Preamble Design and Collision Resolution in a Massive Access IoT System

Ailing Zhong 1,2,3,*, Zhidu Li 1,2,3, Ruyan Wang 1,2,3, Xingjie Li 1 and Boren Guo 4

1 School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; lizd@cqupt.edu.cn (Z.L.); wangry@cqupt.edu.cn (R.W.);
2018210145@stu.cqupt.edu.cn (X.L.)
2 Key Laboratory of Optical Communication and Networks, Chongqing 400065, China
3 Key Laboratory of Ubiquitous Sensing and Networking, Chongqing 400065, China
4 School of Information Engineering and Communication, Beijing University of Posts and Telecommunications, Beijing 100876, China; plume@bupt.edu.cn
* Correspondence: S180101104@stu.cqupt.edu.cn; Tel.: +86-188-8399-3984

Abstract: How to support massive access efficiently is one of the challenges in the future Internet of Things (IoT) systems. To address such challenge, this paper proposes an effective preamble collision resolution scheme to sustain massive random access (RA) for an IoT system. Specifically, a new sub-preamble structure is first proposed to reduce the preamble collision probability. To identify different devices that send the same preamble to the gNB on the same physical random access channel (PRACH), a multiple timing advance (TA) capturing scheme is then proposed. Thereafter, an RA scheme is designed to sustain massive access and the performance of the scheme is studied analytically. Finally, the effectiveness of the proposed RA scheme is validated by extensive simulation experiments in terms of preamble detection probability, preamble collision probability, RA success probability, resource efficiency and TA capturing.

Keywords: random access; preamble collision resolution; timing advance capturing; Internet of Things

1. Introduction

The rapid evolution of the wireless communication technologies are inviting all human beings to the era of the Internet of Everything [1]. The future IoT system is required to guarantee diverse requirements for various applications, such as huge bandwidth, low latency, high reliability, etc. [2,3]. As a result, huge amount of IoT devices need to access to the wireless networks to exchange data, which brings new challenge to the wireless network design [4]. Specifically, convention network scheduling mechanism is mainly designed for human to human communications that cannot sustain a large amount of simultaneous random access requests, since the preamble resources are quite limited. How to support massive random access for IoT devices consequently becomes a critical issue to realize the future Internet of Everything [5,6].

In general, existing research focuses on how to optimize the high-layer performance of an IoT system through resource allocations [7–12]. Such performance metrics include throughput, latency, access capacity, etc. The resource allocation schemes, however, can be realized only if the devices can be access to the IoT system. Hence, random access guarantee is one of the key issues to support massive devices to send data at the same time. Typically, more RA-attempting devices may lead to heavier load and more preamble collisions, which degrades the performance of an IoT system. Thus, it is necessary to design new preamble structure and RA scheme to reduce the congestion for the physical random access channel (PRACH). In the literature, the idea of performance improvement on PRACH can be generalized into three categories, i.e., resource increasing [13], user scheduling [14], and
resource reusing [15]. In fact, resource increasing is costly and impractical, since the total resource of an IoT system is usually limited. In addition, preamble collision probability cannot be reduced through user scheduling since the number of preambles in an IoT system is finite. Differently, the idea of resource reusing can allow multiple devices to send identical preambles on the same PRACH, which is applicable to the massive access scenario. However, how to design a preamble structure and an RA scheme to efficiently reuse the PRACH resources for a massive access IoT system is still an open problem.

Motivated by this, this paper proposes an effective preamble collision resolution scheme to sustain massive random access for an IoT system. A new preamble structure consisting of the shared sub-preamble and the dedicated sub-preamble is first designed. Based on the new designed preamble, a TA capturing scheme is then proposed to identify multiple devices that send identical sub-preambles to the gBN on the same PRACH. Thereafter, we propose a random access scheme to reduce the preamble collision probability. Furthermore, we prove analytically that the proposed RA scheme can guarantee better performance than the conventional one proposed by the 3GPP. In addition, the effectiveness of the proposed scheme is validated by the extensive simulations. The contributions of this paper can be summarized as follows.

- A new preamble structure is designed, where preamble diversity and location diversity are both acquired. The designed preamble is able to guarantee a much lower preamble collision probability than the conventional preamble designed by 3GPP with ignorable detection impairment.
- A multiple TA capturing scheme is proposed based on the designed preamble. With the idea that the sub-preambles sent by a device suffer from the same channel environment, the TA value of each device can be captured, which can be further used to device identification.
- A new random access scheme is proposed and studied analytically. It is verified that the proposed scheme outperforms the conventional scheme in terms of preamble collision probability, RA success probability, etc. Hence, the proposed scheme is more applicable to the IoT systems where massive access is requested.

The remaining of this paper is organized as follows. Section 2 introduces the related works. Section 3 introduces new preamble sequence design. In Section 4, the preamble collision resolution scheme based on the proposed preamble structure is studied. In Section 5, the advantage of the proposed random access scheme is analyzed. Simulation results are presented and discussed in Section 6. Finally, the paper is concluded in Section 7.

2. Related Work

In the literature, performance improvement for massive access IoT systems is usually carried out from two aspects, i.e., cross-layer resource optimization and physical layer design.

On cross-layer resource optimization, researchers usually focused on performance optimization in terms of throughput, latency, and capacity. In [7], a throughput-oriented non-orthogonal random access (NORA) scheme was proposed to improve the access throughput for IoT networks by employing the technique of tagged preambles. With this, multiple devices shared the same physical uplink shared channel (PUSCH) for transmissions by multiplexing in the power domain. Wu et al. proposed a hybrid NORA and data transmission scheme to address the signalling overhead and resource allocation problems of IoT with non-orthogonal multiple access [8]. A relaying threshold-based random access and data transmission scheme for grouped IoT networks was proposed in [9]. Except for throughput optimization in [7–9], Chen et al. proposed an access class barring mechanism to accommodate heterogeneous IoT devices with different quality of service (QoS) requirements based on deep reinforcement learning algorithm in [10]. Additionally, a multi-slot pilot allocation scheme with unequal access latency protection for user equipments (UEs) was studied in [11]. The two solutions proposed in [10,11] devoted to reduce the latency in IoT networks. Moreover, in [12], the receiver is designed for the
uplink transmission of a massive MIMO system to improve network access gain, where human-type communications and IoT communications coexisted.

On physical layer design, researchers usually focus on collision resolution. In [16], the multivariate Bernoulli model which exhibited the distributions of pilot states was proposed. With this, general correlation among device activities and relationship among the states of pilots assigned to one device could be specified. He et al. proposed a protocol which implemented multi-user detection and channel estimation by compressed sensing technology [17]. This protocol ensured that multiple collided packets could be recovered in one single time slot with low signaling overhead. The estimation of round-trip delays of multiple signals in NORA based on the maximum likelihood criterion was studied to assist the detection of collision in an IoT network [18]. Additionally, a contention control method that regulated the number of devices competing simultaneously for the RA channel was studied in a distributed manner [19]. Moreover, an RA scheme was proposed in [20], which introduced virtual preambles and associated them with RA channel indexes to discern multiple low-cost IoT devices. Works [16–20] all devoted to collision probability reduction. However, there was little attention on preamble collision in the physical layer of IoT. Please note that preamble detection is the prime of a successful random access.

Recently, preamble collision resolution was studied under the assumption that the TA values of the IoT devices were fixed [15]. Specifically, the TA value of each device was unique and the colliding devices could be discerned by the gNB. However, as the number of available preambles was finite, the scheme proposed in [15] was not applicable to the massive access scenario. Therefore, how to reduce preamble collision is still a challenge for an IoT system where huge amount of devices request access simultaneously, which motivates this paper.

3. Preamble Design

In this section, the conventional contention-based RA procedure and the structure of conventional preamble sequence are first introduced respectively. To address the drawback of conventional preamble sequence, a new sub-preamble sequence design is then proposed.

3.1. Conventional RA Procedure

The RA procedure includes two manners, i.e., contention-based manner and non-contention-based RA manner [21]. Since preamble collisions only occur in contention-based manner, this paper mainly focuses on contention-based manner. As depicted in Figure 1, a conventional RA procedure consists of four steps [22].

![Figure 1. Conventional RA procedure.](image)

- Step-1: A device randomly selects a preamble from the available preamble sequence set, and then transmits the selected preamble on the available PRACH slot. If different devices transmit identical preambles on the same PRACH resources, RA contention will occur.
• Step-2: The gNB sends Random Access Response (RAR) message to the device after it detects a received preamble and captures the corresponding TA value. The RAR message contains the TA command, an uplink resource grant for MSG3 and a temporary cell-radio network temporary identifier (C-RNTI) for the device.

• Step-3: After receiving the RAR message, the device sends MSG3 to the gNB. MSG3 includes Radio Resource Control (RRC) connection request, scheduling request, and temporary C-RNTI. If preamble collision occurs, the gNB will be unable to decode the MSG3 data correctly.

• Step-4: If the gNB decodes the MSG3 successfully, it sends back the contention resolution message including RNTI acquired from the decoded MSG data. If a device receives its own RNTI within corresponding time window, RA procedure is accomplished successfully. Otherwise, the device restarts a new RA procedure.

Please note that in step-2, the gNB can only detect a specific preamble, but it cannot judge whether the detected preamble is sent by a single device or multiple devices. In addition, the preamble collisions cannot be detected before step-3. Once the collision occurs, it leads to a long RA delay and a waste of PUSCH resource.

3.2. Conventional Preamble

Zadoff-Chu (ZC) sequences are used as the root sequences of preambles due to its desirable periodic correlation. According to [23], a ZC root sequence with index \( u \) is defined as follows.

\[
x_u(n) = \exp\left[-j \frac{\pi un(n+1)}{N_{ZC}}\right], 0 \leq n \leq N_{ZC} - 1,
\]

where \( N_{ZC} \) denotes the length of the ZC root sequence. For the \( u \)th root ZC sequence, the corresponding preamble with zero correlation zones of length \( N_{CS} - 1 \) is defined by cyclic shifts

\[
x_{u,v}(n) = x_u((n + C_v) \mod N_{ZC}),
\]

where \( N_{CS} \) and \( C_v \) are given by [23]. A set of available preambles are generated from one or two root sequences by cyclic shifts.

Then, the received sequence in the gNB can be expressed by

\[
y(n) = \sum_{k=1}^{n_k} \sum_{e=1}^{n_{xe}} h_{ke} x_k[(n + t_{ke}) \mod N_{ZC}] + z(n),
\]

where \( h_{ke} \) denotes the \( e \)th multi-path channel coefficient for the \( k \)th device, \( t_{ke} \) denotes the delay of the \( e \)th multi-path for the \( k \)th device, \( n_k \) denotes the total number of devices which attempt RAs on the same PRACH, \( n_{xe} \) denotes the amount of multi-path for the \( k \)th device, and \( z(n) \) denotes the circular symmetry complex Gaussian noise that can be expressed as

\[
z(n) = z_r(n) + jz_i(n),
\]

where \( z_r(n) \) and \( z_i(n) \) both follow Gaussian distribution.

3.3. Sub-Preamble Sequence Design

To address the drawback of the conventional preamble structure that easily leads to RA collision in a massive connective scenario, this paper proposes a new preamble structure with more diversities to reduce the RA collision probability. Specifically, the new preamble is consists of two sub-preambles and an empty sequence, as depicted in Figure 2. The sub-preambles are known as shared sub-preamble and dedicated sub-preamble respectively. The location of the shared sub-preamble is fixed. The location of the dedicated sub-preamble includes two types, which are denoted by \( \mu \in \{1, 2\} \), as depicted in Figure 2. In addition, we use subscript \( i \in \{s, d\} \) to represent the variables corresponding to the shared sub-preamble and dedicated sub-preamble respectively.
The $n_i$th sub-preamble sequence with root index $u_i$ and cyclic shift index $v_i$ is generated as follows.

$$x_{u_i,v_i}(n_i) = x_{u_i}((n_i + C_{v_i}) \mod N_i), 0 \leq n_i \leq N_i - 1,$$

where $N_i$ denotes the length of the $n_i$th sub-preamble sequence, and $x_{u_i}$ denotes ZC root sequence that is obtained by Equation (2). As a result, the set of available sub-preamble can be obtained as

$$S_i = \{x_{u_i,v_i,0}, x_{u_i,1}, ..., x_{u_i,N_i}\}, i \in \{s, d\},$$

or

$$x_{u_i,v_i,0}^{d}(n) = [x_{u_i,v_i}(n_s), g_1(n_{g_1}), x_0(n_0), g_2(n_{g_2}), x_{u_i,v_i}(n_d)], 0 \leq n_i \leq L_i, 0 \leq n_{g_1} \leq L_{g_1},$$

or

$$x_{u_i,v_i,0}^{s}(n) = [x_{u_i,v_i}(n_s), g_1(n_{g_1}), x_{u_i,v_i}(n_d), g_2(n_{g_2}), x_0(n_0)], 0 \leq n_i \leq L_i, 0 \leq n_{g_1} \leq L_{g_1},$$

where $x_{u_i,v_i}(n_s)$ and $x_{u_i,v_i}(n_d)$ represent the shared sub-preamble and dedicated sub-preamble respectively, $g_m(n_{g_m}) m \in \{1, 2\}$ denotes Guard Interval (GI), $x_0(n_0) = 0^{L_0}$ $(0 \leq n_0 \leq L_0)$ denotes an empty sequence with length $L_0$. Besides, $L_i$ denotes the length of the sub-preambles, $L_{g_i}$ denotes the length of the GI and $\alpha$ represents the transmission power. The preamble sequence generation scheme is summarized as Algorithm 1.

Without loss of reliability, generality and fairness, the following guidelines are considered: (1) The length of each sub-preamble sequence, i.e., $L_i$, should be a prime number to guarantee the periodic correlation of ZC root sequences. (2) The length of each sub-preamble should be selected as long as possible so that ZC root sequences can offer a good performance for TA capturing. (3) The length of the dedicated sub-preamble should be identical to the length of the empty sequence. (4) The total length of sub-preambles and GIs should be no longer than the bandwidth of PRACH that is equal to 839 or 139 according to different scenarios.
### Algorithm 1 Preamble generation scheme

**Require:**
- The sets of available sub-preambles $S_s, S_d$; location index for the dedicated sub-preamble $\mu$;

**Ensure:**
- A preamble sequence for an RA-attempting device $x_{u,v}^{\mu}(n)$;

1. Randomly select a sub-preamble sequence from $S_s$ as a shared sub-preamble, $x_{u,v}^{s}(n)$;
2. Randomly select a sub-preamble sequence from $S_d$ as a dedicated sub-preamble, $x_{u,v}^{d}(n)$;
3. Randomly select a location index for the dedicated sub-preamble, $\mu$;
4. if $\mu = 1$ then
   5. Generating a preamble according to Equation (7);
   6. else
   7. Generating a preamble according to Equation (8);
   8. end if

### 4. Preamble Collision Resolution Scheme

Based on our proposed preamble structure, this section studies how to reduce the preamble collision. The sub-preamble sequence reception scheme is first proposed. Then, we introduce how to capture multiple TA values in a single RA procedure. Thereafter, an integrated preamble collision resolution scheme is studied.

#### 4.1. Sub-Preamble Sequence Reception

After sequence $y(n)$ is received by the gNB, $y(n)$ should be sliced into a shared sub-preamble and a dedicated sub-preamble. Since the sub-preamble structure is available to both IoT devices and gNB, the shared sub-preamble is apt to be obtained in the light of its predetermined length and location. Please note that the gNB is unable to recognize the dedicated sub-preamble, since the location index is random. However, the TA values obtained from the two types of the sub-preamble sequences sent by the same device should be identical, since those sub-preambles suffered from the same channel delay. Hence, for a specific shared sub-preamble, the corresponding dedicated sub-preamble can be identified via the TA values. Accordingly, the sub-preamble sequence reception scheme is summarized as Algorithm 2.

Please note that when multiple devices transmit their preambles on the same PRACH, $y_s$ consists of multiple shared sub-preamble sequences. Likewise, $y_d$ contain multiple dedicated sub-preamble sequences when the corresponding $\mu$ is configured identically. How to identify those overlapping sub-preamble sequences for different devices will be introduced in the next subsection.
Algorithm 2 Preamble sequence reception scheme

Require:
The received preamble sequence in the gNB \( y(n) \);
The length of the sub-preamble \( L_i \);
The length of the GI \( L_g \);
The length of the empty sequence \( L_0 \);
The sets of the local sub-preamble sequences \( S_i \);

Ensure:
The received shared sub-preamble sequence \( y_s(n_s) \);
The received dedicated sub-preamble sequence \( y_d(n_d) \);

1: Intercept the first \( L_s \) elements of \( y(n) \) as \( y_s(n_s) \);
2: Divide the residual sequence in the middle and obtain two sequences with the equal length \( L_d + L_g \), i.e., \( y_{d,\text{candidate}1}(n_d) \) and \( y_{d,\text{candidate}2}(n_d) \);
3: Remove the GI from \( y_{d,\text{candidate}1}(n_d) \) to obtain \( y_{d,\text{candidate}1}(n_d) \);
4: Output \( y_d(n_d) = y_{d,\text{candidate}1}(n_d) \) for \( \mu = 1 \);
5: Remove the GI from \( y_{d,\text{candidate}2}(n_d) \) to obtain \( y_{d,\text{candidate}2}(n_d) \);
6: Output \( y_d(n_d) = y_{d,\text{candidate}2}(n_d) \) for \( \mu = 2 \);

4.2. Multiple TA Values Capturing Method

Based on the sub-preamble sequence reception, the gNB can correlate each sub-preamble with the local sequences respectively. The value of correlation calculation with a delay for the sub-preamble is expressed as

\[
d_i(\tau_i) = \frac{1}{N_{ZC}} \sum_{n_i=0}^{L_i-1} |y_i(n_i) \cdot c_i^*((n_i - \tau_i) \mod L_i)|^2, \quad 0 \leq n_i \leq L_i - 1, \quad i \in \{s, d\} \tag{9}
\]

where \( \tau_i \in \{0, 1, 2, ..., L_i - 1\} \), \( c_i(n_i) \in S_i \) denotes local sub-preamble sequence, and \((\cdot)^*\) denotes the complex conjugate. Additionally, the estimation TA value is related to the peak location of correlation, which can be expressed as

\[
\hat{\tau}_i = \arg \max_{\tau_i} \{d_i(\tau_i) \geq r_{th}\}, \tag{10}
\]

where \( r_{th} \) denotes the pre-determined threshold.

After the gNB detects the received shared sub-preambles, a set of TA values \( \Gamma_s = \{\tau_s^1, \tau_s^2, ..., \tau_s^\theta\} \) can be captured, where \( \theta \) denotes the number of detected shared sub-preambles. Similarly, another set of TA values \( \Gamma_d = \{\tau_d^1, \tau_d^2, ..., \tau_d^\sigma\} \) can be obtained, where \( \sigma \) represents the number of detected dedicated sub-preambles. For ease of description, in the subsequence, the TA values in set \( \Gamma_s \) are defined as shared TA values and those in set \( \Gamma_d \) are defined as dedicated TA values. If \( \tau_s = \tau_d \), the corresponding shared sub-preamble and dedicated sub-preamble are considered to belong to the same devices. This is because the shared sub-preamble and dedicated sub-preamble that sent by the same device always suffer from the same channel delay. Consequently, the gNB can identify preambles sent from different devices through comparing the shared TA value and the dedicated TA value.
Figure 3 shows an example of capturing two TA values for two different devices. In the example, two RA-attempting devices send the shared sub-preambles with the same index $v_s$ and different dedicated sub-preambles on the same PRACH. Therefore, the gNB receives an overlapping preamble sequence. During the detection of the shared sub-preamble, two TA values can be captured by Equation (10) because of the same shared sub-preamble index. Please note that the gNB is unable to distinguish which TA value belongs to which device. However, during the detection of two dedicated sub-preambles, two TA values can be captured respectively. In detail, the TA values of the two shared sub-preambles $\tau_s^1$ and $\tau_s^2$ are captured on an identical detection zone, while the TA values of the two dedicated sub-preambles $\tau_d^1$ and $\tau_d^2$ are captured on two different detection zones. Hence, the gNB can identify a device by finding $\tau_s^1 = \tau_d^1$ or $\tau_s^2 = \tau_d^2$. The scheme of capturing multiple TA values is summarized as Algorithm 3.

Based on Algorithm 3, the TA value of each shared sub-preamble can be obtained by the gNB. If shared sub-preamble collision occurs, the TA value can still be captured with the corresponding dedicated sub-preamble. Only when the shared sub-preamble collision and dedicated sub-preamble collision occur simultaneously, the TA value is unable to be captured. Please note that sub-preamble collision implies more than one devices send dedicated sub-preambles with the same sub-preamble index on the same time-frequency resource. Hence, the probability that shared sub-preamble collision and dedicated sub-preamble collision occur simultaneously is low.
Algorithm 3 Multiple TA value capturing scheme.

Require:
Received shared sub-preamble sequence $y_s(n_s)$;
Received dedicated sub-preamble sequence $y_d(n_d)$;
The sets of the local sub-preamble sequences $S_i$;

Ensure:
Captured TA value for the $k$th device $\tau^k$;
The shared sub-preamble index for the $k$th device $v^k_s$;
The dedicated sub-preamble index for the $k$th device $v^k_d$;
The location index for the $k$th device $\mu^k$;

1: for all $c_s(n_s) \in S_s$ do
2:  Calculate $d_s(\tau_s)$ for the received shared sub-preamble $y_s(n_s)$ according to Equation (9);
3:  if $d_s(\tau_s) \geq r_{bs}$ then
4:     Record the sub-preamble index $v_s$ of $c_s(n_s)$ in $Y_s$;
5:     Calculate $\tilde{\tau}^1_s, ..., \tilde{\tau}^\theta_s$ according to Equation (10), and record them in $\Gamma_s$;
6:  end if
7: end for
8: for all $\mu \in \{1,2\}$ do
9:  for all $c_d(n_d) \in S_d$ do
10:     Calculate $d_d(\tau_d)$ for the received dedicated sub-preamble $y_d(n_d)$ according to Equation (9);
11:     if $d_d(\tau_d) \geq r_{bd}$ then
12:        Record the sub-preamble index $v_d$ of $c_d(n_d)$ in $Y_d$;
13:        Calculate $\tilde{\tau}^1_d, ..., \tilde{\tau}^\sigma_d$ according to Equation (10), and record them in $\Gamma_d$;
14:    end if
15: end for
16: end for
17: for all $\tilde{\tau}^k_s \in \Gamma_s$ do
18:    for all $\tilde{\tau}^k_d \in \Gamma_d$ do
19:        if $\tilde{\tau}^k_s = \tilde{\tau}^k_d$ then
20:           Output $\tau^k = \tilde{\tau}^k_s$ and the corresponding $v^k_s, v^k_d$ and $\mu^k$;
21:        end if
22:    end for
23: end for
4.3. Random Access Scheme

Based on the multiple TA value capturing scheme, the random access scheme for multiple devices is revised as follows.

- **Step-1:** Sub-preambles selection and transmission. An IoT device randomly selects a shared sub-preamble and a dedicated sub-preamble from the available sub-preamble sets $S_s$ and $S_d$ respectively. Applying Algorithm 1, the device randomly selects a location index $\mu$ to generate a preamble sequence and then transmits it to the gNB on an available PRACH slot.

- **Step-2:** Random access response. When the gNB receives the preambles from multiple devices, applying Algorithm 2 to obtain the shared sub-preambles and dedicated sub-preambles respectively. Then, applying Algorithm 3, the gNB captures the TA value for each device and generates the corresponding RAR messages. The RAR message contains the TA command, uplink resource grant for MSG3, the shared sub-preamble index, the dedicated sub-preamble index and the location index $\mu$.

- **Step-3:** MSG3 transmission. Devices can obtain their own RAR messages by confirming the shared sub-preamble index, the dedicated sub-preamble index and the location index $\mu$. If RAR message is decoded correctly, the corresponding device will transmit MSG3 data on the PUSCH resource blocks that are allocated at step-2. As a result, the gNB can recognize that an RA collision occurs in this case.

- **Step-4:** Contention resolution. If the gNB successfully decodes MSG3, it will transmit a contention resolution message to the corresponding device. Such message contains the device ID that includes sub-preamble indexes and location index. Otherwise, nothing will be sent to the device. If the device receives contention resolution message, it will send back an ACK message and complete RA procedure. Otherwise, the device will restart a new RA procedure.

Under the proposed RA procedure, the preamble collision occur only if the shared and dedicated sub-preamble collisions occur at the same time. Additionally, the proposed RA scheme can still work when one type of the sub-preamble collision occur. In the next section, the effectiveness of the proposed RA scheme will be proved analytically.

5. Performance Analysis

In this section, the RA collision probability and the RA success probability under the conventional and proposed schemes are derived analytically.

5.1. Collision Probability

Under the proposed scheme, the preamble collision probability is associated with the probabilities of sub-preamble collisions as well as location collision of the dedicated sub-preambles. In other words, the preamble collision occurs when multiple devices select the same shared sub-preambles, the same dedicated sub-preambles as well as the same location indexes at the same time. Assuming that $n_k$ devices attempt RAs with $n_p$ available preambles. Consider a target device with deterministic preamble selection, the probability that another device users the same preamble holds as $1/n_p$. Since the devices are independent of each other, the preamble collision probability for the target device under the conventional RA [24] scheme holds as

$$P_{\text{collision}}^c = 1 - (1 - \frac{1}{n_p})^{n_k-1}. \quad (11)$$

For the proposed RA scheme, we assume that $n_k$ devices attempt random access with $n_{p,s}$ shared sub-preambles and $n_{p,d}$ dedicated sub-preambles, and the number of available location index is $n_\mu$. For the target device and another device, if the shared sub-preamble, dedicated sub-preamble and location index used by those two devices are all the same, preamble collision will occur for the target device. The probability that another device uses
identical shared sub-preamble, dedicated sub-preamble and location index at the same time holds as

\[ P_\Delta = \frac{1}{n_p,s} \cdot \frac{1}{n_p,d} \cdot \frac{1}{n_p} = \frac{1}{n_p n_p d n_p} \]  

(12)

Hence, the preamble collision probability for the target device under the proposed RA scheme holds as

\[ P_{\text{collision}} = 1 - (1 - P_\Delta)^{n_k - 1} = 1 - (1 - \frac{1}{n_p n_p d n_p})^{n_k - 1} \]  

(13)

Comparing Equation (11) with Equation (13), when the number of available preambles in conventional RA scheme is equal to the number of available sub-preambles in the proposed RA scheme, i.e., \( n_p = n_p s, n_p = n_p d \), the proposed scheme can guarantee lower collision probability than the conventional scheme.

5.2. RA Success Probability

During a complete RA procedure, the gNB allocates PUSCH resources for MSG3 transmission after successful preamble detection and TA capturing. For a conventional RA procedure, there are 4 cases for preamble collision and PUSCH resource allocation:

1. \( p_{c1} = Pr\{\text{preamble collision, PUSCH resource allocated}\} \);
2. \( p_{c2} = Pr\{\text{preamble collision, PUSCH resource unallocated}\} \);
3. \( p_{c3} = Pr\{\text{no preamble collision, PUSCH resource unallocated}\} \);
4. \( p_{c4} = Pr\{\text{no preamble collision, PUSCH resource allocated}\} \);

where \( p_{c4} \) represents the RA success probability in conventional scheme. Typically, the RA success probability in the conventional scheme is defined as the probability that a device uses an exclusive preamble and is allocated an exclusive PUSCH resource. Besides, \( p_{c1} \) represents the probability that the PUSCH resources are wasted. Moreover, the RA failure probability in conventional scheme is equal to \( p_{c1} + p_{c2} + p_{c3} \). Consequently, the RA success probability under conventional scheme can be expressed as

\[ P_{\text{c,suc}} = p_{c4} = (1 - P_{\text{collision}}) \cdot P_{\text{c, res}} = (1 - \frac{1}{n_p})^{n_k - 1} \cdot P_{\text{c, res}} \]  

(14)

where \( P_{\text{c, res}} \) denotes the successful resource allocation rate for RA-step 3 under conventional scheme. Specifically, the PUSCH resources will be allocated to the IoT device in the RA-step 3 if its preamble is detected by the gNB successfully. Therefore, the successful resource allocation rate for RA-step 3 under the conventional scheme holds as

\[ P_{\text{c, res}} = \min\{1, \frac{n_{\text{PUSCH}}}{Pr\{\xi_m \geq 1\} \cdot \min\{n_k, n_p\}}\} \]  

(15)

where \( n_{\text{PUSCH}} \) denotes the number of available PUSCH resources, \( \xi_m \) denotes the number of devices selecting the \( m \)th preamble to perform access request. Since the devices are independent of each other, \( \xi_m \) follows a Binomial distribution. Hence, we have

\[ Pr\{\xi_m \geq 1\} = 1 - Pr\{\xi_m = 0\} = \binom{n_k}{0} (1 - \frac{1}{n_p})^{n_k} = (1 - \frac{1}{n_p})^{n_k}. \]  

(16)
As a result, the RA success probability under conventional scheme can be obtained as

$$P_{\text{suc}}^c = P_4^c$$

$$= (1 - \frac{1}{n_p})^{n_k-1} \cdot P_{\text{res}}^c$$

$$= (1 - \frac{1}{n_p})^{n_k-1} \min\{1, \frac{n_{\text{PUSCH}}}{(1 - \frac{1}{n_p})^{n_k} \cdot \min\{n_k, n_p}\}}$$

(17)

Similarly, there are 6 cases for PA collision and PUSCH resource allocation for the proposed RA procedure:

1. $p_1 = Pr[\text{no shared preamble collision, PUSCH resource allocated}]$;
2. $p_2 = Pr[\text{no shared preamble collision, PUSCH resource unallocated}]$;
3. $p_3 = Pr[\text{shared preamble collision, no dedicated preamble collision, PUSCH resource allocated}]$;
4. $p_4 = Pr[\text{shared preamble collision, no dedicated preamble collision, PUSCH resource unallocated}]$;
5. $p_5 = Pr[\text{shared preamble collision, dedicated preamble collision, PUSCH resource allocated}]$;
6. $p_6 = Pr[\text{shared preamble collision, dedicated preamble collision, PUSCH resource unallocated}]$.

It is easily verified that the RA success probability under the proposed scheme is equal to $p_1 + p_3$. On the contrary, the RA failure probability of the proposed scheme is equal to $p_2 + p_4 + p_5 + p_6$. Besides, $p_5$ represents the PUSCH resources waste probability. Therefore, the RA success probability under the proposed scheme holds as

$$P_{\text{suc}} = p_1 + p_3$$

$$= (1 - P_{\text{collision}}) \cdot P_{\text{res}}$$

$$= (1 - \frac{1}{n_{p,s}n_{p,d}n_{\mu}})^{n_k-1} \cdot P_{\text{res}}$$

(18)

where $P_{\text{res}}$ denotes the successful resource allocation rate under the proposed scheme. Additionally, let $\xi_{s,m}$ denote the number of devices selecting the $m$th shared sub-preamble, $\xi_{d,m}$ denote the number of devices selecting the $m$th dedicated sub-preamble, and $\zeta_{\mu}$ denote the number of devices selecting the $\mu$th index location. The successful resource allocation rate for RA-step 3 under the proposed scheme holds as

$$P_{\text{res}} = \min\{1, \frac{n_{\text{PUSCH}}}{Pr\{\xi_{s,m} \geq 1\}Pr\{\xi_{d,m} \geq 1\}Pr\{\zeta_{\mu} \geq 1\} \cdot \min\{n_k, n_{p,s}n_{p,d}n_{\mu}\}}\}$$

(19)

Here, as $\xi_{s,m}$, $\xi_{d,m}$, and $\zeta_{\mu}$ both follow Binomial distributions, there holds

$$Pr\{\xi_{s,m} \geq 1\} = 1 - Pr\{\xi_{s,m} = 0\} = \binom{n_k}{0}(1 - \frac{1}{n_{p,s}})^{n_k} = (1 - \frac{1}{n_{p,s}})^{n_k}$$

$$Pr\{\xi_{d,m} \geq 1\} = 1 - Pr\{\xi_{d,m} = 0\} = \binom{n_k}{0}(1 - \frac{1}{n_{p,d}})^{n_k} = (1 - \frac{1}{n_{p,d}})^{n_k}$$

$$Pr\{\zeta_{\mu} \geq 1\} = 1 - Pr\{\zeta_{\mu} = 0\} = \binom{n_k}{0}(1 - \frac{1}{n_{\mu}})^{n_k} = (1 - \frac{1}{n_{\mu}})^{n_k}$$

(20)
Thus, the RA success probability under the proposed scheme holds as

\[ P_{\text{suc}} = p_1 + p_3 = (1 - \frac{1}{n_{p,s} n_{p,d} n_{p}})^{n-1} \cdot P_{\text{res}} \]

\[ = (1 - \frac{1}{n_{p,s} n_{p,d} n_{p}})^{n-1} \min\{1, (1 - \frac{1}{n_{p,s}})^n (1 - \frac{1}{n_{p,d}})^n (1 - \frac{1}{n_{p}})^n \cdot \min\{n_{k}, n_{p,s} n_{p,d} n_{p}\}\} \tag{21} \]

### 6. Simulation Results

In this section, the performance of the proposed RA scheme is simulated and analyzed in terms of preamble detection probability, preamble collision probability, RA success probability, resource efficiency and TA capturing. The simulation parameters and their configurations are listed in Table 1. The simulation experiments are carried out on a MATLAB platform. In the simulation, preamble sequences are first generated for the conventional RA scheme and the proposed RA scheme respectively. Then different preambles with distinct TA values are transmitted to the gNB, suffering from a small-scale fading channel and additive white Gaussian noise. In conventional RA scheme, the receive preamble sequence is detected by calculating the correlation value with all the local available preamble sequences. In the proposed RA scheme, the received sequence is divided into three sub-sequences, and each sub-sequence is detected by calculating the correlation values with all the corresponding available sub-preambles. The TA estimation is determined as the peak location of correlation. Once a TA is captured, the PUSCH resource block will be allocated to the user by the gNB in RA-step 3. An RA failure occurs when the TA cannot be captured or the PUSCH resource blocks are used up. To avoid the impact of randomness on the simulation results, we simulate each case for 10,000 times and output the mean results finally.

**Table 1. Simulation parameters.**

| Number of available shared sub-preambles | 60 |
| Number of available dedicated sub-preambles | 60 |
| Length of ZC sequence | 839 |
| Length of shared sub-preamble sequence | 431 |
| Length of dedicated sub-preamble sequence | 199 |
| Length of GI | 5 |
| Number of communication devices on a PRACH | 1–1000 |
| Signal strength (SNR) | −20~15 dB |

#### 6.1. Preamble Detection Probability

To evaluate the impact of sub-preambles on the preamble detection probability, the proposed RA scheme is compared with conventional RA scheme when total \( P \) devices select exclusive sub-preambles to access to a gNB on the same PRACH. The preamble detection performance of the proposed RA scheme may be degraded, since the length of designed preambles is shorter than the conventional ones. To test the negative effect of sub-preambles, we randomly select sub-preamble sequence with different indexes from the available sub-preamble set.

Figure 4 depicts the average preamble detection probability for different case of the number of devices \( P \). The following observations are obtained: (1) When SNR is higher than −14 dB, the preamble detection performance of proposed scheme is a little poorer than conventional RA scheme; (2) The detection performance variation caused by varying amount of RA-attempting devices can be neglected; (3) The difference in the average preamble detection probabilities of proposed RA scheme and conventional RA scheme is negligible. The above observations imply that the sub-preamble structure does not impair
the preamble detection performance. Hence, it is applicable to the IoT scenario where massive devices may request random access at the same time.

**Figure 4.** Comparison of the average preamble detection probabilities.

### 6.2. Comparison of Collision Probability

Figure 5 depicts the relationship between the RA collision probability and the number of RA-attempting devices on a PRACH. In our simulations, the performance of the proposed RA scheme is compared with the conventional RA scheme, E_PACD scheme [14], and PACR scheme [15]. Firstly, the analytical results under the proposed RA scheme and that under the conventional scheme are both close to the simulation results, which verifies the accuracy of our analysis approach. It is also observed that the collision probability of the conventional RA scheme and that of E_PACD scheme are identical, since their collision probabilities only depend on the number of available preambles. Differently, the collision probability of PACR scheme depends on both the number of TA values $N_{TA}$ and the number of available preambles that can be allocated to the devices. Additionally, greater $N_{TA}$ means the gNB can detect more preambles on a PRACH, which can reduce the collision probability. However, $N_{TA}$ depends on the length of zero correlation zone of the detected preambles. Hence, $N_{TA}$ is finite. On the other hand, the RA collision probability of the proposed scheme depends on the size of sub-preamble sequence sets and the number of available locations of a dedicated sub-preamble. From Figure 5, we also observe that: (1) With the increase of the amount of RA-attempting devices, the collision probability of each scheme increases; (2) When the amount of RA-attempting devices is the same, the proposed scheme guarantees a lower RA collision probability than the other schemes, which means it is more suitable for massive connective scenarios.
6.3. Comparison of RA Success Probability

In this subsection, the RA success probability varying with the number of RA-attempting devices on a PRACH is analyzed. In our simulation experiment, we set $n_{p} = n_{p,s} = 60$. Figures 6 and 7 depict the simulation results with configuration $n_{PUSCH} = 20$ and $n_{PUSCH} = 100$ respectively.
Similar to Figure 5, the analytical results under the proposed and conventional schemes are both close to the simulation results, which validates the effectiveness of our theoretical analysis. As observed in Figure 6, the RA success probability under the proposed scheme is approximately 100%, when \( n_k < 20 \). This is because the RA collision probability of the proposed scheme is close to zero. Therefore, all PUSCH resources can be allocated to the collision-free devices. When \( n_k > 20 \), the RA success probabilities under the proposed scheme, the conventional scheme, PACR and E_PACR scheme all decrease rapidly, since the number of available PUSCH resources is insufficient. In this case, PUSCH resources shortage is the main influence on RA success probability. Besides, the RA success probability under the proposed scheme performs better than the other schemes. It is observed that the RA success probability under PACR scheme can be convergent to the proposed scheme only if \( N_{TA} = \infty^+ \) that is impractical in a real-world system. In Figure 7, \( n_{PUSCH} \) is extended to 100. It is found that with the increase of \( n_{PUSCH} \), the proposed scheme can guarantee a high RA success probability even though the number of RA-attempting devices on a PRACH beyond 100. Hence, it is verified that the proposed scheme can guarantee a higher RA success probability as long as the PUSCH resource is enough.

6.4. Comparison of Resource Efficiency

In this subsection, the resource efficiency is analyzed under different schemes. The resource efficiency is defined as follows

\[
\eta = \frac{n_{success}}{n_{PUSCH}},
\]

where \( n_{success} \) denotes the number of PUSCH resources which are allocated to collision-free devices. As shown in Figure 8, the resource efficiency under the proposed scheme is higher than that under other schemes, especially when the amount of RA-attempting devices on the same PRACH is large. When \( n_k > 20 \), the resource efficiency under the proposed scheme and that under PACR scheme both decrease with \( n_k \). This is because the PUSCH resources are not enough to be allocated to the collision-free devices when \( n_k > n_{PUSCH} \). However, the proposed scheme can still guarantee a high resource efficiency even though \( n_k \) is large.
6.5. Analysis of TA Capturing Performance

In this subsection, the performance of TA capturing is evaluated in terms of TA capturing accuracy. Specifically, TA capturing accuracy is defined as

\[ \varepsilon_{TA} = \frac{N_{captured}}{N_{preamble}} \]  

where \( N_{captured} \) denotes the number of TA values which is captured correctly, and \( N_{preamble} \) denotes the number of received preambles on the same PRACH resource. Compared with the conventional preamble, the sub-preamble is shorter. As a result, sub-preamble structure may degrade the performance of TA capturing. For the sake of performance analysis of the proposed multiple TA value capturing scheme, we focus on the collision of the shared sub-preambles. Assume that \( M \) preambles from \( M \) different devices are received on the same PRACH where the shared sub-preambles are set to the same index. Besides, the dedicated sub-preamble indexes and location indexes are randomly selected. The number of the available dedicated sub-preambles is set to 60, the number of available locations is set to 2.

As depicted in Figure 9, the number of preambles sent on the same PRACH is set to \( M = \{2, 4, 6, 8, 10\} \). It is observed that better TA capturing performance can be guaranteed when less devices transmit identical preamble on the same PRACH. For the case of \( M = 2 \), the proposed scheme can guarantee over 80% accuracy even though the SNR is low (i.e., \(-10 \text{ dB}\)). Similar accuracy can be obtained when SNR reaches 6 dB for the case of \( M = 10 \) where interference is severe. Please note that the probability that 10 devices select the same shared sub-preamble is quite low. According to Equation (12), the probability is only 2.59% when 1000 devices request RAs with 60 available preambles.
Figure 9. TA capturing accuracy.

7. Conclusions

In this work, a preamble collision resolution scheme was proposed for the IoT systems. The number of collision-free devices on the same PRACH was expanded by the proposed sub-preamble sequence structure. To obtain multiple TA values of the RA-attempting devices, a multiple TA capturing scheme was also proposed. Simulation results verified that the proposed RA scheme performed well in preamble collision probability, RA success probability, resource efficiency and TA capturing. In particular, the proposed scheme improved the RA success probability with ignorable degradation in preamble detection performance. Hence, the proposed scheme has a great potential for supporting massive access in the future IoT systems. In addition to the preamble design, preamble repetition is also an important way to improve the random access performance. Hence, our future work will focus on the preamble repetition mechanism design for a massive IoT system.

Author Contributions: This work was a collaborative development by all authors. A.Z., Z.L., R.W., X.L., and B.G. proposed the idea, were involved in the theoretical performance analysis, designed and optimized the algorithm, and wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant 61871062; in part by the Natural Science Foundation of Chongqing under grant cstc2020jcyj-zdzmX0024; in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under grant KJQN201900609; in part by the Program for Innovation Team Building at Institutions of Higher Education in Chongqing under Grant CXTDX201601020, and in part by the University Innovation Research Group of Chongqing under grant CXQT20017.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to further study will be carried out.

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Cao, J.; Ma, M.; Li, H.; Ma, R.; Sun, Y.; Yu, P.; Xiong, L. A Survey on Security Aspects for 3GPP 5G Networks. *IEEE Commun. Surv. Tutor.* 2020, 22, 170–195. [CrossRef]

2. Qiu, T.; Chen, N.; Li, K.; Atiquzzaman, M.; Zhao, W. How Can Heterogeneous Internet of Things Build Our Future: A Survey. *IEEE Commun. Surv. Tutor.* 2018, 20, 2011–2027. [CrossRef]

3. Tran-Dang, H.; Krommenacker, N.; Kim, D. Toward the Internet of Things for Physical Internet: Perspectives and Challenges. *IEEE Internet Things J.* 2020, 7, 4711–4736. [CrossRef]

4. Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update. White Paper. 2014–2019. Available online: https://community.cisco.com/t5/wireless-mobility-documents/cisco-visual-networking-index-global-mobile-data-traffic/ta-p/3612708?tid=osscdc000283 (accessed on 1 January 2021).

5. Akpakwu, G.A.; Silva, B.J.; Hancce, G.P.; Abu-Mahfouz, A.M. A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges. *IEEE Access* 2018, 6, 3619–3647. [CrossRef]

6. Shahab, M.B.; Abbas, R.; Shirvanimoghaddam, M.; Johnson, S.J. Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey. *IEEE Commun. Surv. Tutor.* 2020, 22, 1805–1838. [CrossRef]

7. Wang, Y.; Wang, T.; Yang, Z.; Wang, D.; Cheng, J. Throughput-Oriented Non-Orthogonal Random Access Scheme for Massive MTC Networks. *IEEE Trans. Commun.* 2020, 68, 1777–1793. [CrossRef]

8. Wu, Y.; Zhang, N.; Rong, K. Non-Orthogonal Random Access and Data Transmission Scheme for Machine-to-Machine Communications in Cellular Networks. *IEEE Access* 2020, 8, 27687–27704. [CrossRef]

9. Yang, Z.; Wang, Y.; Wang, Z.; Wang, T. Relaying Threshold Based Random Access and Data Transmission Scheme for Grouped Massive MTC Networks. In Proceedings of the 11th International Conference on Wireless Communications and Signal Processing, Xi’an, China, 23–25 October 2019; pp. 1–6.

10. Chen, Z.; Smith, D.B.; Chen, Z. Heterogeneous Machine-Type Communications in Cellular Networks: Random Access Optimization by Deep Reinforcement Learning. In Proceedings of the IEEE International Conference on Communications, Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.

11. Jiao, J.; Xu, L.; Wu, S.; Wang, Y.; Lu, R.; Zhang, Q. Unequal Access Latency Random Access Protocol for Massive Machine-Type Communications. *IEEE Trans. Wirel. Commun.* 2020, 19, 5924–5937. [CrossRef]

12. Kuai, X.; Yuan, X.; Yan W.; Liang, Y. Coexistence of Human-Type and Machine-Type Communications in Uplink Massive MIMO. *IEEE J. Sel. Areas Commun.* 2020. [CrossRef]

13. Jang, H.S.; Kim, S.M.; Ko, K.S.; Cha, J.; Sung, D.K. Spatial Group Based Random Access for M2M Communications. *IEEE Commun. Lett.* 2014, 18, 961–964. [CrossRef]

14. Jang, H.S.; Kim, S.M.; Sung, D.K.; Park, H.S. An Early Preamble Collision Detection Scheme Based on Tagged Preambles for Cellular M2M Random Access. *IEEE Trans. Veh. Technol.* 2017, 66, 5974–5984. [CrossRef]

15. Jang, H.S.; Kim, S.M.; Park, H.S.; Sung, D.K. A Preamble Collision Resolution Scheme via Tagged Preambles for Cellular IoT/M2M Communications. *IEEE Trans. Veh. Technol.* 2018, 67, 1825–1829. [CrossRef]

16. Jiang, D.; Cui, Y. MAP-Based Pilot State Detection in Grant-Free Random Access for mMTC. In Proceedings of the IEEE 21st International Workshop on Signal Processing Advances in Wireless Communication, Atlanta, GA, USA, 26–29 May 2020.

17. He, Q.; Chen, Z.; Quek, T.Q.S.; Chen, Z.; Li, S. A Novel Cross-Layer Protocol for Random Access in Massive Machine-Type Communications. In Proceedings of the IEEE International Conference on Communications Workshops, Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.

18. Bai, L.; Han, R.; Liu, J.; Choi, H.; Zhang, W. Multiple Delay Estimation for Collision Resolution in Non-Orthogonal Random Access. *IEEE Trans. Veh. Technol.* 2020, 69, 497–508. [CrossRef]

19. Park, J.; Lim, Y. Random-Access Control Method for MTC in an LTE System Inspired by the Minority Game. *IEEE Trans. Veh. Technol.* 2018, 67, 9037–9041. [CrossRef]

20. Kim, J.S.; Lee, S.; Chung, M.Y. Efficient Random-Access Scheme for Massive Connectivity in 3GPP Low-Cost Machine-Type Communications. *IEEE Trans. Veh. Technol.* 2017, 66, 6280–6290. [CrossRef]

21. 3GPP TS 36.300, Release 16 10.1.5 Random Access Procedure. December 2016. Available online: https://www.3gpp.org/ftp/Spe cs/archive/38_series/36.300 (accessed on 1 January 2021).

22. 3GPP TS 38.321, Release 16 5.1 Random Access Procedure. October 2020. Available online: https://www.3gpp.org/ftp/Spe cs/archive/38_series/38.321 (accessed on 1 January 2021).

23. 3GPP TS 38.321, Release 16 6.3.3 Physical Random-Access Channel. October 2020. Available online: https://www.3gpp.org/ftp /Specs/archive/38_series/38.211 (accessed on 1 January 2021).

24. Kim, T.; Kim, D.M.; Pratas, N.; Popovski, P.; Sung, D.K. An Enhanced Access Reservation Protocol with a Partial Preamble Transmission Mechanism in NB-IoT Systems. *IEEE Commun. Lett.* 2017, 21, 2270–2273. [CrossRef]