Received December 11, 2018, accepted December 14, 2018, date of publication December 19, 2018, date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2888566

Energy-Delay Evaluation and Optimization for NB-IoT PSM With Periodic Uplink Reporting

HILAL BELLO[1], XIN JIAN[1], (Member, IEEE), YIXIAO WEI[1], AND MIN CHEN[2] (Senior Member, IEEE)

[1]College of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China
[2]School of Computer Science, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Xin Jian (jianxin@cqu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61501065, Grant 61571069, Grant 61701054, Grant 61601067, and Grant 61771080, and in part by the Chongqing Basic Science and Advanced Technology Research under Grant cstc2016jcyjA0021.

ABSTRACT The narrowband Internet of Things (NB-IoT) is a new wireless protocol proposed by the 3rd Generation Partnership Project intending for low data rate IoT applications. The general objectives of the NB-IoT include supporting massive connections, enhanced coverage, reduced cost and complexity, ultra-low power consumption, and flexible delay characteristics. To lower energy consumption while providing reliable connections, extended discontinuous reception and power saving mode (PSM) mechanism are applied in the NB-IoT. To evaluate the energy consumption and delay performance under periodic uplink reporting, which is common among cellular IoT applications, this paper develops a semi-Markov chain with four states, namely, PSM, idle, random access (RACH), and transmission (Tx) states. RACH and Tx states are introduced from the well-known CONNECTED STATE to account for the extra power consumed due to increased access collisions under massive synchronous connections. Furthermore, an optimization model is introduced to find the best PSM duration, which is configured to minimize energy consumption and average delay according to user's preference. The numerical results show that setting higher limits for the number of possible RACH request transmission can make the user equipment (UE) more tolerant to delay and energy consumption in massively deployed concurrent communication UEs. Extending the PSM duration to longer period will cause excessive increase in delay without much impact on energy saving improvement.

INDEX TERMS Narrowband-Internet of Things, extended discontinuous reception, power saving mode, energy consumption, semi-Markov chain.

I. INTRODUCTION

The global Internet of things (IoT) market is expected to have a significant growth of over 25% over the next few years, enabling billions of devices, above ten times that of cellular, to be connected [1]. The IoT enables connections between devices and the internet with the aim of providing ubiquitous connections among different things to accomplish certain objectives [2]. This affects different aspects of life from personal, home and environmental devices [3], to industrial automation. The growth of IoT requires enhanced communication standards that support various use cases with divergent quality of service (QoS) requirements. Within IMT2020, the 3rd Generation Partnership Project (3GPP) classifies IoT applications into two scenarios with distinct use cases and service requirements, namely, massive machine type communications (mMTC) or mIoT [4] and ultra-reliable and low latency communications (uRLLC) [5]. mMTC has relaxed delay profile but a very large number of connected devices, such as smart metering and smart wearable, while uRLLC requires a strict delay profile where ms-level end-to-end latency and nearly 100% reliability need to be guaranteed, such as Internet of vehicles, industrial control, and emergency report [6], [7].

In general, mMTC comprises a wide range of applications ranging from smart cities with millions of sensors and long-term environmental monitoring demanding low energy consumption, to fully wireless factories that require high reliabilities and low latencies in their connection. During the 5G design, these requirements are considered in order to support the emerging and future applications. The mMTC is however, facing new challenges towards a unified radio solution in the various application areas.

Narrowband-Internet of things (NB-IoT) communication technology is a new 3GPP standard introduced to support connectivity of mMTC within the 5G era. The general objectives of NB-IoT includes supporting massive connectivity, enhanced coverage, reduced cost and complexity, ultra-low power consumption and flexible delay characteristics. Since most NB-IoT devices are battery powered and are massively deployed, replacement of battery is costly. Hence, it is essential that the energy consumption of is kept to the minimal for the effective implementation of the technology. To realize the ultra-low power consumption requirement, NB-IoT devices are required to be functional for a period of more than ten years with battery capacity of 5 Wh (Watt-hours), even in locations with adverse coverage conditions [8].

Taking this into account, extended discontinuous reception (eDRX) and power saving mode (PSM) mechanism are developed to realize the power efficiency requirements of NB-IoT devices [9]. eDRX prolongs the cycle length of traditional discontinuous reception (DRX) used by Long Term Evolution (LTE) and LTE-Advanced (LTE-A), which saves user equipment's (UE) power at the cost of increased delay [10], [11]. In practice, a tradeoff has to be made between delay and power saving during the eDRX parameters configuration. PSM is a novel low power deep sleep state, where the UE is still registered online but cannot be reached through signaling [12]. This keeps the UE in deep sleep for a longer time in order to achieve increased power saving.

With PSM and eDRX, an NB-IoT device sleeps most of the time and turn on its RF module only when data transfer occurs [13]. The eDRX process observes the physical downlink control channel (PHDCCH) discontinuously, to monitor if there is an uplink/downlink (UL/DL) communication task when UE is in radio resource control (RRC) connected state. The UE's RRC connection may be released by its serving base station if it has no information to be transmitted or received for a certain period of time and the UE will change to RRC idle state. PSM will be triggered as soon as the UE switch to RRC idle state in order to save energy. When UL/DL communication happens in the course of PSM, the UE will be woken up through paging procedure. Then RRC reconfiguration will retransform the UE into RRC connected state. The PSM process involves receiving deep-sleep status and paging messages periodically without receiving any other message after the periodic monitoring status for a period of time. The energy consumption in PSM is normally lower than other states. The user behavior as well as network traffic affects the timing of UE's state change. Thus, it is important that a practical paging period and eDRX configuration for different traffic types is selected, so that the transmission delay and the power saving ratio are balanced. The transmission delay in this context can be defined as the total time a packet has to wait at the BS while being buffered. The power saving ratio, however, is the ratio of the duration of sleep time to the total running time.

Some recent studies [14]–[17] have provided analytical model of DRX mechanism in LTE/LTE-A. While LTE/LTE-A supports two MS operation modes (RRC_Connected state and RRC_Idle state), NB-IoT supports three operation modes: connected state, idle state and PSM state. Therefore, the proposed models for LTE/LTE-A are inadequate to NB-IoT. Other studies [18], [19] have analyzed the energy consumption efficiency of NB-IoT based on the DL traffic. However, as pointed out in 3GPP, periodic uplink reporting is common for NB-IoT applications, such as smart water meter, smart environment and so on. The UEs periodically collect the node's status and sensed data, format the data uniformly and forward the data to the sink node. Moreover, possibility of collision is always high among massively deployed UEs in the periodic uplink reporting scenarios, which leads to recurrent data retransmissions and extra power consumption [14]. This brings about the need, thus, for studies to be extended to consider this kind of UL traffic.

Most research works evaluate the energy consumption of idle listening or overhearing, and pay little attention to collision in UL transmissions [20]. However, to account for the extra power consumed in UL reporting by increased access collisions under massive synchronous connections, in this paper, we introduce two sub-states under the CONNECTED STATE, namely; the random access (RACH) state and the transmission state. Together with the other two known states, namely, PSM and idle states, we formulate a semi-Markov model that can be used to estimate the energy consumption and delay for a periodic uplink reporting. From the developed performance models, we derive an optimization model that can be used to achieve a uniform dimension for the energy and delay as well as to attain an optimum energy/delay balance. The optimization model aims at evaluating the trade-off between energy saving and delay requirements based on user requirement for energy saving, or the preference for lesser traffic delay.

The rest of this paper is organized as follows. Section II introduces the four states semi-Markov model and highlights each of the states. Section III provides a comprehensive theoretical analysis of the transition conditions as well as the average energy consumption and access delay, which provides the basis for the objective function and formation of the optimization model in Section IV. Section V presents the numerical performance evaluation results of the above models. Section VI concludes the paper.

II. PROPOSED SEMI-MARKOV MODEL

Periodic uplink reporting considered in this paper is commonly triggered from the UE at a regular interval.

Usually, this type of communication has a constant data size and a regular time pattern [21]. In this section, we used the Semi-Markov model to design a statistical process that matches the behavior of the NB-IoT periodic uplink communication. To be more specific, a Semi-Markov chain model [22] with four states is proposed for NB-IoT periodic uplink traffic. The four states, namely, PSM, RACH, Tx and Idle states are shown in figure 1. The RACH and Tx states

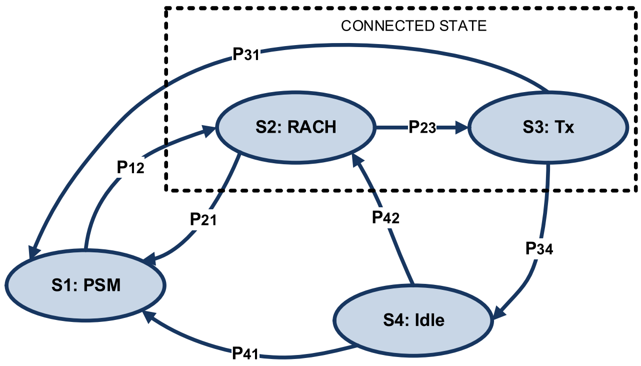

FIGURE 1. Proposed Semi-Markov model.

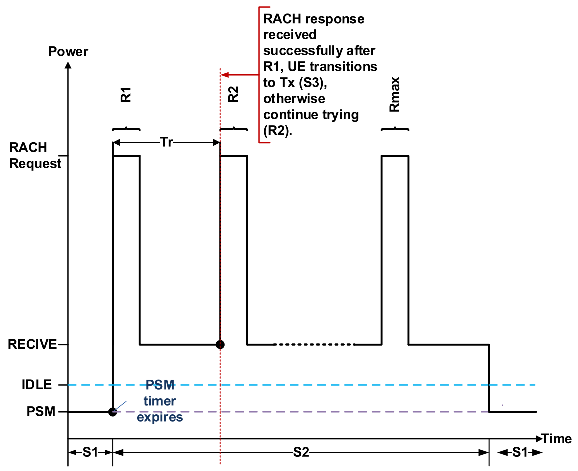

FIGURE 2. Illustration of RACH operation.

are each independent states introduced from the well-known CONNECTED STATE. This system is analyzed for energy consumption and average delay under one complete uplink data transfer in Section III.

Based on the proposed state diagram and semi-Markov chain model, the operation of a UE under periodic uplink reporting can be illustrated and analyzed as follows.

A. POWER SAVING MODE (PSM) STATE (S1)

During PSM state (*S1*), the UE is unreachable from network. It starts a PSM timer with duration T_{PSM} (which is the same as the period of uplink reporting) and keeps sleeping. When PSM timer expires, UE enters RACH state (*S2*) to connect network.

B. RACH STATE (S2)

Figure 2 illustrates the UE's operation during the RACH state (*S2*). During RACH state (*S2*), the UE transfers random access (RA) request to the base station (BS) periodically in T_r cycle. For each T_r cycle, UE transmits the RACH request and then monitors narrow band physical random access channel (NPRACH) to receive RACH response. Maximum number of transmissions of RACH request (T_r cycle) is R_{max}. If the UE receive RACH response and resource is allocated to it after Rth request ($0 < R \leq R_{max}$), it moves to Tx state (*S3*) for data transmission. For instance, as shown in figure 2, the UE will transition to *S3* if the it successfully, receives the RACH response at the first RACH request. Otherwise, if the UE fails to connect to network (receive RACH response) within the period of T_r after R_{max}th request, this suggests that the current condition is not suitable for the communication, and hence, it returns to PSM state (*S1*).

C. TRANSMISSION (TX) STATE (S3)

During Tx state (*S3*), UE transfers data to BS periodically in T_{ACK} cycle. For each of the T_{ACK} cycle, the UE monitors Narrow band physical downlink shared channel (NPDSCH) channel to receive response. Maximum number of data transmissions is N_{max}. If UE receive acknowledgement (ACK) from BS after Nth transmission ($0 < N \leq N_{max}$), it moves to PSM state (*S1*). Otherwise, if the UE still fail to receive ACK in period of T_{ACK} after N_{max}th transmission, it moves to Idle state (*S4*). The Tx operation is illustrated in Figure 3.

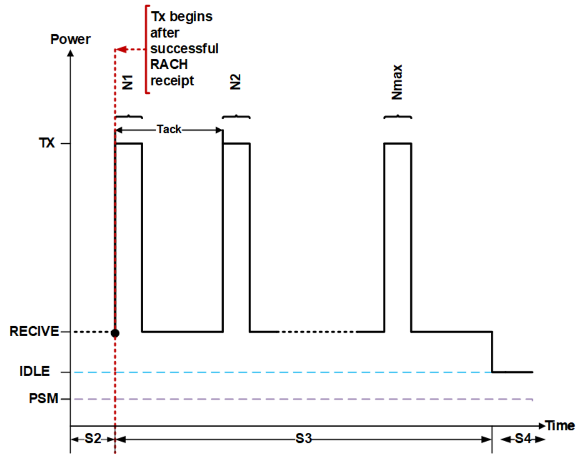

FIGURE 3. Illustration of Tx operation.

D. IDLE STATE (S4)

In Idle state (*S4*), UE releases its allocated resource. It starts an idle timer with duration T_{IDLE} and keep monitoring NPDSCH to wait for respond. If UE have received ACK in the period of T_{IDLE}, it moves to PSM state (*S1*). Otherwise, UE moves to RACH state (*S2*).

III. ENERGY COMSUMPTION AND DELAY ANALYSIS

In this section, we derive the transition and steady state probability of the various states of the proposed semi-Markov model, with which we analytically determine the energy consumption and average delay of the process.

To simplify the analysis, it was assumed that:

1) The concurrent number of UEs and the current transmission environment remains unchanged;
2) The RA request transmission and resource allocation in *S2* is affected by the quality of transmission

environment and the network traffic load given as the ratio between the number of active users and available resources or preambles; while the data transmission in *S3* is only affected by the quality of transmission environment.

3) The inter-arrival time distribution of ACK in *S4* obeys the exponential distribution with parameter λ; thus to some extent $e^{-\lambda}$ can be used to represent the quality of current transmission environment in *S2* and *S3*.

A. STATIONARY PROBABILITY

Denote P_{ij} as the transition probability from S_i to S_j of the semi-Markov chain in Figure 1.

In *S1*, UE transitions to *S2* at the expiry of T_{PSM}. Therefore, $P_{12} = 1$.

In *S2*, denote $p_{r,i}$ as the probability that UE is allocated resource successfully after *i*th request. We assume $p_{r,1} = p_{r,2} = \ldots = p_{r,Rmax} = e^{-\lambda} \cdot e^{-m/n}$, in which m and n represent the concurrent number of UEs and total number of RACH resources respectively. Therefore, P_{21} can be derived as;

$$P_{21} = \prod_{i=1}^{R_{\max}} (1 - P_{r,i}) = (1 - e^{-\lambda - m/n})^{R_{\max}} \tag{1}$$

In *S3*, the probability that UE receives ACK after *i*th data sending trial is denoted as $p_{t,i}$; and $p_{t,1} = p_{t,2} = \ldots = p_{t,Rmax} = e^{-m/n}$. Therefore, P_{31} can be derived as;

$$P_{31} = \sum_{i=1}^{R_{\max}} (1 - e^{-\lambda})^i . e^{-\lambda} = 1 - (1 - e^{-\lambda})^{N_{\max}} \tag{2}$$

In *S4*, P_{41} can be calculated as;

$$P_{41} = \int_0^{T_{IDLE}} \lambda e^{-\lambda t} \mathrm{d}t = 1 - e^{-\lambda T_{IDLE}} \tag{3}$$

From equation (3) and above, the transition probability matrix of the proposed semi-Markov chain model can be given as:

$$P = \begin{bmatrix} 0 & P_{12} & 0 & 0 \\ P_{21} & 0 & 1 - P_{21} & 0 \\ P_{31} & 0 & 0 & 1 - P_{31} \\ P_{41} & 1 - P_{41} & 0 & 0 \end{bmatrix} \tag{4}$$

Now, let $q_i \forall i \in \{1, 2, 3, 4\}$ represent the steady-state probability of state S_i. By using $\sum_i q_i = 1$ and $q_i = \sum_j q_j P_{ji}$, the expressions for steady-state probability can be given as

$$\begin{cases} q_1 = \dfrac{P_{21}+P_{31}+P_{41}-P_{21}P_{31}-P_{21}P_{41}-P_{31}P_{41}+P_{21}P_{31}P_{41}}{3 - P_{21} + P_{41} - P_{21}P_{41} - P_{31}P_{41} + P_{21}P_{31}P_{41}} \\ q_2 = \dfrac{1}{3 - P_{21} + P_{41} - P_{21}P_{41} - P_{31}P_{41} + P_{21}P_{31}P_{41}} \\ q_3 = \dfrac{1 - P_{21}}{3 - P_{21} + P_{41} - P_{21}P_{41} - P_{31}P_{41} + P_{21}P_{31}P_{41}} \\ q_4 = \dfrac{(P_{21} - 1)(P_{31} - 1)}{3 - P_{21} + P_{41} - P_{21}P_{41} - P_{31}P_{41} + P_{21}P_{31}P_{41}} \end{cases} \tag{5}$$

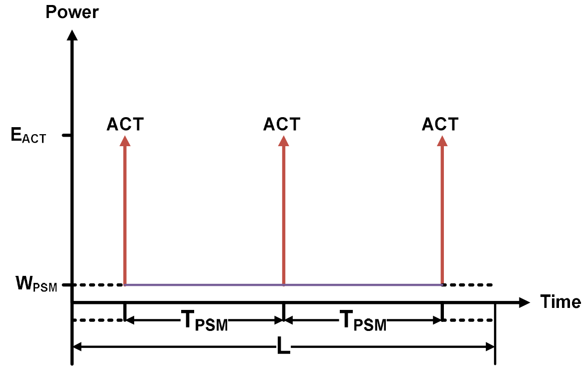

FIGURE 4. Illustration of operation over a period L.

B. POWER CONSUMPTION ANALYSIS

Let L represent a specific duration for which transmission has occurred. The total energy consumption $E(L)$ for the period L, as shown in figure 4, is calculated by finding the total energy consumption during the active period $E_{ACT}(L)$ and during the PSM period $E_{PSM}(L)$ respectively, so that;

$$E(L) = E_{ACT}(L) + E_{PSM}(L) \tag{6}$$

To find $E_{ACT}(L)$; Let $E_{ACT}(i)$, represent the power consumption during the active period of one transmission for state S_i, for $i \in \{1, 2, 3, 4\}$. The total power consumption during the active period of one transmission E_{ACT} can be calculated as;

$$\begin{aligned} E_{ACT} &= \sum_{i=1}^{4} q_i E_{ACT}(i) \\ &= q_1 E_{ACT}(1) + q_2 E_{ACT}(2) + q_3 E_{ACT}(3) + q_4 E_{ACT}(4) \end{aligned} \tag{7}$$

The total energy consumption $E_{ACT}(i)$ for each state S_i, for $i \in \{1, 2, 3, 4\}$ are calculated as follows.

For *S1*,

$$\begin{aligned} E_{ACT}(1) &= W_1 * T_1 \\ &= W_{PSM} * 0 = 0 \end{aligned} \tag{8}$$

where $W_1 = W_{PSM}$ represents the average power during the PSM. T_1 is the elapsed PSM time during the transmission, which is 0.

For *S2*,

$$\begin{aligned} E_{ACT}(2) &= E_{RACH} * \overline{R} \\ &= E_{RACH} \cdot \sum_{i=1}^{R\max} i \cdot \left(1 - p_{r,i}\right)^i \end{aligned} \tag{9}$$

where E_{RACH}, represents the energy consumption of each RACH process and $\overline{R}$ represents average failure number of random access.

For *S3*,

$$\begin{aligned} E_{ACT}(3) &= E_{TR} * \overline{N} \\ &= E_{TR} \cdot \sum_{i=1}^{N\max} i \cdot \left(1 - p_{t,i}\right)^i \end{aligned} \tag{10}$$

where E_{TR}, represents the energy consumption of each transmission (*S3*) process and $\overline{N}$ represents the average number of times the data sent.

For *S4*,

$$E_{ACT}(4) = W_4 * T_{IDLE} \tag{11}$$

where W_4 represents the average power during the Idle period (*S4*) in one transmission. T_{IDLE} is the elapsed active time of Idle state (*S4*).

During the period L, since there are L/T_{PSM} times ACT, therefore,

$$E_{ACT}(L) = \frac{L}{T_{PSM}}.E_{ACT} \tag{12}$$

and the average power consumption during PSM E_{PSM} can be calculated as

$$E_{PSM}(L) = W_{PSM} \cdot L \tag{13}$$

From Equation (12) and (13), the energy consumption E in the period L can be calculated as

$$\begin{aligned} E(L) &= E_{ACT}(L) + E_{PSM}(L) \\ &= L\left(\frac{E_{ACT}}{T_{PSM}} + W_{PSM}\right) \end{aligned} \tag{14}$$

C. DELAY ANALYSIS

Let $D_{ACT}(i)$, $i \in \{1, 2, 3, 4\}$ represent the holding time of state S_i. The average delay during a transmission D_{ACT} can be calculated as;

$$\begin{aligned} D_{ACT} &= \sum_{i=1}^{4} q_i D_{ACT}(i) \\ &= q_1 D_{wake}(1) + q_2 D_{wake}(2) \\ &\quad + q_3 D_{wake}(3) + q_4 D_{wake}(4) \end{aligned} \tag{15}$$

whereby, the individual hold time for each of the states *S1* to *S4* is given by;

$$D_{ACT}(1) = 0 \tag{16}$$

$$D_{ACT}(2) = T_r \cdot \overline{R} \tag{17}$$

$$D_{ACT}(3) = T_{ACK} \cdot \overline{N} \tag{18}$$

$$\begin{aligned} D_{ACT}(4) &= \int_0^{T_{IDLE}} t \cdot \lambda e^{-\lambda} \cdot dt + \int_{T_{IDLE}}^{\infty} T_{IDLE} \cdot \lambda e^{-\lambda} \\ &= \frac{1 - e^{-\lambda \cdot T_{IDLE}}}{\lambda} \end{aligned} \tag{19}$$

However, for each successful transmission, $^1/_{P_{SUC}}$ times of data transmission are required; and for a successful PSM, $\left(^1/_{P_{SUC}} - 1\right)$ times of data transmission is required; where the P_{SUC} is the probability of successful transmission, and is given by;

$$P_{SUC} = 1 - P_{12}P_{21} - P_{12}P_{23}P_{34}P_{42}P_{21} \tag{20}$$

Therefore, the average delay D can be calculated as;

$$D = \frac{1}{P_{SUC}}.D_{ACT} + \left(\frac{1}{P_{SUC}} - 1\right).T_{PSM} \tag{21}$$

IV. OPTIMIZATION MODEL

Energy-saving mechanisms using PSM usually leads to increased latency. It is required that the trade-off between energy saving and delay should be balanced. An optimization model of the PSM parameters that offers maximized energy saving and simultaneously minimizing the communication delay is presented in this section. We used the priori method [23] in a multi-objective optimization technique to formulate a problem that can simultaneously solve the two parameters i.e. delay and energy saving. The priori weight δ is introduced in the function in order to predetermine a bias in the function towards the user's preference for delay or energy saving.

Taking into account the trade-off [14] between power consumption and delay, the following optimization model is presented:

$$\min \delta \frac{E}{E_{MAX}} + (1 - \delta) \frac{D}{D_{MAX}} \tag{22}$$

$$s.t. \begin{cases} E \le E_{MAX} \\ D \le D_{MAX} \\ N_{MAX} \ge 1 \\ R_{MAX} \ge 1 \\ T_{PSM} > 0 \\ T_R > 0 \\ T_{ACK} > 0 \\ T_{IDLE} > 0 \end{cases} \tag{23}$$

where E_{MAX} represents the maximum energy consumption in the period L and D_{MAX} is the average delay of a successful transmission that the operator can accept. N_{MAX} and R_{MAX} are respectively the number of transmission and the number of RACH request sent for each transmission circle. Their values are integers greater than or equal to 1.

The objective will be to obtain the best PSM duration T_{PSM}, so that the optimal T_{PSM} value can be configured in the UE to minimize the energy consumption E or the communication delay, depending on the requirement.

The weight $\delta \in (0 \le \delta \le 1)$ represent the weight of priority for higher power saving over PSM induced delay or lower delay at the expense of low power saving. $\delta = 0$ signifies a minimized delay and maximized power consumption settings. When $\delta = 1$, it signifies a maximized delay settings.

V. PERFORMANCE EVALUATION

In this section, the proposed four state Markov chain model for periodic uplink traffic scheme is validated and evaluated by simulations. We then compare the behavior of energy consumption (E) and average delay (D) curves in terms of different PSM duration (T_{PSM}). From the evaluation results, we obtain the optimum TPSM values achieved from the best configuration parameter values.

We use MATLAB to obtain the numerical results based on the proposed model. The valid 3GPP NB-IoT [2] parameters are used in the setup. Table 1 shows the details of simulation

TABLE 1. Simulation parameters.

Parameter	*Value*
L	*24h*
$1/\lambda$	*10 ms/packet*
T_r, T_{ACK}	*8ms*
T_{IDLE}	*1000ms*
W_1	*0.05 mW*
W_4	*100 mW*
E_{RACH}, E_{TR}	*2000 mW·ms*

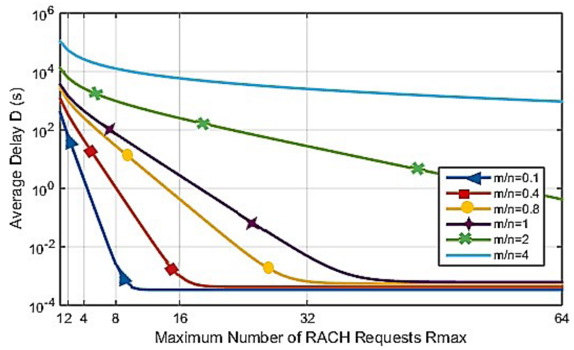

FIGURE 5. Average delay under different maximum number of RACH requests.

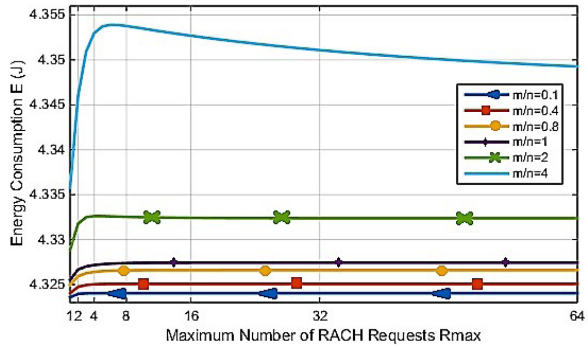

FIGURE 6. Energy consumption under different maximum number of RACH requests.

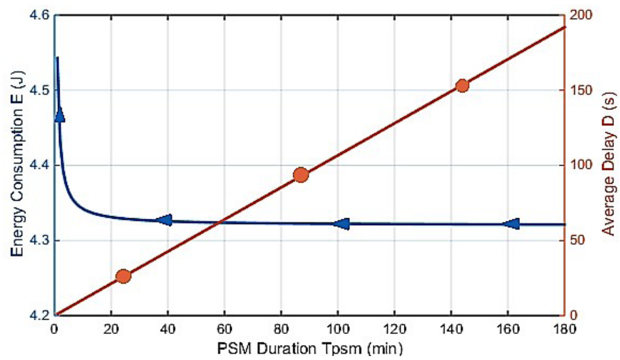

FIGURE 7. Energy consumption and average delay under different PSM duration.

parameters. A reference simulation period (L) of 24 hours is used so that the scheme is evaluated under a complete day period.

Figures 5 and 6 show the graphical representation of E and D relationship with respect to different maximum number of RACH requests (R_{max}). The results are obtained for 6 different network traffic load given as the ratio (m/n) of the number of existing UEs (m) to the number of RACH resources (n), each representing the different graph curves.

In Figure 5, D reduces initially as R_{max} increases. The decline in the value of D is steeper in the scenarios where the network traffic load (m/n) is deteriorated (i.e. higher m/n ratio). When R_{max} reaches a certain point, D stops decreasing and maintains a stable value. The decline in D usually happens alongside increase in energy consumption E. The energy consumption E increases as the number of R_{max} increases as shown in figure 6. The value of E then remains relatively stable after R_{max} reaches a certain point. When the network traffic load is within a conducive range i.e. the number of concurrent active users is not more than the accessed resources ($m/n \leq 1$), E barely changes with R_{max}. However, as the R_{max} is increased in the scenario where the network traffic load is high ($m/n \geq 1$), there is a significant rise in E at the region of lower R_{max} values. This shows that, lower R_{max} values has significant effect in network with congested traffic load. Therefore, setting larger R_{max} values will help improve the UE's tolerability in scenario having active UEs higher than resources.

The effect of maximum number of retransmissions (N_{max}) on E and D is similar to that of R_{max}. Therefore, increasing N_{max} improves the UE's resistance and tolerance to high traffic load network. Since the slopes for E and D with respect to N_{max} shows similar trend with R_{max}, we therefore choose not to repeat the illustration in this section.

Figure 7 shows the variation of energy consumption E and average delay D according to different PSM duration (T_{PSM}). This result is obtained for a normal traffic load of m/n=1. As the T_{PSM} duration is increased from 1 to 180 minutes, D increases while E decreases. The increase in D is linear from beginning of the graph, up to the maximum T_{PSM} point. However, E decreases sharply at shorter T_{PSM} region at the beginning of the graph up to a certain point, where the graph remains almost stable afterwards. This implies that, from a certain point, the effect of increasing T_{PSM} on declining energy consumption is much less than on increasing delay. Therefore, longer T_{PSM} values will only increase the communication delay without much impact on reducing the energy consumption. This further explains why the T_{PSM} is very critical and hence, the need to ensure the best duration is selected.

VI. OPTIMAL TPSM CONFIGURATION

Limiting communication delay is usually at the cost of increased energy consumption. This necessitates that ideal

TABLE 2. m/n=2.0,L=24h,E_{MAX}=5J,D_{MAX}=1000ms, R_{max}=N_{max}=32.

δ	0	0.25	0.5	0.75	1
T_{PSM}^{*} (ms)	1030	2148	3720	6444	98818
E^{*} (J)	4.999	4.646	4.508	4.428	4.327
D^{*} (ms)	10	22	37	65	1000

values are set for the related parameters in order to weigh between the energy consumption and delay effectively. According to Equations (14) and (21), T_{PSM} is the biggest factor that affects both energy consumption and average delay. In this section, we obtain the optimal T_{PSM} using a conventional configuration setup on our optimization model based on the power saving or delay priority.

The optimized PSM duration is obtained by Equation (22). The network traffic load is set as *2* in order to have the number of UEs to be double the number of resources. The value m/n=2.0 is used taking into account the massive number of UEs commonly present in mMTC network [24], which is considered in this work. The reference duration, L, is selected for one day, taken as L=24 hours. Based on the MTC devices power consumption requirement, battery capacity of 5Wh should have battery life of up to 10 years [8]. Therefore, the maximum power consumption E_{MAX}, for the 24 hours' (1 day) reference period is set as;

$$\frac{(5watts * 3600\sec onds)}{10years * 365days} = 5J \tag{24}$$

The appropriate uplink delay requirement for NB-IoT can be 1 second [25], [26], therefore the maximum delay value D_{MAX} is set as 1000ms. R_{max} and N_{max} values are selected based on the simulation result. Since R_{max} and N_{max} values are always incremented in the order of 2^{n} [27], from figure 5 and 6, 32 is selected, being the closest and lowest point where both delay and energy consumption are relatively even. The optimization results are listed in table 2.

Table 2 represents an example of T_{PSM} configuration optimization using equation (22) for different energy consumption or delay requirement. As described alongside the optimization equation in section III, the weight $\delta \in (0 \leq \delta \leq 1)$ determines the priority of energy saving or PSM induced delay. Lower δ values always bias the result towards lower delay at the expense of higher energy consumption while higher δ value tend to produce a bias towards lower energy consumption at the cost of higher delay. It is apparent from table 2 that, in condition of strict delay limitation, i.e. lower δ value, low T_{PSM} value is required to be configured in order to obtain minimal delay with an optimal higher energy consumption. When the condition requires higher power saving, the UE must use higher T_{PSM} value so that UE stays longer in the power saving mode. Although the optimal T_{PSM} values produced the E and D values in table 2, however, beyond certain T_{PSM} *(6444ms)*, the delay value skyrocket from *65ms* to *1000ms* while the corresponding energy consumption only reduced by *0.101J*, showing the superiority of lower T_{PSM} in attaining an ideal Energy-Delay balance.

VII. CONCLUSION

In this paper, we have presented a semi-Markov chain model to evaluate the NB-IoT energy consumption and delay for periodic uplink traffic. PSM mechanisms is used to introduce an optimization method that maintains a tradeoff between energy consumption and delay. Considering the additional energy consumption due to increased collision as a result of the massive concurrently active users in the mMTC network, the "Connected state" is further divided into two states, namely; "RACH state" and "Tx state" to develop a semi Markov model. The four state Markov chain is modeled and analyzed with PSM state, RACH state, Tx state and Idle state as the state variables. Based on the developed model, the computation method of energy consumption and delay in periodic uplink reporting is given.

The obtained simulation results indicated that, reasonable increase of maximum possible number of RACH request transmissions can increase user's energy consumption and delay tolerance in the situation of excessive number of concurrently active users in mMTC. Also from the simulation results, it is demonstrated that, longer T_{PSM} values will only cause increase in the average delay without much impact on the energy saving. An example of optimization process is subsequently developed to generate the ideal PSM duration T_{PSM}, which resulted in optimal energy consumption and average delay according to the user's preference.

REFERENCES

[1] (Nov. 2015). *Ericsson Mobility Report*. Accessed: Aug. 30, 2017. [Online]. Available: https://www.ericsson.com/assets/local/mobility-report/documents/2015/ericsson-mobility-report-june-2015.pdf

[2] A. Whitmore, A. Agarwal, and L. D. Xu, "The Internet of Things—A survey of topics and trends," *Inf. Syst. Frontiers*, vol. 17, no. 2, pp. 261–274, Apr. 2015.

[3] R. Nordin *et al.*, "The world-first deployment of narrowband IoT for rural hydrological monitoring in UNESCO biosphere environment," in *Proc. ICSIMA*, Putrajaya, Malaysia, Nov. 2017, pp. 1–5.

[4] X. Jian, Y. Liu, Y. Wei, X. Zeng, and X. Tan, "Random access delay distribution of multichannel slotted ALOHA with its applications for machine type communications," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 21–28, Feb. 2017.

[5] P. Popovski *et al.*, *D6.6 Final Report on the METIS 5G System Concept and Technology Roadmap*, METIS Version, document ICT-317669-METIS/D6.6, vol. 1, Apr. 2015. [Online]. Available: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D6.6_v1.pdf

[6] *Feasibility Study on New Services and Markets Technology Enablers for Critical Communications*, document 3GPP TR 22.862 V14.1.0, Sep. 2016.

[7] Beijing, China. *IMT2020 White Paper on 5G Wireless Technology Architecture, IMT-2020 (5G) Promotion Group*. Accessed: May 2015. [Online]. Available: www.imt-2020.org.cn/zh/documents/download/25

[8] M. Chen, Y. Miao, X. Jian, X. Wang, and I. Humar, "Cognitive-LPWAN: Towards intelligent wireless services in hybrid low power wide area networks," *IEEE Trans. Green Commun. Netw.*, to be published.

[9] *Non-Access-Stratum (NAS) Protocol for Evolved Packet System (EPS)*, Tech. Specification Group Core Netw. Terminals, document 3GPP TS 24.301, V14.3.0 Release 14, Mar. 2017.

[10] *NarrowBand IOT*, document 3GPP TSG-RAN WG2 Meeting 92, Anaheim, CA, USA, Nov. 2015.

[11] *LTE-M–Optimizing LTE for the Internet of Things*, Nokia Solution Netw., Espoo, Finland, May 2015.
[12] M. Chen, Y. Miao, Y. Hao, and K. Hwang, "Narrow band Internet of Things," *IEEE Access*, vol. 5, pp. 20557–20577, 2017.
[13] *Evolved Universal Terrestrial Radio Access (E-UTRA); NB-IOT; Technical Report for BS and UE Radio Transmission and Reception*, 3GPP TR 36.802, V.13.0.0, Jun. 2016.
[14] A. T. Koc, S. C. Jha, R. Vannithamby, and M. Torlak, "Device power saving and latency optimization in LTE-A networks through DRX configuration," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2614–2625, May 2014.
[15] K. Zhou, N. Nikaein, and T. Spyropoulos, "LTE/LTE-A discontinuous reception modeling for machine type communications," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 102–105, Feb. 2013.
[16] S. Jin and D. Qiao, "Numerical analysis of the power saving in 3GPP LTE advanced wireless networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, pp. 1779–1785, May 2012.
[17] H. Ramazanali and A. Vinel, "Performance evaluation of LTE/LTE-A DRX: A Markovian approach," *IEEE Internet Things J.*, vol. 3, no. 3, pp. 386–397, Jun. 2016.
[18] S. Xu, Y. Liu, and W. Zhang, "Grouping-based discontinuous reception for massive narrowband Internet of Things systems," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1561–1571, Jun. 2018.
[19] S.-M. Oh, K.-R. Jung, M. S. Bae, and J. Shin, "Performance analysis for the battery consumption of the 3GPP NB-IoT device," in *Proc. ICTC*, Oct. 2017, pp. 981–983.
[20] R. Zhang, H. Moungla, J. Yu, and A. Mehaoua, "Medium access for concurrent traffic in wireless body area networks: Protocol design and analysis," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2586–2599, Mar. 2017.
[21] M. Laner, N. Nikaein, P. Svoboda, M. Popovic, D. Drajic, and S. Krco, "Traffic models for machine-to-machine (M2M) communications: Types and applications," in *Machine-to-machine (M2M) Communications: Architecture, Performance and Applications*, 1st ed. Amsterdam, The Netherlands: Elsevier, Jan. 2015.
[22] M. Kijima, *Markov Processes for Stochastic Modeling*. London, U.K.: Chapman and Hall, 2013.
[23] R. V. N. Melnik, "Dynamic system evolution and Markov chain approximation," *Discrete Dyn. Nature Soc.*, vol. 2, no. 1, pp. 7–39, Sep. 1998.
[24] C. Bockelmann *et al.*, "Towards massive connectivity support for scalable mMTC communications in 5G networks," *IEEE Access*, vol. 6, pp. 28969–28992, 2018.
[25] *Cellular System Support for Ultra-Low Complexity and Low Throughput Internet of Things (CIoT)*, document 3GPP TR 45.820, V.13.1.0, Nov. 2015.
[26] M. Chen, Y. Qian, Y. Hao, Y. Li, and J. Song, "Data-driven computing and caching in 5G networks: Architecture and delay analysis," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 70–75, Feb. 2018.
[27] X. Yunzhou, "NB-IoT Key technology," in *Technology Application For Narrow Band Internet Of Things*. Beijing, China: CSPM, 2017, pp. 36–39.

HILAL BELLO received the B.E. degree in electrical engineering from Bayero University, Kano, Nigeria, in 2013, and the M.E. degree in communication and computer engineering from Universiti Kebangsaan Malaysia (National University of Malaysia), Bangi, Malaysia, in 2016. He is currently pursuing the Ph.D. degree with the College of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China. He is also a Lecturer with the Faculty of Engineering, Usmanu Danfodiyo University, Sokoto, Nigeria. His research interests include power saving techniques in wireless communication.

XIN JIAN (M'14) received the B.E. and Ph.D. degrees from Chongqing University, Chongqing, China, in 2009 and 2014, respectively. He is currently an Associate Professor with the College of Microelectronics and Communication Engineering, Chongqing University. His research interests include the next-generation mobile communication, massive machine-type communications, and narrow band Internet of Things.

YIXIAO WEI received the B.E. degree in communication engineering from Jilin University, Jilin, China, in 2016. She is currently pursuing the master's degree with the College of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China. Her current research interests include ultra-low power consumption techniques for narrowband Internet of Things.

MIN CHEN (SM'09) has been a Full Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology, since 2012. His research interests include cyber physical systems, IoT sensing, 5G networks, SDN, and healthcare big data. He received the IEEE Communications Society Fred W. Ellersick Prize, in 2017. His Google Scholars Citations reached over 13 500 with an h-index of 58. He is the Chair of the IEEE Computer Society STC on Big Data.