Things (IoT) can realize the transmission and processing of multiple types and large-scale data [4].

Based on this, some scholars use IoT technology to monitor elevators' operating status data, dramatically improving elevator operation security and effectively reducing equipment operation and maintenance costs. Mao et al. [5] discussed the integration of Internet of Things (IoT) technology to enhance the remote security management of elevators, addressing the associated safety risks. They proposed an IoT-based architecture for elevator fault diagnosis and maintenance. The study established a fault diagnosis management system centered on IoT, outlining maintenance methods to ensure the safety and stability of elevator operations. This approach aims to improve the overall security and efficiency of urban transportation through advanced technology. Lai et al. [6] adopt the more predictive state maintenance method to realize the remote monitoring of highly distributed elevator equipment status, effectively improving the safety and reliability of equipment operation.

IoT devices, ranging from consumer products to industrial components, are becoming ubiquitous, driving the concept of "Smart homes" with enhanced safety and energy efficiency. Wearable fitness and health monitors, network-enabled medical devices, and smart traffic systems contribute to "smart cities" that reduce congestion and energy use. IoT also promises to improve the independence and quality of life for people with disabilities and the elderly. The impact of IoT extends to agriculture, industry, and energy sectors, enhancing information flow along the production value chain. Companies and research organizations predict significant economic effects [7]. A market research report revealed that the global IoT market was valued at $1.90 billion in 2018 and is projected to grow to $11.03 billion by 2026. Additionally, the European Union (EU), the United States (USA), China, and other nations have developed IoT-related action plans. These initiatives include the IoT-An Action Plan for Europe and various IoT development plans for the years 2016–2020 [8].

Song et al. [9] discussed the adoption of smart technologies and networking solutions like the Internet of Things (IoT) by leading cities in China to enhance economic opportunities and global climate resilience. They presented the smart city concept as a complex system integrating sensors, data, applications, and organizational forms to make cities more agile and sustainable. The paper provided a comprehensive assessment of smart city initiatives in China, classifying practices into six key dimensions: energy, agriculture, transport, buildings, urban services, and urban security operations. Chinese smart city policies and practices aim to explore renewable energy, improve public convenience, and enhance urban comfort and citizen friendliness. The study also addressed concerns in areas such as system integration, governance, innovation, and finance. A policy vision was outlined to build public-private collaborative networks, encourage innovation and investment in smart city initiatives, and emphasize smart services.

In practical applications, the infrastructure of the wireless cellular network is relatively perfect, the coverage area is comprehensive, and the security is high, which is one of the leading carrying networks of IoT communication. However, the original intention of traditional wireless cellular network design is to deal with the communication problem between humans and humans (H2H), and there are some differences in the communication characteristics between machine to machine (M2M). Machine-type communication (MTC) devices, integral to Industry 4.0, support smart factories, healthcare, and surveillance by generating data and making policy-based decisions. The demand for these devices is projected to reach 50 billion by 2025. These devices require robust security due to their vulnerability and usage in open environments. Lightweight cryptography is the preferred solution for MTC devices due to their limited computational and memory capacities. This cryptographic approach ensures strong encryption while being efficient and cost-effective, enhancing security for the growing number of IoT devices. MTC devices are autonomous and central to automating IoT frameworks, evolving to support the advancements of Industry 4.0. They form Machine-to-Machine (M2M) communication networks, also known as cyber-physical systems and edge nodes, creating an autonomous system of resource-constrained devices [10].

The six key features of Machine Type Communication (MTC) in 6G are ultra-low latency and high reliability, massive connectivity, energy efficiency, scalable and flexible network architecture, enhanced security and privacy, and advanced AI and machine learning integration. These features ensure instantaneous and reliable data transmission for critical applications, support billions of IoT devices, extend the battery life of remote sensors, allow dynamic resource allocation, protect sensitive data, and optimize network performance through predictive maintenance and anomaly detection. These features collectively create an efficient, reliable, and secure communication environment for the 6G era [11].

Due to the limitation of channel resources, when the IoT at metro stations has a large number of elevators, and other equipment data access, the time delay indicator of the system is higher, and the throughput will decrease significantly. Therefore, there is a great demand for a large-scale terminal access algorithm tailored to the communication characteristics of the IoT at large metro stations to ensure the reliable transmission of information data of crucial equipment. In response to the above issues, Chou et al. [12] used Bayesian theory to estimate the number of access applications, preamble code conflict rate, and the number of following time-slot applications at the current time-slot. Furthermore, the optimal ACB control parameters are discussed by judging the number of applications for the subsequent access time slot through quantitative prediction methods. The scheme is based on the premise that the current time-slot access conflict makes direct rebleeding at the next time-slot, with some error from the system of refeeding in the actual access process.

Zhang et al. [13] addressed the growing need for improved communication content and quality in the context of advancing network and communication technologies. This research concerns the optimal data collection and path planning of multi-unmanned aerial vehicle (UAV) to achieve extensive terminal accessibility in IoT scenarios. The novelty of the approach consists of integrating sensor area partitioning with the flight trajectory planning of multiple UAVs with the main objectives of load balancing while the overall completion time for the tasks at hand is minimized. A novel k-means algorithm has been developed to balance the quantity of data in each cluster. Accordingly, the flight trajectories of the UAVs were represented discretely by an enhanced genetic algorithm including the 2-opt optimization operator for solving the multiple traveling salesman problem (MTSP) problem, improving the computational effectiveness. Extensive simulations have validated the efficiency of the suggested approach in smoothing out the imbalances in the distribution of tasks among UAVs and significantly reducing the duration of tasks. The convergence rate for this methodology was higher than the conventional genetic algorithm; hence, this proved that it was computationally efficient. Equipped with a new, efficient methodology for multi-UAV-assisted IoT terminal data gathering, it brings balance and efficiency in task distribution, unfolding the full power of professional algorithm solutions when acquiring optimal results in more complicated engineering scenarios.

Varsha et al.[14] proposed an innovative intelligent traffic management system for wireless cellular networks to enhance M2M connections, pivotal for IoT. They focused on improving Access Class Barring (ACB), a method traditionally relying on a static factor to manage machine-type communication device (MTCD) traffic. The study introduced a Bayesian inference-based learning automatons (BI-LA) approach that dynamically adjusts the ACB factor. This system leverages learning automata's self-adaptive learning to estimate and manage M2M traffic more effectively. By framing the problem around collision probability and using Bayesian inference to adapt the ACB factor, the proposed method was tested using network simulator-3 (NS3). The performance metrics—average access delay, access attempts, access success rate, and access success—demonstrated that the BI-LA ACB technique outperformed traditional and contemporary ACB methods, achieving minimal access delays of approximately 1876 ms and 27.6 ms.

The main problem arises due to a large amount of UEs present in the RA techniques, as discussed by Piao and Lee [15], where increased collisions and delays arise. They propose a new RA scheme that combines four-step RA with two-step RA, based on the 3rd Generation Partnership Project Release 16. This work tries to avoid a conflict with the available RA resource, then achieves a better performance of efficiency and brings down the average RA delay. This solution aims to optimize the two-step RA probability and thus provides a resource configuration and parameter setting algorithm that allows the UEs to carry out both RA methods simultaneously. Then, the authors proved further that the proposed approach is valid using a Markov chain model. The proposed approach also has its potential confirmed in extensive comprehensive simulations on supporting RA procedures in the case of massive and heterogeneous device access for 5G and 6G communication applications.

Yu et al. [16] investigated the performance of massive machine-type communications (mMTC) in status update systems, where numerous machine-type communication devices (MTCDs) send status packets to a base station (BS) for system monitoring. The authors identified that packet collisions due to massive MTCDs negatively impact status update performance. To address this, they proposed a joint access control, frame division, and subchannel allocation scheme. They first analyzed access control, packet collisions, and packet errors, deriving a closed-form expression of the average age of information for all MTCDs as a performance metric. Their proposed scheme was shown through simulations and numerical results to achieve near-optimal performance, comparable to exhaustive search methods, and outperformed benchmark schemes. Bui et al. [17] present an access protocol based on distributed queue (DQ) mechanisms to deal with M2M communication large-scale access problems for cellular networks. To maximize the DQ mechanism performance, first of all, the base station in the random-access opportunities is roughly the number of conflict detection equipment to avoid excessive division of DQ. Then based on the probing results, the base station randomly divides the device into a determined number of groups and "pushes" these groups to the end of the logical access queue. Finally, the validity and feasibility of the proposed protocol are verified by simulation.

Congestion control and optimization methods overview in IoT applications-the methodologies, the datasets used, the results, and the limitations are represented in Table 1. This comparison identifies the gaps that this paper will address with the proposed hybrid access method.

Table 1: Summary of related works on congestion control and access optimization Methods in IoT applications highlighting limitations and positioning the hybrid access method as a novel solution

Study	Method	Datasets	Key Results	Limitations
Mao et al. [5]	IoT-based architecture for fault diagnosis	Elevator operational data	Improved safety and stability of elevator operations through IoT monitoring	Limited scope to fault diagnosis only
Lai et al. [6]	Predictive maintenance with IoT integration	Distributed elevator equipment data	Enhanced safety and reliability of elevator systems	Focused only on maintenance, lacks scalability analysis
Chou et al. [12]	Bayesian theory-based ACB optimization	Simulated data	Improved ACB parameters, reduced conflict rate	Errors in real-time predictions
Zhang et al. [13]	Multi-UAV data collection and path optimization	Simulated IoT scenarios	Balanced task distribution, reduced completion time	High computational overhead
Varsha et al.[14]	Learning Automaton-based ACB scheme (LA-ACB)	Cellular Base Station data	Controlled M2M data, reduced H2H interference	High implementation complexity
Piao and Lee [15]	Integrated 2-4 step Random Access (RA) methods	Cellular network simulations	Reduced collisions and delays	Limited to specific RA configurations

Bui et al. [17]	Distributed Queue (DQ)-based access protocol	LTE/LTE-A network data	Reduced congestion, improved success rate	Requires precise group partitioning

This paper proposes a hybrid access methodology that combines Slot ALOHA with Access Class Barring for large-scale IoT scenarios in metropolitan transit stations. The proposed methodology, by dynamically changing ACB control parameters and implementing predictive modeling on access requests, should be able to provide high QoS for important applications like elevator monitoring under different traffic conditions. This novel strategy overcomes some fundamental limitations of the previous approaches by providing a scalable, reliable, and economic solution to congestion management in IoT systems with complex networks. Therefore, the key contributions of the paper are as follows:

The key contributions of the paper are as follows:

1. **Congestion control method**: Developed a method for managing large-scale elevator terminal data access in metro stations using IoT, addressing low access success rates and delays.
2. **Data categorization**: Divided business data based on volume and latency requirements, using Slot ALOHA (SA) for delay-insensitive data and Access Class Barring (ACB) for time-sensitive data.
3. **Dynamic ACB adjustment**: Proposed dynamically adjusting ACB control parameters by estimating access requests to optimize terminal access.
4. **Performance evaluation**: Demonstrated through simulations that the hybrid access method improves access success rates and reduces delays, especially with high access requests.
5. **Application in IoT environments**: Ensured Quality of Service (QoS) for high-priority services in large-scale IoT environments in metro stations.
6. **Predictive access application**: Developed a method to predict access applications for better access control.
7. **Experimental validation:** Validated the method in a Shanghai metro station, showing practical advantages over traditional methods.

2 Systems model and custom MAC layer protocol for IoT communication in large metro stations

2.1 Systems model

Based on the practical application, a metro station communication model is built with large-scale MTCD to simulate the congestion caused by frequent network access by communication devices. Illustration of IoT communication model for large metro stations in Fig. 1 shows how the MTCDs will be sending their data to the server via the eNB.

The evolved Node B (eNB) receives, controls, and allocates up/down dynamic resources. MTCD data is transmitted to a fixed gateway through the narrowband IoT, which forwards the data to the server. In the IoT model, when two or more MTCDs use the same preamble code simultaneously, it indicates that the decision is in conflict and the device access fails.

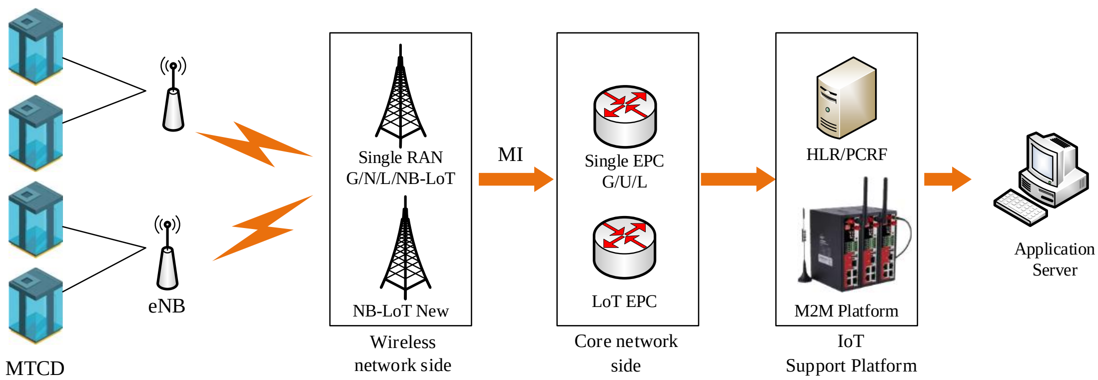

Figure 1: Illustration of the IoT communication model for large metro stations, showcasing the flow of data from MTCDs to servers via the eNBCustom MAC Layer Protocol

Given that complex signaling can reduce the success rate of device access, the network employs the Media Access Control (MAC) protocol [18]. The MAC layer protocol combines Selective Acknowledgement (SA) and Access Class Barring (ACB) controls to adapt to various types of business data and enhance access speed and success. For services with small amounts of valid data and low sensitivity to delay, SA direct access is utilized. Conversely, ACB random access is applied to delay-sensitive and data-intensive services. Fig. 2 illustrates the hybrid MAC layer protocol diagram, where T_i is the i-th access timeslot.

The hybrid MAC layer protocol divides each incoming data packet into four parts:

1. Broadcasting data access information and ACB control parameters for the current timeslot.
2. Assigning preamble codes to randomly accessed services.
3. Handling SA direct access business.
4. Conducting data transmission.

Using the hybrid MAC layer protocol for the classified transmission of different business data effectively reduces signaling consumption, accelerates data access, and ensures the Quality of Service (QoS) demands of high-priority business services.

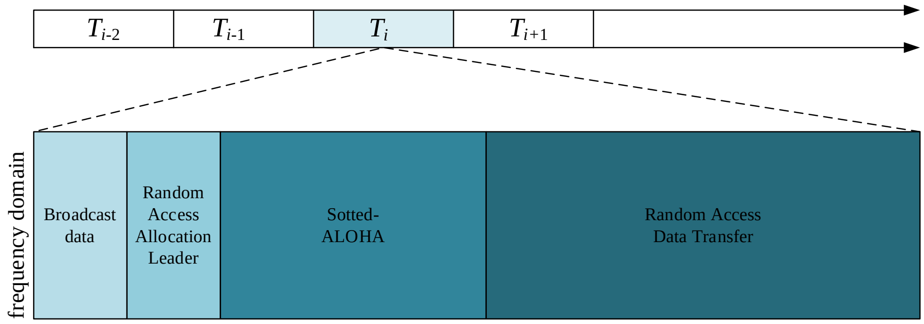

Figure 2: Hybrid MAC layer protocol

3 Design of hybrid access method

3.1 SA method and improvement

The SA transmits data by speaking first. Signal overlap is likely to occur during concurrent operations, leading to network congestion. Therefore, a random wait period is introduced before attempting to resend the data. The data transmission process is illustrated in Fig. 3.

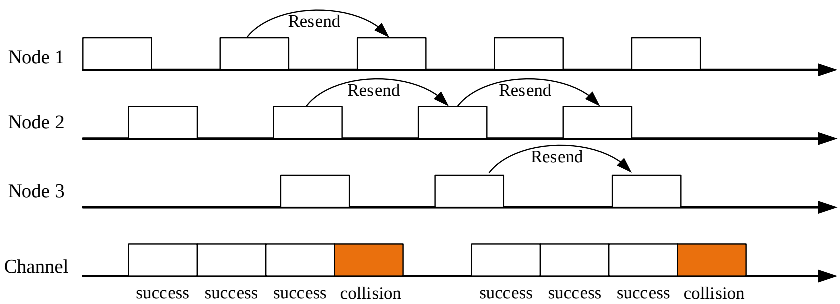

Figure 3: Data sending process for traditional SA method

In the traditional Slot ALOHA (SA) method, the time for retransmission is random, leading to a high probability of complete or partial collisions. This randomness reduces the efficiency of information utilization and decreases system throughput.

To address these issues, the data transmission process has been improved. The transmission period is divided into several time slots, and data can only be sent at the initial point of a time slot. By ensuring that nodes transmit information within their designated time slots, the likelihood of collisions is significantly reduced, as nodes are not transmitting simultaneously. This structured approach allows for more efficient use of the available bandwidth and improves overall system throughput.

The improved SA data-sending process, which mitigates collisions and enhances throughput, is illustrated in Fig. 4. This method ensures that each node's transmission is independent of others, leading to more reliable and orderly communication within the network.

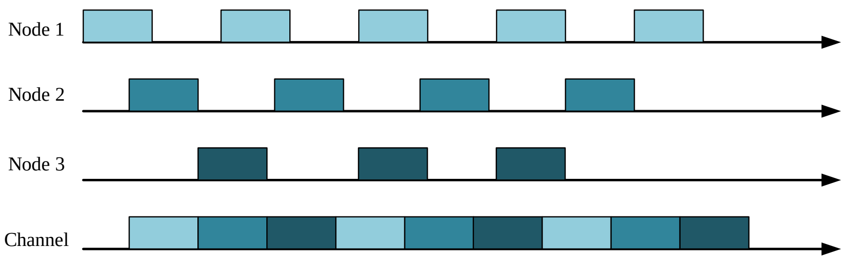

Figure 4: Data sending process for improving the SA method

The relationship between the throughput rate Q and the sent packet quantity G can be expressed as Eq. (1):

$$Q = Ge^{-G} \tag{1}$$

When two nodes transmit within the period T', the data transmission delay function is given in Eq. (2):

$$T_Y = 2T' + t_d + [\varphi T' + (B+1)T'](e^G - 1) \tag{2}$$

Where φ represents the waiting time for a response, t_d represents the propagation duration and B represents the maximum value of the backoff time slot.

The fixed transmission channel and the number of inherent node parameters determine the transmission delay of SA. Therefore, the improved method is only suitable for processing delay-insensitive and small data volume services. Otherwise, the transmission error will increase, and the availability of information will be reduced.

3.2 Estimation of access applications based on time series prediction

3.2.1 Estimation of current timeslot access applications

For services using the ACB (Access Class Barring) random access mode, the application amount of the service should be estimated based on the occupation of the preamble code [19, 20]. Assume that w_i represents the state of the i -th preamble code. The states are defined as follows:

- When $w_i = 1$, the preamble code is not selected and is idle.
- When $w_i = 1$, an MTCD (Machine-Type Communication Device) has selected the preamble code and it is busy.
- When $w_i \geq 2$, two or more MTCDs have selected the preamble code, resulting in a conflict status [21].

The probabilities of the i -th preamble being in these three states is given by the following Eq. (3):

$$P(w_i) = \begin{cases} \left(1-\frac{1}{N_p}\right)^{n_a}, & w_i = 0 \\ \frac{n_a}{N_p}\cdot\left(1-\frac{1}{N_p}\right)^{n_a-1}, & w_i = 1 \\ 1-\left(1-\frac{1}{N_p}\right)^{n_a} - \frac{n_a}{N_p}\cdot\left(1-\frac{1}{N_p}\right)^{n_a-1}, & w_i \geq 2 \end{cases} \tag{3}$$

Where N_p represents the number of available preamble codes in the current timeslot, n_a indicates the number of access requests for the current timeslot.

Assume that the number of preamble codes satisfying $\mathrm{w_i} = 0$, $\mathrm{w_i} = 1$, and $\mathrm{w_i} \geq 2$ in the current timeslot are $\mathrm{L_1}$, $\mathrm{L_2}$, and $\mathrm{L_3}$, respectively. Then, the maximum likelihood estimation of the number of access applications in the current timeslot is expressed as Eq. (4):

$$P = P(w_i = 0|N_a)^{n_1} \cdot P(w_i = 1|N_a)^{n_2} \cdot P(w_i \geq 2|N_a)^{n_3} \tag{4}$$

The principle is to ensure that the number of access requests in the next time slot is optimal. The estimated number $\widehat{N}_a$ of access requests in the current timeslot can be obtained by setting N_a to the maximum value. The expression is given in Eq. (5):

$$\widehat{N}_a = arg\ \max \sum_j^J ln\, P\,(w_j|N_a) \tag{5}$$

After ACB, the comparison results between the maximum likelihood estimate and the actual application amount are shown in Fig. 5. It can be seen from the figure that the trend changes of the two lines are relatively consistent, indicating that the estimated value aligns well with the actual value.

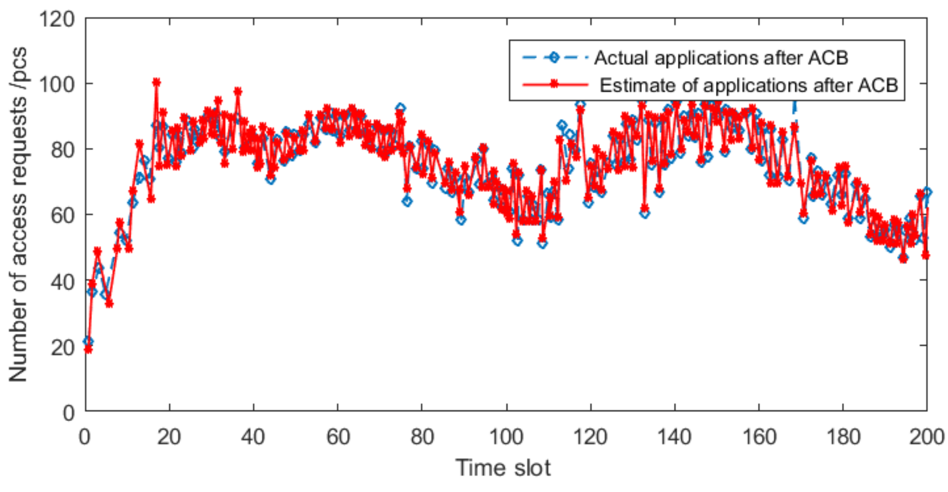

Figure 5: Comparison results of maximum likelihood estimation and actual application volume (After passing the ACB)

According to the maximum likelihood estimation after passing the ACB, the actual number of access applications can be calculated as $\hat{N}=\hat{N}_a/a$, where a is the ACB control parameter of the current timeslot. Before passing the ACB, he comparison between the maximum likelihood estimates and the actual number of applications is shown in Fig. 6.

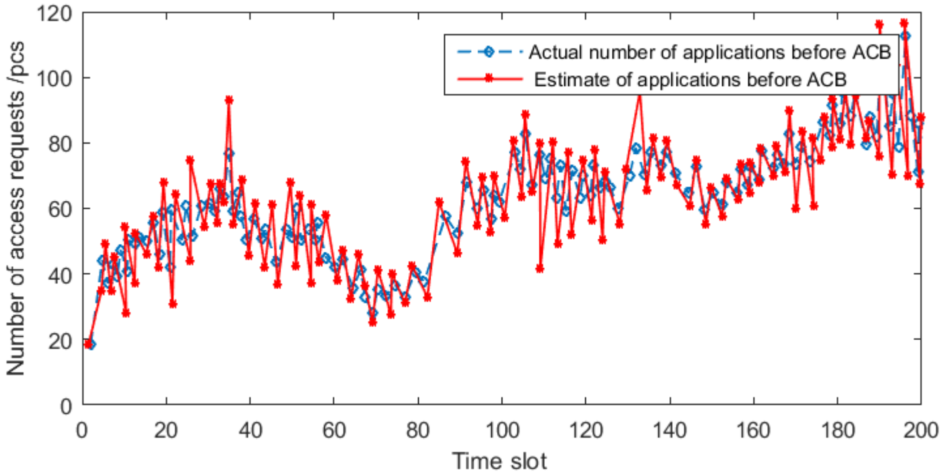

Figure 6: Comparison results of maximum likelihood estimation and actual application volume (Before passing the ACB)

For services accessed in SA mode, the estimation is based on the physical resource block status of the current time slot [22]. Assuming that the total number of available resource blocks is U_s, and the number of idle rate blocks in the current timeslot is $\tilde{U}_{k,i}$. The actual idle rate is $\tilde{P}_{k,i}=\frac{\tilde{U}_{k,i}}{U_s}$, the theoretical idle rate is $P_{k,i}=\left(\frac{U_s-1}{U_s}\right)^{c_i}$, where C_i is the access application volume of the current timeslot. y equating the theoretical idle rate to the actual idle rate, $\tilde{P}_{k,i}=P_{k,i}$, the number of access requests in the current time slot is obtained as shown in Eq. (6):

$$\hat{C}_i=\frac{log(\tilde{P}_{k,i})}{log(N_i(N_i-1)} \tag{6}$$

3.2.2 Estimation of next timeslot access applications

Assume that the estimated number of access applications in the i -th time slot is $\hat{N}_i$, the number of access successes is W_i, the number of newly arrived access applications in the $i+1$ time slot is $T_{i_{+1}}$, and the number of access applications that need to be retransmitted is $H_{i_{+1}}$. Then the estimated number of access applications in the $i+1$ time slot can be shown as Eq. (7):

$$\hat{N}_{i+1}=\begin{cases}\hat{N}_i-W_i+H_{i+1}+T_{i+1}, i\le I_D\\ \hat{N}_i-W_i+H_{i+1}, \qquad i>I_D\end{cases} \tag{7}$$

Where, I_D represents the last timeslot.

Since the access request volume is a time series, the weighted sum of historical increments is used as an increment in the next time slot. The newly arrived access applications in the $i+1$ time slot, $T_{i_{+1}}$ can be expressed as shown in Eq. (8):

$$T_{i_{+1}}=\frac{3}{5}T_i+\frac{3}{10}T_{i-1}+\frac{1}{10}T_{i-2} \tag{8}$$

Because $T_i=\hat{N}_i-\hat{N}_{i-1}-H_i+W_{i-1}$, the Eq. (9) is as follows:

$$\begin{aligned}T_{i_{+1}}&=max\left\{0,\ \left(\frac{3}{5}T_i+\frac{3}{10}T_{i-1}+\frac{1}{10}T_{i-2}\right)\right\}\\&=max\left\{0,\ \begin{pmatrix}\frac{3}{5}\hat{N}_i-\frac{3}{10}\hat{N}_{i-1}-\frac{2}{10}\hat{N}_{i-2}-\frac{1}{10}\hat{N}_{i-3}\\-\frac{3}{5}H_i-\frac{3}{10}H_{i-1}-\frac{1}{10}H_{i-2}+\\\frac{3}{5}W_{i-1}+\frac{3}{10}W_{i\text{-}2}+\frac{1}{10}W_{i-3}\end{pmatrix}\right\}\end{aligned} \tag{9}$$

After transformation, the estimated amount of access requests for the next time slot can be obtained. The expression is given in Eq. (10):

$$\widehat{N}_{i+1} = \begin{cases} max\left\{\widehat{N}_i, \begin{pmatrix} \frac{3}{5}\widehat{N}_i - \frac{3}{10}\widehat{N}_{i-1} - \frac{2}{10}\widehat{N}_{i-2} - \frac{1}{10}\widehat{N}_{i-3} \\ -\frac{3}{5}H_i - \frac{3}{10}H_{i-1} - \frac{1}{10}H_{i-2} + \\ \frac{3}{5}W_{i-1} + \frac{3}{10}W_{i\text{-}2} + \frac{1}{10}W_{i-3} \end{pmatrix}\right\} - W_i, i \le I_D \\ \widehat{N}_i - W_i + H_{i+1}, \quad i > I_D \end{cases} \quad (10)$$

The comparison between the predicted application amount and the actual application amount of the time series is shown in Fig. 7. It can be seen from the figure that the curve change trend of the estimated value and the actual value is relatively consistent, indicating that the predicted result of the access application volume aligns well with the actual value.

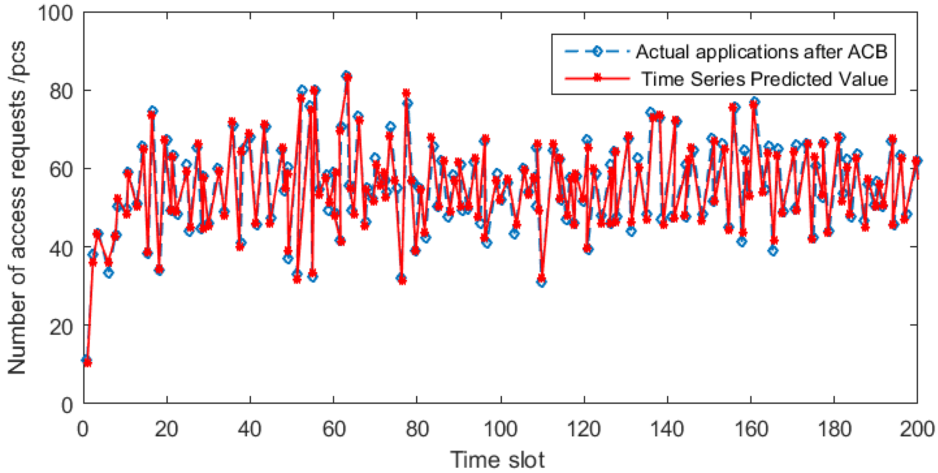

Figure 7: Comparison results of predicted and actual application volumes of time series

3.2.3 Parameter adjustment of predicted values

Update the packet parameter L_1 and ACB control parameter a of the dynamic preamble code according to the prediction value of the service arrival to ensure the access success rate of the next timeslot. Since $w_i = 1$ indicates the successful transmission of the preamble code, the estimated value of the preamble code that can transmit successfully is given in Eq. (11):

$$M[N_s|N_a = n_a] = \sum_{i=1}^{N_p} P(w_i = 1|N_a = n_a) = N_p \cdot C_{n_a}^1 \cdot \frac{1}{N_p} \cdot (1 - \frac{1}{N_p})^{n_a-1} = n_a \cdot (1 - \frac{1}{N_p})^{n_a-1} \quad (11)$$

N_s represents the number of preamble codes successfully transmitted, and N_a represents the number of services filtered by ACB. Suppose the system contains N MTCDs, and N_a MTCDs pass the screening. The probability is given in Eq. (12):

$$P(N_a = n_a|N = n) = C_n^{n_a} \cdot a^{n_a} \cdot (1 - a)^{n-n_a} \quad (12)$$

Then the estimated value of success access is given in Eq. (13):

$$M[N_s|N = n] = n \cdot a \cdot (1 - \frac{a}{N_p})^{n_a-1} \quad (13)$$

Deriving from a, the optimal control parameter is given in Eq. (14):

$$a' = \frac{J}{n} \quad (14)$$

From Eq (14), the access success rate is highest when the number of access requests matches the number of currently available preamble codes. The effect is optimal when L_1 equals the number of high-priority access requests in the current timeslot.

Fig. 8 shows the relationship between the number of access requests and access successes when the number of preamble codes is 35, 60, and 76, further verifying the correctness of the above conclusions.

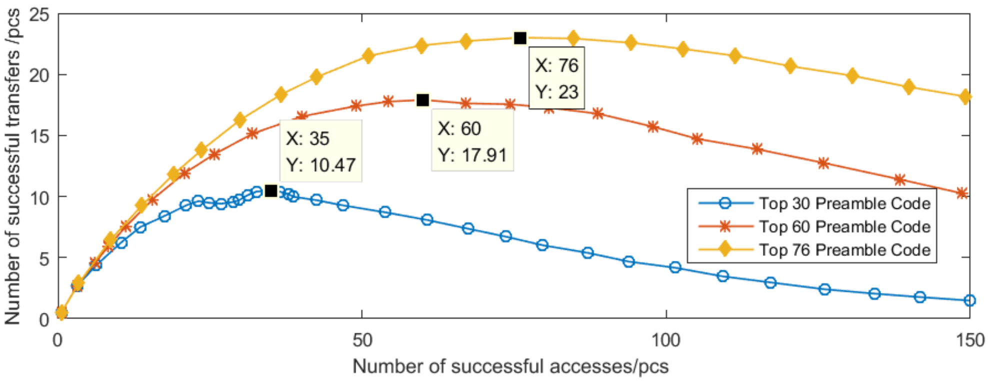

Figure 8: Relationship between access success and access requests

3.3 Hybrid access process

The access process is outlined in Figure 9, illustrating the steps involved in managing access requests for high-priority and low-priority services using the hybrid access method. Here's a detailed explanation of the process:

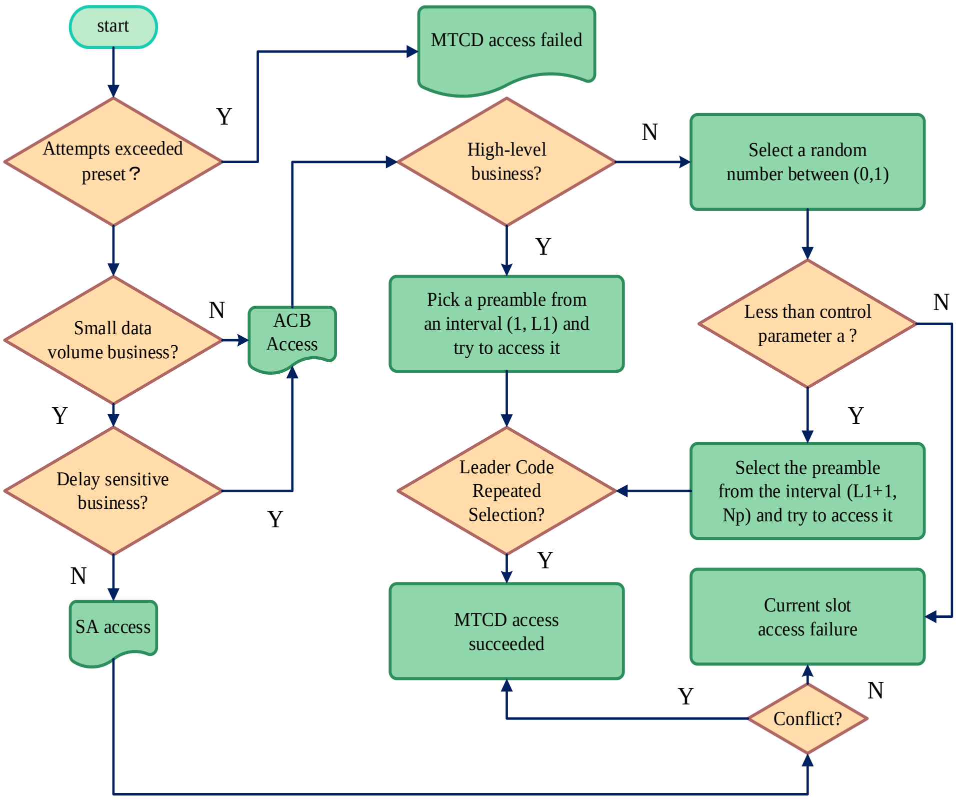

Figure 9: Access flow of the hybrid method

1. **Initial collection and setup:**

❖ The evolved Node B (eNB) collects access data from the previous timeslot, counts the usage of preamble codes, completes channel resource allocation, and sets parameters such as ACB control and backoff parameters.

2. **Random access phase:**

❖ Determine the priority of the application access business:

➢ For high-priority services, the system directly selects a preamble code from the set $K_1[1,\ L_1]$ reserved for high-priority services and proceeds to the access link.

➢ For low-priority services, a random number p is selected from the interval [0,1]. If p is less than the ACB control parameter a of the current timeslot, a preamble is selected from the set $K_2[L_1+1,\ N_p]$ designated for low-priority services. If $p \geq a$, the access is terminated.

3. **Direct access phase:**

❖ Services with small data volumes proceed with direct access.

4. **Data transmission phase:**
   - MTCDs that have successfully obtained a transmission opportunity begin data transmission.

This structured approach ensures that high-priority services are given precedence and that low-priority services are managed in a way that minimizes conflicts and optimizes resource use. The hybrid access method dynamically adjusts parameters based on historical data, improving overall system throughput and efficiency.

4 Experiments

4.1 Experimental preparation

The experimental site for the study is a large metro station in Shanghai, equipped with a significant number of IoT terminals. The configuration of the parameters used in the experiments including the number of preambles, maximum transmission attempts, conflict resolution time, and escape time, providing a baseline for evaluating the hybrid access method are detailed in Table 2.

Table 2: Key parameters used in the simulation experiments, including preambles and conflict resolution time, forming the baseline for evaluating the hybrid access method

Parameter	Value
Number of preambles	60
Maximum transmission times of preamble code	8
Conflict resolution time	24 ms
Escape time	15 ms

These parameters were utilized to simulate and analyze the performance of the hybrid access method under various traffic conditions, including uniform and beta distribution models, to verify its effectiveness in managing access congestion and ensuring timely data transmission in large-scale IoT environments.

The uniform and beta distribution models are employed to verify the feasibility of the hybrid access method by simulating various types of business data, including periodic and sudden data as well as random and irregular data, in elevator monitoring. To ensure comparability, ACB access and LA-ACB with different parameters are also used as benchmarks in the experiments. These experiments aim to count and compare the average access delay and access success rate of different services [23].

Given that the hybrid access method assigns different ranges of preamble codes according to the priority of services, while the ACB method shares all access resources uniformly, a direct comparison would be unfair. Therefore, the success rate of preamble code access is redefined for a fair assessment. The success rate, P_T, is calculated as the ratio of the number of successfully accessed services (N_c) to the total number of preamble codes used in the access process (N_{all}). This redefinition allows for a more accurate comparison of the efficiency and effectiveness of the hybrid access method against traditional ACB methods.

4.2 Experimental results and analysis

4.2.1 Simulation results and analysis of uniform distribution model

This section discusses the simulation results and analysis using a uniform distribution model to evaluate the performance of the hybrid access method compared to traditional methods such as ACB (Access Class Barring) and LA-ACB (Learning Automata ACB).

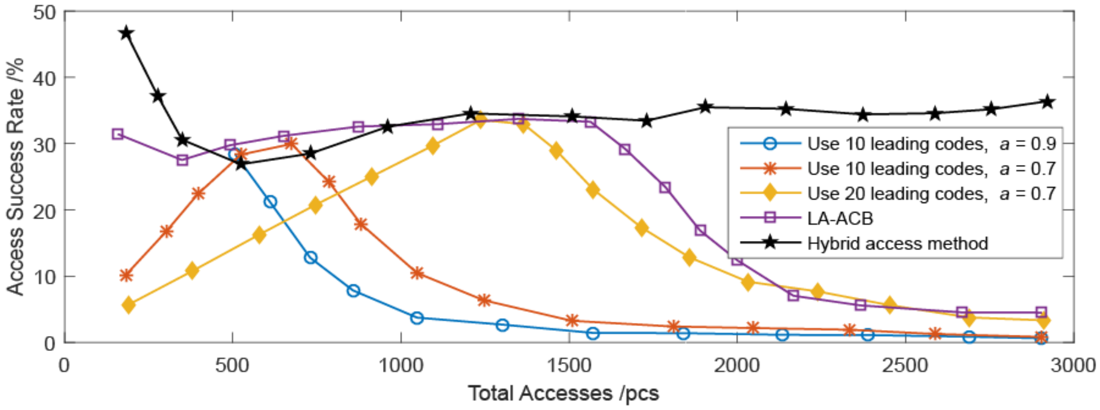

Figure 10: Comparison of access success rates for high-priority services using the hybrid access method, ACB, and LA-ACB under the uniform distribution model

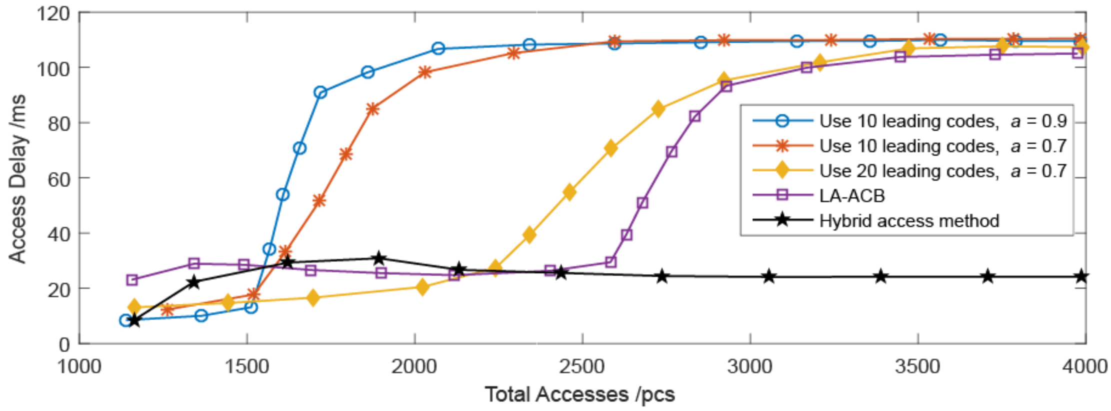

Figure 11: Comparison of average access delay for high-priority services under the uniform distribution model

The access success rate of high-priority services is demonstrated in Fig. 10. When the number of access applications is small, the LA-ACB method performs excellently. However, as the number of applications increases, LA-ACB causes resource wastage, and its performance gradually declines. The hybrid access method initially shows lower success rates and higher delays due to high estimation errors but improves significantly as the number of access applications increases. Precisely, the hybrid method demonstrates a higher success rate as access requests increase, reaching 52.43% at 4000 applications. Fig. 11 shows the comparison of average access delay for high-priority services. With an increase in access applications, the average access delay for the hybrid access method remains relatively stable, indicating higher resource utilization and meeting high-priority service requirements more effectively than LA-ACB. In other words, the hybrid access method achieves a lower delay (76.72 ms at 4000 requests) compared to ACB and LA-ACB, ensuring QoS for time-sensitive applications.

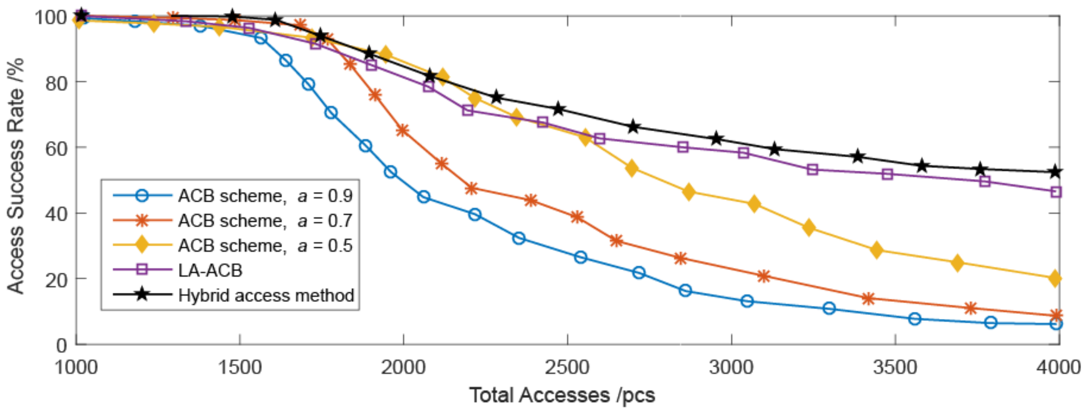

Figure 12: Comparison of access success rates for concurrent services in the uniform model, with the hybrid method outperforming ACB and LA-ACB by reducing collisions and improving resource use

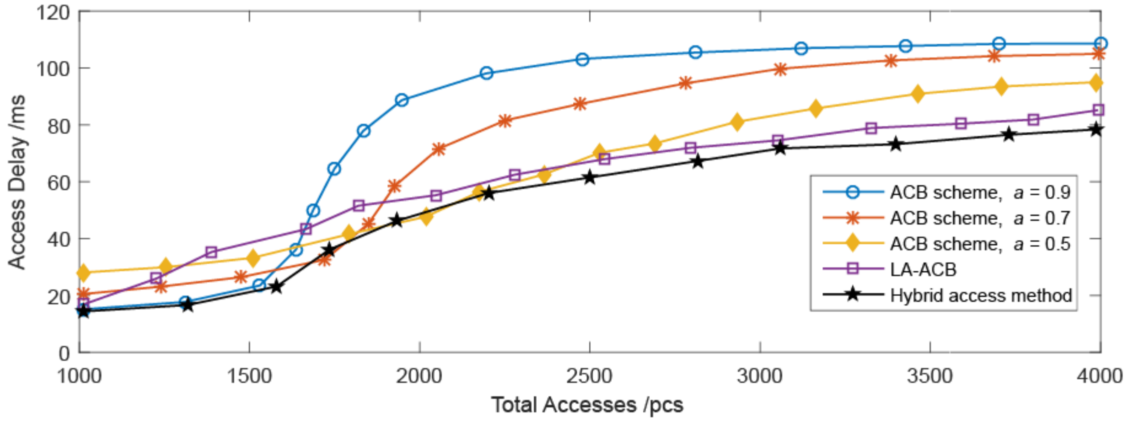

Figure 13: Comparison of average access delays for concurrent services in the uniform model, showing the hybrid method's lower delays (76.72 ms), ensuring timely transmission

The comparison of access success rates for multiple types of concurrent services is illustrated in Fig. 12, while Figure 13 shows the comparison of average access delay for these concurrent services. The hybrid access method outperforms ACB and LA-ACB, showing a higher success rate and lower delay, especially when the number of access applications reaches 4000. At this point, the hybrid method achieves a 52.43% success rate and an average delay of 76.72 ms, demonstrating undeniable advantages in efficiency and effectiveness.

These results indicate that the hybrid access method, especially under a uniform distribution model, significantly improves the system's access success rate and average access delay, thereby meeting the QoS (Quality of Service) needs for high-priority services in large-scale IoT terminal access scenarios.

4.2.2 Simulation results and analysis of beta distributed access model

When the beta distribution model is adopted, the performance of the hybrid access method is evaluated in terms of the access success rate and average access delay for high-priority services.

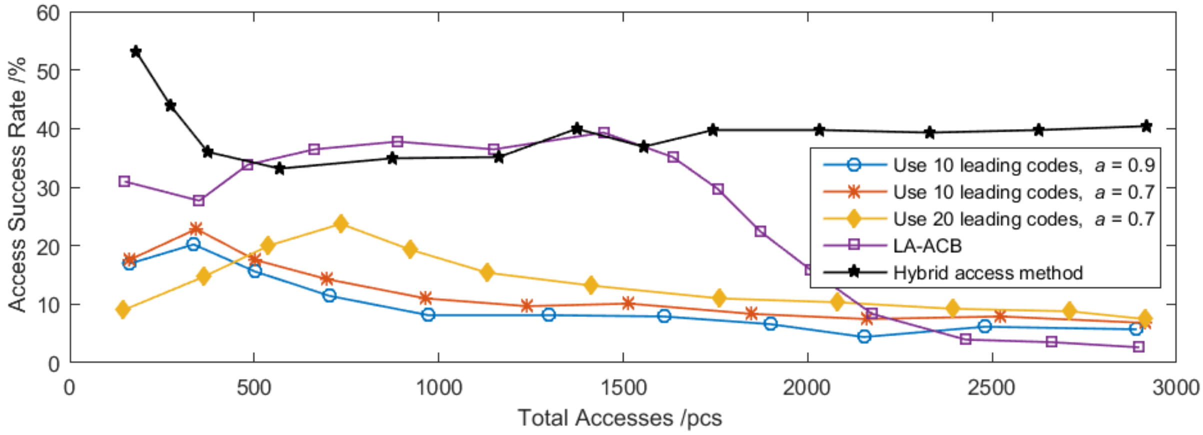

Figure 14: Comparison of access success rates for high-priority services in the beta distribution model, with the hybrid method excelling (42.07% at 4000 applications) through dynamic adjustments and efficient resource use

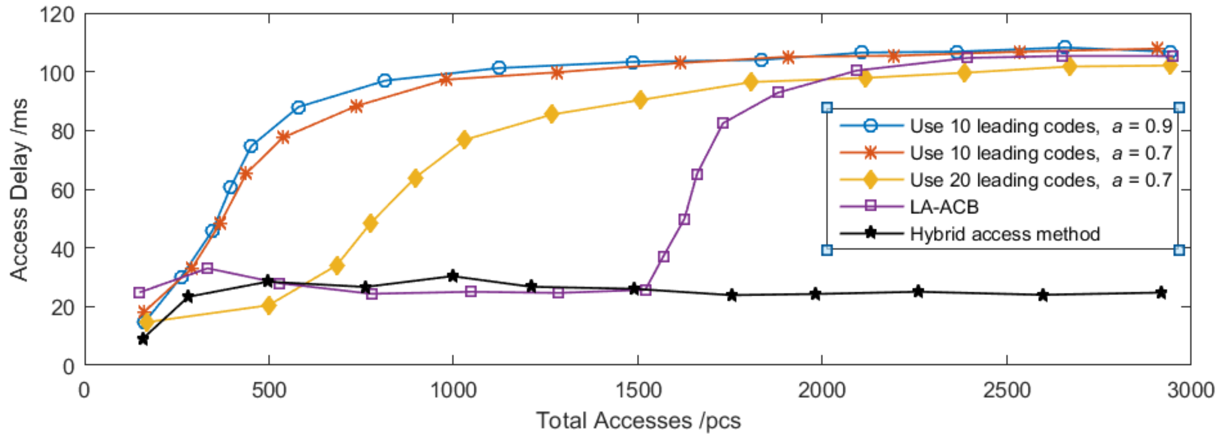

Figure 15: Average access delays for high-priority services in the beta distribution model, with the hybrid method achieving a lower delay (82.02 ms at 4000 applications) than ACB and LA-ACB

Fig. 14 illustrates the access success rate of high-priority services under the beta distribution model. The results indicate that the hybrid access method achieves a higher access success rate compared to the ACB and LA-ACB methods. This improvement is due to the dynamic adjustment of access application amounts and access parameters in the next timeslot, which optimizes the allocation of resources for high-priority services.

Fig. 15 presents the comparison of average access delay for high-priority services using the beta distribution model. The hybrid access method demonstrates a lower average access delay compared to ACB and LA-ACB methods. This reduction in delay is attributed to the method's ability to better predict and manage access requests, thereby minimizing the waiting time and improving overall efficiency.

These results highlight the advantages of the hybrid access method in managing high-priority service requests, ensuring higher access success rates, and reducing average access delays under the beta distribution model. This demonstrates the method's effectiveness in handling dynamic and bursty traffic patterns in large-scale IoT environments.

The total number of system preamble codes is 60. When high-priority services are concurrent with low-priority services, the access success rate is shown in Fig. 16, and the average access delay is shown in Fig. 17.

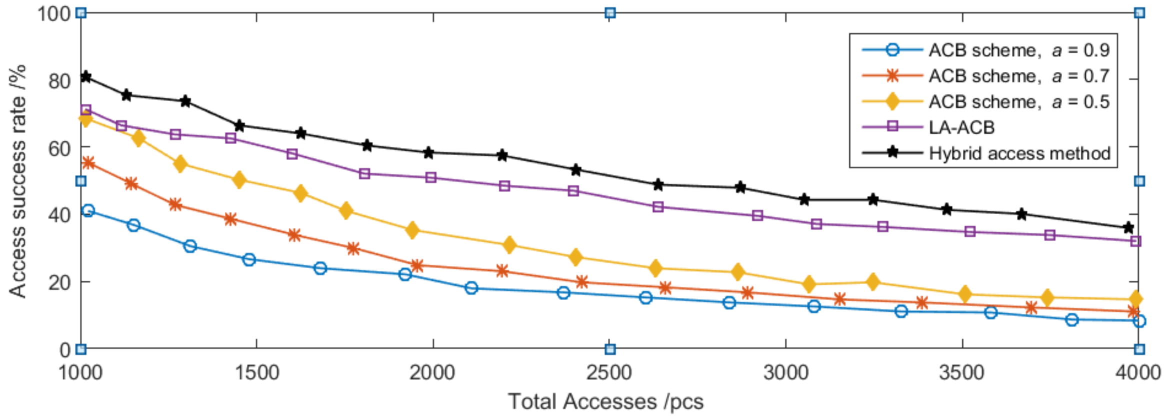

Figure 16: Access success rate for concurrent services in the beta distribution model, with the hybrid method achieving 42.07% at 4000 applications, surpassing ACB and LA-ACB

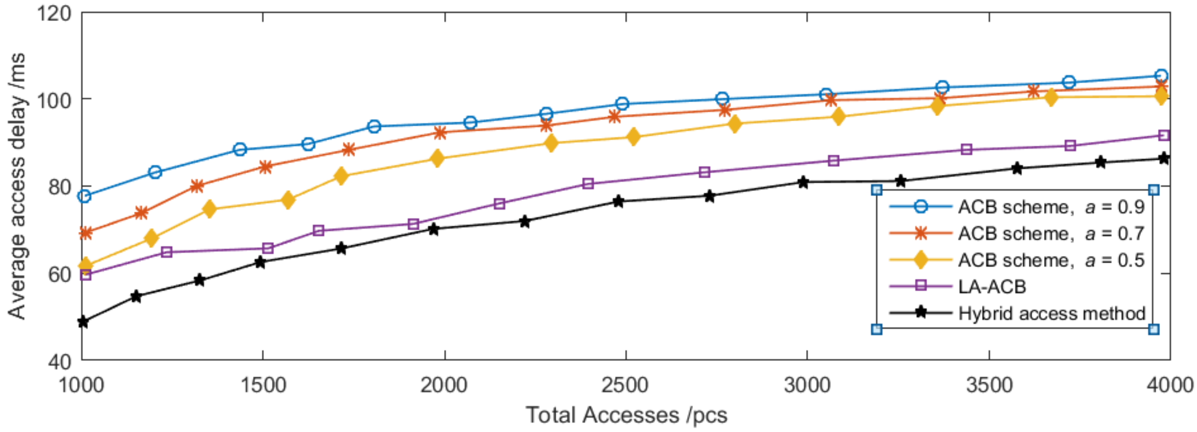

Figure 17: Average access delay for concurrent services in the beta distribution model, with the hybrid method achieving 82.02 ms, outperforming ACB and LA-ACB

Figure 16 shows access success rate for concurrent services under the beta distribution model. The hybrid access method outperforms ACB and LA-ACB methods, achieving a success rate of 42.07% at 4000 applications, demonstrating robust handling of burst traffic. Figure 17 illustrates average access delay for concurrent services under the beta distribution model. The hybrid access method reduces delay to 82.02 ms at 4000 applications, ensuring better performance for high-priority and time-sensitive services. In fact, it is these very measures of performance that represent important favorable points for the proposed hybrid model over conventional algorithms like ACB and LA-ACB.

The experimental results also reveal that the access success rate and average access delay are significantly improved by the proposed hybrid access method. In addition, it well satisfies the requirements brought by the Quality of Service of high-priority traffic for periodic and bursty large-scale terminal access requests. It enables the method to predict the volume of the access application effectively in the next timeslot in a dynamic way by taking advantage of the historical state of the preamble code, without assuming anything about the quantity of access applications.

The predictability allows for the tailoring of the hybrid access method to the various characterizations of different services, hence optimality in the choice of access methodologies. This leads to a substantial increase in system throughput that ensures reliable and efficient communications over large-scale IoT topologies.

In summary, the hybrid access method enhances the performance of the system and also responds to robustness and scalability challenges; hence, it is the best against all the complexities in communications in IoT at a metro railway station. Dynamic adaptability and predictive accuracy make this tool indispensable to maintain the optimum service level and meet the stringently demanding QoS of critical infrastructure.

5 Discussion

The proposed hybrid access scheme constitutes one of the key improvements in congestion management schemes over large-scale IoT networks, especially in highly populated areas such as in metro stations. In the process, SA-ACB merging is targeted at the solution of fundamental issues like low access success rates and high delays in a network. Higher performance indices are promised to be exhibited compared with the existing methodologies LA-ACB and traditional ACB. For instance, under the uniform distribution model, the maximum access success rate reaches 52.43% at 4000 requests, which is far beyond the limitation of LA-ACB owing to the inefficiency of resource utilization when requests are too many. Besides, this approach ensures an average latency of no more than 76.72 ms for high-priority services that strictly meet the QoS requirement. Under correspondence, within the beta distribution model, robustness exposed to bursty traffic by the hybrid approach achieved 42.07% in success rate and 82.02 milliseconds average delay.

Those advantages come forth due to novelty in resource allocation and predictive adjustments that this hybrid method will implement. The method dynamically adapts the ACB control parameters in view of historical data and real-time estimation to optimize channel utilization with minimum collision. It efficiently spreads the network load in a dual-access approach wherein small data services are managed by SA and large delay-sensitive services are overseen by ACB. This flexibility is a key ingredient for achieving high scalability and reliability, especially under scenarios that exhibit diversified traffic patterns where high-priority applications must coexist with low-priority ones.

The practical implications of these findings are huge. Hybrid should guarantee environments like metro stations with very low latency and high access success ratios, dependably surveilling the very important equipment of elevators and escalators, while improving operational safety and efficiency. Besides, this solution also provides a scalable and economically feasible way to handle congestion in IoT networks, thus making it suitable for smart city, industrial automation, and, generally speaking, high-traffic IoT systems. Future works may further optimize the proposed approach for energy efficiency and extend its applicability to realistic traffic for further generalization. These results have established the hybrid access method as a robust and practical solution to handle congestion in large-scale IoT networks.

6 Conclusion

The paper proposed an IoT-based congestion management strategy for mass data access from the elevator terminals at the metro station. This method categorized the business data by volume and latency requirements and adopted SA for delay-tolerant services and ACB for real-time services. Besides, in the proposed methodology, dynamically adjusting ACB control parameters was adopted to optimize the access efficiency for terminals. The effectiveness of the approach is corroborated by the simulation results: from a uniform distribution model, based on 4000 access requests, the hybrid method can achieve an access success rate of 52.43% and an average access delay of 76.72 ms. From the Beta distribution model, 42.07% with an average access delay of 82.02 ms can be achieved. It is presented that the Hybrid Access Method increases the access success rate greatly and decreases the delay hence fulfilling the QoS requirements for high-priority services in a large-scale IoT environment. Future investigations ought to encompass practical implementation and examine more extensive traffic models, sophisticated prediction methodologies, and scalability to further substantiate and augment the applicability and dependability of the method. Nevertheless, the suggested congestion control approach, primarily corroborated through simulations, may not entirely reflect the intricacies of real-world scenarios and the diverse traffic patterns encountered. Therefore, even the refined uniform and Beta distribution models need further refinement and validation in order to ensure their accuracy against different scenarios. The scalability of the method, especially above 4000 access requests, was not deeply analyzed, as was the application of the method to other IoT applications. It has to be implemented on-site, considering variations in traffic models, advance prediction methods using machine learning techniques, and scalability analysis for performance evaluation. Extension of the method to other IoT applications, investigation of energy efficiency, and incorporating robust security will ensure its sustainability, hence reliable in different IoT environments.

Acknowledgment

Thanks to our families and colleagues who supported us morally.

Funding statement

Not applicable

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Authorship contribution statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Availability of data and materials

On Request

Declarations

Not applicable

References

[1] ShuangChang F, Jie C, Yanbin Z, Zheyi L (2020). Discussion on improving safety in elevator management. In: *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). IEEE*, pp 195–198. https://doi.org/10.1109/MLBDBI51377.2020.00043.

[2] Ushakov D, Dudukalov E, Kozlova E, Shatila K (2022). The Internet of Things impact on smart public transportation. *Transportation Research Procedia*, 63:2392–2400. https://doi.org/10.1016/j.trpro.2022.06.275.

[3] Wang C, Feng S (2020). Research on big data mining and fault prediction based on elevator life cycle. In: *2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). IEEE*, pp 103–107. https://doi.org/10.1109/ICBASE51474.2020.00030.

[4] Yao W, Jagota V, Kumar R, et al (2022). Study and application of an elevator failure monitoring system based on the internet of things technology. *Sci Program*, 2022:2517077. https://doi.org/10.1155/2022/2517077.

[5] Mao J, Chen L, Cheng H, Wang C (2023). Elevator fault diagnosis and maintenance method based on Internet of Things. In: *Proc.SPIE*, p 1279305. https://doi.org/10.1117/12.3006383.

[6] Lai CTA, Jiang W, Jackson PR (2019). Internet of Things enabling condition-based maintenance in elevators service. *J Qual Maint Eng*, 25:563–588. https://doi.org/10.1108/JQME-06-2018-0049.

[7] Mouha RARA (2021). Internet of things (IoT). *Journal of Data Analysis and Information Processing*, 9:77. http://www.scirp.org/journal/Paperabs.aspx?PaperID=108574.

[8] Wang J, Lim MK, Wang C, Tseng M-L (2021). The evolution of the Internet of Things (IoT) over the past 20 years. *Comput Ind Eng*, 155:107174. https://doi.org/10.1016/j.cie.2021.107174.

[9] Song T, Cai J, Chahine T, Li L (2021). Towards Smart Cities by Internet of Things (IoT)—a Silent Revolution in China. *Journal of the Knowledge Economy*, 12:1–17. https://doi.org/10.1007/s13132-017-0493-x.

[10] Ullah S, Radzi RZ, Yazdani TM, et al (2022). Types of Lightweight Cryptographies in Current Developments for Resource Constrained Machine Type Communication Devices: Challenges and Opportunities. *IEEE Access*, 10:35589–35604. https://doi.org/10.1109/ACCESS.2022.3160000

[11] Mahmood NH, Alves H, López OA, et al (2020). Six Key Features of Machine Type Communication in 6G. In: *2020 2nd 6G Wireless Summit (6G SUMMIT),*. pp 1–5. https://doi.org/10.1109/6GSUMMIT49458.2020.9083794.

[12] Chou CM, Huang CY, Chiu C-Y (2013). Loading prediction and barring controls for machine type communication. In: *2013 IEEE International Conference on Communications (ICC). IEEE*, pp 5168–5172. https://doi.org/10.1109/ICC.2013.6655404.

[13] Zhang L, He C, Peng Y, et al (2023). Multi-UAV Data Collection and Path Planning Method for Large-Scale Terminal Access. *Sensors*, 23:8601. https://doi.org/10.3390/s23208601.

[14] Varsha V, Prakash SPS, Krinkin K (2024). An Intelligent Bayesian Inference Based Learning Automaton Approach for Traffic Management in Radio Access Network. *Wirel Pers Commun*, 135:233–260. https://doi.org/10.1007/s11277-024-10943-5.

[15] Piao Y, Lee T-J (2024). Integrated 2–4 Step Random Access for Heterogeneous and Massive IoT Devices. *IEEE Transactions on Green Communications and Networking*, 8:441–452. https://doi.org/10.1109/TGCN.2023.3322539

[16] Yu B, Cai Y, Wu D (2021). Joint Access Control and Resource Allocation for Short-Packet-Based mMTC in Status Update Systems. *IEEE Journal on Selected Areas in Communicati*ons, 39:851–865. https://doi.org/10.1109/JSAC.2020.3018801.

[17] Bui A-TH, Nguyen CT, Thang TC, Pham AT (2019). A comprehensive distributed queue-based random-access framework for mMTC in LTE/LTE-A networks with mixed-type traffic. *IEEE Trans Veh Technol*, 68:12107–12120. https://doi.org/10.1109/TVT.2019.2949024.

[18] Cui Y, Liu F, Jing X, Mu J (2021). Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges. *IEEE Netw*, 35:158–167. https://doi.org/10.1109/MNET.010.2100152.

[19] Zhao L, Xu X, Zhu K, et al (2018). QoS-based dynamic allocation and adaptive ACB mechanism for RAN overload avoidance in MTC. In: *2018 IEEE Global Communications Conference (GLOBECOM). IEEE*, pp 1–6. https://doi.org/10.1109/GLOCOM.2018.8647599

[20] Sari RF, Harwahyu R, Cheng R-G (2020). Load Estimation and Connection Request Barring for Random Access in Massive C-IoT. *IEEE Internet Things J*, 7:6539–6549. https://doi.org/10.1109/JIOT.2020.2968091.

[21] He H, Ren P, Du Q, Sun L (2015). Estimation based adaptive ACB scheme for M2M communications. In: *Wireless Algorithms, Systems, and Applications: 10th International Conference, WASA 2015, Qufu, China, August 10-12, 2015, Proceedings 10. Springer*, pp 165–174. https://doi.org/10.1007/978-3-319-21837-3_17.

[22] Zhai D, Lu Y, Shi R, Ji Y (2022). Large-Scale Micro-Power Sensors Access Scheme Based on Hybrid Mode in IoT Enabled Smart Grid. In: *2022 7th International Conference on Signal and Image Processing (ICSIP). IEEE*, pp 719–723. https://doi.org/10.1109/ICSIP55141.2022.9886684.

[23] Liu G, Jiang X, Li H, et al (2022). Adaptive access selection algorithm for large-scale satellite networks based on dynamic domain. *Sensors*, 22:5995. https://doi.org/10.3390/s22165995.