# Modeling and Performance Analysis of 5G RRC Protocol with Machine-Type Communications

Yuanhui Mo*, Weiwen Cai†, Wen Zhan*, Qiming Chen†, Ying Yin* and Xinghua Sun*

*School of Electronics and Communication Engineering, Shenzhen Campus of Sun Yat-sen University
† China Mobile Group Guangdong Co., Ltd.
{moyh7, yiny36}@mail2.sysu.edu.cn, {zhanw6, sunxinghua}@mail.sysu.edu.cn, {caiweiwen, chenqiming}@gd.chinamobile.com

***Abstract*—5G New Radio (NR) introduces a new Radio Resource Control (RRC) state, i.e., RRC INACTIVE, for providing the efficient service for massive Machine Type Communications (mMTC). To release the full potential of the new RRC state, it is of great importance to properly model the new RRC state transition process and reveal the effect of system parameters on the network performance.**

**To address the above issue, this paper proposes a novel 5G RRC analytical model based on discrete-time vacation queuing theory, where the time period of the device in RRC INACTIVE state is regarded as the vacation period of the server in the queueing system. By leveraging this novel model, key performance metrics, such as the random access rate and the RRC resource utilization ratio, are explicitly characterized and obtained as functions of system parameters, including packet arrival rate, service rate and inactivity timer. The analysis reveals that to reduce the random access rate, the system should increase the inactivity timer, packet arrival rate or decrease the service rate. On the other hand, to improve the RRC resource utilization ratio, the inactivity timer should be cut down especially when the arrival rate is small or the service rate is large. The analysis is verified by simulations and sheds important light on practical 5G network design for supporting mMTC.**

***Index Terms*—Machine type communications, 5G RRC, discrete-time vacation queue, performance analysis.**

## I. Introduction

The 5G New Radio (NR) networks are tailored to support three types of service: Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC) and massive Machine Type Communications (mMTC) [1], among which mMTC has captured significant attention as it serves as the foundation for many emerging Internet of Things services, such as environmental monitoring and e-Health [2]. In MTC, Machine Type Devices (MTDs), such as sensors and actuators, process and exchange information packets without human intervention. The MTC traffic characteristics are different from the traditional human-type communications, such as video streaming, in that it contains a large number of MTDs while each MTD transmits small packets sporadically.

The work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2020B0101120003, in part by The Shenzhen Science and Technology Program (No.2021A04, No.RCBS20210706092408010), in part by National Natural Science Foundation of China under Grant 62001524. *(Corresponding author: Wen Zhan)*

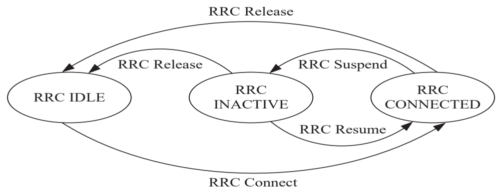


Fig. 1. RRC state machine in 5G.

To provide efficient support for MTC traffic, 5G NR has made many upgrades from the system point of view, compared to 4G. One key revision lies in the RRC layer, where a new RRC state, i.e., RRC INACTIVE, is introduced [3].

### A. Background of 5G RRC States

5G RRC state machine is shown in Fig. 1, which contains three states: RRC IDLE, RRC CONNECTED and RRC INACTIVE according to 5G standards [4]. The former two RRC states are inherited from legacy 4G. The RRC INACTIVE state is the newly introduced one for MTC for reducing signaling overhead, energy consumption and access delay [5]–[8].

When the device is in RRC IDLE state, there is no RAN/CN (Random Access Network/Core Network) connection and User Equipment (UE) Access Stratum (AS) context in the 5G Core Network (5GCN) and gNB. In this state, when a data packet arrives in the buffer, the device performs the random access procedure for shifting into the RRC CONNECTED state. A connection between the MTD, gNB and 5GCN is established, via which the device transmits data packets in a contention-free manner.

Note that the gNB would configure an inactivity timer, denoted by $T_{in}$, for each device in RRC CONNECTED state [4]. If the gNB detects the device which does not have any packet to send and receive during the inactivity timer $T_{in}$, then the gNB releases the connection and places the device in RRC INACTIVE state. When the device is in RRC INACTIVE state, only the on-the-air connection with gNB is released while the UE context is kept at the 5GCN and gNB. Therefore, compared with the state transition from RRC IDLE to RRC CONNECTED, the state transition from RRC INACTIVE to RRC CONNECTED incurs less signaling overhead and energy consumption.

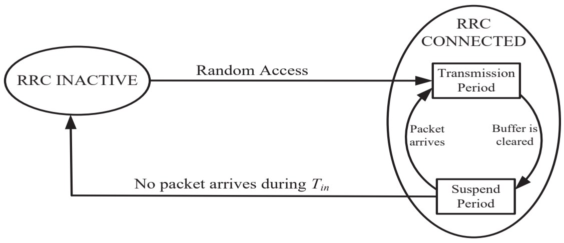


Fig. 2. State transition diagram of the model.

It is clear that with such a new RRC state, i.e., RRC INACTIVE, 5G NR can provide more efficient support for MTC traffic, compared to 4G. Yet, the performance gain still depends on the network configuration, such as the inactivity timer $T_{in}$ [6], [8]. To evaluate the effect of system settings on the network performance, lots of related works have been done.

### *B. Related Work*

So far, a large body of existing works focused on evaluating the performance of 5G RRC protocol with MTC via simulations. For instance, authors in [5] revealed that the RRC INACTIVE state enables a quick and lightweight transition from inactive to active data transmission, and the state transition process becomes highly configurable. Simulation results in [6]–[8] indicated that to achieve a trade-off between UE power consumption, latency and signaling overhead, the inactivity timer should be carefully selected, e.g., a small inactivity timer is preferable in the case of light traffic load. In [9] and [10], authors redesigned the random access procedure such that one small packet can be delivered in random access procedure and the signaling overhead can be reduced.

Although the above simulation results are extensive, there is still a lack of proper model of 5G RRC protocol with MTC and therefore no insight can be provided for performance analysis and optimization. To address this issue, in our recent work [11], an RRC state transition model was proposed to analyze the performance metrics such as resource utilization ratio and mean time length of each connection, based on which, a utility-based analytical framework was formulated for the optimal connection management of mMTC in 5G networks. However, due to iterative calculation, the analytical mode in [11] becomes computationally expensive, when the inactivity timer is large, limiting its implementation to practical scenarios.

In summary, while the aforementioned developments have been substantial, they either evaluated the network performance via simulations without analytical analysis, i.e., [5]–[10], or proposed analytical models which are computationally expensive, i.e., [11]. As a result, how to properly model and analyze 5G NR with MTC from the RRC layer point of view is still an open issue.

### *C. Our Contributions*

To address the above issue, in this paper, we propose a novel analytical model based on the discrete-time vacation queuing framework for 5G RRC protocol with MTC. Specifically, by focusing on the state transition between RRC INACTIVE and RRC CONNECTED, the time period when the MTD is in RRC INACTIVE state and that in RRC CONNECTED state is regarded as the vacation period and the busy period of the queuing system, respectively.

To facilitate the analysis, we further divide each RRC CONNECTED state into two sub-periods, i.e., transmission period and suspend period, and each RRC INACTIVE state into random access period and silence period. The mean time lengths of each sub-period are derived, based on which the random access rate (i.e., the frequency that the MTD initiates the random access procedure) and utilization ratio (i.e., the percentage of slots that the MTD is in the transmission period when it is in RRC CONNECTED state) are obtained as explicit functions of system parameters. By the virtue of the vacation queuing model, the computational complexity can be significantly reduced, which is in sharp contrast to [11], where the computational complexity of the model rapidly grows with the inactivity timer.

For the performance evaluation of 5G RRC RRC protocol with MTC, our analysis reveals that the network performance is crucially determined by the arrival rate, service rate and inactivity timer. It is found that the utilization ratio of RRC resource and random access rate decrease with the increase of the inactivity timer, indicating that the inactivity timer is a key system parameter that determines the tradeoff between the signaling overhead and the resource utilization ratio. As the arrival rate of each device grows, the random access rate can be reduced while the utilization ratio can be improved.

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 5G networks are obtained. The analysis is verified by the simulation results in Section IV. Finally, concluding remarks are summarized in Section V.

## II. System Model

Consider a single-cell 5G network serving $n$ registered MTDs. We assume each MTD is either in RRC INACTIVE state or RRC CONNECTED state[1], as shown in Fig. 2. Define an RRC cycle, denoted by $T_R$, as the time interval between two consecutive instants that the MTD enters RRC INACTIVE state. We divide an RRC cycle into the following subperiods: silence period, random access period, transmission period, and suspend period, as shown in Fig. 3.

### *A. Behavior Characterization in One RRC Cycle*

The arrival of data packets at each MTD follows a Bernoulli process with parameter $\lambda$, and the buffer of each MTD is infinite. With a busy buffer, the MTD in RRC INACTIVE state would access gNB for data transmission via random access procedure

[1]Here we ignore the RRC IDLE state because we assume all MTDs have registered at the gNB. The UE context information remains at the 5GCN such that the device does not move into to RRC IDLE state. This assumption can be applied to the stationary MTC use cases, where the locations of MTDs remain unchanged for a long time.

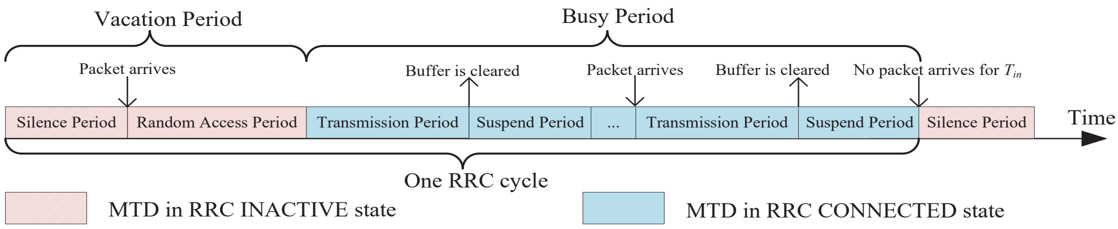


Fig. 3. Illustration of inactive period, random access period, transmission period, and suspend period in an RRC cycle.

and enters the random access period, as shown in Fig. 3. Upon a successful random access procedure, the MTD transfers into RRC CONNECTED. As shown in Fig. 2, we divide the RRC CONNECTED state into two sub-states: Transmission period and suspend period. In the transmission period, the MTD is busy transmitting data packets with service rate $\mu$. When the buffer is cleared, the MTD enters suspend period.

In the suspend period, although there is a connection between the MTD and gNB, the MTD does not transmit packets since the buffer is empty. If a packet arrives during the suspend period, then the MTD would start the transmission by delivering a scheduling request to the gNB and then enter the transmission period. If there is no packet arrival during $T_{in}$ slots in suspend period, then the connection is released and the MTD shifts back to the silence period which is in RRC INACTIVE state, where $T_{in}$ is referred to as the inactivity timer [4].

### B. Performance Measures

It is clear that the network performance crucially depends on the inactivity timer $T_{in}$, the number of MTDs $n$ and the arrival rate $\lambda$, etc. Yet, how to evaluate the effect of those parameters on the network performance still remains unexplored due to the lack of a proper model. For performance evaluation, this paper focus on two key performance measures:

- Random access rate, denoted by $\gamma$, refers to the frequency that the MTD initiates the random access procedure. This metric is an important performance indicator for practical MTC scenarios. A large $\gamma$ indicates that the MTD has to perform the random access procedure often, which incurs extensive signaling overhead and energy consumption.
- Utilization ratio, denoted by $\phi$, refers to the percentage of slots that the MTD is in the transmission period when it is in RRC CONNECTED state. A large $\phi$ is preferable for the 5G system.

Table I summarizes notations and corresponding definitions used in this paper. In the following, a new analytical model will be established, based on which the above two performance measures are derived.

TABLE I
NOTATIONS AND DEFINITIONS

| Notations | Definitions |
|---|---|
| $\lambda$ | Arrival rate |
| $\mu$ | Service rate |
| $T_{in}$ | Inactivity timer |
| $n$ | The number of MTDs |
| $\gamma$ | Random access rate |
| $\phi$ | Utilization ratio |
| $q$ | Access request transmission probability (ACB factor) |
| $p$ | Probability of successful access |
| $M$ | The number of preambles |
| $T_R$ | Length of an RRC cycle |
| $T_{Access}$ | Length of one random access period |
| $T_{Con}$ | Length of RRC CONNECTED state |
| $T_{Trans}$ | Length of one transmission period |
| $T_{Sus}$ | Length of one suspend period |
| $T_{Sil}$ | Length of one silence period |
| $Q$ | Queue length at the beginning of the transmission period |
| $N$ | The number of transmission periods or suspend periods in an RRC cycle |

## III. Modeling and Performance Analysis

In this section, we establish a novel analytical model based on discrete-time vacation queuing and analyze the performance of 5G NR in the massive MTC scenario.

### A. Discrete-time Vacation Queueing Model

By definition, a vacation queue is a queueing system in which the server is available only a portion of the time. At other times it is busy serving other stations or just not available may be due to maintenance activities [12]. For 5G RRC with MTC, the behavior of MTD and gNB is in line with the behavior between customer and service in the vacation queueing system.

Specifically, it can be seen from Fig. 3 that transmission period and suspend period alternatively appear in the time domain. With the expiry of the inactivity timer, the MTD in suspend period will transfer to the RRC INACTIVE where there is no data connection between the device and the gNB. Such a state transition process can be properly regarded as a vacation queueing system, as shown in Fig. 3. The time away from the primary service center is called a vacation. It is the time period when the MTD is in the RRC INACTIVE state, which is constituted by two sub-periods: the silence period and random access period. The time with the server in service is the time period when the MTD is in the RRC CONNECTED state, which is constituted by the transmission period and the suspend period.

By modeling the 5G RRC with MTC as a vacation queueing system, the extensive analytical results in the queueing theory can be applied, which simplifies the network performance analysis. In the following, we will first derive the mean time length of the busy period of the queueing system, i.e., the time period when the MTD is in the RRC CONNECTED state, and then derive the mean time length of the vacation period. Performance measures, i.e., random access rate $\gamma$ and utilization ratio $\phi$, will be obtained in subsection III-C.

## B. Preliminary Analysis

*1) Mean Length of RRC CONNECTED State:* As shown in Fig. 2, we divide the period of the MTD in RRC CONNECTED state into two sub-periods: Transmission period and suspend period. Let $T_{Con}$ denote the time length of RRC CONNECTED state. Accordingly, we have

$$T_{Con} = N(T_{Trans} + T_{Sus}), \tag{1}$$

where $T_{Trans}$ denotes the time length of one transmission period and $T_{Sus}$ denotes the time length of one suspend period, and $N$ denotes the number of transmission periods and suspend periods in the RRC CONNECTED state.

**Derivation of $T_{Trans}$:** Note that there are two cases of entering the transmission period: one is the transition from random access period to the transmission period, the other is the transition from the suspend period to the transmission period which are shown in Fig. 2. Let us denote the mean queue length in the buffer at the beginning of the transmission period as $Q$. We can consider the following two cases according to the way that the MTD enters the transmission period:

- Transition from the random access period to the transmission period: $Q$ is equal to the number of packets arriving during the random access period plus the one packet in the buffer. Note that the MTD shifts to RRC INACTIVE state because it has no packet arrival in $T_{in}$ slots when it is in the suspend period of the last RRC cycle. Accordingly, the probability of this event is given by

$$P\{\text{No packet arrives in suspend period}\} = (1-\lambda)^{T_{in}}. \tag{2}$$

- Transition from the suspend period to the transmission period: In this case, $Q = 1$. The probability of this event is given by

$$P\{\text{Packets arrive in suspend period}\} = 1-(1-\lambda)^{T_{in}}. \tag{3}$$

To sum up, the Probability Generating Function (PGF) and mean value of $Q$ are as follows:

$$Q(z) = (1-(1-\lambda)^{T_{in}})z + z(1-\lambda)^{T_{in}} T_{Access}[\lambda(z)], \tag{4}$$

$$E[Q] = Q(z)^{'}|_{z=1} = 1 + \lambda(1-\lambda)^{T_{in}} E[T_{Access}], \tag{5}$$

where $T_{Access}(z)$ is the PGF of the length of random access period, $\lambda(z) = \lambda z + 1 - \lambda$ is the PGF of the number of packet arrival in one time slot, and $T_{Access}[\lambda(z)]$ denotes the PGF of the number of packets arrival in the random access period. $E[T_{Access}]$ will be derived in the following subsection III-B3.

Let $B$ denote the length of transmission period with $Q = 1$. The vacation queueing theory has revealed that [12]

$$E[B] = \frac{1}{\mu - \lambda}, \tag{6}$$

and the mean length of transmission period[2]

$$E[T_{Trans}] = E[Q]E[B] = \frac{E[Q]}{\mu - \lambda}. \tag{7}$$

**Derivation of $T_{Sus}$:** Recall that if there is no packet arrival during $T_{in}$ slots, then the connection is released and the MTD leaves the suspend period. Since the packet arrivals in each time slot with probability $\lambda$, the Probability Mass Function (PMF) of $T_{Sus}$ can be obtained as

$$P\{T_{Sus} = i\} = \begin{cases} \lambda(1-\lambda)^{i-1} & i < T_{in}, \\ (1-\lambda)^{T_{in}-1} & i = T_{in}. \end{cases} \tag{8}$$

Thus, the mean length of one suspend period is

$$E[T_{Sus}] = \sum_{i=1}^{T_{in}} iP\{T_{Sus} = i\} = \frac{1-(1-\lambda)^{T_{in}}}{\lambda}. \tag{9}$$

**Derivation of $N$:** Note that when no packet arrives during the suspend period, i.e., no packet arrives in consecutive $T_{in}$ slots, the MTD leaves the RRC CONNECTED state. The probability of this event is obtained by (2), with which $N$ can be obtained by

$$E[N] = \frac{1}{(1-\lambda)^{T_{in}}}. \tag{10}$$

Finally, by combining (1), (7), (9), (10), we can have the mean length of RRC CONNECTED state, i.e., the busy period of the vacation queueing system, as

$$E[T_{Con}] = \frac{1-(1-\lambda)^{T_{in}}}{(1-\lambda)^{T_{in}}\lambda} + \frac{E[Q]}{(\mu-\lambda)(1-\lambda)^{T_{in}}}. \tag{11}$$

When no packet arrives in the consecutive $T_{in}$ slots in suspend period, the MTD transfers from the RRC CONNECTED state to the RRC INACTIVE state. Equivalently, the server of the queueing system is on vacation. Each RRC INACTIVE state is constituted by two sub-periods: the random access period and the silence period. The mean length of each sub-period is given below.

*2) Mean Length of the Silence Period:* Let $T_{Sil}$ denote the time length of the silence period. Note that the silence period comes to the end when a packet arrives in the buffer. Thus, with Bernoulli arrival, we can have the mean time length of the silence period as

$$E[T_{Sil}] = \frac{1}{\lambda}. \tag{12}$$

With data packets in the buffer, the device performs the random access procedure and therefore enters the random access period, as shown in Fig. 3.

*3) Mean Length of the Random Access Period:* The mean length of the random access period, denoted by $E[T_{Access}]$, is

[2]Iterative algorithms are required in [11] for calculation, with which the computational complexity would be larger than that of the analysis in this paper.

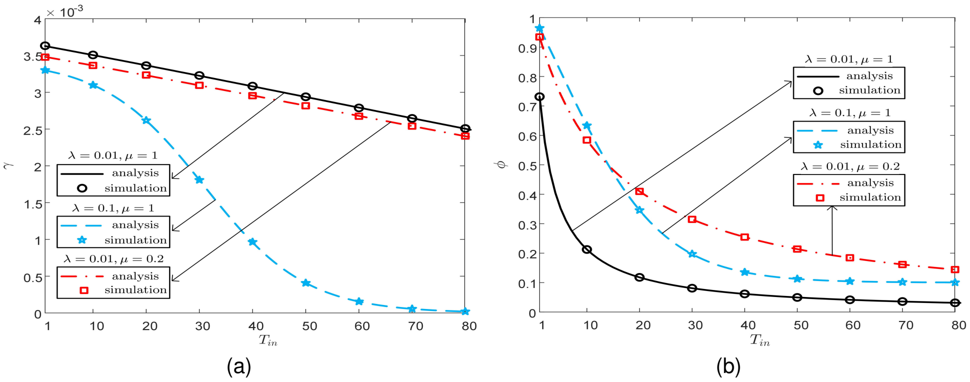


Fig. 4. Random access rate $\gamma$ and utilization ratio $\phi$ versus the inactivity timer $T_{in}$ (in unit of slots). $n = 100$. (a) Random access rate $\gamma$ versus $T_{in}$. (b) Utilization ratio $\phi$ versus $T_{in}$.

defined as the time interval from the moment that MTD starts the random access procedure until the random access procedure is successful. Note that to efficiently handle the massive number of access requests and avoid the collision[3], the gNB can tune the access request transmission probability of each MTD, denoted by $q \in (0, 1]$, which is also referred to as the Access Classing Barring (ACB) factor in the 5G standards [4].

The optimal setting of the transmission probability of each MTD for minimizing $E[T_{Access}]$ has been given in [13] as

$$q = \begin{cases} \dfrac{\hat{\lambda}}{n(\hat{\lambda} - e^{-1})} & \text{if} \quad \hat{\lambda} > \hat{\lambda}_0, \\[2ex] \dfrac{4\mathbb{W}_{-1}^2\left(-\frac{\sqrt{\hat{\lambda}}}{2}\right)}{n\left(-2\mathbb{W}_{-1}\left(-\frac{\sqrt{\hat{\lambda}}}{2}\right)-1\right)} & \text{otherwise,} \end{cases} \tag{13}$$

where $\hat{\lambda} = n\lambda$, $\hat{\lambda}_0 \approx 0.48$ and $\mathbb{W}_{-1}(\cdot)$ is one real-valued branch of the Lambert W function [14]. The corresponding minimum $E[T_{Access}]$ is given by

$$E[T_{Access}] = \begin{cases} ne - \frac{n}{\hat{\lambda}} & \text{if } \lambda > \frac{\hat{\lambda}_0}{n}, \\[2ex] \dfrac{n\left(-2\mathbb{W}_{-1}\left(-\frac{\sqrt{\hat{\lambda}}}{2}\right)-1\right)}{4M\mathbb{W}_{-1}^2\left(-\frac{\sqrt{\hat{\lambda}}}{2}\right)p} & \text{otherwise,} \end{cases} \tag{14}$$

where $p$ denotes the probability of successful access.

To sum up, by combining (11), (12), (14), the mean time length of an RRC cycle is given by

$$\begin{aligned} E[T_R] &= E[T_{Access}] + E[T_{Con}] + E[T_{Sil}] \\ &= \frac{1 + \lambda(1-\lambda)^{T_{in}} E[T_{Access}]}{(1-\lambda)^{T_{in}} \lambda (1 - \frac{\lambda}{\mu})}, \end{aligned} \tag{15}$$

where $E[T_{Access}]$ is given in (14).

[3]In the random access procedure, each MTD randomly selects one out of $M \geq 1$ orthogonal preambles and transmits it to the gNB. If more than one MTDs transmit the same preamble simultaneously, then a collision occurs and all of them fail. The access request is successful as long as there is no collision. 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 [13].

### C. Performance Metrics

So far, based on the vacation queueing model, we have obtained the mean time lengths of each sub-period in Fig. 3. Let us now derive the performance measures, i.e., random access rate $\gamma$ and utilization ratio $\phi$.

Note that with each random access procedure, the device enters the RRC CONNECTED state once. Accordingly, the random access rate $\gamma$, i.e., the frequency that the MTD initiates the random access procedure, can be written as

$$\gamma = \frac{1}{E[T_R]} = \frac{\lambda(1 - \frac{\lambda}{\mu})(1-\lambda)^{T_{in}}}{E[Q]}. \tag{16}$$

The utilization ratio $\phi$ refers to the percentage of slots that the MTD is in the transmission period when it is in RRC CONNECTED state, which can be obtained as

$$\phi = \frac{E[T_{Trans}]}{E[T_{Trans}] + E[T_{Sus}]} = \frac{\frac{E[Q]}{\mu - \lambda}}{\frac{E[Q]}{\mu - \lambda} + \frac{1-(1-\lambda)^{T_{in}}}{\lambda}}. \tag{17}$$

Therefore, by combining (16) and (17), the relationship between $\phi$ and $\gamma$ is obtained as follows:

$$\phi = \frac{\lambda^2(1-\lambda)^{T_{in}}}{\lambda^2(1-\lambda)^{T_{in}} + \mu[1-(1-\lambda)^{T_{in}}]\gamma}. \tag{18}$$

## IV. Results and Discussions

### A. Simulation

In this section, simulation results are presented to verify the preceding analysis. The simulation setting is the same as the system model in Fig. 2. Specifically, each RRC cycle starts from the silence period. The packet arrival rate at each MTD follows the Bernoulli process with parameter $\lambda$, and the buffer of each MTD is infinite. With a busy buffer, MTD will perform the random access procedure. The simulation of the random access procedure follows the settings in [13]. With the successful random access procedure, the MTD's data buffer will be cleared and the service time for each packet follows the geometric distribution with parameter $\mu$. The device then transfers to suspend period. Only when the inactivity timer expires can MTD switch back to the silence period from the suspend period.

The simulation of the system is on a time-slot basis and is based on MATLAB. Each simulation is carried for $10^8$ time slots. We count the time length of each RRC cycle, each transmission period and each suspend period. The utilization ratio $\phi$ is obtained by calculating the ratio of the total slots of transmission period to the total slots of the case in which the MTD is in RRC CONNECTED state. The random access rate $\gamma$ is the reciprocal of the mean length of the RRC cycle.

## B. Discussions

Fig. 4a demonstrates how the random access rate $\gamma$ varies with the inactivity timer $T_{in}$ with $\lambda = 0.1$ or 0.01 and $\mu = 1$ or 0.2. Recall that the MTD leaves the RRC CONNECTED state if it is in the suspend period and does not have packets to send in $T_{in}$ slots. Therefore, it is intuitively clear that with a large inactivity timer $T_{in}$, a heavy traffic input rate $\lambda$ or a small service rate $\mu$, the MTD would remain in RRC CONNECTED state long and less likely to initiate the random access procedure. Accordingly, we can see from Fig. 4a, the random access rate decreases with the increase of $T_{in}$, $\lambda$ and the decrease of $\mu$. Particularly, with a larger arrival rate, i.e., $\lambda = 0.1$, the random access rate $\gamma$ sharply declines as the inactivity timer $T_{in}$ grows, implying that in this case, once the MTD initiates a connection with the gNB, the connection would remain for a long time.

Fig. 4b demonstrates how the utilization ratio $\phi$ varies with the inactivity timer $T_{in}$ with the arrival rate $\lambda = 0.1$ or 0.01 and service rate $\mu = 1$ or 0.2. Intuitively, given $\mu$, the utilization ratio $\phi$ should grow with arrival rate $\lambda$ and decrease with $T_{in}$. Accordingly, we can see from Fig. 4b that with $\lambda = 0.01$ and $\mu = 1$, the utilization ratio $\phi$ is the lowest among the three cases, especially when the inactivity timer $T_{in}$ is large. If $\mu$ deceases to be 0.2, then the utilization ratio $\phi$ with $\lambda = 0.01$ increases and may even larger than that with $\lambda = 0.1$, which reveals that to boost the utilization ratio, a smaller $\mu$ is preferable, particularly when the traffic is light.

By comparing Fig. 4a and Fig. 4b, we can see that a larger inactivity timer $T_{in}$ can reduce the frequency that the device initiates the random access procedure, leading to less signaling overhead. While it also reduces the utilization ratio of channel resources. Such a tradeoff between the random access rate $\gamma$ and the utilization ratio $\phi$ in terms of $T_{in}$ has been given in (18). A closer look at this tradeoff is further presented in Fig. 5. We can see from Fig. 5 that as $T_{in}$ decreases from 40 to 1, both the random access rate $\gamma$ and the utilization ratio $\phi$ are enlarged. To reduce $\gamma$ while improving $\phi$, the system should either increase the traffic input rate $\lambda$ or cut down the service rate $\mu$. Finally, we can see from Figs. 4-5 that the analysis results match well with the simulation results, which verifies the analytical framework proposed in this paper.

# V. Conclusion

Based on the discrete-time vacation queuing theory, this paper proposes a novel analytical framework for 5G RRC protocol in the massive MTC scenario. Key performance metrics, such as the random access rate and the utilization ratio, are obtained as explicit functions of packet arrival rate, service rate and inactivity timer, with which the effect of system parameter

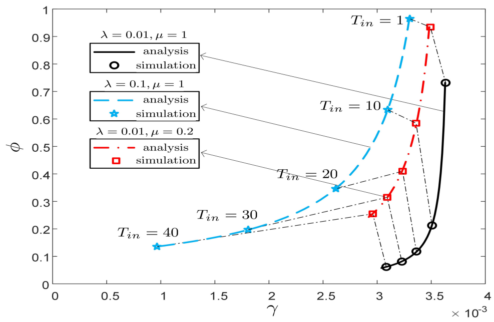


Fig. 5. The utilization ratio $\phi$ and random access rate $\gamma$ versus $T_{in}$. $n = 100$.

settings on the 5G network performance from the RRC layer point of view is evaluated.

The analysis sheds important light on the performance optimization of the 5G networks. It reveals that the inactivity timer crucially determines the tradeoff between the random access rate and the utilization ratio. Optimizing one metric always leads to the degradation of the other unless the packet arrival rate and service rate are jointly tuned. Based on this novel RRC model, we can analyze and optimize the relationship between various network performances. How to properly tune them for optimizing 5G network performance in the context of massive machine-type communications is an important issue that will be addressed in our future work.

# References


[1] V. W. S. Wong, R. Schober, D. W. K. Ng and L. C. Wang, *Key technologies for 5G wireless systems*, Cambridge University Press, 2017.

[2] F. Ghavimi and H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 525-549, Secondquarter 2015.

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

[4] 3GPP TS 38.331, v. 16.4.1, "5G;NR;Radio Resource Control (RRC);Protocol specification," Apr. 2021.

[5] I. L. Da Silva, G. Mildh, M. Säily and S. Hailu, "A novel state model for 5G radio access networks," in *Proc. IEEE ICC*, May 2016.

[6] A. Khlass, D. Laselva and R. Jarvela, "On the flexible and performance-enhanced radio resource control for 5G NR networks," in *Proc. IEEE VTC-Fall*, Sept. 2019.

[7] W. Zhao, H. Shen, F. Zhang and H. Tan, "Adaptive power optimization for mobile traffic based on machine learning," in *Proc. IEEE CSCWD*, May 2019.

[8] G. Foddis, R. G. Garroppo, S. Giordano, G. Procissi, S. Roma and S. Topazzi, "LTE traffic analysis for signalling load and energy consumption trade-off in mobile networks," in *Proc. IEEE ICC*, Jun. 2015.

[9] S. Ryoo, J. Jung and R. Ahn, "Energy efficiency enhancement with RRC connection control for 5G new RAT," in *Proc. IEEE WCNC*, Apr. 2018.

[10] P. Andres-Maldonado, P. Ameigeiras, J. Prados-Garzon, J. J. Ramos-Munoz and J. M. Lopez-Soler, "Reduced M2M signaling communications in 3GPP LTE and future 5G cellular networks," in *Proc. WD*, Mar. 2016.

[11] W. Zhan, C. Xu, X. Sun and J. Zou, "Towards optimal connection management for massive machine type communications in 5G system," to appear in *IEEE Internet Things J.*, 2021.

[12] A. S. Alfa, *Queueing theory for telecommunications: Discrete time modelling of a single node system*, Springer. 2010.

[13] 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.

[14] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert w function," *Adv. Comput. Math.*, vol. 5, pp.329-359, 1996.