

Article

# Performance Analysis of a Novel Hybrid S-ALOHA/TDMA Protocol for Beta Distributed Massive MTC Access

Nannan Sui <sup>1</sup>, Youyun Xu <sup>2</sup>, Cong Wang <sup>1,\*</sup>  and Wei Xie <sup>1</sup>

<sup>1</sup> College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China; nannansui@foxmail.com (N.S.); edifier@163.com (W.X.)

<sup>2</sup> National Engineering Research Center of Communication and Network Technologies, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; yyxu@vip.sina.com

\* Correspondence: dolphin6488@163.com; Tel./Fax: +86-138-5158-9658

Received: 2 October 2017; Accepted: 5 December 2017; Published: 15 December 2017

**Abstract:** Simultaneous random access of massive machine type communications (MTC) devices are expected to cause congestion in the radio access network. Not only the performance of MTC, but the coexisting human to human (H2H) communications would also degrade dramatically without an appropriate medium access control (MAC) protocol. However, most existing solutions focus on the random access procedure without dealing with the subsequent data transmission procedure. In this paper, we firstly derive a packet size threshold based on the capacity analysis of slotted ALOHA (S-ALOHA) and time division multiple access (TDMA) protocols. Then a novel hybrid S-ALOHA/TDMA MAC protocol (HSTMAC) is presented for massive MTC access, in which the resources are separated for beta distributed machine to machine (M2M) traffic with small size packets and high priority H2H traffic with large size packets. Considering access class barring (ACB) scheme as an overload control method, the system equilibrium under arbitrary retransmission limit is analyzed rigorously, which can provide insights on quality of service (QoS) guarantee. Finally, a dynamic pre-backoff (DPBO) algorithm is designed for load balance by adaptively scattering the highly synchronized M2M traffic over the transmission interval. Numerical and simulation results validate our analysis and show that the HSTMAC protocol is superior to pure S-ALOHA protocol and pure TDMA protocol. The proposed DPBO algorithm can achieve a higher success probability and resource utilization ratio with a much reduced average delay than that of uniform pre-backoff (UPBO) scheme.

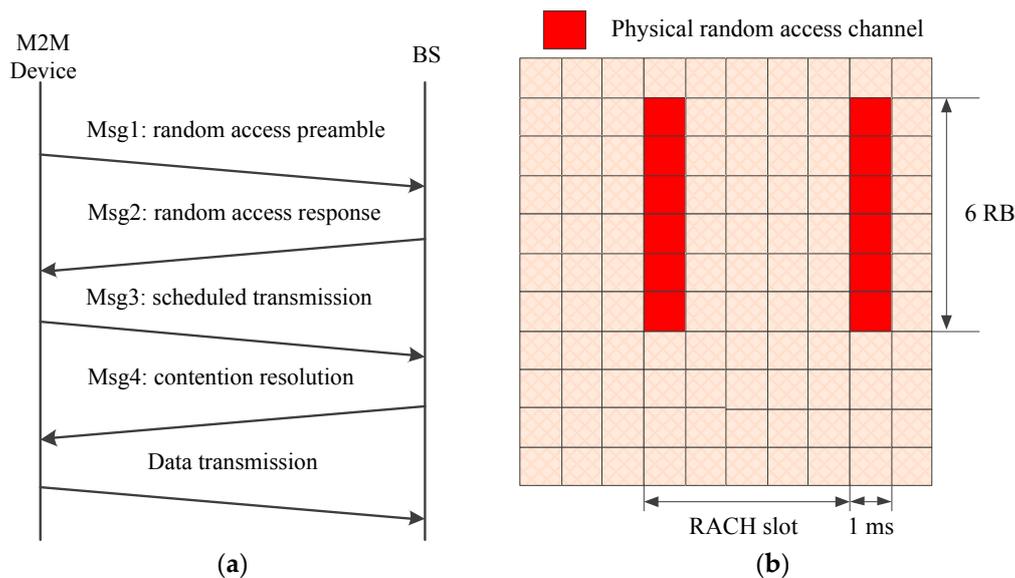
**Keywords:** machine to machine (M2M); human to human (H2H); random access; hybrid MAC; dynamic pre-backoff (DPBO); equilibrium analysis

## 1. Introduction

Machine type communication (MTC), also known as machine to machine (M2M) communication, is envisioned as a main enabler for the Internet of Things (IoT) [1,2]. Featuring no (or little) human intervention, M2M applications have been experiencing rapid growth in various domains, such as intelligent transport systems, smart cities, e-Health, smart grids, industry automation, monitoring and control systems, etc. [3,4]. It is expected that there will be more than 50 billion M2M devices at 2020 [5]. Unlike traditional human to human (H2H) communications, M2M traffic is characterized by massive devices, low mobility, small data transmissions and large uplink-to-downlink traffic ratio [6]. Therefore, the long term evolved (LTE) and LTE-Advanced (LTE-A) networks have to be enhanced to accommodate such differentiated M2M services and the coexisting H2H traffic [7–9].

A simplified model of the random access procedure in LTE/LTE-A networks is depicted in Figure 1a, which shows that four handshake messages are needed before a successful data

transmission [10]. If two or more devices select the same preamble in Msg1, they will receive the same random access response (RAR) message from the base station (BS) and then send Msg3 on the same granted uplink channel. However, the BS cannot decode the superposed Msg3 accurately. Hence the colliding devices fail in this random access attempt and would send retransmission requests on the next random access channel (RACH). As a result, the main concern of most works is the RACH congestion problem caused by simultaneous random access requests from massive devices [11–16]. The existing overload control mechanisms can be classified as push-based schemes (such as access class barring (ACB), dynamic RACH resources allocation, backoff, and slotted access), pull-based scheme (group paging) and a combination of these methods.



**Figure 1.** Contention-based random access procedure and physical random access channel: (a) Contention based random access procedure; (b) Physical random access channel.

The packet sizes of many M2M applications are usually small, some of them even on the order of bits or bytes (e.g., health or environment monitors and alarm devices) [17]. As to the random access procedure, shown in Figure 1b, just Msg1 occupies six resource blocks (RBs, a time-frequency resource unit which consists of 180 KHz in frequency domain and a time slot, i.e., 0.5 ms in time domain) in frequency and 1 ms in time (that is 12 RBs in total). Not to mention Msg3 and the retransmissions caused by collisions which can result in much more resource consumption. Thus the ratio of signaling overhead to effective payload may be unacceptable in practice [18]. However, a comprehensive investigation of both RACH congestion and the following data transmission is still lacking attention.

A few works have been aware of this problem [13,14,19–21]. The authors in [19] demonstrated that the supportable arrival rate of one-stage protocol, like slotted-ALOHA (S-ALOHA), is higher than that of two-stage protocol, like time division multiple access (TDMA), if the payload size is small. In [20], the LTE/LTE-A random access procedure was tailored exclusively for M2M traffic to reduce the signaling overhead. Along with the surveys of various MAC protocols for M2M communications in [13,14], we present a novel hybrid S-ALOHA and TDMA MAC protocol (HSTMAC) for M2M and H2H coexisting scenario in [21]. This paper differs from [21] in the following aspects: (i) we assume that H2H traffic follows a Poisson distribution while M2M traffic follows a Beta distribution, which represents the worst case for random access procedure; (ii) we rigorously derive the system equilibrium under any retransmission limit ( $Q \geq 1$ ) instead of large value assumption ( $Q \geq 9$ ) in [21]; (iii) based on the equilibrium analysis, a dynamic pre-backoff (DPBO) algorithm is designed for load balance of M2M traffic and quality of service (QoS) guarantee of H2H traffic. As a summary, the main contributions of this paper are as follows:

- The framework of the proposed HSTMAC protocol is detailed in this paper. The reserved uplink RBs are divided into TDMA part for H2H traffic with large size packets and S-ALOHA part for M2M traffic with small size packets. The foundation of HSTMAC protocol is proved through system capacity analysis.
- Considering the impact of ACB factor and backoff scheme, this paper derives the analytical performance of HSTMAC iteratively. Then the system equilibrium under arbitrary retransmission limit is analyzed rigorously, which can provide an insight on QoS provision.
- Based on the equilibrium analysis, a dynamic pre-backoff algorithm is designed for QoS guarantee by adaptively scattering new M2M arrivals over the transmission interval according to the highly synchronized traffic load.
- The superiority of HSTMAC in terms of success possibility, resource utilization ratio, delay and average access request number, is validated through numerical and simulation results. Compared to uniform pre-backoff (UPBO) algorithm, the proposed DPBO algorithm can achieve a higher success probability with a much reduced delay.

The remainder of this paper is organized as follows: Section 2 gives a brief review of the related works. Section 3 presents the frame structure of HSTMAC and its fundamental motivation. We exploit an iterative method to analyze the performance of HSTMAC in Section 4. Then we derive a more rigorous equilibrium analysis in Section 5, in which we also design a dynamic pre-backoff scheme. In Section 6, the performance of HSTMAC is analyzed through numerical and simulation results. We conclude this paper in Section 7.

## 2. Related Works

The differentiated characteristics of M2M traffic have been identified in [6], and are evaluated by field measurements in a large scale network in [17]. The most prominent features of M2M traffic are massive number of devices and uplink small data transmission. Therefore, the prime challenge faced with LTE/LTE-A networks is RACH congestion due to bursts of M2M access requests. Comprehensive surveys of load control mechanisms have been made in [12–15]. Third generation partnership project (3GPP) classifies these proposals into two main categories, namely push-based and pull-based methods [11]. As two of the push-based methods, ACB scheme aims at restricting the contending levels of services with different priorities, and dynamic resource allocation scheme aims at providing sufficient access opportunities according to traffic load. In ACB, BS firstly broadcasts an access barring factor  $p$  ( $0 \leq p \leq 1$ ). Then each device generates a random number and proceeds with the following random access procedure if the number is not larger than  $p$ , otherwise the device has to wait for the next available access opportunity. Thus the determination of optimal ACB factor is the main concern.

In [22], both fixed and dynamic preamble allocation schemes are proposed. The ACB factor is adjusted based on load estimation without prior information. In [23], in order to maximize success probability and to satisfy the delay requirement, a Markov chain is exploited to jointly optimize the extended access barring factor and RACH resource allocation. In [24], the throughput is maximized by dynamically partitioning preambles for delay-sensitive and delay tolerant classes. In order to reduce the collision probability, the authors in [25] propose to reserve multiple uplink channels for Msg3 for a single detected preamble.

Another solution for congestion avoidance is to reuse the available preambles and uplink resources. In [26], the devices are grouped according to their spatial locations. Then for each group, a single root can generate more preamble sequences due to the reduced radius. Moreover, if the distance between two groups is larger than a threshold, the devices from different groups are allowed to reuse the same uplink channel for Msg3 [27]. The BS can decode the overlapped information through successive interference cancellation.

As a pull-based overload control method, the performance of group paging is analyzed by an iterative method in [28]. Compared to uniform distributed traffic, the devices in group paging are

usually activated simultaneously in the first access slot, which results in an extremely high system load. Due to the similar reason, surge of beta distributed M2M traffic can also cause severe performance reduction [11,29,30]. Therefore, backoff based load balance schemes for synchronized M2M traffic have been widely adopted. The authors in [31] adjust the backoff indicator dynamically to ensure the delayed devices would not enter into the next paging cycle. In [32,33], uniform pre-backoff scheme has been applied to group paging. The authors also investigate the impact of backoff window size on the system performance. Further in [34], based on the stability analysis, an optimal pre-backoff scheme has been proposed for group paging to reduce energy consumption. In [35], the UPBO mechanism is also used for beta distributed traffic.

The above works mainly focus on the M2M traffic issues. However, the H2H traffic is still taking an important role in current cellular networks. Then how to accommodate both H2H and M2M traffics at the same time becomes an urgent issue. One of the possible solutions is resources splitting. The authors in [36] analyze the impact of M2M traffic on H2H traffic. Simulation results show that H2H traffic has a better performance with RACH resource separation scheme while the M2M traffic prefers ACB scheme. A preamble set partition scheme is proposed in [37]. The impact of backoff indicator and retransmissions has been investigated by simulations. In [38,39], the RACH resources are divided into M2M section, H2H section and a sharing section. Then a mixed strategy game method in [38] and a continuous time Markov chain in [39] are used to determine the optimal allocation parameters. The authors in [40] introduce a power prioritized random access scheme. The H2H devices send preambles in a higher power level while the M2M devices in a lower power level. The BS can distinguish the requests from H2H devices from M2M devices based on two detection levels and a timing alignment matching mechanism.

Obviously, most existing works draw their attentions to RACH congestion problem without considering the following data transmission requirement. Taking signaling overhead into consideration, a theoretical analysis about supportable arrival rate against packet size is derived in [19]. The conclusion is that one stage protocol can accommodate more traffics with small size packets than two stage protocol. The advantages and drawbacks of various MAC protocols have been discussed in [13,14]. Due to the differentiated QoS requirements between M2M traffic and H2H traffic, the traditional four-way handshake random access procedure may be not applicable to such a complex scenario. Therefore, it's a reasonable idea to design a hybrid MAC protocol to combine the strength of different MAC protocols. In [20], M2M devices can send data on the granted channels for Msg3 immediately after preamble transmissions. The authors in [41] divide the resources into contention only period and transmission only period. In contention only period, the M2M devices contend for the collision free access opportunities in the transmission only period in the way of carrier sense multiple access (CSMA) mechanism. In order to reduce the signaling overhead, the authors split preamble set into H2H part and M2M part [42]. The reserved preamble sequences for M2M devices are used in a CSMA/CA manner. However, the H2H traffic is not considered in [20,41] while the data transmission is not considered in [42]. The HSTMAC proposed in [21] divides reserved uplink resources into S-ALOHA part for M2M small data transmission and TDMA part for traditional H2H traffic. A joint dynamic ACB and resource allocation algorithm has been designed to maximize the resource utilization under a fairness constraint. In this paper, we further analyze the equilibrium performance of HSTMAC under arbitrary retransmission limit. Based on the equilibrium analysis, a dynamic pre-backoff algorithm is designed for load balance of beta distributed M2M traffic and QoS provision for H2H traffic.

### 3. The Proposed HSTMAC Framework and System Model

We first demonstrate the basis of the proposed HSTMAC through capacity analysis. Then the details of HSTMAC framework and the essential assumptions are presented.

### 3.1. Foundation of the HSTMAC Protocol

We evaluate the performance of a MAC protocol in the view of resource utilization maximization. The concept of *capacity* is defined as the maximum amount of resource blocks used for successful data transmissions. We will show how packet size affects the system capacity.

#### 3.1.1. System Capacity of S-ALOHA Protocol

Let  $d$  denote the packet size and  $L$  denote the total available resource blocks. If the number of active devices is  $X$ , the average resource utilization  $R$  is calculated as:

$$R = \delta \cdot X e^{-\frac{X}{L\delta}} \quad (1)$$

where the available data channels is  $L/\delta$ . Then the well-known capacity of S-ALOHA system  $C_{S-ALOHA}$  is obtained when  $X = L/\delta$ :

$$\begin{aligned} C_{S-ALOHA} &= L e^{-1} \\ X_{S-ALOHA}^* &= L/\delta \end{aligned} \quad (2)$$

where  $X_{S-ALOHA}^*$  is the maximum supportable devices of the S-ALOHA system.

#### 3.1.2. System Capacity of TDMA Protocol

The resources in TDMA system are divided into two parts for contention based random access and the following contention free data transmission. The random access procedure in TDMA can be modeled as a multi-channel S-ALOHA system [43], where each preamble is regarded as an access slot. If the RACH occupies too many RBs, the reserved data channels would be insufficient for the successful devices. On the contrary, if the reserved RBs for data channels are too many, the devices may fail due to lack of random access opportunities, which may also cause underutilization of the reserved data channels. Therefore, a tradeoff must be made for the resource allocation between random access procedure and data transmission procedure.

In order to model the uplink signaling overhead in the random access procedure, we assume that a resource block can bear  $\gamma$  random access opportunities. Let  $\theta$  denote the proportion of RBs reserved for RACH, the resource utilization of TDMA is:

$$R = \min \left\{ \delta \cdot X e^{-\frac{X}{L\theta\gamma}}, L(1-\theta) \right\} \quad (3)$$

where  $L\theta\gamma$  is the reserved random access opportunities and  $L(1-\theta)$  is the reserved RBs for data channels. According to Equations (2) and (3), the resource utilization is maximized when:

$$\begin{aligned} X &= L\theta\gamma \\ \delta X e^{-\frac{X}{L\theta\gamma}} &= L(1-\theta) \end{aligned} \quad (4)$$

Then the capacity of TDMA system is obtained by solving Equation (4):

$$\begin{aligned} C_{TDMA} &= \delta \cdot L\theta^* \gamma \cdot e^{-1} = \frac{L\delta\gamma}{e+\delta\gamma} \\ \theta^* &= \frac{1}{1+\delta\gamma/e} \\ X_{TDMA}^* &= L\theta^* \gamma \end{aligned} \quad (5)$$

where  $X_{TDMA}^*$  is the maximum supportable devices of the TDMA system.

### 3.1.3. Packet Size Threshold

Solving the equation  $C_{S\text{-ALOHA}} = C_{TDMA}$ , we can get the packet size threshold:

$$\bar{\delta} = \frac{1}{1 - e^{-1}} \cdot \frac{1}{\gamma} \quad (6)$$

**Remark 1.** According to Equations (2) and (5), the S-ALOHA system is superior to the TDMA system if  $\delta < \bar{\delta}$ . On the contrary, the TDMA system is superior to the S-ALOHA system if  $\delta > \bar{\delta}$ . Considering that M2M traffic is usually in a small packet size while traditional H2H traffic is in a large packet size, it would be a feasible idea to design a hybrid S-ALOHA/TDMA MAC protocol for M2M and H2H coexisting scenario.

### 3.2. Optimal ACB Parameter

For both S-ALOHA system and TDMA system, the resource utilization is a concave function with the number of active devices. In order to maximize the resource utilization, all the devices should be allowed to pass the ACB check if  $X \leq X^*$ , while only  $X^*$  out of  $X$  devices can be allowed to pass the ACB check if  $X > X^*$ . Therefore, the optimal ACB parameter for both S-ALOHA and TDMA systems is:

$$p = \min(X^*/X, 1) \quad (7)$$

**Remark 2.** According to the above discussions, we can derive the concept of saturation as follows. For a certain number of available resource blocks  $L$ , the maximum number of supportable devices  $X^*$  can be computed by Equations (2) and (5). We define the system is in a saturation state if the actual number of active devices is  $X \geq X^*$ , where some devices are barred to proceed the random access procedure. We define the system is in a non-saturation state if  $X < X^*$ , where all the devices can participate in the contention of random access opportunities. Accordingly, the resource utilization of S-ALOHA and TDMA systems in different state can be computed as:

$$R_{S\text{-ALOHA}} = \begin{cases} Le^{-1}, & \text{if } X \geq X_{S\text{-ALOHA}}^* \\ \delta X e^{-\frac{X}{L\bar{\delta}}}, & \text{if } X < X_{S\text{-ALOHA}}^* \end{cases} \quad (8)$$

$$R_{TDMA} = \begin{cases} L(1 - \theta^*), & \text{if } X \geq X_{TDMA}^* \\ \delta X e^{-\frac{X}{L\bar{\theta}\gamma}}, & \text{if } X < X_{TDMA}^* \end{cases} \quad (9)$$

s.t.  $\delta X e^{-\frac{X}{L\bar{\theta}\gamma}} = L(1 - \theta)$

### 3.3. Frame Structure of the Proposed HSTMAC Protocol

As shown in Figure 2, in HSTMAC the BS schedules resources on a cycle-by-cycle basis in time domain and each access cycle spans  $\tau$  ms.

In order to utilize the resources more efficiently, the HSTMAC protocol mainly divides the reserved resources into three sections: broadcast period (BP) for essential downlink control information, S-ALOHA period (SP) for M2M traffic with small size packets and TDMA period (TP) for H2H traffic with large size packets. The TDMA period is further split into random access period (RAP) for contention based random access and data transmission period (DTP) for contention free packet transmission. The optimal tradeoff between RAP and DTP is made according to Equation (4).

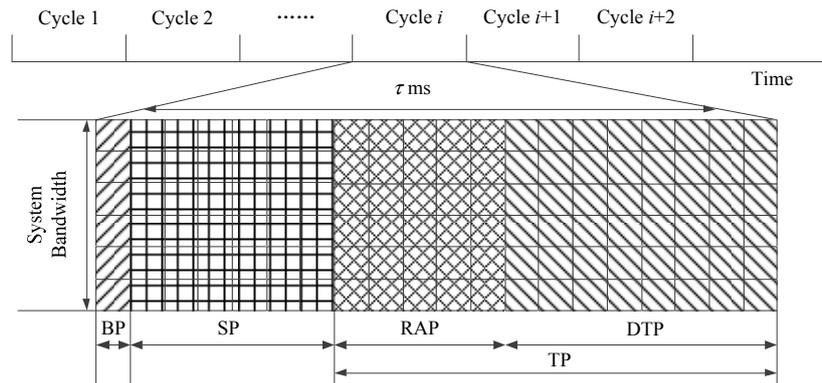


Figure 2. Frame structure of the proposed hybrid S-ALOHA/TDMA MAC protocol.

In the broadcast period of each cycle, BS estimates the number of active devices based on the obtained statistics and calculates the optimal resource allocation and ACB parameters for M2M and H2H traffics respectively. Then the BS broadcasts a notification message which contains these parameters to all the devices for synchronization. In this paper, we mainly focus on the uplink transmission in a single cellular cell. Let  $L$  denote the reserved uplink resource blocks in each cycle and let  $L_{SP,i}, L_{TP,i}, L_{RAP,i}, L_{DTP,i}$  denote the allocated RBs for the corresponding part in the  $i$ -th cycle, we can have:

$$\begin{aligned} L &= L_{SP,i} + L_{TP,i} \\ &= L_{SP,i} + L_{RAP,i} + L_{DTP,i} \end{aligned} \tag{10}$$

In order to make the analysis more tractable, we assume that all the active devices, new and backlog, arrive at the beginning of a cycle and complete the transmission procedure in current cycle (no matter success or fail). Packet loss on the wireless channel is not considered in this paper yet. Hence, a packet transmission request may fail due to ACB check failure, contention collision, or insufficient data channels. The failed device would select a random backoff interval between 0 and  $W_{BO}$  and start a retransmission attempt after the backoff counter expires. A packet will be dropped if it has not been sent successfully within the maximum  $Q$  transmission attempts.

In order to evaluate the performance of HSTMAC protocol in an extreme scenario, we assume that the arrival rate of H2H packets follows a Poisson distribution with mean  $\lambda$  in each access cycle and the arrival rate of M2M devices follows a Beta distribution  $g(t)$  defined in [11]:

$$g(t) = \frac{t^{\alpha-1}(T-t)^{\beta-1}}{T^{\alpha+\beta-1}B(\alpha, \beta)}, \quad t \in [0, T] \tag{11}$$

where  $B(\alpha, \beta)$  is the Beta function and  $T$  is the activation time interval of M2M devices. The packet sizes of H2H traffic and M2M traffic are  $\delta_1, \delta_2$  respectively. Let  $N^{M2M}$  denote the total number of M2M devices, then the number of new arrivals in cycle  $i$  is:

$$N_i^{M2M} = \int_{t_i}^{t_{i+1}} N^{M2M} \frac{t^{\alpha-1}(T-t)^{\beta-1}}{T^{\alpha+\beta-1}B(\alpha, \beta)} dt, \quad t \in [0, T], i \in [1, \left\lceil \frac{T}{\tau} \right\rceil] \tag{12}$$

where  $t_i$  is the time of the  $i$ -th access cycle.

#### 4. A General Model for Analytical Performance Analysis

##### 4.1. Analytical Performance Analysis

Due to the beta distribution of M2M traffic, the total transmission time interval is:

$$I_{\max} = \left\lceil \frac{T}{\tau} \right\rceil + (Q - 1)(W_{BO} + 1) \quad (13)$$

For the readability of this paper, we use superscripts  $(\cdot)^{M2M}$  and  $(\cdot)^{H2H}$  to denote the variables for M2M traffic and H2H traffic respectively. As a general analytical model used in this paper, we usually omit the superscripts for brevity without causing ambiguity. Let  $X_{i,j}[q]$  denote the average number of devices which first arrive at the  $j$ -th cycle and perform their  $q$ -th transmission attempts in the  $i$ -th cycle. Then the new arrivals in each cycle are:

$$\begin{aligned} X_{i,j}^{M2M}[1] &= \begin{cases} N_i^{M2M}, & i = j, i \leq \left\lceil \frac{T}{\tau} \right\rceil \\ 0, & \text{otherwise} \end{cases} \\ X_{i,j}^{H2H}[1] &= \begin{cases} \lambda, & i = j \\ 0, & i \neq j \end{cases} \end{aligned} \quad (14)$$

Let  $X_{ij}[q]$  denote the average number of devices which perform their  $q$ -th transmission attempts in the  $i$ -th cycle and let  $X_i$  denote the total number of devices which perform their transmission attempts in the  $i$ -th cycle. Considering the backoff indicator, we can derive that:

$$X_i[q] = \sum_{j=\max(i-(q-1)(W_{BO}+1),1)}^{i-(q-1)} X_{i,j}[q] \quad (15)$$

$$X_i = \sum_{q=1}^Q X_i[q] \quad (16)$$

According to the discussions in Section 3, optimal ACB parameters should be adjusted according to the actual number of devices in each cycle. For M2M traffic, the number of reserved data channels in the  $i$ -th cycle is:

$$M_i^{M2M} = L_{SP,i} / \delta_2 \quad (17)$$

As to the number of reserved random access opportunities for H2H traffic, we should first estimate whether the system is in a saturation state or not. According to Equations (4) and (5) and Remark 2, we can have:

$$\begin{aligned} M_i^{H2H} &= \begin{cases} L_{TP,i} \theta^* \gamma, & X_i^{H2H} \geq L_{TP,i} \theta^* \gamma \\ L_{TP,i} \theta \gamma, & X_i^{H2H} < L_{TP,i} \theta^* \gamma \end{cases} \\ \text{s.t.} \quad & \delta_1 X_i^{H2H} e^{-\frac{X_i^{H2H}}{L_{TP,i} \theta \gamma}} = L_{TP,i} (1 - \theta) \end{aligned} \quad (18)$$

Therefore, the optimal ACB parameter can be computed as:

$$p_i = \min(M_i / X_i, 1) \quad (19)$$

So the average number of devices which pass the ACB check is  $X_i p_i$  and the average number of successful devices is:

$$Z_i = X_i p_i e^{-\frac{X_i p_i}{M_i}} \quad (20)$$

Because each device has the same success probability, if we let  $Z_{i,j}[q]$  denote the average number of devices which first arrive at the  $j$ -th cycle and succeed in their  $q$ -th transmission attempts in the  $i$ -th cycle, we can derive that:

$$Z_{i,j}[q] = \frac{X_{i,j}[q]}{X_i} Z_i \quad (21)$$

$$\begin{aligned} Z_i &= \sum_{q=1}^Q Z_i[q] \\ &= \sum_{q=1}^Q \sum_{j=\max(i-(q-1)(W_{BO}+1),1)}^{i-(q-1)} Z_{i,j}[q] \end{aligned} \quad (22)$$

where  $Z_i[q]$  is the number of devices which succeed in their  $q$ -th access attempts in the  $i$ -th cycle.

In each cycle, the failed device which uses up its retransmission opportunities would drop the packet and experiences a final transmission failure. Otherwise, the failed device would randomly select a backoff interval  $\omega_{BO} \in [0, W_{BO}]$  and retry access to the network in the  $i+\omega_{BO}+1$ -th cycle. So the backlogged devices due to backoff can be computed as follows:

$$X_{i,j}[q] = \frac{1}{W_{BO}+1} \sum_{k=\max(i-(W_{BO}+1),1)}^{i-1} (X_{k,j}[q-1] - Z_{k,j}[q-1]), \quad 2 \leq i \leq I_{\max}, 2 \leq q \leq Q \quad (23)$$

$$\begin{aligned} X_i[q] &= \sum_{j=\max[i-(q-1)(W_{BO}+1),1]}^{i-(q-1)} X_{i,j}[q] \\ &= \sum_{j=\max[i-(q-1)(W_{BO}+1),1]}^{i-(q-1)} \frac{1}{W_{BO}+1} \sum_{k=\max(i-(W_{BO}+1),1)}^{i-1} (X_{k,j}[q-1] - Z_{k,j}[q-1]) \\ &= \frac{1}{W_{BO}+1} \sum_{k=\max(i-(W_{BO}+1),1)}^{i-1} \sum_{j=\max[i-(q-1)(W_{BO}+1),1]}^{i-(q-1)} (X_{k,j}[q-1] - Z_{k,j}[q-1]) \\ &= \frac{1}{W_{BO}+1} \sum_{k=\max(i-(W_{BO}+1),1)}^{i-1} (X_k[q-1] - Z_k[q-1]) \end{aligned} \quad (24)$$

#### 4.2. Performance Metrics

Four performance metrics, which are average success probability, resource utilization ratio, the statistics of delay and the number of transmission request, are chosen to evaluate the performance of the HSTMAC protocol. These four performance metrics are defined as follows.

The average success probability, i.e.,  $P_S$ , is defined as the ratio of the total number of successful devices to the total number of newly arrived devices during  $I_{\max}$  cycles:

$$\begin{aligned} P_S^{H2H} &= \frac{\sum_{i=1}^{I_{\max}} Z_i^{H2H}}{\lambda I_{\max}} \\ &= \frac{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q Z_i^{H2H}[q]}{\lambda I_{\max}} \end{aligned} \quad (25)$$

$$\begin{aligned} P_S^{M2M} &= \frac{\sum_{i=1}^{I_{\max}} Z_i^{M2M}}{N^{M2M}} \\ &= \frac{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q Z_i^{M2M}[q]}{N^{M2M}} \end{aligned} \quad (26)$$

The resource utilization ratio  $U$  is defined as the ratio of the total number of RBs used for successful packet transmissions to the total number of reserved RBs:

$$U^{H2H} = \frac{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q \delta_1 Z_i^{H2H}[q]}{\sum_{i=1}^{I_{\max}} L_{TP,i}} \quad (27)$$

$$U^{M2M} = \frac{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q \delta_2 Z_i^{M2M}[q]}{\sum_{i=1}^{I_{\max}} L_{SP,i}} \quad (28)$$

The delay of a successful device is defined as the time intervals (granularity in cycle) between its first arrival time and the time when it sends its packet successfully. So the statistical average delay  $D$  is computed as:

$$D = \frac{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q \sum_{j=\max[i-(q-1)(W_{BO}+1),1]}^{i-(q-1)} (i-j) Z_{i,j}[q]}{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q Z_i[q]} \quad (29)$$

The average request number  $B$  is defined as the average number of access attempts that a successful device has conducted:

$$B = \frac{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q q Z_i[q]}{\sum_{i=1}^{I_{\max}} \sum_{q=1}^Q Z_i[q]} \quad (30)$$

## 5. System Equilibrium Analysis under Arbitrary Transmission Limit

### 5.1. System Balance Equation

The authors in [44] analyzed the system equilibrium point of the random access procedure without considering the following data transmission procedure and ACB scheme. In [21], the system equilibrium is derived asymptotically for joint optimization of random access and data transmission under a large transmission limit ( $Q \geq 9$ ). Compared to [21], this section would provide a more rigorous system equilibrium analysis under arbitrary transmission limit  $Q(Q \geq 1)$ .

Apparently, when the system with uniform distributed traffic converges to the equilibrium state, there must be  $X_i = X_j, \forall i, j$  and  $Z_i = Z_j, \forall i, j$ . Let  $f$  denote the one-shot success probability in each cycle, according to Equation (24) we can derive that:

$$\begin{aligned} X_i[2] &= \frac{1}{W_{BO}+1} \sum_{k=i-(W_{BO}+1)}^{i-1} (X_k[1] - Z_k[1]) \\ &= \frac{1}{W_{BO}+1} (W_{BO}+1)(X_{i-1}[1] - Z_{i-1}[1]) \\ &= X_{i-1}[1] - Z_{i-1}[1] \\ &= \lambda(1-f) \end{aligned} \quad (31)$$

where  $\lambda$  is the average number of new arrivals. Therefore, in a recursive way, we can have the following equation:

$$\begin{aligned} X_i[q] &= X_{i-1}[q-1] - Z_{i-1}[q-1] \\ &= \lambda(1-f)^{q-1}, \quad 2 \leq q \leq Q \end{aligned} \quad (32)$$

Omitting the time subscript  $i$ , the system balance equation is derived as follows:

$$\begin{aligned} X &= \sum_{q=1}^Q X[q] \\ &= \sum_{q=1}^Q \lambda(1-f)^{q-1} \\ &= \lambda \frac{1-(1-f)^Q}{f} \end{aligned} \quad (33)$$

**Remark 3.** According to Equations (31)–(33), we can find that the components of devices in each cycle are the same with and without backoff scheme. Therefore, the uniform backoff scheme has no impact on the system performance except delay. Then according to the definitions in Subsection 4.2, the performance metrics in equilibrium state are computed as:

$$\begin{aligned} P_S &= 1 - (1-f)^Q \\ U &= \frac{\delta \sum_{q=1}^Q \lambda(1-f)^{q-1} f}{L} \\ &= \frac{\delta \lambda P_S}{L} \\ B &= \sum_{q=1}^Q q \cdot \frac{(1-f)^{q-1} f}{1-(1-f)^Q} \\ D &= \sum_{q=1}^Q (q-1) \frac{1+(W_{BO}+1)}{2} \cdot \frac{(1-f)^{q-1} f}{1-(1-f)^Q} \end{aligned} \quad (34)$$

where  $L$  is the reserved resource blocks and  $\frac{1+(W_{BO}+1)}{2}$  is the average delay caused by backoff.

## 5.2. Equilibrium Analysis for S-ALOHA System

We define the system load as  $\rho = \frac{\lambda}{L/\delta}$ . When ACB is enabled, the average number of contending devices on a data channel is defined as  $s = \frac{Xp}{L/\delta}$ . Let  $p = 1$ , i.e., disable the ACB scheme, we can have the same system balance equation as Equation (11) in [44]:

$$\begin{aligned} f &= e^{-\frac{X}{L/\delta}} = e^{-s} \\ \frac{X}{\lambda} &= \frac{s}{\rho} = \frac{1-(1-e^{-s})^Q}{e^{-s}} \end{aligned} \quad (35)$$

That is to say if the ACB is not considered, we can obtain all the system equilibrium points by solving Equation (35). Then according to [44], there may be multiple solutions and some of which are on  $s \in (1, +\infty)$ . However, when ACB scheme is enabled, there must be  $s = \frac{Xp}{L/\delta} \leq 1$ . Therefore, we have to discuss the system equilibrium under different system load and retransmission limit case by case.

### 5.2.1. Equilibrium Analysis in the Saturation State

If the S-ALOHA system is in a saturation state when it converges to the equilibrium state, there must be  $s = 1$  according to Section 3.2. A successful device should first pass the ACB check and then succeed in the data transmission procedure without collision. Therefore, we can rewrite the system balance equation as follows:

$$\begin{aligned} f &= p e^{-\frac{Xp}{L/\delta}} \\ &= p e^{-1} \\ X &= \frac{L/\delta}{p} \\ &= \lambda \frac{1-(1-f)^Q}{f} \end{aligned} \quad (36)$$

Then we can derive that:

$$\begin{aligned} p &= e \left[ 1 - \left( 1 - \frac{L/\delta}{e\lambda} \right)^{1/Q} \right] \\ f &= 1 - \left( 1 - \frac{L/\delta}{e\lambda} \right)^{1/Q} \\ P_s &= \frac{L/\delta}{e\lambda} \end{aligned} \quad (37)$$

Due to ACB scheme, there must have  $0 \leq p \leq 1$ , so we can derive the following inequation:

$$0 \leq L \leq e\lambda\delta \left[ 1 - \left( 1 - e^{-1} \right)^Q \right] \quad (38)$$

In this paper, we define the saturation threshold in the equilibrium state as:

$$L_{th}^{S-ALOHA} = e\lambda\delta \left[ 1 - \left( 1 - e^{-1} \right)^Q \right] \quad (39)$$

When  $L = L_{th}^{S-ALOHA}$ , the system load  $\rho_{th}$  at the saturation threshold is obtained as:

$$\rho_{th} = \frac{\lambda}{L_{th}^{S-ALOHA}/\delta} = \frac{e^{-1}}{\left[ 1 - \left( 1 - e^{-1} \right)^Q \right]} \quad (40)$$

**Remark 4.** Based on the above analysis, there must be  $\rho \geq \rho_{th} \left( L \leq L_{th}^{S-ALOHA} \right)$  if the system converges to a saturated equilibrium state. Until this step we just have demonstrated that  $\rho \geq \rho_{th} \left( L \leq L_{th}^{S-ALOHA} \right)$  is the necessary condition for a saturated equilibrium state. In the following subsections, we still need to prove that  $\rho \geq \rho_{th} \left( L \leq L_{th}^{S-ALOHA} \right)$  is also the sufficient condition to ensure the system is saturated and to prove that there is only one equilibrium point on  $s \in (0,1]$  if  $\rho < \rho_{th}$ .

### 5.2.2. Equilibrium Analysis in the Non-Saturation State for $Q \leq 8$

Let us define function  $h(s)$  as follows:

$$h(s) := \frac{1 - (1 - e^{-s})^Q}{e^{-s}} \quad (41)$$

Obviously, the solutions for Equation (35) are the cross points of line  $s/\rho$  and curve  $h(s)$ . Function  $h(s)$  has exactly one convex piece on  $(0, s_0)$  and one concave piece on  $(s_0, +\infty)$  for  $Q \geq 1$ . Especially for  $Q \leq 8$ , there is only a single solution for Equation (35) on  $(0, +\infty)$  (the proof is not a contribution of this paper, readers may refer to [44] for more details). Apparently,  $s/\rho_{th}$  has a cross point with  $h(s)$  at point  $s = 1$ . Therefore, as shown in Figure 3, there must be precisely a single equilibrium point on  $(0,1)$  if  $\rho < \rho_{th}$ . On the other hand, there must be precisely a single solution for Equation (35) on  $(1, +\infty)$  if  $\rho > \rho_{th}$ , which means that the system is in a saturation state.

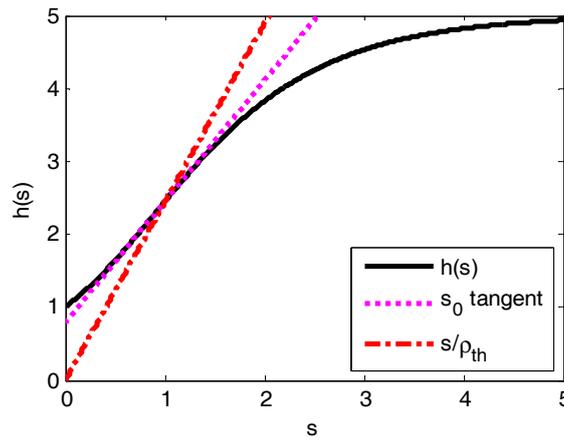


Figure 3. Function  $h(s)$  for  $Q = 5$ ,  $s_0$  is the inflexion point of  $h(s)$ .

5.2.3. Equilibrium Analysis in the Non-Saturation State for  $Q \geq 9$

For  $Q \geq 9$ , there are two load boundaries  $0 < \rho_1 < \rho_2 < \infty$ . Actually, as shown in Figure 4a,  $s/\rho_1$  is the tangent of  $h(s)$  on  $(s_0, +\infty)$  and  $s/\rho_2$  is the tangent of  $h(s)$  on  $(0, s_0)$ . Let  $s_2$  denote the point of tangency of  $s/\rho_2$  and  $h(s)$ , we can derive that:

- (i) for  $\rho > \rho_2$ , there is only a single equilibrium point on  $(s_2, +\infty)$ ;
- (ii) for  $\rho < \rho_2$ , there is only a single equilibrium point on  $(0, s_2)$  and there are two equilibrium points on  $(s_2, +\infty)$  for  $\rho_1 > \rho > \rho_2$  and one equilibrium point on  $(s_2, +\infty)$  for  $\rho > \rho_1$ .

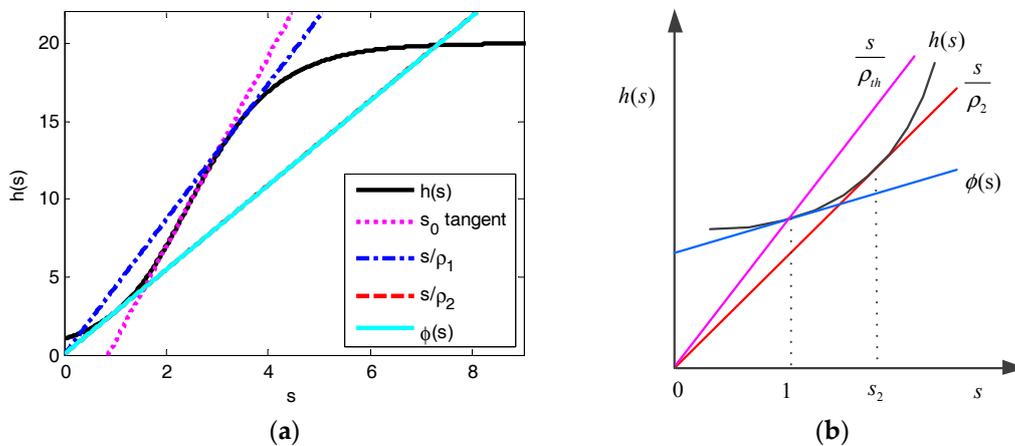


Figure 4. Function  $h(s)$  for  $Q = 20$ ,  $s_0$  is the inflexion point of  $h(s)$ : (a) Function  $h(s)$ ; (b) Illustration for proof of  $s_2 \geq 1$ .

As shown in Figure 4b, if we can prove that  $s_2 \geq 1$ , there must be precisely a single equilibrium point on  $(0,1)$  for  $\rho < \rho_{th}$ .

As shown in Figure 4b, we can have the tangent  $\phi(s)$  of  $h(s)$  at  $s = 1$ :

$$\phi(s) = h'(1)s - h'(1) + h(1) \tag{42}$$

The vertical intercept of  $\phi(s)$  is computed as:

$$\begin{aligned} \phi(0) &= -h'(1) + h(1) \\ &= Q(1 - e^{-1})^{Q-1} \\ &> 0 \end{aligned} \tag{43}$$

So we must have  $s_2 \geq 1$ . That is to say, there is only a single equilibrium point on  $(0,1)$  if  $\rho < \rho_{th}$ . For  $\rho > \rho_{th}$ , all the solutions for Equation (35) are on  $(1,+\infty)$ , which means that the system is in a saturation state.

To sum up, if the ACB scheme is enabled, we can derive the following conclusions for S-ALOHA system under  $Q \geq 1$

1. If  $L \leq L_{th}^{S-ALOHA}$ , i.e.,  $\rho \geq \rho_{th}$ , the system would surely converge to a saturated equilibrium state. Therefore, the equilibrium point cannot be computed by Equation (35) directly and the ACB scheme would ensure that  $s = 1$ . Then the average success probability and resource utilization ratio can be computed as:

$$\begin{aligned} P_S &= \frac{L/\delta}{e\lambda} \\ U &= \frac{\delta\lambda P_S}{L} = e^{-1} \end{aligned} \quad (44)$$

2. If  $L > L_{th}^{S-ALOHA}$ , i.e.,  $\rho < \rho_{th}$ , the ACB scheme would ensure the system converges to an equilibrium state which is non-saturate. The single unique equilibrium point  $s \in (0,1)$  can be computed by Equation (35). In the non-saturation state we can have:

$$\begin{aligned} P_S &= 1 - (1 - e^{-s})^Q \\ U &= \frac{\delta\lambda P_S}{L} \end{aligned} \quad (45)$$

### 5.3. Equilibrium Analysis for TDMA System

For the TDMA system, we can define the system load as  $\rho = \lambda/L\theta\gamma$ . Considering the ACB factor, the average number of contending devices on a random access opportunity is defined as  $s = Xp/L\theta\gamma$ . Substituting  $\rho, s$  into Equation (35), we can have the balance equation for the TDMA system. The following condition must be satisfied:

$$\begin{aligned} \delta X p e^{-\frac{Xp}{L\theta\gamma}} &= \delta X p e^{-s} \\ &= L(1 - \theta) \end{aligned} \quad (46)$$

The capacity of TDMA system is achieved when  $\theta = \theta^*$ . Hence, in a similar way to the S-ALOHA system, we can obtain the saturation threshold in the equilibrium state as:

$$L_{th}^{TDMA} = \frac{e\lambda \left[ 1 - (1 - e^{-1})^Q \right]}{\theta^* \gamma} \quad (47)$$

Therefore, we can derive the following conclusions for TDMA system under  $Q \geq 1$ :

1. If  $L < L_{th}^{TDMA}$ , the system would surely converge to a saturated equilibrium state. Therefore, the ACB scheme would ensure that  $s = 1$ . Then the average success probability and resource utilization ratio can be computed as:

$$\begin{aligned} p &= e \left[ 1 - \left( 1 - \frac{L\theta^*\gamma}{e\lambda} \right)^{1/Q} \right] \\ f &= 1 - \left( 1 - \frac{L\theta^*\gamma}{e\lambda} \right)^{1/Q} \\ P_S &= \frac{L\theta^*\gamma}{e\lambda} \\ U &= \frac{\delta\lambda P_S}{L} = 1 - \theta^* \end{aligned} \quad (48)$$

2. If  $L > L_{th}^{TDMA}$ , the ACB scheme would ensure the system converges to an equilibrium state which is non-saturate. The single unique equilibrium point  $s \in (0,1)$  and the corresponding  $\theta$  can be computed by Equations (35) and (46). In the non-saturation state we can have:

$$\begin{aligned} P_S &= 1 - (1 - e^{-s})^Q \\ U &= \frac{\delta \lambda P_S}{L} = 1 - \theta \end{aligned} \quad (49)$$

#### 5.4. QoS Guaranteed Resource Allocation and Dynamic Pre-Backoff Algorithm

For Poisson distributed H2H traffic, a backoff scheme can only increase delays without improving the other performance metrics. Hence, in this paper the backoff scheme is not used for H2H traffic. We will first satisfy the resource requirement of H2H traffic due to its high priority. Then according to the prior equilibrium analysis, the reserved resource blocks  $L_{TP}$  for a given QoS requirement can be computed by Equations (34), (48) and (49).

As to the M2M traffic, the maximum supportable arrival rate for a given available resources  $L_{SP}$  can be computed by Equation (39):

$$\lambda_{th}^{M2M} = \frac{L_{SP}}{\delta_2 e \left[ 1 - (1 - e^{-1})^Q \right]} \quad (50)$$

The pre-backoff scheme should be enabled to balance the system load for beta distributed M2M traffic if  $\lambda \geq \lambda_{th}^{M2M}$ . Therefore in this paper, we design an aggressively increase conservatively decrease based dynamic pre-backoff scheme as stated in Algorithm 1, in which  $\eta$  is a pre-defined time parameter to control the pre-backoff window size. In the view of the base station,  $X_i[1]$  is the actual rate of new arrivals. Each new arrival in the  $i$ -th cycle would randomly select a backoff interval  $\omega_{PBO} \in [0, W_{PBO}]$  and send its first access request in the  $i + \omega_{PBO}$ -th cycle.

---

#### Algorithm 1: Dynamic pre-backoff algorithm (DPBO)

---

1. **Input:**  $L, \lambda, N_{M2M}, T, \alpha, \beta, \theta, \gamma, \tau, \delta_1, \delta_2, \eta, Q, W_{BO}$ , QoS requirements of H2H traffic
  2. **Initialization:**
  3. Solve Equation (13) and return  $I_{max}$ ;
  4.  $X_i^{M2M}[1] = 0, \quad \forall i \in [1, I_{max}]$ ;
  5.  $W_{PBO} = 0$ ;
  6. **Step1:** Solve Equations (34), (48) and (49) and return  $L_{TP}$ ;
  7. Solve Equation (10) and return  $L_{SP}$ ;
  8. Solve Equation (50) and return  $\lambda_{th}^{M2M}$ ;
  9. **Step2:** for each cycle  $i = 1 : I_{max}$  **do**
  10. Solve Equation (12) and return  $N_i^{M2M}$ ;
  11. # Newly arrived M2M devices execute pre-backoff procedure
  12. **for**  $k = 1 : W_{PBO} + 1$  **do**
  13.  $X_{i+k-1}^{M2M}[1] = X_{i+k-1}^{M2M}[1] + \frac{1}{W_{PBO}+1} N_i^{M2M}$ ;
  14. **end for**
  15. # Adjust the pre-backoff window size dynamically
  16. **if**  $X_i^{M2M}[1] \geq \lambda_{th}^{M2M}$
  17.  $W_{PBO} = W_{PBO} + 1$ ;
  18. **else if**  $X_i^{M2M}[1] < \lambda_{th}^{M2M}$  continuously in time interval  $[i - \eta, i]$
  19.  $W_{PBO} = W_{PBO} - 1$ ;
  20. **end if**
  21.  $I_{max} = \max[i + W_{PBO} + (Q - 1)(W_{BO} + 1), I_{max}]$
  22. **end for**
-

## 6. Numerical and Simulation Results

In this paper, we consider a TDD cellular cell and the corresponding system parameters are defined in Table 1 [11]. The simulations are performed using Matlab and the Monte-Carlo results are obtained after 1000 runs. As to the QoS provision for H2H traffic, we assume that the reserved RBs would ensure H2H traffic converges to the equilibrium state which is precisely at the saturation threshold. The corresponding success probability at the saturation threshold under different transmission limit is listed in Table 2. In order to compare the performance of the proposed HSTMAC protocol with pure S-ALOHA protocol and pure TDMA protocol, we assume that the resources in S-ALOHA and TDMA protocols are also split for M2M traffic and H2H traffic separately.

**Table 1.** Performance Evaluation Parameters for Proposed HSTMAC Protocol.

Parameter	Settings
Cell bandwidth	10 MHz (50 RBs in frequency domain)
Cycle interval $\tau$	5 ms
Reserved RBs for uplink transmission $L$	200
Transmission limit $Q$	5, 10
H2H traffic model	Poisson distribution
M2M traffic model	Beta distribution
Signaling overhead parameter for H2H traffic $\gamma$	2
Packet size of H2H traffic $\delta_1$	2
Packet size of M2M traffic $\delta_2$	0.5
Number of M2M devices $N^{M2M}$	10,000–60,000
Activation time of M2M devices $T$	10 s
Beta function parameters $\alpha, \beta$	3, 4
Backoff window size for M2M traffic $W_{BO}$	20
DPBO time control parameter for M2M traffic $\eta$	10

**Table 2.** Success Probability at the Saturation Threshold under Different Transmission Limit.

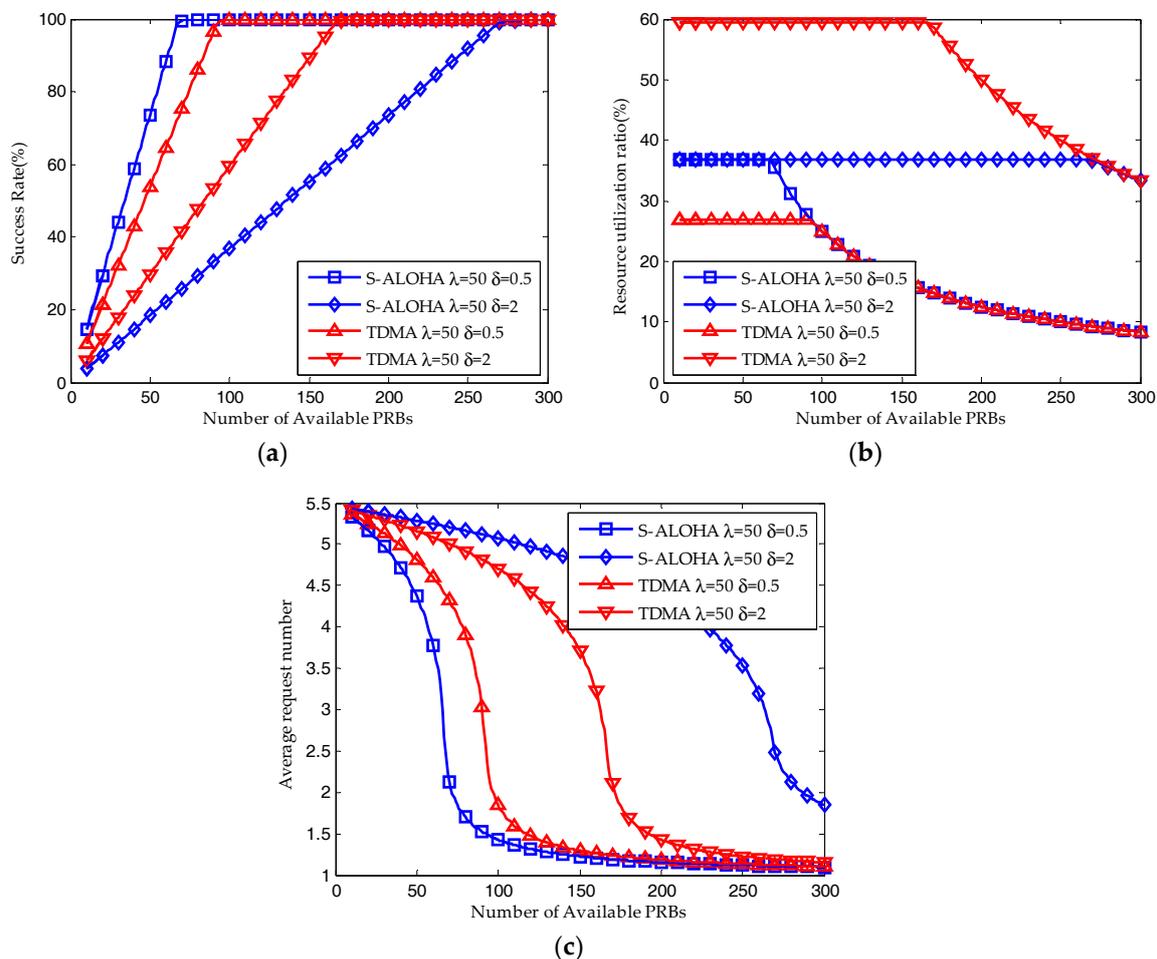
$Q$	1	2	3	4	5	6	7	8	9	10
$P_S$	0.3679	0.6004	0.7474	0.8403	0.8991	0.9362	0.9597	0.9745	0.9839	0.9898

### 6.1. System Performance versus Packet Size

In this and the next subsections, we will validate the analysis of system capacity and system equilibrium for the pure S-ALOHA protocol and pure TDMA protocol. For this purpose, we assume that the arrival rate of new packets follows a Poisson distribution and the evaluation time is set to 500 cycles.

Shown in Figure 5, the system performance of S-ALOHA and TDMA protocols with different packet size are derived by the iterative method. For small packet size  $\delta = 0.5$ , as shown in Figure 5a,b, when the available resource blocks are insufficient, both the average success probability and resource utilization ratio of S-ALOHA protocol is higher than that of TDMA protocol. At the same time, as shown in Figure 5c, the average request number of S-ALOHA protocol is lower than that of TDMA protocol. When the number of available resource blocks is large enough, the success probability and resource utilization ratio of S-ALOHA protocol are almost the same with that of TDMA protocol while the average request number of S-ALOHA protocol is still lower than that of TDMA protocol. For large packet size  $\delta = 2$ , on the contrary, when the available resource blocks are insufficient, both the average success probability and resource utilization ratio of S-ALOHA protocol is lower than that of TDMA protocol and the average request number of S-ALOHA protocol is higher than that of TDMA protocol. When the number of available resource blocks is large enough, the success probability and resource utilization ratio of S-ALOHA protocol are almost the same with that of TDMA protocol while the average request number of S-ALOHA protocol is still higher than that of TDMA protocol.

Therefore, we can validate the analysis in Section 3.1 and make a conclusion that the S-ALOHA protocol is superior to the TDMA protocol for small packet transmission while the TDMA protocol is superior to the S-ALOHA protocol for large packet transmission.

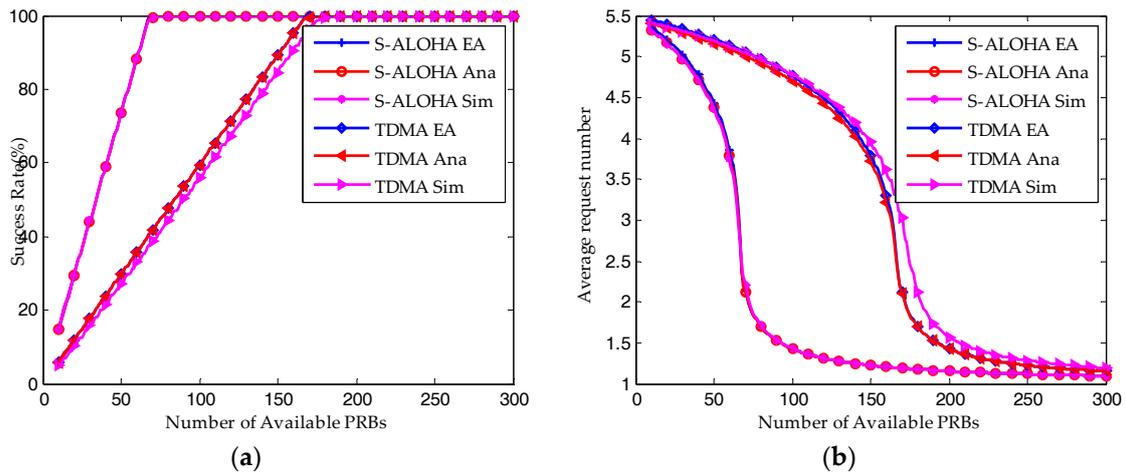


**Figure 5.** Performance results of pure S-ALOHA protocol and pure TDMA protocol with different packet size ( $Q = 10, W_{BO} = 0$ ): (a) Average success probability; (b) Resource utilization ratio; (c) Average request number.

## 6.2. Equilibrium Analysis

As shown in Figure 6, the simulation results (Sim in the figure) match well with the equilibrium analysis (EA in the figure) and analytical results (Ana in the figure). For S-ALOHA protocol, the maximum difference is found in the average request number, which is 5.54%. For TDMA protocol, though the simulation results are shifting to the right, they have the same tendencies with the equilibrium analysis and analytical results.

As shown in Figures 7 and 8, when the available RBs are insufficient, the ACB scheme would ensure the system converges to a saturated equilibrium state. Then according to Equations (44) and (48), for both the S-ALOHA and TDMA protocols, the average success probability is increasing linearly with the number of available RBs and the resource utilization ratio remains constant. The transmission limit has no impact on the success probability and resource utilization ratio under such a condition. However, as shown in Figure 8, the average request number for  $Q = 10$  is much higher than that for  $Q = 5$ .

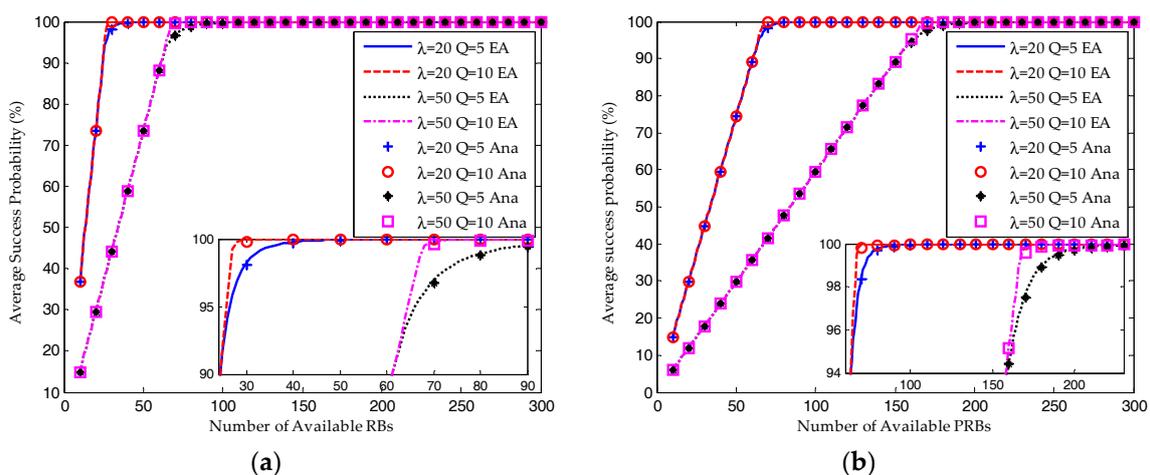


**Figure 6.** Comparison of simulation results with equilibrium analysis and iterative analytical results ( $\lambda = 50, Q = 10, W_{BO} = 0$ ): (a) Average success probability; (b) Average request number.

As shown in Figure 7, when the number of available RBs is around the saturation threshold, the average success probability is still increasing but the growth is decreasing. The success probability for  $Q = 10$  is a little higher than that for  $Q = 5$ . And the request number for  $Q = 10$  is still higher than that for  $Q = 5$ . This demonstrates that the device can increase its success probability by trying more times.

When the number of RBs is large enough, almost all the devices can succeed in their transmission attempts. However, the resource utilization ratio is decreasing due to the underutilization of RBs. On the other hand, as shown in Figures 7–9, the performance results for  $Q = 10$  and  $Q = 5$  are almost the same.

As a conclusion, the system performance can be improved dramatically by increasing the available resources when it is in the saturation state. However, the performance improvement would be slight when the system is in the non-saturation state. Increasing the transmission limit can impose a positive impact on the success probability at a cost of increasing in the average request number, which means more energy consumption and higher delay.



**Figure 7.** Comparison of average success probability between equilibrium analysis and iterative analytical results ( $W_{BO} = 0$ ): (a) S-ALOHA protocol; (b) TDMA protocol.

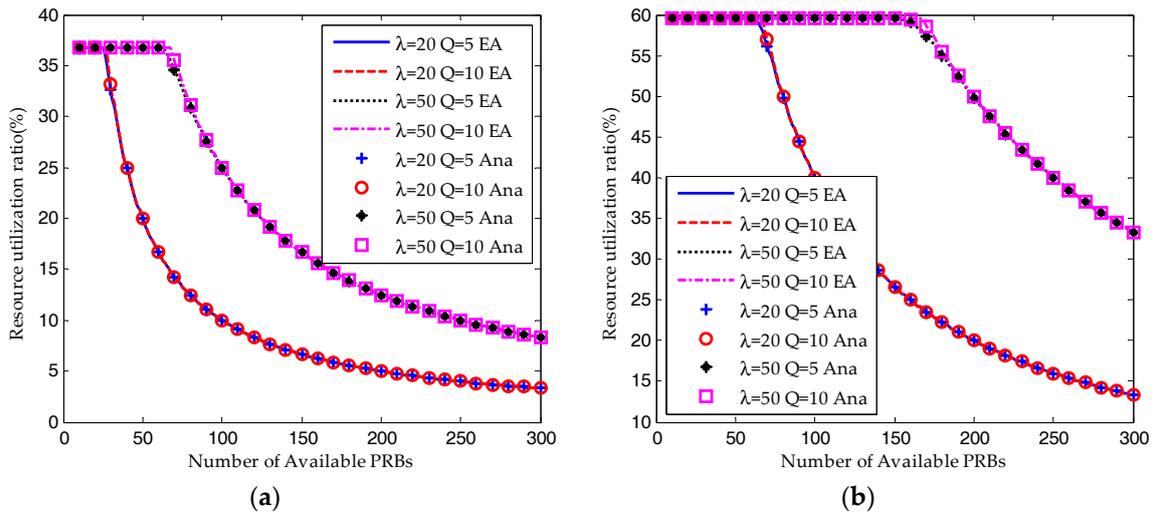


Figure 8. Comparison of resource utilization ratio between equilibrium analysis and iterative analytical results ( $W_{BO} = 0$ ): (a) S-ALOHA protocol; (b) TDMA protocol.

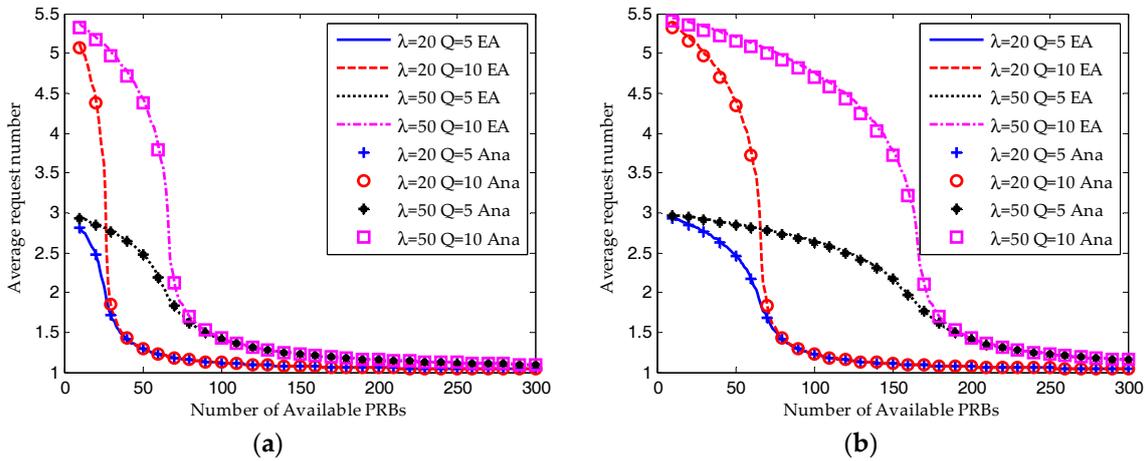
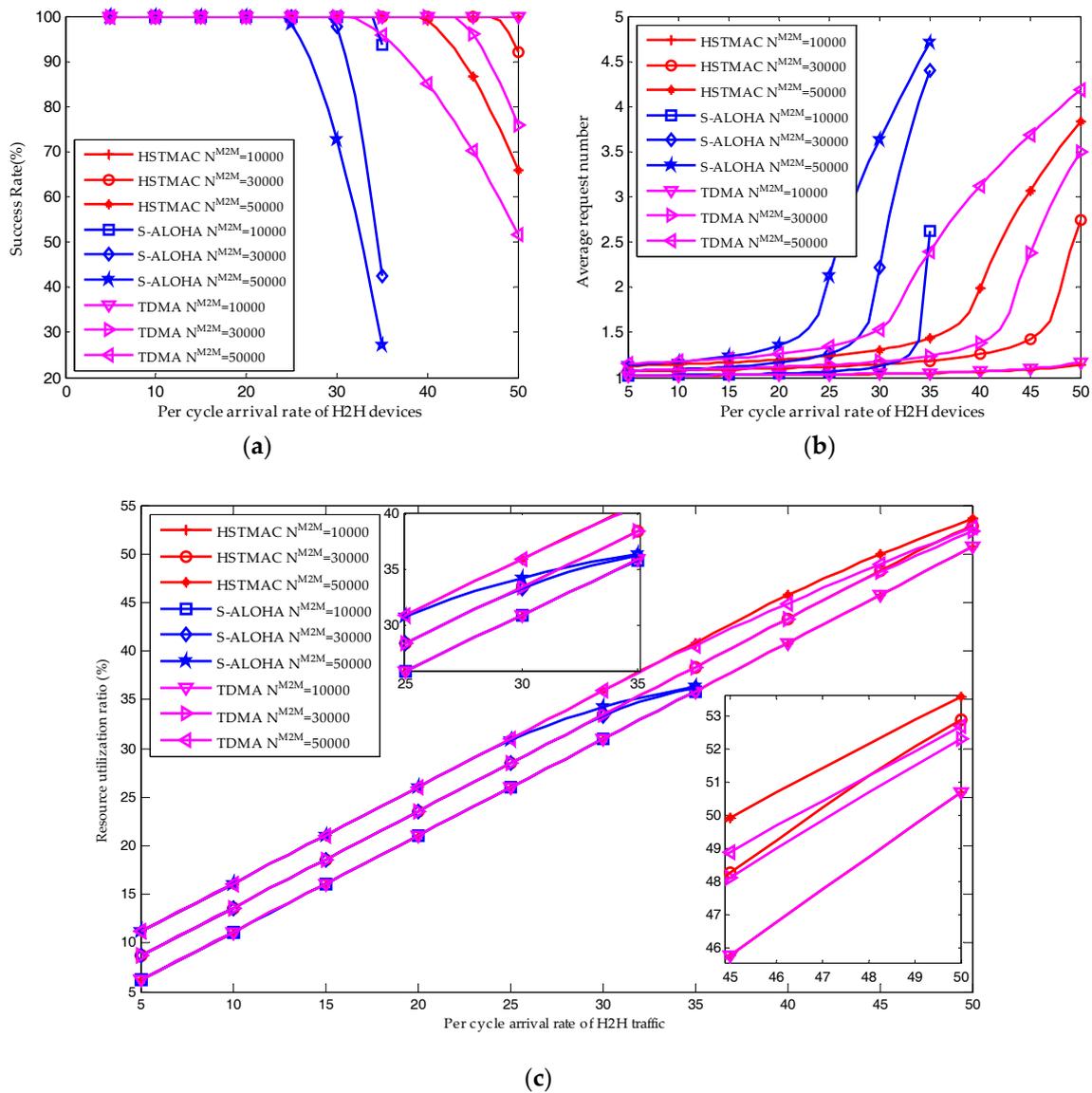


Figure 9. Comparison of request number between equilibrium analysis and iterative analytical results ( $W_{BO} = 0$ ): (a) S-ALOHA protocol; (b) TDMA protocol.

### 6.3. Performance of the HSTMAC Protocol without Backoff Scheme

As shown in Figure 10a, the average success probability of M2M traffic is always decreasing with the arrival rate of H2H packets. On the other hand, shown in Figure 10c, the average resource utilization ratio is increasing with the arrival rate of H2H packets. That is because the number of reserved RBs for M2M traffic is decreasing and the reserved resources are insufficient to satisfy the QoS requirement of M2M traffic. Under such a condition, according to the equilibrium analysis in Section 5, the resource utilization ratio is increasing with the system load.



**Figure 10.** Performance comparison of the proposed HSTMAC protocol with pure S-ALOHA protocol and pure TDMA protocol ( $Q = 10$ ,  $W_{BO} = 0$ ): (a) Average success probability of M2M traffic; (b) Average request number of M2M traffic; (c) Resource utilization ratio.

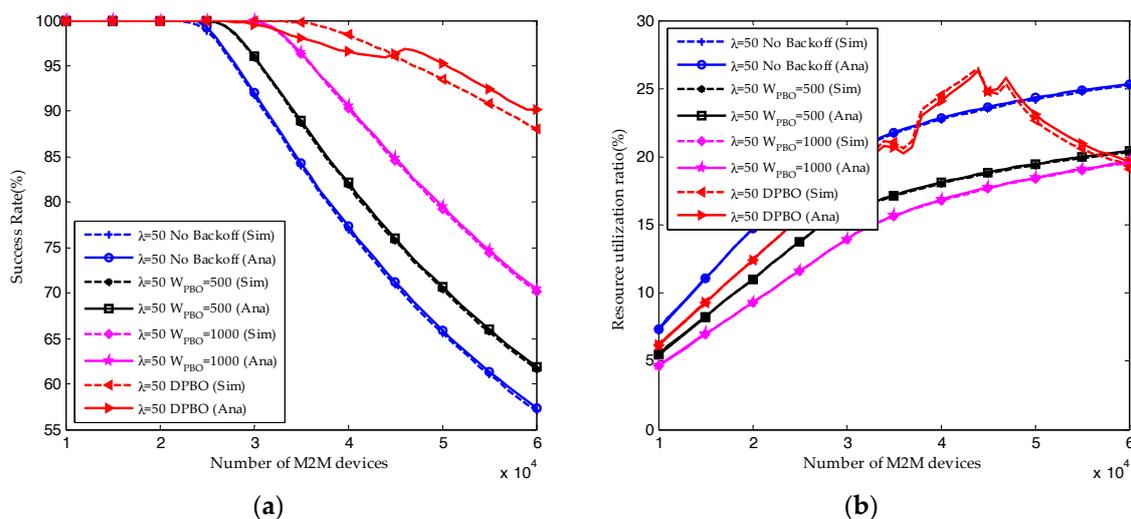
As the number of M2M devices is increasing from 10,000 to 50,000, there is a growth in the system load, which results in a higher resource utilization ratio. The reserved RBs are insufficient to satisfy the QoS requirement of M2M traffic. Hence, as shown in Figure 10a,b, there is a remarkable decrease in the average success probability and there is a remarkable increase in the average request number.

Apparently, Figure 10 shows that the average success probability and resource utilization ratio of the proposed HSTMAC protocol are always higher than that of S-ALOHA and TDMA protocols while the average request number is always lower than that of S-ALOHA and TDMA protocols. The reason is that S-ALOHA is appropriate for small size packet transmission while TDMA is appropriate for large size packet transmission. In order to satisfy the QoS requirement of H2H traffic, the number of reserved RBs for H2H traffic in S-ALOHA protocol is much higher than that in TDMA protocol. As shown in Figure 10, for a high H2H traffic load in the S-ALOHA protocol, the QoS cannot be satisfied even if H2H traffic occupies all the RBs. Therefore, comparing the HSTMAC protocol to the TDMA protocol, the number of reserved RBs for M2M traffic are the same. However, comparing the

HSTMAC protocol to the S-ALOHA protocol, the number of reserved resource blocks for M2M traffic is much higher. As a result, the proposed HSTMAC protocol is superior to the pure TDMA protocol and pure S-ALOHA protocol especially in a high system load scenario.

#### 6.4. Performance of the HSTMAC Protocol with Dynamic Pre-Backoff Scheme

As shown in Figure 11, we validate the analytical analysis of DPBO algorithm through simulation results. For DPBO algorithm, jitters can be found in the resource utilization ratio and average request number. The reason is that the pre-backoff window size of DPBO is changing dynamically according to the traffic load, which results in the dynamically changing of the transmission interval  $I_{\max}$  and the distribution of devices in different transmission stages.

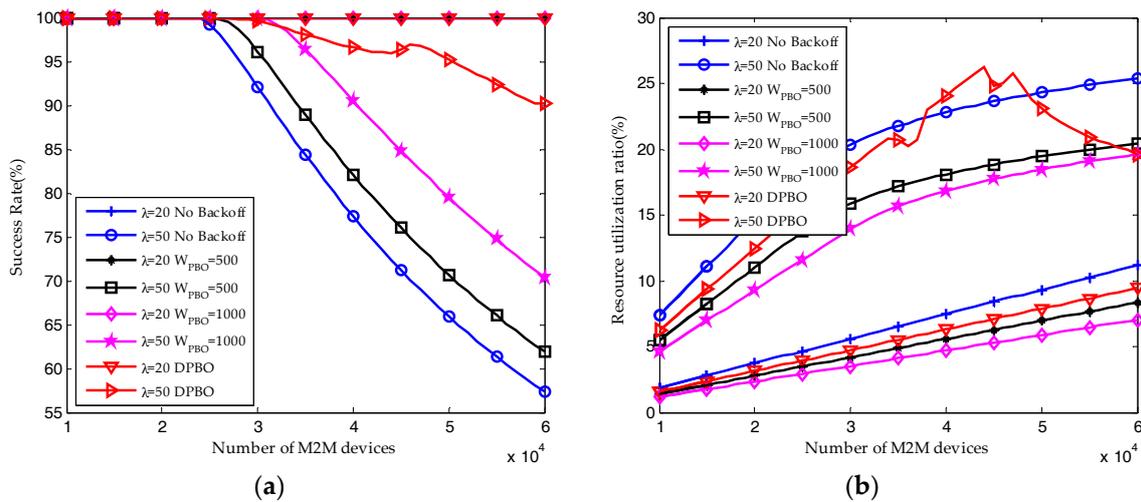


**Figure 11.** Performance comparisons of HSTMAC protocol between analytical and simulation results ( $\lambda = 50$ ,  $Q = 10$ ,  $W_{BO} = 20$ ): (a) Average success probability of M2M traffic; (b) Average resource utilization ratio of M2M traffic.

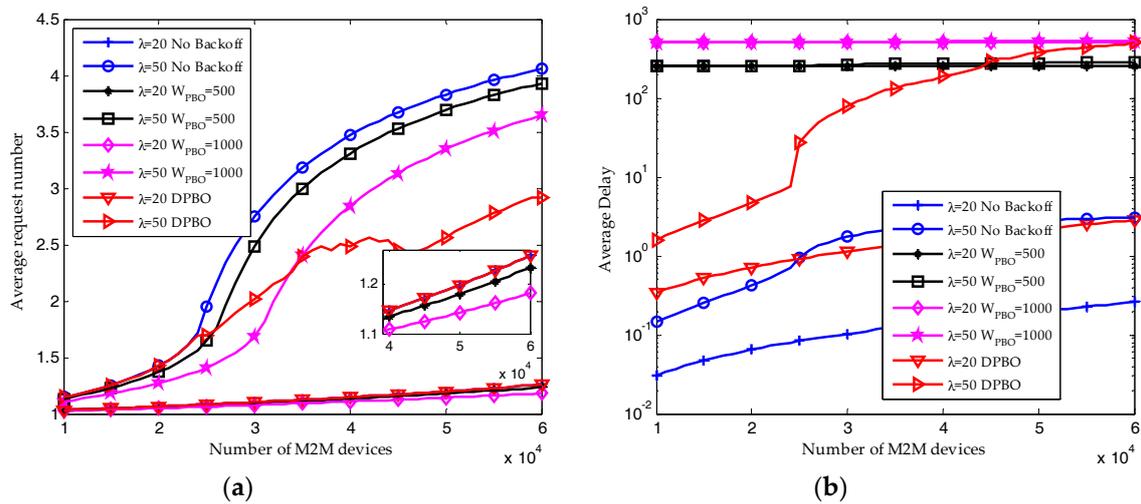
When the arrival rate of H2H packets is  $\lambda = 20$ , the reserved RBs are sufficient to satisfy the QoS requirement of M2M traffic. Shown in Figure 12a, almost all the M2M devices can send their packets successfully, and the resource utilization ratio of M2M traffic shown in Figure 12b is increasing linearly with the number of M2M devices. Under this condition, the pre-backoff window size in DPBO algorithm would remain close to zero. The time interval  $I_{\max}$  is increasing with the pre-backoff and backoff window size. Hence, as shown in Figures 10 and 11, the resource utilization ratio and the average request number in the DPBO algorithm is higher while the average delay is much lower than that in the UPBO algorithm.

When the arrival rate of H2H packets is  $\lambda = 50$ , the system quickly becomes congested due to the growth in the number of M2M devices. Shown in Figures 12a and 13a, the performance results without backoff scheme are the worst. On the other hand, though the average success probability is decreasing with the number of M2M devices, the results in DPBO algorithm is better than that in scenarios with UPBO algorithm. In a scenario with high system load, as shown in Figures 12b and 13a, the resource utilization ratio in DPBO algorithm is higher while the average request number is lower than that in UPBO algorithm. Moreover, as shown in Figure 13b, the average delay in DPBO algorithm is lower or in the same order with that in UPBO algorithm.

As a conclusion, we can derive that pre-backoff scheme can improve the success probability at cost of lowering the resource utilization ratio and increasing the delay. However, the proposed DPBO algorithm can improve the system performance at a much lower cost when compared to the UPBO algorithm.



**Figure 12.** Performance results of HSTMAC protocol with dynamic pre-backoff scheme ( $Q = 10$ ,  $W_{BO} = 20$ ): (a) Average success probability of M2M traffic; (b) Average resource utilization ratio of M2M traffic.



**Figure 13.** Performance results of HSTMAC protocol with dynamic pre-backoff scheme ( $Q = 10$ ,  $W_{BO} = 20$ ): (a) Average request number of M2M traffic; (b) Average delay of M2M traffic (granularity in cycles).

## 7. Conclusions

In order to jointly optimize the random access procedure and the following data transmission procedure, we propose a hybrid S-ALOHA/TDMA MAC protocol to address the radio access network congestion problem caused by massive M2M devices. We prove that there exists a packet size threshold  $\bar{\delta}$  when considered the signaling overhead in the random access procedure. The S-ALOHA protocol has superiority in small data transmission when  $\delta < \bar{\delta}$  and the TDMA protocol has superiority in large data transmission when  $\delta > \bar{\delta}$ . Then a more rigorous system equilibrium analysis is derived and the results demonstrate that the system performance can be improved dramatically by allocating more resource blocks when it is in a saturation state. However, the improvement becomes negligible when the system is in a non-saturation state. Therefore, the equilibrium analysis could provide the operator with more insights on QoS provision. At last, for the load balance of beta distributed M2M traffic, we design a dynamic pre-backoff scheme. The performance results show that the proposed HSTMAC protocol is superior to both the pure S-ALOHA protocol and pure TDMA protocol in terms

of average success probability and resource utilization ratio, especially when the system load is high. The results also demonstrate that the proposed DPBO algorithm can improve the access success probability and resource utilization ratio at a much lower cost in delay when compared to the uniform pre-backoff algorithm.

**Acknowledgments:** This work is supported by National Natural Science Foundation of China (No. 61771486, No. 61371123, No. 61301165) and Major Research Plan of National Natural Science Foundation of China (No. 91438115).

**Author Contributions:** N.S. conceived the idea and wrote the paper; Y.X. proposed the dynamic pre-backoff algorithm; C.W. conducted the analysis; W.X. performed the experiments.

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

## References

1. Andreev, S.; Galinina, O.; Pyattaev, A. Understanding the IoT connectivity landscape: A contemporary M2M radio technology roadmap. *IEEE Commun. Mag.* **2015**, *53*, 32–40. [[CrossRef](#)]
2. Palattella, M.R.; Dohler, M.; Grieco, A. Internet of things in the 5G era: Enablers, architecture, and business models. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 510–527. [[CrossRef](#)]
3. Kartsakli, E.; Lalos, A.; Antonopoulos, A. A survey on M2M systems for mHealth: A wireless communications perspective. *Sensors* **2014**, *14*, 18009–18052. [[CrossRef](#)] [[PubMed](#)]
4. Chen, M.; Wan, J.; Gonzalez, S.; Liao, X.; Leung, V.C.M. A survey of recent developments in home M2M networks. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 98–114. [[CrossRef](#)]
5. OECD Report. *Machine-to-Machine Communications: Connecting Billions of Devices*; OECD Digital Economy Papers; OECD: Paris, France, 2012; Available online: <http://dx.doi.org/10.1787/5k9gsh2gp043-en> (accessed on 26 August 2017).
6. 3GPP TS 22.368 V13.1.0. Service Requirements for Machine-Type Communications (MTC). December 2014. Available online: [http://www.3gpp.org/ftp/Specs/2014-12/Rel-13/22\\_series/](http://www.3gpp.org/ftp/Specs/2014-12/Rel-13/22_series/) (accessed on 8 December 2017).
7. Shariatmadari, H.; Ratasuk, R.; Iraj, S.; Laya, A.; Taleb, T.; Jäntti, R.; Ghosh, A. Machine-type communications: Current status and future perspectives toward 5G systems. *IEEE Commun. Mag.* **2015**, *53*, 10–17. [[CrossRef](#)]
8. Ghavimi, F.; Chen, H.-H. M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 525–549. [[CrossRef](#)]
9. Rico-Alvarino, A.; Vajapeyam, M.; Xu, H.; Wang, X.; Blankenship, Y.; Bergman, J.; Tirronen, T.; Yavuz, E. An overview of 3GPP enhancements on machine to machine communications. *IEEE Commun. Mag.* **2016**, *54*, 14–21. [[CrossRef](#)]
10. 3GPP TS 36.211 V12.8.0. Evolved Universal Terrestrial Radio Access (E-UTRA): Physical Channels and Modulation. December 2015. Available online: [http://www.3gpp.org/ftp/Specs/2015-12/Rel-12/36\\_series/](http://www.3gpp.org/ftp/Specs/2015-12/Rel-12/36_series/) (accessed on 8 December 2017).
11. 3GPP TR 37.868 V11.0.0. Study on RAN Improvements for Machine-Type Communications. September 2011. Available online: [http://www.3gpp.org/ftp/Specs/2011-09/Rel-11/37\\_series/](http://www.3gpp.org/ftp/Specs/2011-09/Rel-11/37_series/) (accessed on 8 December 2017).
12. Islam, M.; Taha, A.-E.; Akl, S. A survey of access management techniques in machine type communications. *IEEE Commun. Mag.* **2014**, *52*, 74–81. [[CrossRef](#)]
13. Laya, A.; Alonso, L.; Alonso-Zarate, J. Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 4–16. [[CrossRef](#)]
14. Rajandekar, A.; Sikdar, B. A survey of MAC layer issues and protocols for machine-to-machine communications. *IEEE Internet Things J.* **2015**, *2*, 175–186. [[CrossRef](#)]
15. Soltanmohammadi, E.; Ghavami, K.; Naraghi-Pour, M. A survey of traffic issues in machine-to-machine communications over LTE. *IEEE Internet Things J.* **2016**, *6*, 865–884. [[CrossRef](#)]
16. Shih, M.-J.; Wei, H.-Y.; Lin, G.-Y. Two paradigms in cellular internet-of-things access for energy-harvesting machine-to-machine devices: Push-based versus pull-based. *IET Wirel. Sens. Syst.* **2016**, *6*, 121–129. [[CrossRef](#)]
17. Shafiq, M.Z.; Ji, L.; Liu, A.X.; Pang, J.; Wang, J. Large-scale measurement and characterization of cellular machine-to-machine traffic. *IEEE/ACM Trans. Netw.* **2013**, *21*, 1960–1973. [[CrossRef](#)]

18. Durisi, G.; Koch, T.; Popovski, P. Toward massive, ultrareliable, and low-latency wireless communication with short packets. *Proc. IEEE* **2016**, *104*, 1711–1726. [[CrossRef](#)]
19. Dhillon, H.S.; Huang, H.; Viswanathan, H.; Valenzuela, R.A. Fundamentals of throughput maximization with random arrivals for M2M communications. *IEEE Trans. Commun.* **2014**, *62*, 4094–4109. [[CrossRef](#)]
20. Wiriaatmadja, D.T.; Choi, K.W. Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 33–46. [[CrossRef](#)]
21. Sui, N.; Wang, C.; Xie, W.; Xu, Y. Hybrid S-ALOHA/TDMA protocol for LTE/LTE-A networks with coexistence of H2H and M2M traffic. *KSII Trans. Internet Inf. Syst.* **2017**, *11*, 687–708. [[CrossRef](#)]
22. Duan, S.; Shah-Mansouri, V.; Wang, Z.; Wong, V.W.S. D-ACB: Adaptive congestion control algorithm for bursty M2M traffic in LTE networks. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9847–9861. [[CrossRef](#)]
23. Oh, C.-Y.; Hwang, D.; Lee, T.-J. Joint access control and resource allocation for concurrent and massive access of M2M devices. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 4182–4192. [[CrossRef](#)]
24. Vilgelm, M.; Guersu, M.; Kellerer, W.; Reisslein, M. LATMAPA: Load-adaptive throughput-maximizing preamble allocation for prioritization in 5G random access. *IEEE Access* **2017**, *5*, 1103–1116. [[CrossRef](#)]
25. Kim, J.S.; Munir, D.; Hasan, S.F.; Chung, M.Y. Enhancement of LTE RACH through extended random access process. *Electron. Lett.* **2014**, *50*, 1399–1400. [[CrossRef](#)]
26. 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](#)]
27. Jang, H.S.; Park, H.; Sung, D.K. A non-orthogonal resource allocation scheme in spatial group based random access for cellular M2M communications. *IEEE Trans. Veh. Technol.* **2017**, *66*, 4496–4500. [[CrossRef](#)]
28. Wei, C.-H.; Cheng, R.-G.; Tsao, S.-L. Performance analysis of group paging for machine-type communications in LTE networks. *IEEE Trans. Veh. Technol.* **2013**, *62*, 3371–3382. [[CrossRef](#)]
29. Arouk, O.; Ksentini, A. General model for RACH procedure performance analysis. *IEEE Commun. Lett.* **2016**, *20*, 372–375. [[CrossRef](#)]
30. Koseoglu, M. Lower bounds on the LTE-A average random access delay under massive M2M arrivals. *IEEE Trans. Commun.* **2016**, *64*, 2104–2115. [[CrossRef](#)]
31. Chen, J.; Lin, Y.-T.; Cheng, R.-G. A delayed random access speed-up scheme for group paging in machine-type communications. In Proceedings of the 2015 IEEE International Conference on Communications (ICC 2015), Taipei, Taiwan, 23–25 April 2015; pp. 623–627.
32. Jiang, W.; Wang, X.; Deng, T. Performance analysis of a pre-backoff based random access scheme for machine-type communications. In Proceedings of the 2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG 2014), Taipei, Taiwan, 23–25 April 2014; pp. 1–4.
33. Harwahyu, R.; Wang, X.; Sari, R.F.; Cheng, R.-G. Analysis of group paging with pre-backoff. *EURASIP J. Wirel. Commun. Netw.* **2015**, *2015*, 1–9. [[CrossRef](#)]
34. Arouk, O.; Ksentini, A.; Taleb, T. Group paging-based energy saving for massive MTC accesses in LTE and beyond networks. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 1086–1102. [[CrossRef](#)]
35. Jian, X.; Liu, Y.; Wei, Y.; Zeng, X.; Tan, X. Random access delay distribution of multichannel Slotted ALOHA with its applications for machine type communications. *IEEE Internet Things J.* **2017**, *4*, 21–28. [[CrossRef](#)]
36. De Andrade, T.P.C.; Astudillo, C.A.; da Fonseca, N.L.S. The impact of massive machine type communication devices on the access probability of human-to-human users in LTE networks. In Proceedings of the 2014 IEEE Latin-America Conference on Communications (LATINCOM), Cartagena de Indias, Colombia, 5–7 November 2014; pp. 1–6.
37. Foddis, G.; Garroppo, R.G.; Giordano, S.; Procissi, G.; Roma, S.; Topazzi, S. On RACH preambles separation between human and machine type communication. In Proceedings of the 2016 IEEE International Conference on Communications (ICC 2016), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6.
38. Pang, Y.C.; Chao, S.L.; Lin, G.Y.; Wei, H.Y. Network access for M2M/H2H hybrid systems: A game theoretic approach. *IEEE Commun. Lett.* **2014**, *18*, 845–848. [[CrossRef](#)]
39. Sidhu, P.; Woungang, I.; Carvalho, G.H.S.; Anpalagan, A.; Dhurandher, S.K. An analysis of machine-type-communication on human-type-communication over wireless communication networks. In Proceedings of the 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, Gwangju, Korea, 24–27 March 2015; pp. 332–337.

40. Kim, T.; Ko, K.S.; Sung, D.K. Prioritized random access for accommodating M2M and H2H communications in cellular networks. In Proceedings of the 2015 IEEE Globecom Workshops (GC Workshops), San Diego, CA, USA, 6–10 December 2015; pp. 1–6.
41. Liu, Y.; Yuen, C.; Cao, X.; Hassan, N.U.; Chen, J. Design of a scalable hybrid MAC protocol for heterogeneous M2M networks. *IEEE Internet Things J.* **2014**, *1*, 99–111. [[CrossRef](#)]
42. Mistic, J.; Mistic, V.; Khan, N. Sharing it my way: Efficient M2M access in LTE/LTE-A networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 696–709. [[CrossRef](#)]
43. Wei, C.-H.; Cheng, R.-G.; Tsao, S.-L. Modeling and estimation of one-shot random access for finite-user multichannel Slotted ALOHA systems. *IEEE Commun. Lett.* **2012**, *16*, 1196–1199. [[CrossRef](#)]
44. Tyagi, R.R.; Aurzada, F.; Lee, K.-D.; Kim, S.G.; Reisslein, M. Impact of retransmission limit on preamble contention in LTE-Advanced network. *IEEE Syst. J.* **2015**, *9*, 752–765. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).