2023 International Conference on Future Communications and Networks (FCN) | 979-8-3503-9603-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/FCN60432.2023.10543535

Closed-Loop Modeling and Performance Analysis of B5G RRC with Processor Sharing Discipline

Hongzhi Guo[1,2], Wen Zhan[1] and Xiang Chen[2,3]
[1]School of Electronics and Communication Engineering, Sun Yat-sen Univ., Shenzhen 518107, China
[2]Research Institute of Tsinghua University in Shenzhen (RITS), Shenzhen 518075, China
[3]School of Electronics and Information Technology, Sun Yat-sen Univ., Guangzhou 510006, China
{guohzh9}@mail2.sysu.edu.cn, {zhanw6, chenxiang}@mail.sysu.edu.cn

***Abstract*—Massive devices are expected to access Beyond fifth-generation (B5G) networks, establish connections and then request data transmission resources. Both the access efficiency and the data transmission efficiency should be analyzed and optimized. Yet, existing models considered the access process and the data transmission process as two independent processes that do not affect each other, which does not hold in practical networks. To address this open issue, this paper proposes an analytical model for a closed-loop B5G system by establishing a Markov renewal process for characterizing the Radio Resource Control (RRC) state transition behavior from the device's point of view. The processor sharing queueing model is embedded to approximate the variation of data transmission performance regarding the number of connected devices, based on which the mutual effect between the two processes can be further revealed. Our analysis derives key performance metrics (such as access throughput and energy efficiency) as functions of the system parameters (such as traffic rate, access classing barring factor and inactivity timer) in both processes of the B5G system, and verified via extensive simulations.**

***Index Terms*—Modeling, B5G RRC, performance analysis, random access, data transmission, processor sharing.**

I. Introduction

The forthcoming Beyond fifth-generation (B5G) networks are growing to support various applications related to multiple vertical industries such as e-health, intelligent transportation, manufacturing and agriculture [1]. Massive devices will access the gNB, establish connections and then request data transmission resources. How to characterize the access efficiency and the data transmission efficiency of B5G network is of practical significance.

Over 40 years, the cellular system has evolved from 1G to 5G. Each generation of the cellular system has brought about radical innovations and new features that reformed physical layer design, such as code division multiple access in 3G and orthogonal frequency-division multiple access (OFDMA) in 4G [2]. In sharp contrast, basic principles of Radio Resource Control (RRC) layer design evolve slowly in relative. The connection-based random access scheme has been a long-standing principle since 2G, that is, a connection would first be established via a contention-based access procedure before contention-free data packet transmission [3]. From the perspective of the RRC layer, a state machine has been used to characterize the behavior of the device and identify the operation that the Radio Access Network (RAN) should take.

The RRC layer state machine gradually evolves with each generation of the cellular system, while two basic components remain, i.e., idle state and connected state (For instance, both 4G and 5G have RRC Connected state and RRC Idle state. [4]), which correspond to the two steps in the connection-based random access scheme. For the device in the idle state, it should successfully pass the random access procedure first and moves into the connected state, and then obtain granted uplink resources from base station for data transmission. Both the access efficiency and the data transmission efficiency are fundamentally important in cellular systems.

To evaluate and improve access efficiency, a wealth of work has been done. For instance, in [5], the throughput in the random access channel is analyzed and then maximized via the optimal selection of the channel access probability. In [6], the access delay performance of the connection-based slotted-Aloha is analyzed based on a vacation queueing model. Moreover, novel massive access schemes, such as successive interference cancellation based non-orthogonal random access scheme [7] and cluster-based congestion-mitigating access scheme [8], are proposed to mitigate the access collision and improve access capability. While the aforementioned developments have been substantial, the data transmission process has either been neglected [7] or simplified by assuming that the data buffer can be cleared in a single time slot [5],[8], or that the transmission time for each packet is fixed [6].

On the other hand, to enhance data transmission efficiency, numerous time-frequency resource scheduling algorithms are proposed for B5G. In [9], deep deterministic policy gradient (DDPG) is applied for radio resource scheduling in the 5G RAN, considering multiple users, varied channel conditions, and random traffic arrivals. Packet loss rate and data throughput are jointly considered and improved. In [10], an uplink scheduler that incorporates Orthogonal Multiple Access (OMA) and Non-orthogonal Multiple Access (NOMA) is formulated for overloaded scenarios, where the sum rate of the network is maximized through a heuristic algorithm. Although significant improvement in data transmission performance can be achieved, access process is ignored [10] or simplified by assuming that the arrival of the user equipments (UEs) follows

The work is supported in part by the Shenzhen Natural Science Foundation under Grant JCYJ20200109143016563, in part by National Natural Science Foundation of China under Grant 62001524 and in part by Shenzhen Science and Technology Program (Grant No. RCBS20210706092408010).*(Corresponding author: Wen Zhan.)*

Poisson distribution [9].

In short, existing literature considers the access process and data transmission process as two independent processes that do not affect each other. This starting point simplifies the analysis, yet, does not hold in practical networks due to the fundamental limitation in time-frequency resources, where both processes share. Considering the internal relation between these two processes from the RRC layer, for instance, greater access efficiency causes more devices to swarm into the connected state, reducing the average data transmission rate that each device may share. On the other hand, with an improved data transmission efficiency, more devices shift back into the idle state, which may exacerbate the contention in the random access channel. It can be seen that for the performance analysis of cellular systems, the access process and data transmission process should be carried out within a unified theoretical framework, based on which the close relationship between them can be explicitly characterized. Yet, to the best of our knowledge, how to formulate such a theoretical framework is still an open question.

To solve the above problem, this paper proposes an analytical model for B5G RRC protocol that integrates both the access process and the transmission process. Specifically, the Markov renewal process for each user equipment (UE) is established by further dividing each 3GPP-defined state into multiple sub-states, which allows granular characterization of the behavior of UEs in both RRC INACTIVE state and RRC CONNECTED state.

The proposed model focuses on the performance analysis of a closed-loop B5G system, where a fixed number of UEs cyclically move among B5G RRC states. Key input parameters of the RRC layer such as traffic rate, inactivity timer, Access Class Barring (ACB) factor, data transmission rate, et al, are taken into consideration. By virtue of the proposed model, the access throughput (i.e., the average number of successfully transmitted access requests per time slot) and energy efficiency (i.e., the ratio of the energy consumed for data transmission to the total energy consumption) can be calculated. Our analysis reveals key insights on the interplay between the access process and the data transmission process. For instance, increasing the inactivity timer would enlarge the sojourn time that UEs stay in the data transmission process, which further reduces the access throughput as fewer UEs join the random access channel contention. Moreover, to optimize energy efficiency, the inactivity timer should be carefully selected. In addition, the optimal inactivity timer crucially depends on the power consumption level in the access process.

The paper is organized as follows. In Section II, the system model is presented. In Section III, the analytical model is formulated and key performance metrics of the B5G network are obtained. In Section IV, the analysis is verified by the simulation results and technical insights are revealed. Finally, concluding remarks are summarized in Section V.

II. System Model

Consider a single-cell B5G network serving n UEs. It is assumed that each UE is either in RRC CONNECTED state or RRC INACTIVE state. RRC CONNECTED state includes Transmission state and Suspend state, while RRC INACTIVE state can be divided into Silence state and Access state. Fig.1 demonstrates the state transition process of each UE. We would further characterize the state transition process of each UE in the following Section II-A.

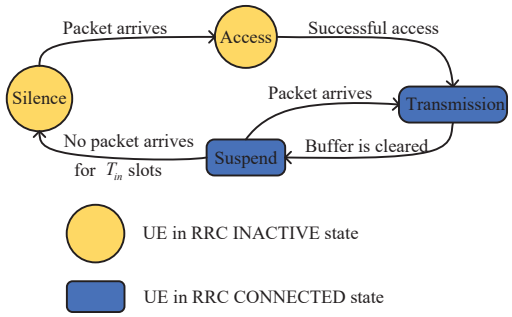


Fig. 1. State transition diagram of each UE

A. State Transition Process of Each UE

In the random access procedure, UE generates a random number between 0 and 1, and compares it with the ACB factor $q \in (0, 1]$. If the number is less than q, the UE proceeds to transmit the access request. Otherwise, it is barred temporarily. Then, the UE which passes ACB check randomly selects one out of $M \in \{1, 2, \dots\}$ orthogonal preambles and transmits via the Physical Random Access CHannel (PRACH) to the gNB[1]. If more than one UE transmits the same preamble simultaneously, then a collision occurs and all of them fail. The access request is successful if and only if there are no concurrent transmission of the same preamble.

Upon a successful request, UE shifts from RRC INACTIVE state to RRC CONNECTED state, more specifically, from Access state to Transmission state. In Transmission state, UE is busy with data transmission. When the buffer is cleared, it enters Suspend state. If no packet arrives for T_{in} consecutive time slots, the UE leaves Suspend state and turn back to the Silence state, where T_{in} is referred to as the inactivity timer [3]. Otherwise, the UE in Suspend state enters Transmission state again, starting the transmission by delivering a scheduling request to the gNB.

We assume that the arrival of data packets at each UE follows a Bernoulli process with parameter $\lambda \in (0, 1)$. While, no packet arrival when UE is in Transmission state. UE stays in Silence state if the buffer is empty. With the arrival of packet, UEs in Silence state would enter Access state, accessing the gNB for data transmission via contention-based access procedure.

B. Processor Sharing Queueing Model

Note that the time-frequency resources of the network are finite. It has long been observed that as the number of devices in the transmission period increases, each device may receive less resource blocks and suffers a diminishing data rate.

[1]For simplicity, we only consider the single-preamble case $M = 1$ and the extension to the multi-preamble case can be implemented based on the multi-group model in [5].

Denote n_t the number of UEs in the transmission period and μ_i the data rate of UE $i, i \in \{1, 2, \ldots, n_t\}$. It's clear that data rate μ_i of each UE i should be inversely proportional to n_t, for which we adopt the processor sharing queueing model for characterization. The processor sharing queueing model considers that in the queueing system, the service rate each customer receives is inversely proportional to the number of customers in service [13]. Service capacity of the gNB, denoted by $\hat{\mu}$, is equal to the sum of data rates of all n_t UEs, i.e., $\hat{\mu} = \sum_{i=1}^{n_t} \mu_i$. Assume the resource blocks are fairly shared among n_t UEs and the expected data rate of each UE is the same[2].

C. Performance Measures

In this paper, we aims at evaluating the throughput performance and the energy efficiency of the B5G network.

There are two kinds of throughput: data throughput and access throughput. Data throughput is defined as average number of successfully transmitted data packets per time slot. Access throughput, denoted by $\hat{\lambda}^a_{out}$, is defined as average number of successfully transmitted access requests per time slot, which evaluates the access efficiency of the network. Note that when service capacity is higher than the aggregate arrival rate, the data throughput is equal to the aggregate input data rate, because all the arrived packets can be transmitted in a long-term perspective [14]. Accordingly, the throughput we aim to analyse is access throughput $\hat{\lambda}^a_{out}$.

Each UE's access request queue can be modeled as the $Geo/G/1/1$ queue. Based on the $Geo/G/1/1$ model of each request queue, the access throughput can be obtained as

$$\hat{\lambda}^a_{out} = E[n_a]\lambda(1-\alpha), \tag{1}$$

where $E[n_a]$ is the average number of UEs in RRC INACTIVE state, which will be derived in Section III, and α denotes the probability that each request queue is nonempty, which is given by [14]

$$\alpha = \frac{\lambda}{\lambda + pq}, \tag{2}$$

where pq is the service rate of each access request queue, p is the successful transmission probability of access requests and $q \in (0, 1]$ is the access request transmission probability of each UE, which is also known as the ACB factor in the 3GPP standards. The analysis in [5] has also revealed that the probability of successful transmission of access requests is

$$p = \exp\left(\frac{-E[n_a]\lambda q}{\lambda + pq}\right). \tag{3}$$

To evaluate the energy efficiency, we let P_{Sil}, P_{Access}, P_{Trans} and P_{Sus} respectively denote the average power of Silence state, Access state, Transmission state and Suspend state. The power model of each sub-state is set according to 3GPP protocol [15]. Define the energy efficiency η as the ratio of the energy consumed in transmission period to the total energy consumption, which is given by

$$\eta = \frac{P_{Trans}\tilde{\pi}_{Trans}}{P_{Sil}\tilde{\pi}_{Sil} + P_{Access}\tilde{\pi}_{Access} + P_{Trans}\tilde{\pi}_{Trans} + P_{Sus}\tilde{\pi}_{Sus}}, \tag{4}$$

[2]The data rate of each device depends on various factors, such as the modulation and coding scheme, the channel fading and the signal power. How to extend the analysis to incorporate other factors, not just the resource scheduling, is an interesting issue that we will address in our future work.

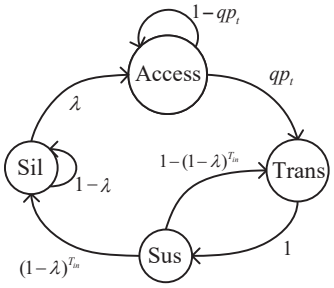


Fig. 2. Embedded Markov chain $\{X_j\}$ of the state transition process

where $\tilde{\pi}_{Sil}$, $\tilde{\pi}_{Access}$, $\tilde{\pi}_{Trans}$ and $\tilde{\pi}_{Sus}$ respectively denote the probability of the UE being in Silence state, Access state, Transmission state and Suspend state.

To derive the expression of access throughput and energy efficiency, we firstly derive the probability of the UE being in each state according to the discrete-time Markov renewal process, as shown in Section III.

III. Performance Analysis

A. State Characterization of UE

To model the behavior of each UE, a discrete-time Markov renewal process $(\boldsymbol{X}, \boldsymbol{Y}) = \{(X_j, V_j), j = 0, 1, \cdots\}$ is established according to the state transition process in Section II-A, where X_j denotes the state of a UE at the j-th transition and V_j denotes the epoch at which the j-th transition occurs. Fig. 2 shows the embedded Markov chain $\boldsymbol{X} = \{X_j\}$, which has four states: 1) Silence state (State Sil), 2) Access state (State Access), 3) Transmission state (State Trans), 4) Suspend state (State Sus). As shown in Fig. 2, it shifts from State Sil to State Access with packet arrival. Otherwise, it remains in State Sil. It leaves State Access for State Trans if it passes the ACB check and the access request is successfully transmitted. Otherwise, it stays in State Access. It shifts from State Trans to State Sus when the data transmission finishes. If no packet arrives for T_{in} consecutive time slots, the UE leaves State Sus and turn back to State Sil. Otherwise, the UE enters State Trans again. The state transition probabilities are given in Fig. 2. The steady-state probability distribution of the embedded Markov chain in Fig. 2 can be derived as

$$\begin{cases} \pi_{Sil} = \left(1 + \frac{2\lambda}{(1-\lambda)^{T_{in}}} + \frac{\lambda}{pq}\right)^{-1}, \\ \pi_{Sus} = \pi_{Trans} = \frac{\lambda}{(1-\lambda)^{T_{in}}}\pi_{Sil}, \\ \pi_{Access} = \frac{\lambda}{pq}\pi_{Sil}, \end{cases} \tag{5}$$

where $p = \lim_{t\to\infty} p_t$ is the steady-state probability of successful transmission of access requests and p_t is the probability of successful transmission of access requests at time slot $t, t = 1, 2, \cdots$.

The interval between successive transitions, i.e., $V_{j+1} - V_j$, is called the holding time in State $X_j, j = 1, 2, \cdots$. Let τ_i denote the mean holding time in State i, where $i \in \{\text{Sil}, \text{Access}, \text{Trans}, \text{Sus}\}$. The holding time in State Sil and that in State Access are both one time slot, i.e.,

$$\tau_{Sil} = \tau_{Access} = 1. \tag{6}$$

The mean holding time of State Sus τ_{Sus} is given by [12]

$$\tau_{Sus} = \frac{1-(1-\lambda)^{T_{in}}}{\lambda}. \tag{7}$$

Finally, the limiting state probability of the Markov renewal process $(\boldsymbol{X}, \boldsymbol{Y})$ is given by

$$\tilde{\pi}_i = \frac{\pi_i \tau_i}{\sum_{j \in \mathbb{S}} \pi_j \tau_j}, \tag{8}$$

$i \in \mathbb{S}$, where $\mathbb{S}$ is the state space of $\boldsymbol{X}$. Accordingly, the probability of the UE being in State Trans can be obtained by combining (5)-(8) as

$$\begin{aligned} \tilde{\pi}_{Trans} &= \frac{\pi_{Trans}\tau_{Trans}}{\pi_{Trans}\tau_{Trans}+\pi_{Sus}\tau_{Sus}+\pi_{Sil}+\pi_{Access}} \\ &= \frac{\tau_{Trans}}{\tau_{Trans}+\frac{(1-\lambda)^{T_{in}}}{qp}+\frac{1}{\lambda}}, \end{aligned} \tag{9}$$

where τ_{Trans} is the mean holding time of State Trans, which will be derived in Section III-C.

B. Average Number of UEs in Each State

Denote $E[n_t]$ the average number of UEs in State Trans, which is given by

$$\begin{aligned} E[n_t] &= n\Pr\{\text{UE in State Trans}\} = n\tilde{\pi}_{Trans} \\ &= n\frac{\tau_{Trans}}{\tau_{Trans}+\frac{(1-\lambda)^{T_{in}}}{qp}+\frac{1}{\lambda}}. \end{aligned} \tag{10}$$

By converting (10), the mean holding time of transmission period can be obtained as

$$\tau_{Trans} = \frac{\frac{\lambda(1-\lambda)^{T_{in}}}{qp}+1}{\frac{n\lambda}{E[n_t]}-\lambda}. \tag{11}$$

Similarly, since RRC INACTIVE state includes State Sil and State Access, the average number of UEs in RRC INACTIVE state is given by

$$\begin{aligned} E[n_a] &= n\Pr\{\text{UE in RRC INACTIVE state}\} \\ &= n(\tilde{\pi}_{Access}+\tilde{\pi}_{Sil}) \\ &= n\frac{\frac{1}{pq}+\frac{1}{\lambda}}{\frac{1}{pq}+\frac{1}{\lambda(1-\lambda)^{T_{in}}}+\frac{\tau_{Trans}}{(1-\lambda)^{T_{in}}}}, \end{aligned} \tag{12}$$

by combining (11) and (12), we have

$$E[n_a] = \frac{\left(\frac{1}{qp}+\frac{1}{\lambda}\right)(n-E[n_t])}{\frac{1}{qp}+\frac{1}{\lambda(1-\lambda)^{T_{in}}}}. \tag{13}$$

C. Mean Holding Time of State Trans

Note that there are two cases for entering the transmission period: case 1 is the transition from State Access, case 2 is the transition from State Sus, as shown in Fig. 2. We derive the mean holding time of State Trans by considering the following two cases according to the way that UEs enter transmission:

1) Transition from State Access to State Trans: Denote $\hat{n_t}$ as the average number of UEs in transmission period under the precondition that at least one UE is in transmission period, which is given by

$$\hat{n}_t = \frac{E[n_t]}{1-\rho}, \tag{14}$$

where ρ is the probability that no UE is in transmission period. For one UE, the probability of not being in transmission period is given by

$$\Pr\{\text{UE not being in State Trans}\} = 1-\tilde{\pi}_{Trans}. \tag{15}$$

Accordingly, the probability that all n UEs are not in transmission period is obtained by combining (10) and (15) as

$$\rho = (1-\tilde{\pi}_{Trans})^n = (1-\frac{E[n_t]}{n})^n, \tag{16}$$

where n is the total number of the UEs.

The queue length in the buffer at the beginning of the transmission period in case 1, denoted as Q_1, is equal to the number of packets arriving during the random access process plus one packet that triggers the random access process. Therefore, the Probability Generating Function (PGF) and mean value of Q_1 are as follows:

$$Q_1(z) = zT_{Access}[\lambda(z)], \tag{17}$$

$$E[Q_1] = Q_1(z)^{'}|_{z=1} = \lambda E[T_{Access}]+1, \tag{18}$$

where T_{Access} is defined as the time interval from the moment that UE starts the random access procedure until its success, $T_{Access}(z)$ is the PGF of the length of random access procedure, $\lambda(z) = \lambda z + 1 - \lambda$ is the PGF of the number of packet arrival in a time slot, and $T_{Access}[\lambda(z)]$ denotes the PGF of the number of packets arrival in the random access period. The Probability Mass Function (PMF) of T_{Access} is obtained based on Eq. (l) in [11] as

$$P(T_{Access}=l) = (1-qp)^{l-1}qp \qquad l \geq 1. \tag{19}$$

Since the packets arrival of each UE follows the Bernoulli process with parameter λ, the PMF of Q_1 can be obtained as

$$P(Q_1 = m+1) = \sum_{l=m}^{\infty} \binom{l}{m} \lambda^m (1-\lambda)^{l-m} P\{T_{Access}=l\}. \tag{20}$$

The buffer can be cleared in one time slot when $0 < Q_1 \leq \frac{\hat{\mu}}{\hat{n_t}}$, where $\frac{\hat{\mu}}{\hat{n_t}}$ is the data rate of each UE. Similarly, k time slots are needed for transmission when $(k-1)\frac{\hat{\mu}}{\hat{n_t}} < Q_1 \leq k\frac{\hat{\mu}}{\hat{n_t}}, k \geq 1$. Accordingly, the mean holding time of State Trans in case 1 can be obtained as

$$\tau_{Trans}|_{case1} = \sum_{k=1}^{\infty} kP\left\{(k-1)\tfrac{\hat{\mu}}{\hat{n_t}} < Q_1 \leq k\tfrac{\hat{\mu}}{\hat{n_t}}\right\}, \tag{21}$$

by combining (14), (19), (20), (21), we have

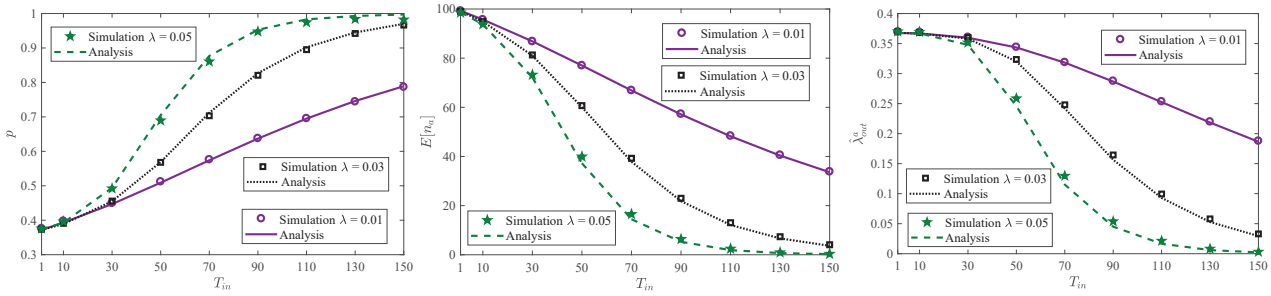
$$\tau_{Trans}|_{case1} = 1 + \frac{(1-qp)^{\frac{\hat{\mu}(1-\rho)}{\lambda E[n_t]}-\frac{1}{\lambda}}}{1-(1-qp)^{\frac{\hat{\mu}(1-\rho)}{\lambda E[n_t]}}}. \tag{22}$$

Note that when no packet arrives for T_{in} consecutive slots in suspend period, the UE transfers from the RRC CONNECTED state to the RRC INACTIVE state. Accordingly, the probability of case 1 is given by

$$\Pr\{\text{From State Access to State Trans}\} = (1-\lambda)^{T_{in}}. \tag{23}$$

2) Transition from State Sus to State Trans: In this case, the queue length in the buffer at the beginning of transmission is fixed to 1, i.e., Q_2 = 1. We make an approximation that the transmission of case 2 only takes one time slot, which means

$$\tau_{Trans}|_{case2} = 1. \tag{24}$$



(a) (b) (c)

Fig. 3. (a) Probability of successful transmission of access requests p, (b) the average number of UEs in RRC INACTIVE state $E[n_a]$ and (c) Access throughput $\hat{\lambda}^a_{out}$ versus inactivity timer T_{in} with the input rate $\lambda = 0.01$, 0.03 or 0.05. $n = 100$. $\hat{\mu} = 10$.

It's a reasonable approximation when the network service capacity $\hat{\mu}$ is large. The probability of case 2 is given by

$$\Pr\{\text{From State Sus to State Trans}\} = 1 - (1-\lambda)^{T_{in}}. \quad (25)$$

To sum up, the mean holding time of State Trans is given by

$$\begin{aligned}\tau_{Trans} &= (1-\lambda)^{T_{in}} \tau_{Trans}|_{case1} \\ &\quad + \left[1-(1-\lambda)^{T_{in}}\right] \tau_{Trans}|_{case2} \\ &= 1 + (1-\lambda)^{T_{in}} \frac{\frac{\hat{\mu}(1-\rho)}{\lambda E[n_t]} - \frac{1}{\lambda}}{1-(1-qp)\frac{\hat{\mu}(1-\rho)}{\lambda E[n_t]}}.\end{aligned} \quad (26)$$

D. Numerical Solutions of Equations

By combining (3), (11), (13), (16), (26), the probability of successful transmission p can be obtained. Algorithm 1 demonstrates how to get the numerical solution of p, where two while loops are used to find values that satisfy the constraints of the above equations and $\epsilon = 0.001$ is the iteration step size. By substituting p into (3) and (13), $E[n_a]$ and $E[n_t]$ can be obtained. Finally, access throughput $\hat{\lambda}^a_{out}$ can be obtained by combining (1) and (2) as

$$\hat{\lambda}^a_{out} = E[n_a]\frac{\lambda}{\lambda/(pq)+1}. \quad (27)$$

τ_{Trans} can be obtained by substituting $E[n_t]$ and p into (11). With τ_{Trans} and p, the energy efficiency η can be obtained by combining (4)-(8) as

$$\eta = \frac{P_{Trans}\tau_{Trans}}{P_{Sil}\frac{(1-\lambda)^{T_{in}}}{\lambda} + P_{Access}\frac{(1-\lambda)^{T_{in}}}{pq} + P_{Trans}\tau_{Trans} + P_{Sus}\frac{1-(1-\lambda)^{T_{in}}}{\lambda}}. \quad (28)$$

IV. Results And Discussions

A. Simulation and Discussion

In this section, simulation results are presented to verify the preceding analysis. Specifically, each UE starts from the Silence state. The packets arrival at each UE follows the Bernoulli process with parameter λ, and the buffer of each UE is infinite. With a busy buffer, UE performs the random access procedure. The simulation of the random access procedure follows the settings in [11]. With a successful random access procedure, the UE could transmit at most $\frac{\hat{\mu}}{n_t}$ packet for each time slot, where n_t is the number of UEs in Transmission state at each time slot. With an empty buffer, UE enters Suspend state. Only when the inactivity timer expires can UE switches back to Silence state from Suspend state. The simulation of the system is on a time-slot basis and is based on

Algorithm 1 Calculation of p

Input: ACB factor, q; node number, n; service capacity, $\hat{\mu}$; inactivity timer, T_{in}; iteration step size, ϵ;
Output: Probability of successful transmission, p;

```
E[n_t] ← 0, ε ← 0.001
while E[n_t] < n do
    E[n_t] ← E[n_t] + ε, P ← 1
    while P > 0 do
        Calculate E[n_a] based on Eq.(13)
        Calculate ρ based on Eq.(16)
        Calculate τ_Trans based on Eq.(26) denoted as T_1
        Calculate τ_Trans based on Eq.(11) denoted as T_2
        if Eq.(3) is satisfied and |T_1 − T_2| < ε then
            p ← P
        end if
        P ← P − ε
    end while
end while
```

Table I
Power Level of Each State [15]

Sub-state	Power level averaged per time slot
Silence state	0.08mW
Access state	0.68mW
Transmission state	1mW
Suspend state	0.18mW

MATLAB. The power consumption model of each subperiod is set according to the 3GPP protocol [15], as shown in Table I. Each simulation is carried for 10^6 time slots and the optimal ACB factor q for minimizing the access delay is set according to Eq. (l4) in [11].

Fig. 3a, Fig. 3b and Fig. 3c respectively demonstrate how the probability of successful transmission of access requests p, the average number of UEs in RRC INACTIVE state $E[n_a]$ and the access throughput of the network $\hat{\lambda}^a_{out}$ varies with the inactivity timer T_{in} when the input rate λ = 0.01, 0.03 or 0.05, the node number $n = 100$ and the service capacity $\hat{\mu} = 10$. As shown in Fig. 3, as the inactivity timer increases, less UEs remain in RRC INACTIVE state, the probability of successful transmission of access requests increases and the access throughput decreases. Actually, the increase of the inactivity timer makes less UEs stay in RRC INACTIVE state, which brings the reduction of the aggregate input rate of access requests, finally resulting in the decrease of access throughput. Simulation results also corroborate the theoretical derivation of (3), (13), (26) and (27).

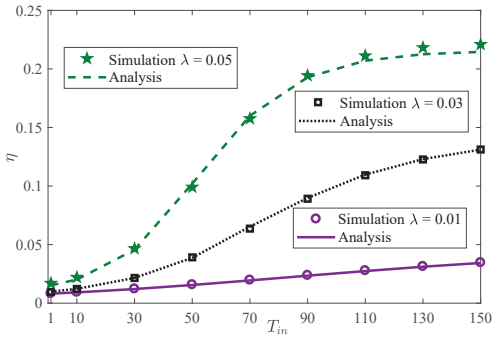


Fig. 4. Energy efficiency η versus inactivity timer T_{in} with $n = 100$, $\hat{\mu} = 10$, $\lambda = 0.01$, 0.03 or 0.05

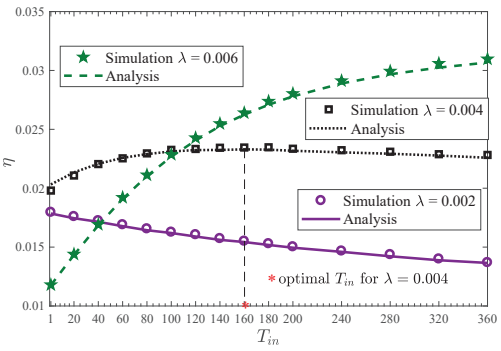


Fig. 5. Energy efficiency η versus inactivity timer T_{in} with $n = 100$, $\hat{\mu} = 10$, $\lambda = 0.002$, 0.004 or 0.006

Fig. 4 demonstrates how the energy efficiency η varies with the inactivity timer T_{in} when the input rate $\lambda = 0.01$, 0.03 or 0.05, the node number $n = 100$ and the service capacity $\hat{\mu} = 10$. As shown in Fig. 4, energy efficiency monotonically increases with the inactivity timer when the traffic rate is large. The increase of energy efficiency is mainly because the decrease of total energy consumption. Firstly, as the inactivity timer increases, the reduction of UEs in RRC INACTIVE state brings less competition for random access, resulting in less energy consumption for random access. Although the energy consumed in suspend period increases with inactivity timer, the energy saved from the random access period is higher than the energy wasted in suspend period. Secondly, the energy consumed for transmission is fixed since the data throughput doesn't change with the inactivity timer. Accordingly, the total energy consumption decreases with the inactivity timer when the traffic rate is large, making energy efficiency higher.

Fig. 5 demonstrates how the energy efficiency η varies with the inactivity timer T_{in} when the input rate $\lambda = 0.002$, 0.004 or 0.006, the node number $n = 100$ and the service capacity $\hat{\mu}$ = 10. As shown in Fig. 5, energy efficiency η is crucially determined by the traffic rate and the inactivity timer. Energy efficiency decreases with inactivity timer when the traffic rate is low, but increases with inactivity timer when the traffic rate is large. That means maintaining a large inactivity timer is a waste of energy for light-traffic scenarios while it is preferred for heavy-traffic scenarios. For light-traffic scenarios, energy saved from random access competition is no longer higher than the energy wasted in Suspend state, which causes energy efficiency decreases with T_{in}. What's more, for input rate $\lambda =$ 0.004, the energy efficiency can be optimized by properly setting the inactivity timer. Simulation results also corroborate the theoretical derivation of (28).

V. Conclusion

This paper proposes a closed-loop analytical framework for B5G RRC protocol, based on which key performance metrics such as access throughput and energy efficiency can be calculated.

The analysis sheds important light on the performance optimization of the B5G networks. It reveals that in high-traffic scenarios, increasing the inactivity timer improves energy efficiency at the cost of lower access throughput. However, in light-traffic scenarios, maintaining a small timer can achieve high energy efficiency and access throughput. That means the optimal inactivity timer should be carefully selected. Moreover, the upper limit of energy efficiency achieved after tuning inactivity timer mainly depends on the traffic rate. The energy efficiency gains from tuning inactivity timer are more significant in high-traffic scenarios. Whether the energy efficiency can be improved mainly depends on the balance between energy saved from the random access competition and energy wasted in connected state, in other words, the optimal inactivity timer crucially depends on the power consumption level in the access process.

References

[1] M. Z. Chowdhury, M. Shahjalal, S. Ahmed and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957-975, 2020.

[2] T. Kebede, Y. Wondie, J. Steinbrunn, H. B. Kassa and K. T. Kornegay, "Multi-carrier waveforms and multiple access strategies in wireless networks: Performance, applications, and challenges," *IEEE Access*, vol. 10, pp. 21120-21140, 2022.

[3] *NR; Radio Resource Control (RRC) protocol specification (Release 17)*, document TS 38.331 V17.0.0, 3GPP, May 2022.

[4] S. Hailu, M. Saily and O. Tirkkonen, "RRC state handling for 5G," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 106-113, Jan. 2019.

[5] W. Zhan and L. Dai, "Massive random access of machine-to-machine communications in LTE networks: Modeling and throughput optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2771-2785, Apr. 2018.

[6] H. Huang, T. Ye, T. T. Lee and W. Sun, "Delay and stability analysis of connection-based slotted-Aloha," *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 203-219, Feb. 2021.

[7] Y. Liang, X. Li, J. Zhang and Z. Ding, "Non-orthogonal random access for 5G networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4817-4831, July 2017.

[8] L. Liang, L. Xu, B. Cao and Y. Jia, "A cluster-based congestion-mitigating access scheme for massive M2M communications in Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2200-2211, June 2018.

[9] S. -C. Tseng, Z. -W. Liu, Y. -C. Chou and C. -W. Huang, "Radio resource scheduling for 5G NR via deep deterministic policy gradient," in *Proc. IEEE ICC Workshops*, May 2019.

[10] A. Ramadan, N. Zorba and H. S. Hassanein, "Uplink cluster-based radio resource scheduling for HetNet mMTC scenarios," in *Proc. IEEE Globecom'* 2022.

[11] W. Zhan and L. Dai, "Access delay optimization of M2M communications in LTE networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1675-1678, Dec. 2019.

[12] Y. Mo, W. Cai, W. Zhan, Q. Chen, Y. Yin and X. Sun, "Modeling and performance analysis of 5G RRC protocol with machine-type communications," in *Proc. IEEE PIMRC'* 2022.

[13] M. Zukerman, *Introduction to queueing theory and stochastic teletraffic models*, 2014.

[14] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*. Hoboken, NJ, USA: Wiley, 1998.

[15] *Study on User Equipment (UE) power saving in NR (Release 16)*, document TS 38.840 V16.0.0, 3GPP, June 2019.