2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)

Accurate Performance Modeling of Uplink Transmission in NB-IoT

Huikang Li, Gonglong Chen, Yihui Wang, Yi Gao, and Wei Dong*
College of Computer Science, Zhejiang University
Email: {lihk, chengl, wang1hui, gaoyi, dongw}@zju.edu.cn

***Abstract*—With the development of LPWA (Low Power Wide Area) technology, the emerging NB-IoT (Narrowband Internet of Things) has attracted much attention and enabled a wide range of applications. An uplink of NB-IoT is a link from a user equipment (UE) to a base station (BS). Uplink transmission is a key component of NB-IoT, accomplishing the sensor data collection task for many applications. However, the performance of uplink transmission has not been rigorously analyzed in the current literature, while uplink performance degradation like long latency could be harmful to many applications with strict uplink performance requirements. In this work, we show a way of mathematically analyzing the performance of uplink transmission for NB-IoT systems, concerning the transmission latency and transmission reliability. Our model is accurate with consideration of the protocol details and the new features of NB-IoT, including link quality, packet size, channel access contention, and etc. We validate the analytical results through detailed simulations. Results show that our analytical model can achieve 83% accuracy for latency calculation and 96% accuracy for reliability calculation. Moreover, we demonstrate that the analytical results can be used to aid protocol design for performance optimization, e.g., repetition number tuning for reducing the transmission latency.**

Index Terms*—uplink transmission, performance analysis, random access, Markov chain, NB-IoT.

I. Introduction

Recently, Narrowband Internet of Things (NB-IoT), one of the most promising Low Power Wide Area (LPWA) technologies, has attracted much attention. Industries, including Ericsson, Nokia, and Huawei, have shown great interests in NB-IoT as part of 5G systems and spent a lot of efforts on the standardization of NB-IoT, which has been widely considered as a main technique for next-generation wireless communications.

In the recent 3rd Generation Partnership Project (3GPP) specifications [1], NB-IoT is expected to possess many advantages over existing cellular technologies, such as ultra-low power consumption, wide-area coverage (e.g., >10km), support of massive number of devices, and low device complexity. Although characterized by the above attractive features, NB-IoT may not guarantee the satisfactory performance in some user-concerned aspects, e.g., long transmission delay due to the extended coverage. In fact, many applications have strict requirements on the network performance, especially for the uplink transmissions. For example, in real-time structural monitoring the major concern is typically the data transfer delay, and domain experts also require a certain reliability in delivering sensed data [2]. Therefore, it is very necessary to efficiently and accurately model the uplink transmission performance of NB-IoT, which could be useful for performance evaluation, protocol design (e.g., parameter tuning) and optimization in further versions of NB-IoT specifications.

The problem of fine-grained analysis of the NB-IoT uplink transmission performance, however, has not been addressed sufficiently in the current literature. The analysis of [3] only applies to the evaluation of coverage and capacity performance. The analysis of [4] is inaccurate because it does not model the performance of the random access procedure, which is an indispensable step for the data packet transmission. The performance analysis in [5] also suffers from the same issues because it does not consider the packet retransmissions (due to the bad link quality) and the protocol details (i.e., the initial data packet is segmented into a number of fixed-sized blocks by the coding scheme), which is a common technique for data transmission.

In this work, we show a way of mathematically analyzing the performance of uplink data transmission for NB-IoT systems. More specifically, we model the transmission latency and transmission reliability from a UE to a BS. Notice that random access and data transmission have a great effect on the uplink transmission performance, we explicitly analyze both of the key procedures. In particular, we provide a simple but efficient model that accounts for the channel access contention and all the details of random access procedure, including a Markov chain for the random backoff time of unsuccessful channel access attempts. In the data transmission procedure, we analyze the uplink resource allocation and data segmentation based on the current link quality, and packet payload length. Considering the new feature of repeating transmission data in NB-IoT, we also model the impact of repetition number on the transmission performance.

We validate the analytical results through detailed simulations. In addition, we demonstrate that the analytical results can be used to aid protocol design for performance optimization, e.g., repetition number tuning for reducing the transmission latency.

The contributions of this paper are summarized as follows:

- We propose a general and accurate model for uplink transmission in NB-IoT, with consideration of the channel access contention, link quality, and packet payload length during the data delivery.
- We validate the analytical results through extensive simulations. Results show that our analytical model can achieve 83% latency calculation accuracy, and 96% reliability calculation accuracy.
- We demonstrate how our model can be used to aid protocol design and parameter settings for performance optimization.

The rest of this paper is structured as follows. Section II discusses the related work. Section III gives an overview of our analytical model. Section IV introduces the preliminaries,

978-1-5386-7308-9/18/$31.00 ©2018 IEEE
DOI 10.1109/ICPADS.2018.00122

including some important uplink transmission behaviors in NB-IoT. Section V presents the details of our analytical model. Section VI numerically illustrates the performance of NB-IoT uplink transmission in various cases. Finally, Section VII concludes this paper.

II. Related Work

A. NB-IoT technology

LPWA technologies have recently gained a great deal of attention as a topic of research, with a wide range of systems and applications being explored [6]. Following the most recent 3GPP activities on cellular IoT (CIoT) [7], the work on NB-IoT technology has been approved. The NB-IoT is a novel radio-access technology specifically designed for IoT, which can be directly integrated into existing GSM or LTE networks to reduce the deployment cost [8]. In particular, NB-IoT reuses the LTE design extensively, including the numerologies, channel coding, rate matching, etc. Unlike the existing cellular technologies, the bandwidth of both downlink and uplink is further narrowed down to 180 kHz, which thus can provide a 20 dB higher gain in coverage enhancement.

Nevertheless, the narrowed bandwidth is still sufficient to offer connectivity to thousands of connected UEs. Specifically, the band can be divided into 12 sub-bands with each of 15 kHz or 48 sub-bands with each of 3.75 kHz. For the data transmissions, the total data rate can be up to 250 Kbit/s in the uplink and 170 Kbit/s in the downlink, respectively. Moreover, the lower bandwidth also contributes to the simplification of the radio part in end-user devices, which consequently leads to lower UEs costs and enables truly massive deployments of connected devices.

B. Modeling of NB-IoT

Although theoretical analysis of NB-IoT is a relatively new area, there is a growing interest and new types of analysis are continuously developed. Lee *et al*. [9] design a novel mechanism to improve energy consumption based on the predictive resource allocation. Based on the prediction of resource consumption in different uplink sessions, it simplifies the resource allocation procedures of BS by pre-assigning radio resources. The transmission time is thus reduced and the energy is saved. It is orthogonal to our work because we focus on modeling the uplink transmission performance at the UE side.

Later, Yu *et al*. [4] investigate an uplink link adaptation scheme to adjust the modulation and coding scheme (MCS) and repetition number of data transmissions. Unfortunately, it only considers the impact of data transmissions on the network performance. In contrast, we also analyze the performance of random access procedure, which is indispensable for transmitting the data packet. Moreover, note that Petrov *et al*. [5] develop an analytical model for the device-centric performance indicators, such as the message loss probability, latency, and energy-efficiency. However, a key enabler of this model is the assumption of error-free channel conditions, which is rather impractical in the realistic scenarios. Differently, our analytical approach is applicable to the performance computation under any channel condition.

III. Overview

In this section, we formally give the problem that we are addressing and present our approach towards solving it.

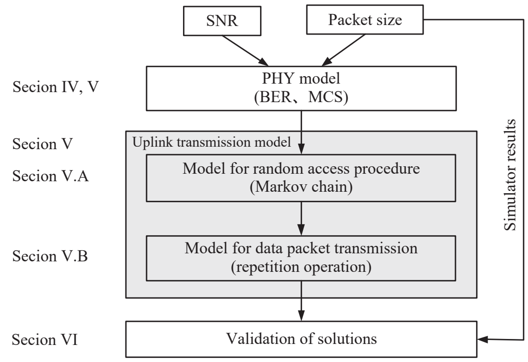

Fig. 1. Our approach overview for the performance analysis of uplink transmission in NB-IoT.

A. Problem Formulation

We are interested in characterizing the key performance metrics of real-world applications [2, 10]: uplink transmission latency L, and uplink transmission reliability R in NB-IoT. More formally, in a network with n NB-IoT UEs, when a UE has S_d bytes packet payload to be delivered, we need to estimate the probability that the UE delivers this packet to BS successfully, and the time it takes. Noticing that data and the control signal transmission is vulnerable to channel fading and wireless interference [11, 12], we also consider the channel conditions and link qualities, which can be represented by the Signal-to-Noise Ratio (SNR).

In NB-IoT systems, the uplink transmission is composed of registration/attachment, cell search, system synchronization, random access, data transmission and etc [1]. Among these procedures, some are indispensable for the real-world applications (e.g., random access and data transmission), some are conditionally executed. For example, a UE conducts cell search only when it is powered on for the first time. Therefore, our goal is to provide an efficient and accurate model to compute the performance of NB-IoT uplink transmission, involving the random access and data transmission procedures.

B. Approach Overview

In this paper, we provide a fine-grained analysis of the key performance measures of NB-IoT uplink transmissions in an error-prone channel condition. A high-level block diagram of our approach is shown in Figure 1, with pointers to sections where different parts are described in this paper. The centerpiece is an uplink transmission model for NB-IoT. The uplink transmission model is divided into two distinct parts. First, we study the random access behaviors of a single UE with a Markov model. We obtain the stationary probability τ that the UE transmits a preamble in a randomly chosen slot time. Then, by studying the events that can occur within a slot time, we express the latency and the reliability of random access procedure as a function of the computed value τ. In addition, we study the new features of packet transmission in NB-IoT, i.e., each packet can be continuously transmitted for several times. One original packet will be retransmitted only when all of its repetitions are not delivered successfully. Based on probability theory, we express the performance metrics of packet transmission as a function of the repetition number. Essentially, the uplink transmission behaviors depend on the physical-level model of the UE.

The physical-level model captures the bit error rate (BER)

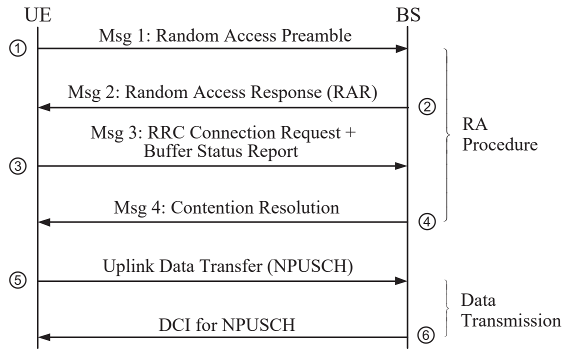

Fig. 2. Uplink transmission procedure in NB-IoT, including the random access procedure and the data packet transmission procedure.

and the MCS selection in the uplink transmission. More specifically, given the packet payload length S_d that need to be transmitted, we select suitable MCS and the optimal number of resource units (RUs) to minimize the packet on-air time. Then we can estimate BER based on the current link qualities (i.e., SNR). Note that this model does not require the same packet length S_d and SNR for all UEs in the network, since each UE will occupy a proper number of uplink resources after a successful random access. Conversely, our approach can be applied to model the uplink transmission performance of a UE with any size of packet and SNR.

We validate the entire approach by comparing the performance metrics estimated via this modeling approach with those produced by the simulator.

IV. Preliminaries of Uplink Transmission Behaviors

In this section, we provide some preliminaries on the behaviors of random access and packet transmission in NB-IoT, which are different from the traditional LTE.

A. Random Access

In NB-IoT, random access (RA) serves multiple purposes such as the initial access when establishing a radio link and requesting the radio resource for uplink transmission. Further, the random access procedure is always contention-based and starts with the transmission of a preamble. As shown in Figure 2, the contention-based random access procedure in NB-IoT consists of four steps:

(1) Random Access Preamble: At start, UE transmits a random access preamble (Msg 1) on the Narrowband Physical Random Access Channel (NPRACH). One NPRACH preamble is composed of four symbol groups, and each symbol group is composed of one cyclic prefix (CP) and five symbols. Unlike LTE, the value of each symbol is fixed to 1, and the preamble transmission is based on SC-FDMA (single-carrier frequency-division multiple-access) with the subcarrier spacing of 3.75 kHz. Upon transmission of the preamble, the UE first calculates its RA-RNTI (random access radio network temporary identifier) from the transmission time.

(2) Random Access Response: After BS detects the preamble transmission from a UE, it transmits a random access response (RAR or Msg 2) corresponding to the UE's RA-RNTI. With the RAR, the UE gets the timing advance command and the uplink grant for Msg 3 transmission. Note that the RAR transmission is based on OFDMA (orthogonal frequency-division multiple-access) with the same 15 kHz subcarrier spacing as LTE. If the preamble transmission was not successful (i.e., the associated RAR was not received), the UE transmits another one.

(3) Random Access Message: It is possible that multiple UEs send the random access preamble with the same uplink resources. An access collision occurs if multiple UEs transmit Msg 1 in the same random access slot and frequency band. Thus, after the associated response from BS, a scheduled message, RRC (Radio Resource Control) connection request (Msg 3) is transmitted in order to start the contention resolution process. Msg 3 contains the UE specific identifier (C-RNTI).

(4) Random Access Contention Resolution: To resolve the possible contention, BS sends back the C-RNTI of a UE (Msg 4). The UE, on seeing its own C-RNTI echoed back, concludes that the random access procedure is successful and proceeds to the data packet transmission. Otherwise, the contention resolution is considered to be failed, and the UE will try to access the channel again by waiting for a random time, i.e., performing the random backoff procedure. Note that multiple backoff procedures may be needed until the UE successfully accesses the channel.

Table 1: Uplink transportation block size table in Rel. 14 [13]

I_{TBS}	I_{RU}							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	16	32	56	88	120	152	208	256
1	24	56	88	144	176	208	256	344
2	32	72	144	176	208	256	328	424
3	40	104	176	208	256	328	440	568
4	56	120	208	256	328	408	552	680
5	72	144	224	328	424	504	680	872
6	88	176	256	392	504	600	808	1032
7	104	224	328	472	584	680	968	1224
8	120	256	392	536	680	808	1096	1352
9	136	296	456	616	776	936	1256	1544
10	144	328	504	680	872	1032	1384	1736
11	176	376	584	776	1000	1192	1608	2024
12	208	440	680	904	1128	1352	1800	2280
13	224	488	744	1032	1256	1544	2024	2536

B. Data Packet Transmission

As shown in step (3) of Figure 2, the UE sends the buffer status report through Narrowband Physical Uplink Shared Channel (NPUSCH). From the uplink transportation block size Table 1, we can find that the maximum block size is limited to 2536 bits. I_{TBS} and I_{RU} denote the index of transportation block size (TBS) and resource unit respectively. If the arrived data size is larger than the maximum block size, BS will split the whole data into sub-data to ensure the completeness of the data and follow the definition of 3GPP Rel. 14 [13]. Then, BS sends the scheduled NPUSCH resources along with the contention resolution message in Msg 4 to the UE (as shown in step (4)). The scheduled NPUSCH resources include many parameters that ensure the quality of uplink transmissions, such as MCS and the subcarrier assignment.

If the uplink data is successfully received at BS, BS will send back a DCI (Downlink Control Indicator) message with the toggled flag of NDI (New Data Indicator) (as shown in step (6)). Otherwise, the NDI flag is not toggled in the DCI message, and UE will retransmit the data packet.

V. Performance Analysis

In this section, we present the details of our analytical approach. First, we use a two-dimensional Markov chain to model the random access behaviors of a single UE. For the data packet transmission, we then study the new features of NB-IoT, such as MCS selection and data segmentation. Finally, we obtain the overall uplink transmission performance.

A. Random Access Procedure

For simplicity, we assume a timely and reliable feedback response (i.e., Msg 2 and Msg 4) transmission from BS. The reason is that uplink transmission for NB-IoT systems is considerably more complicated than the downlink transmission [4]. In the following, we will analyze the random access performance in detail, including the Msg 1 transmission and the Msg 3 transmission. To save space, a few explicit expressions in this paper are omitted, which can be found in our technical report [14].

1) Msg 1 Transmission: In NB-IoT systems, Msg 1 is modulated by BPSK. Given the current link quality (captured by SNR) between the UE and BS, we can estimate the BER of the transmitted Msg 1 [15], B_{msg1}. By assuming the independent bit error model, the probability P_{msg1} for Msg 1 to be transmitted successfully can be computed as

$$P_{msg1} = (1 - B_{msg1})^{S_{msg1}} \tag{1}$$

where S_{msg1} is the length of Msg 1.

Accordingly, the average elapsed time L_{msg1} for the UE to deliver a Msg 1 to BS can be given by

$$L_{msg1} = ETX_{msg1} \cdot (T_{pr} + T_{int} + T_{sl}) + (T_{pr} + T_{int}) \tag{2}$$

where T_{pr} denotes the time for transmitting a Msg 1, T_{int} denotes the interval time for the start of Msg 2 monitoring, and T_{sl} denotes the time window for monitoring Msg 2. The expected number of failed transmissions ETX_{msg1} depends on P_{msg1} and the Msg 1 retransmission threshold N_p [14].

2) Msg 3 Transmission: Upon receiving the random access response (Msg 2), UE sends the Msg 3 to BS to finalize the random access procedure. In the following time of T_{res}, UE consistently monitors the Random Access Contention Resolution (Msg 4) from BS. When T_{res} times out, UE employs *the binary exponential backoff rules* and tries to access the channel again after some backoff time chosen randomly within a backoff window. It is possible for the UE to conduct multiple consecutive backoff processes, bounded by the random access attempt threshold N. Hence, we need to determine the average latency L_{ra} of random access from the moment in which the backoff procedure is initiated until the UE successfully access the channel. In this subsection, we use a discrete Markov chain model to analyze the performance of the whole random access procedure.

Formally, the state of each UE is represented by the two-dimensional Markov chain, which is depicted in Figure 3. A current state (i, k) of a UE is determined by the current value of its backoff timer $k \in (0, W_i - 1)$ after it suffered i previous unsuccessful random access attempts (i.e., row i in Figure 3). Considering the low device complexity and the possible long period of uplink transmissions, NB-IoT introduces an uplink transmission gap for the start of random access procedure [16]. To this end, starting with the very first random access attempt (backoff stage $i = 0$), the initial value of the backoff timer (i.e., the gap length) is uniformly chosen in the range between 0 and $W_0 - 1$ (slots). After the UE enters backoff stage i, its backoff timer is reinitialized to a random value between 0 and $W_i - 1$ (slots). After N failed random access attempts, the UE will stop accessing the channel and report a failure to RRC. Moreover, until the m-th random access attempt, the maximal backoff timer W_i increases by a factor of 2, after which it is frozen to W_m until the N-th random access attempt, i.e.,

$$W_i = \begin{cases} W \cdot 2^i, & 0 \le i \le m \\ W \cdot 2^m, & m+1 \le i \le N \end{cases} \tag{3}$$

where W is the initial contention window.

It is paramount to note that, due to the long distance to the BS (e.g., $>$ 10 km), it is not feasible for the UEs to perform the carrier sensing before an uplink transmission as legacy radio access technologies (e.g., WLAN). Thus, the backoff timer is decremented by 1 at the beginning of each slot, regardless of the channel conditions. When the backoff timer reaches zero, the UE makes a random access attempt (i.e., sends Msg 1). Here, the slot time size σ is set equal to the duration of NB-slot in the uplink.

We denote the transition probability from one stage to another one (e.g., from row $i-1$ to row i in Figure 3) by p. It is also the probability of an unsuccessful random access attempt experienced by a UE. We assume that the value of p is constant and independent of the number of unsuccessful attempts already suffered. In NB-IoT system, an unsuccessful random access attempt can happen due to: collision of this UE with at least one of the $n-1$ remaining UEs, occurring with probability p_c [17], and/or an errored Msg 3, occurring with probability P_e (due to the channel fading, noise, etc). Specifically, the collision probability p_c for a UE is given by [17]

$$p_c = 1 - e^{-(\alpha \cdot \tau \cdot n)/\beta} \tag{4}$$

where α is the number of slots contained in one second, β is the total number of random access opportunities per second, τ is the probability that a UE transmits a Msg 1 in a generic slot time, and n is the number of UEs that attempt to access the channel. Then $\alpha \cdot \tau \cdot n$ represents the overall random access intensity.

Further, the error probability of Msg 3 is

$$P_e = 1 - (1 - B_{msg3})^{S_{msg3}} \tag{5}$$

where B_{msg3} is the BER of Msg 3 (note that Msg 3 is modulated by QPSK), and S_{msg3} is the Msg 3 length.

Since both events are independent, the probability p for a failed random access attempt can be expressed as

$$p = 1 - (1 - p_c)(1 - P_e) = p_c + P_e - p_c P_e \tag{6}$$

As shown in Figure 3, in case of an unsuccessful random access attempt, after backoff timer expiry in state $(i-1, 0)$, the UE moves to any state (i, k) on row i with probability p/W_i. Following a successful random access attempt (occurring with probability $1-p$) when the UE is in stage $i \in (0, N)$, UE returns to backoff stage 0 for the next uplink transmission scheduling, and its backoff timer uniformly selects any integer value in the range $(0, W_0 - 1)$ with probability $(1-p)/W_0$. If the UE reaches backoff stage N, and once its backoff timer reaches 0, the random access attempt can be successful or unsuccessful. In both cases, the random access procedure will be ceased. Meanwhile, UE will return to backoff stage 0 for a

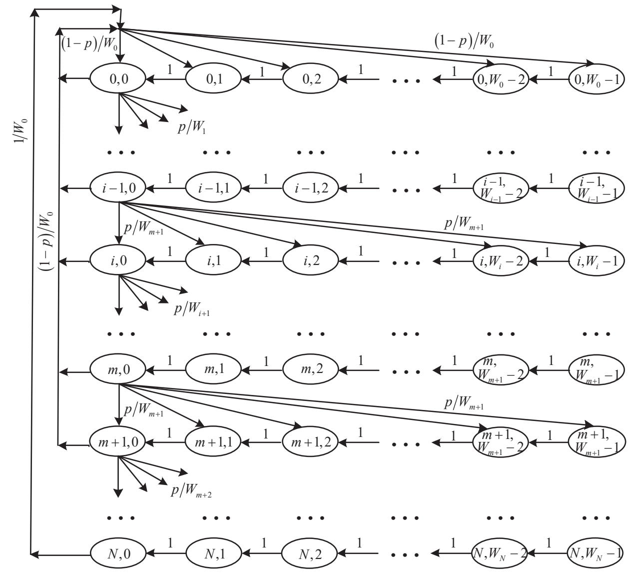

Fig. 3. Markov chain model for random access procedure in NB-IoT.

new scheduling request, and its backoff timer will be uniformly chosen in the range $(0, W_0 - 1)$ with probability $1/W_0$.

Let $b_{i,k}$ be the stationary distribution of this chain, which denotes the probability for a UE to be in state (i, k). The probability for UE to be in state $(i, 0)$ can be derived from the probability in state $(i-1, 0)$ as

$$b_{i,0} = b_{i-1,0} \cdot p = p^i \cdot b_{0,0}, \quad 0 < i \leq N \tag{7}$$

The probability $b_{i,k}$ for each $k \in (0, W_i - 1)$ can then be given simply as

$$b_{i,k} = \frac{W_i - k}{W_i} p \cdot b_{i-1,0}, \quad 0 < i \leq N, \; 0 \leq k \leq W_i - 1 \tag{8}$$

and $b_{0,k}$ is

$$\begin{aligned} b_{0,k} &= (1-p)\frac{W_0 - k}{W_0}\sum_{j=0}^{N-1} b_{j,0} + \frac{W_0 - k}{W_0} b_{N,0} \\ &= (1-p)\frac{W_0 - k}{W_0}\sum_{j=0}^{N-1} p^j b_{0,0} + \frac{W_0 - k}{W_0} p^N b_{0,0} \\ &= \frac{W_0 - k}{W_0} b_{0,0} \end{aligned} \tag{9}$$

From Eq. (8) and Eq. (9), $b_{i,k}$ can be rewritten as

$$b_{i,k} = \frac{W_i - k}{W_i} b_{i,0}, \quad 0 \leq i \leq N, \; 0 \leq k \leq W_i - 1 \tag{10}$$

Thus, by normalizing the stationary distribution of the Markov chain to 1, and using Eq. (3) and Eq. (10), we have

$$\begin{aligned} 1 &= \sum_{i=0}^{N}\sum_{k=0}^{W_i - 1} b_{i,k} = \sum_{i=0}^{N} b_{i,0}\sum_{k=0}^{W_i - 1}\frac{W_0 - k}{W_0} = \sum_{i=0}^{N} b_{i,0} \cdot \frac{W_i + 1}{2} \\ &= \frac{b_{0,0}}{2}\left[\sum_{i=0}^{m} p^i (W \cdot 2^i + 1) + \sum_{i=m+1}^{N} p^i (W \cdot 2^m + 1)\right] \end{aligned} \tag{11}$$

Finally, we attain the probability τ that a single UE transmits a random access preamble (i.e., Msg 1) in a randomly chosen slot as

$$\begin{aligned} \tau &= \sum_{i=0}^{N} b_{i,0} = \sum_{i=0}^{N} p^i \cdot b_{0,0} = \frac{1 - p^{N+1}}{1-p} \cdot b_{0,0} \\ &= \frac{2(1-2p)(1-p^{N+1})}{(1-2p)(1-p^{N+1}) + W[1 - p - p(2p)^m(1 + p^f - 2p^{f+1})]} \end{aligned} \tag{12}$$

Eq. (6) and Eq. (12) represent a nonlinear system with the two unknowns τ and p, which can be solved efficiently by *Mathematica*. It is obvious that there is a unique solution of τ for each n, W, m, N and P_e, i.e., $\tau = f(n, W, m, N, P_e)$.

3) Comprehensive Analysis: Now let us concentrate on a single UE to determine the average reliability L_{ra}, as well as the delay L_{ra} from the moment the backoff procedure is initiated until the final successful random access attempt.

Random Access Latency. In each backoff stage $i \in (0, N)$, the initial value of the backoff timer has a mean of $(W_i - 1)/2$, so the average time between two consecutive random access attempt transmissions of an observed UE is $(W_i - 1)/2$ slots. Hence, the average elapsed time $T_{tct,i}$ before the observed UE makes its $(i+1)$-th random access attempt (corresponding to row $i+1$ in Figure 3) is

$$T_{tct,i} = \sum_{k=0}^{i}\left(\frac{W_k - 1}{2}\right)\sigma + iT_{coe} = \frac{1}{2}\sum_{k=0}^{i}\sigma W_k - \frac{(i+1)\sigma}{2} + iT_{coe} \tag{13}$$

where T_{coe} is the average time duration of each unsuccessful random access attempt for the observed UE. We define T_c and T_e as the average times of a failed random access due to a collision and due to the transmission error of Msg 3 [14].

$$T_{coe} = \frac{p_c T_c + (1 - p_c) P_e T_e}{p} \tag{14}$$

We next substitute Eq. (3) into Eq. (13), which leads to

$$\begin{aligned} T_{tct,i} &= iT_{coe} - \frac{(i+1)\sigma}{2} \\ &+ \frac{\sigma W}{2} \cdot \begin{cases} 2^{i+1} - 1, & 0 \leq i \leq m \\ 2^{m+1} - 1 + 2^m (i - m), & m+1 \leq i \leq N \end{cases} \end{aligned} \tag{15}$$

Finally, the average latency L_{ra} until the final successful random access attempt is then

$$L_{ra} = \sum_{i=0}^{N}(1-p) \cdot p^i \cdot (T_{tct,i} + T_s) \tag{16}$$

where $T_s = L_{msg1} + T_{rrc}$ is the average time spent for the final successful random access attempt (T_{rrc} is the time to send a Msg 3). By integrating Eq. (15) and Eq. (16), we can solve L_{ra} in closed-form.

Random Access Reliability. The reliability of random access is actually the probability of occurrence of at least one successful random access attempt. In order to access the channel successfully and acquire transmission resources, a UE not only needs to deliver Msg 1 without error and collision (with probability $1 - P_{ftx}^{N_p+1}$), but also deliver Msg 3 without error (with probability $1-p$). For the random access reliability R_{ra}, we have

$$R_{ra} = 1 - \left[1 - (1 - P_{ftx}^{N_p+1})(1-p)\right]^{N+1} \tag{17}$$

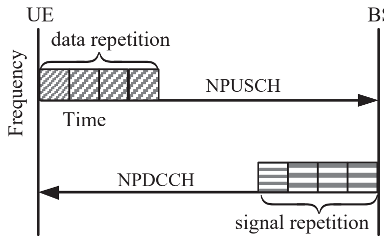

Fig. 4. Illustration of repetition during one transmission in NB-IoT. In detail, a block is repeated four times.

B. Packet Transmission Procedure

When the size of data packet is larger than the maximum TBS shown in Table 1, the UE segments the whole data and put each segmented data into a block accordingly. Note that the meta transmission unit in NB-IoT systems contains only one block, which is different from that in LTE [13]. There are two steps for UEs to encode the block before feeding into the PHY-layer modulation component (e.g, BPSK or QPSK). 1) Appending 24 bits CRC check to the end of the block. 2) Encoding the block with κ (e.g., 1/3 in NB-IoT) rate of turbo code. Formally, given S_d as the length of initial data packet payload to be transmitted, we can estimate the total size of the coded block $n_c = \kappa(S_d/n_{bk} + a + b)$. Here, a is the number of bits for appending CRC (i.e., a=24), b is the number of bits for turbo code feedback (i.e., b=4) [13], and n_{bk} is the number of segmented blocks. To determine n_{bk}, the following problem is solved:

$$\begin{aligned} &\arg\min_{n_{bk}}[TBS(i,j) - n_c] \\ &s.t., TBS(i,j) - n_c \geq 0 \end{aligned} \tag{18}$$

where $TBS(i,j)$ denotes the TBS selected from Table 1, i denotes the index I_{TBS} , and j denotes the index I_{RU}. In turbo codes, a larger code block size contributes to a lower block error rate (BLER) [18]. To reduce the block error rate and retransmission number, the UE first segment the whole data into as few code blocks as possible. Moreover, the UE will select the smallest number of resource frame j_m from the above resulted TBS set to minimize the packet transmission time. The smallest i_m is chosen to minimize the transmission power, and then the size of each block $x_{bk} = TBS(i_m, j_m)$. Thus, the transmission time of each block can be produced as

$$T_{bk} = j \cdot T_{RU} \tag{19}$$

It is known that repeating transmission data and the associated control signaling several times has been utilized as a promising approach to achieve coverage enhancement for NB-IoT [19], since more repetition number will enhance the reliability. In particular, repetition for NB-IoT can only be selected among {1, 2, 4, 8, 16, 32, 64, 128}, which means the allowable repetition number of the same transmission block. Figure 4 shows a simple illustration of repetition in NB-IoT, where both NPDCCH (Narrowband Physical Downlink Shared Channel) and NPUSCH transmission blocks with the same content are repeated four times during one transmission. In order to derive the total latency of data transfer, we need to compute the time taken for each block.

Given the SNR value, the coding rate κ, and the block size x_{bk}, we can estimate BLER of each transmission $F(snr, \kappa, x_{bk})$ based on the theory proposed in [18, 20].

Packet Transmission latency. Under the paradigm of repeating transmission (denoting repetition number by r), we determine the latency L_{bk} for a successful block transmission as

$$L_{bk} = ETX_{bk} \cdot (rT_{bk} + T_{out}) + rT_{bk} \tag{20}$$

where T_{bk} is the time to send a block (given in Eq.(19)), and T_{out} is the timer length for monitoring the response (i.e., DCI) from BS. ETX_{bk} is the expected number of failed transmissions per block, which is a function of the retransmission threshold N_d of each block and the failed transmission probability P_{fbk} for a block with r repetitions ($P_{fbk} = (F(snr, \kappa, x_{bk}))^r$) [14].

Following the sequential transfer pattern [19], a new block can be sent only when the previous block is transmitted successfully. As a result, the total elapsed time L_d for n_{bk} blocks is

$$L_d = L_{bk} \cdot n_{bk} \tag{21}$$

Packet Transmission Reliability. The reliability of data packet transmission is the probability for the transmitted packets to be received by BS correctly. That is, all the segmented blocks of the packet require to be delivered. The data packet transmission reliability R_d is then

$$R_d = (1 - P_{fbk}^{N_d+1})^{n_{bk}} \tag{22}$$

C. Overall Performance Measures

For each UE in real-world applications, the major concerns are typically the overall performance from its first channel access attempt to the final packet reception at BS.

Uplink Transmission Latency. In this regard, we can easily obtain the aggregate latency L as follows.

$$L = L_{ra} + L_d \tag{23}$$

Here, the expected random access latency L_{ra} and the expected data transmission latency L_d can be calculated according to Eq. (16) and Eq. (21), respectively.

Uplink Transmission Reliability. The uplink transmission reliability R is the expected fraction of packets delivered from a UE to BS. Thus, R is the product of random access reliability and packet transmission reliability, i.e.,

$$R = R_{ra} \cdot R_d \tag{24}$$

Here, random access reliability R_{ra} and packet transmission reliability R_d can be calculated by Eq. (17) and Eq. (22), respectively.

VI. Numerical Results and Simulations

In this section, we perform extensive simulations to validate our analytical results. We first introduce the simulation methodology. Then we present the numerical results.

A. Methodology

To comprehensively validate our analytical approach, we compare the characteristics obtained from the proposed analytical approach with those produced by system-wide simulations. The said verification is conducted by employing our custom-made system-level simulator. Specifically, our simulator is an event-driven simulation program, developed in MATLAB, that follows the NB-IoT protocol flow and emulates the transmission behaviors of UEs and BS based on the recent 3GPP specifications [19]. In particular, to simulate wireless

Table 2: Relevant system parameters used to obtain numerical results

Parameter	Value
Msg 1 length S_{msg1}	20 symbols
Msg 3 length S_{msg3}	11 bytes
Initial data packet payload size S_d	[50, 200] bytes
Slot time σ	2 ms
Number of slots per second α	500
RAOs per second β	4.8×10^3
Time to transmit a Msg 1 T_{pr}	6.4 ms
Time interval T_{int} before monitoring Msg 2	12 ms
Time window T_{sl} for Msg 2 listening	10 ms
Time to transmit a Msg 3 T_{rrc}	32 ms
Time window T_{res} for Msg 4 listening	10 ms
Time window T_{out} for DCI listening	50 ms

transmissions in real-world scenarios with obstacles, we also incorporate an existing path loss model [21] into the simulator. It is a statistical path loss model derived from the experimental data collected across the United States in 95 macrocells. The characterization used is a linear curve fitting the path loss to the decibel-distance, with a Gaussian random variation about that curve due to shadow fading. Generally, simulators allow us to tune many different parameters and provide a fairly good resemblance of the real environment. Compared to detailed simulation, theoretical analysis provides an alternative method of testing and designing systems with lower cost.

The fixed parameters used throughout this section are summarized in Table 2. These parameters (except the data packet length S_d) are deterministic and defined by NB-IoT standardization [19]. To evaluate the uplink transmission capability of NB-IoT, we vary the initial packet payload length on a UE from 50 to 200 bytes. Unless otherwise specified, we obtain the numerical results with $n = 500$ UEs, the initial contention window $W = 8$, the random access attempt threshold $m = 8$ (after which the backoff window is frozen), the preamble retransmission threshold $N_p = 16$, the maximum number of backoff stages $N = 16$, the data packet retransmission threshold $N_d = 16$, and the repetition number $r = 4$.

B. Model Validation

We first illustrate the performance measures of our analytical approach. To test the impact of link qualities on uplink transmission performance, we also consider the uplink transmission in various cases of SNR, from -4 to 1 dB. Figure 5 reports the error rate of our analytical results relative to the simulation results. We see that our analytical approach can achieve high calculation accuracies for both latency and reliability, regardless of the packet size and SNR value. Specially, with our analytical approach, more than 90% of transmissions have a relative error of latency computation lower than 17% (i.e., 83% accuracy), and more than 90% of transmissions have a relative error of reliability computation lower than 4% (i.e., 96% accuracy). Moreover, Figure 6 shows the fine-grained numerical results for uplink transmissions with different SNR and packet sizes. As expected, as the SNR value increases, the overall latency decreases, and the overall reliability increases. In particular, we can nearly achieve 95% transmission reliability with a latency less than 2s when SNR is more than 0 dB.

Interestingly, we also observe that the transmission latency of packet with 100 or 150 bytes is relatively smaller than that

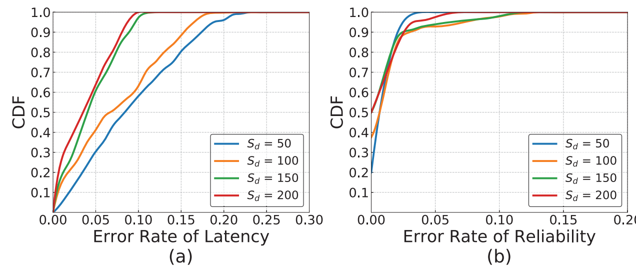

Fig. 5. Error rate of the analytical results relative to the simulation results: (a) latency error rate; (b) reliability error rate.

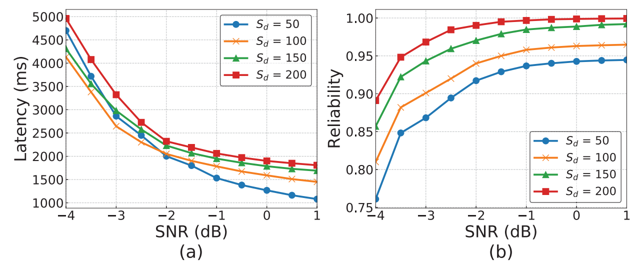

Fig. 6. Overall uplink transmission performance under different link qualities and data packet sizes: (a) transmission latency; (b) transmission reliability.

of 50 bytes, when the link quality is bad. It is because that with a larger packet size, the code block size is larger and the turbo coding achieves a smaller BLER. As a result, the larger packet requires a fewer number of retransmissions, resulting in a lower latency. However, as packet size further becomes larger, the average latency will increase due to transmitting the code block itself is the major time consumer, which is shown in Figure 6(a).

C. Impact of Various Parameters

We now turn to evaluate the key performance metrics of NB-IoT uplink transmission under various settings, which can be used to aid protocol design for performance optimization in the future.

To investigate the dependency of the latency and transmission reliability, on the maximum number of random access attempts (i.e, the allowable backoff stages) N, we have reported the performance metrics of 50 byte packet payload versus different values of N. Figure 7 shows four cases of link qualities, i.e., SNR = -5, -3, -1, and 1 dB. It is reasonable to expect that the average transmission latency and reliability will increase as we allow for a larger number of random access attempts. We can see that the dependence of uplink transmission performance on the threshold N is marginal when the link quality becomes relatively better. Accordingly, we can decrease the random access attempt threshold to support a higher number of connections when the link quality is good.

In addition, Figure 8 displays the transmission performance under different repetition numbers r for the same block. The results show that the latency of uplink transmission highly depends on repetition number r, and the optimal value of r depends on the SNR (i.e., the current link quality). For example, an high value of R (e.g. 32) gives better latency performance in the case of SNR = -5 dB, while it drastically penalizes the latency in the case of better link quality (e.g., SNR = 1 dB). Therefore, a small repetition number is preferable for the good

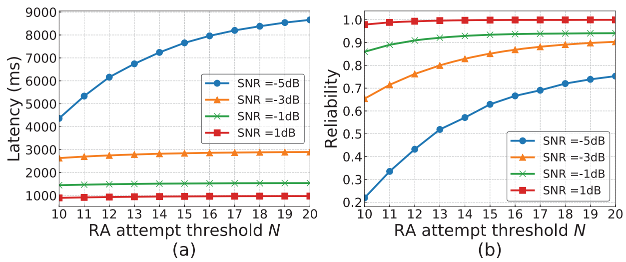

Fig. 7. Influence of random access attempt threshold N over uplink transmission performance: (a) transmission latency; (b) transmission reliability.

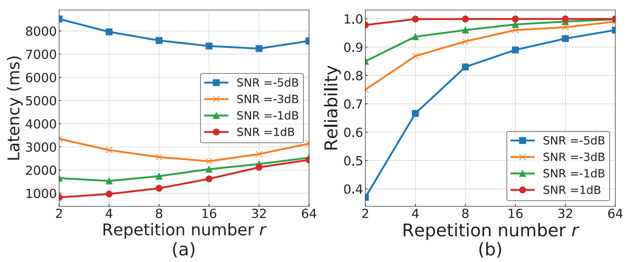

Fig. 8. Influence of repetition number r over uplink transmission performance: (a) transmission latency; (b) transmission reliability.

link qualities.

Finally, let us add some considerations regarding the impact of UE population sizes on the transmission performance. Figure 9 illustrates the latency and reliability in the system with different number of UEs. It is often stated that NB-IoT supports a massive number of IoT devices, e.g,. 50K UEs/cell. However, the results in Figure 9 show that the increase of UE numbers may cause a severe degradation in the uplink transmission performance, especially when the link quality is bad. In fact, for the NB-IoT systems, a huge amount of connections is achieved only when most of the connected devices are in idle mode and do not occupy channel resources. This also implies a need for more efficient schemes to enhance the performance of NB-IoT system with many active UEs.

VII. Conclusion

In this paper, we show a way of mathematically analyzing the performance of uplink transmission for NB-IoT systems, concerning the transmission latency and transmission reliability. Our model is accurate with consideration of the protocol details and the new features of NB-IoT, including link quality, packet size, channel access contention, and etc. We validate the analytical results through extensive simulations. Results show that our analytical model achieves 83% latency calculation accuracy, and 96% reliability calculation accuracy. Moreover, we demonstrate that the analytical results can be used to aid protocol design for performance optimization, e.g., repetition number tuning for reducing the transmission latency.

Acknowledgement

This work is supported by the National Science Foundation of China (No. 61772465, No. 61872437, and No. 61472360), Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Zhejiang Provincial Key Research and Development Program (No. 2017C02044), and the Fundamental Research Funds for the Central Universities (No. 2017FZA5013 and No. 2018FZA5013). Wei Dong is the corresponding author.

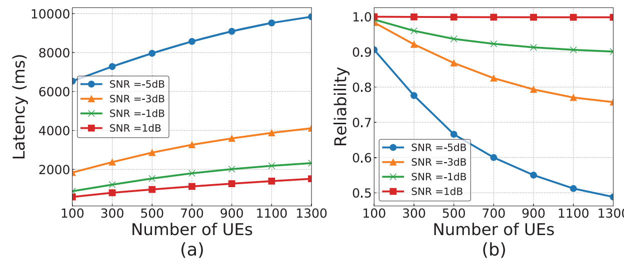

Fig. 9. Influence of the total number of UEs in the system over uplink transmission performance: (a) transmission latency; (b) transmission reliability.

References

[1] 3GPP, "Standardization of NB-IoT completed," June 2016. [Online]. Available: http://www.3gpp.org/snews-events/3gpp-news/1785-nbiotcomplete

[2] M. Ceriotti et al, "Monitoring heritage buildings with wireless sensor networks: the torre aquila deployment," in *Proc. of ACM/IEEE IPSN*, 2009.

[3] M. Pennacchioni, M. G. Di Benedetto, T. Pecorella, C. Carlini, and P. Obino, "NB-IoT system deployment for smart metering: evaluation of coverage and capacity performances," in *Proc. of IEEE AEIT*, 2017.

[4] C. Yu, L. Yu, Y. Wu, Y. He, and Q. Lu, "Uplink scheduling and link adaptation for narrowband internet of things systems," *IEEE Access*, vol. 5, pp. 1724–1734, 2017.

[5] V. Petrov et al, "Vehicle-based relay assistance for opportunistic crowdsensing over narrowband iot (nb-iot)," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1–13, 2017.

[6] R. Sanchez-Iborra and M. D. Cano, "State of the art in LP-WAN solutions for industrial IoT services," *Sensors*, vol. 16, no. 5, pp. 1–12, 2016.

[7] 3GPP, "Cellular system support for ultra-low complexity and low throughput internet of things (CIoT)," in *3GPP TR 45.820 v13.1.0*, December 2015.

[8] A. R. Alvario and M. Vajapeyam, "An overview 3GPP enhancements on machine to manchine communications," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 14–21, 2016.

[9] J. Lee, "Prediction-based energy saving mechanism in 3GPP NB-IoT networks," *Sensors*, vol. 17, no. 9, pp. 1–22, 2017.

[10] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in *Proc. of ACM SenSys*, 2004.

[11] C. Liang, N. Priyantha, J. Liu, and A. Terzis, "Surviving WiFi interference in low power ZigBee networks," in *Proc. of ACM SenSys*, 2010.

[12] R. Musaloiu-E and A. Terzis, "Minimizing the effect of WiFi interference in 802.15.4 wireless sensor networks," *Int. J. Sensor Networks*, vol. 3, no. 1, pp. 43–54, 2008.

[13] 3GPP, "Multiplexing and channel coding (release 14)," in *3GPP TS 36.212 V14.3.0*, June 2017.

[14] "Accurate performance modeling of uplink transmission in nb-iot [technical report]," 2018. [Online]. Available: https://www.dropbox.com/s/1nz4flznhr2dfvn/nbmodeling-techrep.pdf

[15] S. Thoen et al, "Performance analysis of combined transmit-sc/receive-mrc," *IEEE Trans. on Communications*, vol. 49, no. 1, pp. 5–8, 2001.

[16] 3GPP TS 36.211, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation," 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/36_series/36.211/36211-d20.zip

[17] 3GPP, "RAN improvements for machine-type communications,," in *3GPP TR 37.868 v1.0.0*, December 2011.

[18] C. E. Shannon, "Probability of error for optimal codes in a gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.

[19] 3GPP TSG RAN Meeting #69, "Narrowband IoT (NB-IoT)," Sept. 2015. [Online]. Available: http://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_69/Docs/RP-151621.zip

[20] S. D. Lembo *et al.*, "Modeling bler performance of punctured turbo codes," 2011.

[21] V. Erceg and L. J. Greenstein et al, "An empirically based path loss model for wireless channels in suburban environments," *IEEE JSAC*, vol. 17, no. 7, pp. 1205–1211, 1999.