Optimizing the LTE Discontinuous Reception Mechanism Under Self-Similar Traffic

Ke Wang, Xi Li, Hong Ji, *Senior Member, IEEE*, and Xiaojiang Du, *Senior Member, IEEE*

***Abstract*—The discontinuous reception (DRX) mechanism has been adopted in a Long-Term Evolution (LTE) system as a core technology to prolong the user equipment's (UE's) battery lifetime. With the development of Mobile Internet, there is an increasingly urgent need to optimize the DRX performance to accommodate emerging applications. To describe the self-similarity exhibited by the applications, the truncated-Pareto distributed arrival traffic model is introduced into the LTE-DRX analytical framework. Following this premise, an online power-saving strategy (OPSS) has been designed on the basis of our previously proposed DRX analytical model. The OPSS, for the purpose of improving the UE's energy efficiency, is composed of an estimation phase and an optimization phase. In the first phase, several derived statistical estimators are deployed to capture the fluctuations of traffic conditions and DRX operations. It is shown that these estimators could estimate the statistics within just 1 s. In the second phase, the DRX configuration is optimized by considering the tradeoff between packet delay and power savings. Through simulations, we study the impact of various traffic conditions and the efficiency of OPSS. The results have demonstrated that the proposed OPSS could outperform the conventional LTE DRX mechanism in terms of both energy conservation and packet delay.**

***Index Terms*—Discontinuous reception (DRX), Long-Term Evolution (LTE), multimedia, power saving, self-similar traffic.**

NOMENCLATURE

t_I	Threshold of the inactive timer.
t_{SD}	Length of short discontinuous reception (DRX) cycle.
t_{LD}	Length of long DRX cycle.
t_{ON}	Threshold of On-duration timer.
N_{SD}	Maximum number of short DRX cycle.
α	Distribution shape parameter of truncated-Pareto distribution.
x_m	Location parameter of truncated-Pareto distribution.
m	Truncated value of truncated-Pareto distribution.
h_i	Sojourn time distribution of state i.
U_i	Mean sojourn time in state i.
t_{busy}	Length of packets' continuous arriving during **Active period**.
f_{ij}	Conditional distribution of the sojourn time in state j, given that the previous state transition is from state i.
q_{ij}	Discrete-time semi-Markov kernel of state j, given that the previous state transition is from state i.
PS	Power-saving factor.
D_s	Expected wake-up delay in **Light sleep period**.
D_l	Expected wake-up delay in **Deep sleep period**.
$\mathbb{E}(D)$	Expected wake-up delay.
$\hat{X}_m$	Unbiased estimator of x_m.
α^*	Unbiased estimator of α.
$\hat{U}_i(M)$	Estimator of U_i.
$\hat{\nu}(i)$	Estimator of ν_i.

Manuscript received December 19, 2014; accepted December 20, 2014. Date of publication December 29, 2014; date of current version December 14, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61271182 and Grant 61302080 and in part by the National Science Foundation under Grant CNS-1065444. The review of this paper was coordinated by Prof. Y. Fang.

K. Wang, X. Li, and H. Ji are with the Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: wangke@bupt.edu.cn; lixi@bupt.edu.cn; jihong@bupt.edu.cn).

X. Du is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: dxj@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2386352

I. INTRODUCTION

WITH the rapid development of smart mobile devices, wireless communication technologies and various Mobile Internet applications such as multimedia services have emerged and gained huge popularity. This calls for larger system capacity, higher transmission rate, and better user experience, to drive the evolution of new technologies such as high-order modulation, beamforming, channel equalization, and multiple-input–multiple-output antennas. Although these physical-layer techniques could significantly improve spectrum efficiency, they also increase the complexity of receivers' computational circuitry and thus consume more user equipment (UE) battery power. Substantial improvements in power-saving mechanisms are necessary for UE design [1]. For this purpose, the 3GPP Long-Term Evolution (LTE) standard has specified new discontinuous reception (DRX) based on the Universal Mobile Telecommunications System (UMTS) DRX mechanism, namely, LTE-DRX. It is worth noting that another important standard WiMAX has also introduced an improved Sleep/Idle transmission mechanism to prolong the UE's battery lifetime.

The basic idea of DRX is to allow a UE device to turn off its wireless transceiver for the sake of saving unnecessary energy consumption. In LTE, the evolved NodeB (eNB) controls the UE's DRX operation by radio resource control (RRC) entity via the physical downlink control channel (PDCCH). With DRX, the UE is allowed to discontinuously monitor the PDCCH on which the downlink transmission grants were assigned. LTE defines two modes of DRX, depending on whether the UE has an active session, rather than defining power-saving classes corresponding to different kinds of traffic in WiMAX [2]. When there are no active sessions, LTE DRX works in `RRC_Idle` mode, and the UEs perform similarly as other DRX mechanisms [3].

During the active sessions, the RRC_Connected-mode LTE DRX is activated. Different from other DRXs, LTE DRX in RRC_Connected mode allows the UE to enter sleep while it is registered with eNB. According to the Nokia Siemens Networks technical report [4], LTE DRX is one of the most effective power-saving mechanisms and is already supported by the latest LTE phones.

However, the present LTE DRX still has a lot of room for improvement. The LTE DRX mechanism specified in 3GPP Release 8 [5] is based on static sleep mode, which leads to inevitable performance degradation. To provide a *universal and easy-to-implement* method, it abandons the mechanism that is based on traffic types (used by WiMAX [2]) and adopts a static sleep mode that is expected to meet most of the traffic quality-of-service requirements. Nevertheless, with the development of Mobile Internet, a *universal* method that fits all traffic does not exist. It is necessary to consider different applications in LTE DRX. Recently, in 3GPP Release 11, work item "RAN Enhancements of Diverse Data Applications" [6] has been conducted to study battery drainage due to emerging Mobile Internet applications, which also confirmed the aforementioned observation. Therefore, how to optimize LTE DRX based on emerging Mobile Internet applications could significantly influence the energy efficiency of LTE systems.

The main approach to existing works on optimizing LTE DRX is to adjust the *DRX parameters*. The 3GPP LTE standard has specified several DRX parameters to precisely control the UE's behavior, such as when to sleep, when to monitor the PDCCH, when to activate, and how long each state will take. In fact, the early discussion on LTE DRX parameters described in 3GPP technical reports [7], [8] had already shown that 1) flexible configurable DRX parameters could obtain better performance than fixed parameters, and 2) different traffic conditions could influence the LTE DRX performance. Therefore, the DRX parameters should be adaptively optimized with changing traffic conditions.

Over the past few years, this area has attracted intensive attention. Related research can be divided into two directions: optimization based on both concrete traffic types [7], [9], [10] and general traffic analytical models [3], [11], [12], respectively. Studies in the first direction are always with the aid of extensive simulations and realistic measurements. They focused on traffic with specific operating patterns, such as Voice over Internet Protocol [9], [10], Hypertext Markup Language [7], [13], and Hypertext Transfer Protocol [14]. In these works, the tradeoff between power saving and delay is achieved by adjusting the DRX parameters. Meanwhile, the best DRX parameter configurations for different arrival rates were also given. They have the advantages of precisely fitting the specific traffic and are practical to deploy. The drawback is also obvious: The scope of the solutions is narrow, and they cannot deal with various Mobile Internet applications simultaneously running on UEs. Hence, the second direction attracts more attention these days.

Generally speaking, the efforts in the second direction focus on DRX analytical modeling and analytical-model-based optimization algorithms. The DRX analytical model is the foundation of optimization. All of the existing LTE DRX models [11], [12], [15] follow the method and assumptions in [3], in which Yang *et al.* provided an analytical model of UMTS DRX for the first time. All these models assume that the packet arrival intervals and transmission times follow exponential and general distributions, respectively. Furthermore, they all adopted the semi-Markov chain as the main mathematical tool, and most of the statistics were derived by utilizing the memoryless property of exponential distribution. These models can predict power savings performance and expected delay of LTE DRX under nonrealtime traffic. Most existing analytical studies for optimizing traffic-based power-saving mechanisms were conducted in the WiMAX system. In [16], Yu and Feng applied the sleep cycle optimization method of WiMAX in [17] to LTE and adopted the LTE DRX model proposed in [11]. There are two common drawbacks of existing works in the second direction: First, as shown in [18] and [19], the Poisson traffic model cannot accurately reflect the multimedia/mixed traffic in Mobile Internet; second, the traffic estimation algorithm is required to observe and record traffic conditions for a long time, tending to consume much more of eNB's resources.

The complexity of multimedia traffic is a natural consequence of integrating a diverse range of traffic resources such as video, voice, and data, which significantly differed in their traffic patterns and requirements [19]. In [20] and [21], it is demonstrated that broadband multimedia services, as well as Internet traffic, exhibit properties of self-similarity and long-range dependence (LRD). In general, self-similar traffic shows identical statistical characteristics over a wide range of time scales, which may have a significant impact on network performance. With the emergence of Mobile Internet, we believe that considering the self-similar traffic in DRX studies is necessary.

Following this idea, we introduced the self-similar traffic model into DRX modeling and constructed a complete analytical model in our recently published work [22]. In our model, the self-similarity of traffic is described by a truncated-Pareto distributed arrival process, and the LTE DRX is modeled based on the discrete-time semi-Markov process (DTSMP). It is shown that our model is more suitable for analyzing the LTE DRX's performance under self-similar traffic, compared with existing Poisson-based DRX models. Furthermore, the energy efficiency and the delay of LTE DRX under a fixed configuration could be precisely evaluated by using our model. However, how to deploy it in practice has still not been discussed. Therefore, in this paper, we extend our previous works by developing the method of optimizing the LTE DRX.

In this paper, based on the previously established analytical model [22], we have designed a novel and practical online power-saving strategy (OPSS), which included the design of online estimators for both traffic conditions and LTE DRX operations. Specifically, by using the maximum-likelihood method and the properties of Gamma distribution, we derived an unbiased estimator for traffic condition and have shown its effectiveness. Based on the recent progress in estimating DTSMP parameters [24], we propose an unbiased estimator for DRX operations. Simulation results and numerical analysis show that the estimators can make unbiased estimations based on the random samples collected in just 1 s. Compared with other estimation algorithms, our methods could further reduce the eNB's resource consumption. Furthermore, we divided the estimation procedure

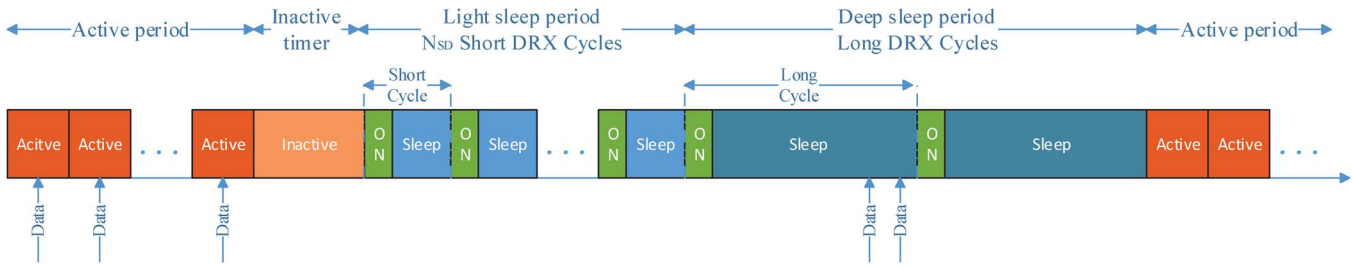

Fig. 1. The 3GPP LTE DRX mechanism.

between eNB and UE, where eNB estimates the traffic, whereas the UE estimates the statistics of the DRX operations. It is worth noting that the idea of a UE-assisted power-saving mechanism has been already adopted in 3GPP Release 11 [6].

The rest of this paper is organized as follows: In Section II, we will introduce the background of LTE DRX. In Section III, we will briefly introduce our analytical LTE DRX model. In Section IV, we will present the online statistical estimators used in our optimization strategy. In Section V, the design of OPSS will be given in detail. In Section VI, the simulation results are presented. Section VII concludes this paper.

II. Background

Here, we discuss two important fundamental concepts: LTE `RRC_Connected`-mode DRX and self-similar traffic, which are the bases of the analysis in the following sections.

A. `RRC_Connected`-Mode DRX

As mentioned in Section I, the difference between LTE DRX and previous DRX is mainly on `RRC_Connected` mode. By this mode, the DRX parameters are allowed to be modified during data transmission. Hence, the `RRC_Connected` DRX can provide quick adaption to traffic status changes and obtain higher energy efficiency [11]. Therefore, in this paper, we will only focus on the `RRC_Connected`-mode DRX. For simplicity, the "DRX" mentioned in the rest of this paper stands for the LTE DRX in the `RRC_Connected` mode.

In LTE, any data transmission requires the UE be in a "High Power" RRC connected state. With all data applications, there are often idle periods when no data is transmitted or received, and during those periods, DRX can save energy. DRX frequently wakes up and shuts down the receiver circuits to save energy. As shown in Fig. 1, the whole data transmission process can be divided into three states, namely, **Active period**, **Light sleep period**, and **Deep sleep period**. In **Active period**, the packet arrival intervals are very short; hence, the UE should stay awake to monitor the PDCCH and receive data as a non-DRX system. The other two states constitute the DRX, and the 3GPP LTE standard [5] specifies the following parameters to control the UE's behavior in these states.

- *Inactive timer* (t_I) denotes the maximum duration after the last transmission that a UE device shall remain turning on the receiver to monitor the PDCCH. This timer is reset to zero and immediately enabled after successful reception of PDCCH (resource grant or allocation).
- *Short DRX cycle* (t_{SD}) specifies the number of consecutive subframes that the UE shall follow in one DRX cycle during the **Light sleep period**.
- *DRX short cycle timer* (N_{SD}) is expressed in an integer number of short DRX cycles. N_{SD} indicates the maximum number of short DRX cycles in **Light sleep period** before transition to the **Deep sleep period**.
- *Long DRX cycle* (t_{LD}) specifies the number of consecutive subframes that the UE shall follow in one DRX cycle during the **Deep sleep period**. A typical long DRX cycle is at least ten times longer than a short DRX cycle.
- *DRX Long cycle timer* (N_{LD}) is expressed in an integer number of long DRX cycles. N_{LD} indicates the maximum number of long DRX cycles in **Deep sleep period** before the eNB indicates the release of UE's RRC connection; in other words, initiating the `RRC_Idle`-mode DRX.
- *On-duration timer* (t_{ON}) denotes the time period for which the UE shall stay awake in each DRX cycle. The On-duration is part of the DRX cycle and has the same length for both short and long DRX cycles by default. To facilitate the later discussion on t_{ON}, we define two new parameters η and η' to represent the proportion of t_{ON} in t_{SD} and t_{LD}, respectively. Hence, we can get $t_{\mathrm{ON}} = \eta t_{\mathrm{SD}} = \eta' t_{\mathrm{LD}}$ and $\eta' = \eta t_{\mathrm{SD}}/t_{\mathrm{LD}}$.

All these adjustable parameters are communicated by means of higher layer (RRC) signaling to each UE during bearer setup and when parameter updates are needed. With the help of these parameters, we can describe the `RRC_Connected`-mode DRX precisely. The inactive timer is equal to the time interval between packets received during **Active period**. At the moment when this value is larger than t_I, the UE turns into the **Light sleep period**, during which only the short DRX cycle can be enabled. In each short DRX cycle, the UE first opens its receiver to monitor the PDCCH for t_{ON} subframes, and then, regardless if packet arrival is detected or not, the UE falls asleep for the remaining $t_{\mathrm{SD}} - t_{\mathrm{ON}}$ subframes. If the PDCCH indicates a downlink transmission, UE transitions from **Light sleep period** into **Active period** and cancels its DRX short cycle timer. Otherwise, the UE goes into the next short DRX cycle and increases its DRX short cycle timer by 1. At the moment that the DRX short cycle timer is equal to $t_{\mathrm{SD}} + 1$, the UE turns into **Deep sleep period**, and the long DRX cycle is activated. Barring the time duration after t_{ON} taking $t_{\mathrm{LD}} - t_{\mathrm{ON}}$ subframes and the maximum value of the DRX long cycle timer being N_{LD}, the UE's behavior during the long DRX cycle is exactly the same as that during the short DRX cycle.

Since t_{SD} is smaller than t_{LD}, the short DRX cycle can be employed in higher frequency, which makes it able to fit bursty traffic. However, when there exist services with an aperiodic silent period or occasional packet arrivals, the long DRX cycle is more suitable [15]. In fact, the two-level DRX concept is a compromise, which considers system complexity, because the more the DRX cycle levels, the more flexibility could be gained to fit a different traffic pattern. To be practical, the DRX mechanism always employs in a semistatic manner, which keeps UE's DRX parameters static for a long duration and updates when needed. Tuning the timer length to reach a suitable and functional length is a tough task. In this paper, we utilized the nature of semistatic by optimizing the DRX parameters based on the observation of the UE's traffic condition and DRX periodical performance.

B. Self-Similar Traffic

The traffic condition, particularly the interarrival time distribution, could significantly impact the performance of the DRX mechanism. As mentioned in Section I, multimedia traffic explicitly exhibits *self-similarity* properties. Thus, we need to figure out which properties that self-similar traffic has and how we could use them to model the interarrival time distribution.

The self-similar traffic, in general, implies LRD [19], [20]. Mathematically, LRD is defined as the autocorrelations of hyperbolical sequence decaying with increasing lag. That is, a covariance stationary process $X = (X_t : t = 0, 1, 2, \ldots)$, with mean μ and variance σ^2, is an LRD if it has an autocorrelation function $r(k), k \geq 0$ of the form

$$r(k) \sim k^{-\beta} L(t) \qquad \text{as } k \to \infty \tag{1}$$

where $0 < \beta < 1$, and $L(t)$ is slowly varying at infinity, i.e., $\lim_{t\to\infty} L(tx)/L(t) = 1$ for all $x > 0$. This definition implies that for such a sequence, the autocorrelations have no finite sum, i.e., $\sum_{k=0}^{\infty} r(k) = \infty$. The parameter β of the autocorrelation function is related to the Hurst parameter by $H = 1 - \beta/2$. The Hurst parameter, i.e., H, is taken to be a measure of the self-similarity of the series. For self-similar traffic, the Hurst parameter should be greater than 1/2: typically $H \approx 0.7$ or 0.8.

In this paper, we are concerned with *stochastic* self-similarity, which implies the random sequences that remain statistically the same when scaled. The pioneering work on self-similarity and its application to communications was by Mandelbrot [21], who argued that an interarrival time distribution of the form $\Pr(\text{interval} \geq x) = Cx^{-k}$ would produce a self-similar arrival process. Here, $0 < k < 1$ is a fixed parameter of the distribution, and x is defined between limits that approach 0 and ∞. The distribution is normalized by scaling constant C. This distribution has a *heavy tail* so that arbitrarily long interarrival times may occur with finite probability. It also has an asymptote at zero, allowing (nearly) zero interarrival times to occur, providing bursts of closely spaced arrivals. This means that the arrival count process is invariant when scaled up from even the smallest initial time intervals. This combination of bursts and spaces is characteristic of the self-similar series.

The best known heavy-tailed distribution is the Pareto distribution, which also fits the packet interval time distribution of self-similar traffic [18]. The probability density function (pdf) of type-I Pareto distribution takes the following form:

$$f(x) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}, \qquad \text{for } x \geq x_m \tag{2}$$

where α is the distribution shape parameter, x_m is the location parameter, and both are positive numbers. Specifically, if $\alpha \leq 2$, then the distribution has infinite variance; if $\alpha \leq 1$, then the distribution has an infinite mean. For self-similar traffic, α should be between 1 and 2. Then, the distribution has slowly decaying variances, and the correlated Hurst parameter is $H = (3-\alpha)/2$.

In the real world, the packet interarrival time cannot be arbitrarily long, since there is always an upper bound of the probability tail. Based on this, the truncated-Pareto distribution, which has been adopted in the 3GPP traffic model [25], is used in this paper. The definition of the pdf of truncated-Pareto distribution is as follows:

$$f(x) = \begin{cases} 0, & x < x_m \\ \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}, & x_m \leq x < m \\ (\frac{x_m}{m})^{\alpha}\delta(x-m), & x \geq m \end{cases} \tag{3}$$

where m is the value of truncated point that is equal to the maximum value of packet interarrival time, $\delta(\cdot)$ is the Dirac delta function, and other parameters are the same with (2). The cumulative distribution function is then defined as

$$F_X(x) = \begin{cases} 0, & x < x_m \\ 1 - \left(\frac{x_m}{x}\right)^{\alpha}, & x_m \leq x < m \\ 1, & x \geq m. \end{cases} \tag{4}$$

It can be seen that the probability of m is a constant, and we could treat other points in the field of definition, following type-I Pareto distribution. It is worth noting that this degradation slightly brings inaccuracy at the probability of m, because its value is brutally equal to the summation of the probability of values greater than or equal to m. However, fortunately, the packet interarrival time in real traffic has a large variation range. Hence, we can assume that the value of m is much larger than the location parameter x_m, which will be shown in the subsequent analysis that the assumption makes the inaccuracy at m affect the analytical results insignificantly.

III. Analytical Model for the RRC_Connected-Mode Discontinuous Reception (DTRX) Mechanism

Here, we first set up the self-similar traffic model and then briefly introduce the DRX analytical model proposed in [22]. Different from our previous works, the procedures of how to generate the traffic trace were given in detail. To make the basis of OPSS clear (proposed in Section IV), we will introduce the necessary steps and the results of DRX analytical modeling, and for more details, [22] could provide complete procedures.

A. Self-Similar Traffic Model

The homogeneous Poisson process is most frequently used to illustrate the packet interarrival times property in former

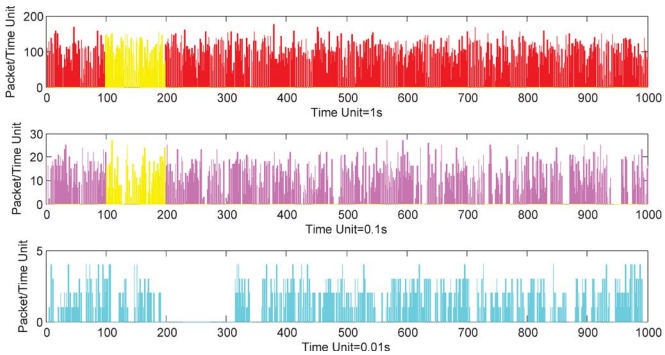

Fig. 2. Self-similar traffic trace ($H = 0.8$) generated by (5) on three different time scales. Different gray levels indicate the same segments of traffic on the different time scales.

research of DRX modeling (i.e., [11], [12], and [15]). To better fit the actual multimedia traffic, we adopted the truncated-Pareto distribution to illustrate the packet interarrival times. This choice had two main advantages in general: First, as discussed in Section II, it could well fit the self-similar property of multimedia traffic; second, it unified different types of packet interarrival times defined by the Poisson process framework; hence, modeling complexity is reduced. Specifically, such distribution could generate bursts of closely spaced arrivals, which corresponds to the concept of "interpacket call idle time," defined in [11] and [15], and it also has unignorable probability to generate sparse-spaced arrivals, which corresponds to the concept of "intersession idle time" defined in [11] and [15] as well.

Random samples that have the pdf of (3) can be generated by using *inverse transform sampling* [26]. Given a random variate U drawn from the uniform distribution on the interval $((x_m/m)^{\alpha}, 1]$, the formula to generate the truncated-Pareto distribution is

$$x = \frac{x_m}{U^{1/\alpha}}. \tag{5}$$

Fig. 2 shows the different scaling behavior of the self-similar traffic trace generated by (5). These traces show the property of LRD, which exhibits fluctuations over a wide range of time scales.

To study the effect of different DRX parameter configurations on the performance directly, we simplify the traffic model under the following assumptions.

- *Exhaustive service*: We assume that the buffer is emptied as soon as the UE wakes up. This assumption is feasible in a practical LTE system, since the peak data rate could achieve more than 120 Mb/s for a specific UE. Therefore, we could ignore the transmission time and focus on the delay that was only caused by the DRX mechanism. This leads to the fact that packet length is *meaningless* in this work. Accordingly, the *packet interarrival time* is defined as the time between instants that the *entire packet* enters eNB's buffer.
- *Neglecting impact of the scheduling and resource allocation algorithms*: As we know, the scheduling and resource allocation algorithms could significantly affect the transmission time and the waiting time as well. For the sake of providing an insight on the effect of different DRX parameter configurations on the DRX performance directly, we neglect the impact of scheduling and resource allocation algorithms. We will take them into account in future works.
- *Single eNB-UE model*: The cell load and the interference among UEs will affect the data transmission; therefore, we only study the single eNB-UE model in this work for the same reasons previously mentioned.
- *Downlink traffic only*: Since in this work we concentrate on the DRX mechanism, only the downlink traffic has been considered. However, as mentioned in [14], the modeling procedures for the DTX is similar with DRX modeling. Hence, the analytical model proposed in this work is also effective in DTX modeling, which considers uplink traffic only.

B. Semi-Markov-Process-Based DRX Analytic Model

This section gives a general overview of our previously proposed DRX analytical model in [22]. Notice that only the procedures and the results that are relevant to the following sections are provided here (for details, see our published paper). As mentioned, our DRX analytical model is built based on DTSMP. The main function of this model is to evaluate the *power-saving factor* and the *expectation of wake-up delay*, which have been considered as the two most important performance output measurements of DRX [3].

The DTSMP can be considered as a generalization of discrete-time Markov chains and the discrete-time renewal process. Therefore, we denote the space of the UE's DRX state as $E = \{1, 2, 3\}$, whose elements correspond to the **Active period**, **Light sleep period**, and **Deep sleep period**, respectively. Assuming that the UE's DRX state during the nth transmission time interval is denoted as Z_n, then the stochastic process $Z = (Z_n \in E; n \in \mathbb{N})$ can be considered as a semi-Markov process. Let the time points of the nth state transitions be denoted as S_n and the stochastic process of S_n as $S = (S_n; n \in \mathbb{N})$. Then, let us denote the UE's DRX state during $[S_{n-1}, S_n]$ as J_n and its stochastic process as $J = (J_n \in E; n \in \mathbb{N})$. Now, following the definitions in [30], we could know that $(J, S) = ((J_n, S_n); n \in N)$ is a discrete-time Markov renewal process; moreover, J is a discrete-time Markov chain, which is also called the embedded Markov chain (EMC). To make the description easier to follow, we denote the sojourn time of J_n as D_n and its stochastic process as $D = (D_n; n \in \mathbb{N}^*)$ with state space $\mathbb{N}^* = \mathbb{N} - \{0\}$. Thus, we have for all $n \in \mathbb{N}^*$, $D_n = S_n - S_{n-1}$.

To deduce the *power-saving factor* and the *expectation of wake-up delay*, we need to calculate the following necessary statistical parameters: the transition probability and the stationary distribution of EMC and, therefore, the mean sojourn time of each state and the limit distribution of DTSMP.

Based on the mechanism of DRX, we can obtain its EMC as shown in Fig. 3 [22]. $S1$, $S2$, and $S3$ here correspond to each DRX state in E. Denoting the probability p_{ij} where $(i, j \in E)$ presents the state transition probability of state i to state j in

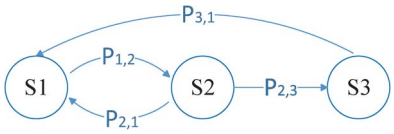

Fig. 3. DTSMP for the RRC_Connected-mode DRX analysis.

EMC, namely, $p_{ij} = \mathbb{P}(J_{n+1} = j | J_n = i)$. Then, the transition probability matrix $\mathcal{P}$ of EMC can be given as

$$\mathcal{P} = \begin{bmatrix} 0 & 1 & 0 \\ p_{21} & 0 & p_{23} \\ 1 & 0 & 0 \end{bmatrix}. \tag{6}$$

In accordance with our thorough consideration of the DRX mechanism as described in [22], we obtain $p_{23} = [t_I/(t_I + (N_{\mathrm{SD}} - 1 + \eta)t_{\mathrm{SD}})]^{\alpha}$, and based on the balance equation of every state, we could also obtain $p_{21} = 1 - p_{23}$.

Let the row vector $\nu = (\nu_n; n \in \{1, 2, 3\})$ be the stationary distribution of the EMC. By applying $\sum_{n=1}^{3} \nu_n = 1$ and the balance equations $\nu_i = \sum_{j=1}^{3} \nu_j p_{ji}$, we can solve the stationary distribution and obtain

$$\begin{cases} \nu_1 = \frac{1}{2+p_{23}} \\ \nu_2 = \frac{1}{2+p_{23}} \\ \nu_3 = \frac{p_{23}}{2+p_{23}}. \end{cases} \tag{7}$$

Now, all the relevant statistical parameters of EMC have been obtained. It is easy to find that the EMC is irreducible and positive recurrent. According to [30, Lemma 6.8.1], the expected holding interval at each state exists WP1 (with probability 1).

Then, we derive the mean sojourn time of each state, denoted as U_i where $i \in E$. According to the method provided in [28], U_i can be calculated as

$$U_i = \mathbb{E}(S_1 | J_0 = i) = \sum_{k \geq 1} k h_i(k) \tag{8}$$

where $h_i(\cdot)$ is the sojourn time distribution in state i and is defined as

$$h_i(k) = \mathbb{P}(D_{n+1} = k | J_n = i) = \sum_{j \in E} q_{ij}(k), \qquad k \in \mathbb{N}. \tag{9}$$

The $q = \{q_{ij}(k), k \in \mathbb{N}\}$ in (9) is called the discrete-time semi-Markov kernel, which is defined as

$$q_{ij}(k) = \mathbb{P}(J_{n+1} = j, D_{n+1} = k | J_n = i) \tag{10}$$

where $i, j \in E$, and $q_{ij}(0) = 0$. It satisfies the relation

$$q_{ij}(k) = p_{ij} f_{ij}(k), \qquad i, j \in E, \quad k \in \mathbb{N} \tag{11}$$

where $f_{ij}(k)$ is the conditional distribution of the sojourn time and is given by

$$f_{ij}(k) = \mathbb{P}(D_{n+1} = k | J_n = i, J_{n+1} = j) \qquad i, j \in E, \quad k, n \in \mathbb{N}. \tag{12}$$

By applying the fact that, during S1, the interarrival times between packets are independent and identically distributed (i.i.d.) and thus the increment for the time of new packet arrivals follows the geometric distribution, we could obtain

$$\begin{aligned} U_1 &= \sum_{c=1}^{\infty} c \psi_c \mathbb{E}(\hat{x}) \\ &= \left[\frac{\alpha x_m^{\alpha} t_I^{1-\alpha} - \alpha x_m}{1 - \alpha} \right] \left[\left(\frac{t_I}{x_m} \right)^{\alpha} - 1 \right]. \end{aligned} \tag{13}$$

Following the aforementioned definitions, U_2 is given by

$$U_2 = \sum_{l=1}^{N_{\mathrm{SD}}} l t_{\mathrm{SD}} q_{21}(l t_{\mathrm{SD}}) + N_{\mathrm{SD}} t_{\mathrm{SD}} q_{23} \tag{14}$$

where the corresponding discrete-time semi-Markov kernels can now be obtained by $q_{21} = p_{21} f_{21}$ and $q_{23} = p_{23} f_{23}$. Specifically, f_{21} and f_{23} are given by

$$\begin{aligned} f_{21}(l t_{\mathrm{SD}}) &= \mathbb{P}(D_{n+1} = l t_{\mathrm{SD}} | J_n = 2, J_{n+1} = 1) \\ &\qquad l \in \{1, 2, \ldots, N_{\mathrm{SD}}\} \\ &= \begin{cases} \frac{\mathbb{P}[t_I \leq x \leq \eta t_{\mathrm{SD}} + t_I]}{\mathbb{P}[x \geq t_I]}, & l = 1 \\ \frac{\mathbb{P}[t_I \leq x \leq (l-1) t_{\mathrm{SD}} + \eta t_{\mathrm{SD}}]}{\mathbb{P}[x \geq t_I]}, & 1 < l \leq N_{\mathrm{SD}} \end{cases} \\ &= \begin{cases} 1 - (\eta t_{\mathrm{SD}} + t_I)^{-\alpha} t_I{}^{\alpha}, & l = 1 \\ \left\{ [(l - 2 + \eta) t_{\mathrm{SD}} + t_I]^{-\alpha} \right. \\ \left. - [(l - 1 + \eta) t_{\mathrm{SD}} + t_I]^{-\alpha} \right\} t_I{}^{\alpha}, & 1 < l \leq N_{\mathrm{SD}} \end{cases} \end{aligned} \tag{15}$$

$$\begin{aligned} f_{23}(N_{\mathrm{SD}} t_{\mathrm{SD}}) &= \mathbb{P}(D_{n+1} = N_{\mathrm{SD}} t_{\mathrm{SD}} | J_n = 2, J_{n+1} = 3) \\ &= \left[\frac{x_m}{(N_{\mathrm{SD}} - 1 + \eta) t_{\mathrm{SD}}} \right]^{\alpha}. \end{aligned} \tag{16}$$

Similarly, U_3 can be obtained by calculating

$$U_3 = \sum_{l=1}^{n_m} l t_{\mathrm{LD}} q_{31}(l t_{\mathrm{LD}}) \tag{17}$$

where the semi-Markov kernel is calculated by $q_{31} = p_{31} f_{31} = f_{31}$. f_{31} is given by (18), shown at the bottom of the page, where

$$\begin{aligned} f_{31}(l t_{\mathrm{LD}}) &= \mathbb{P}(D_{n+1} = l t_{\mathrm{LD}} | J_n = 3, J_{n+1} = 1) = \begin{cases} \frac{\mathbb{P}[t_I + N_{\mathrm{SD}} t_{\mathrm{SD}} \leq x \leq t_I + N_{\mathrm{SD}} t_{\mathrm{SD}} + \eta' t_{\mathrm{LD}}]}{\mathbb{P}(x > t_I + N_{\mathrm{SD}} t_{\mathrm{SD}})}, & l = 1 \\ \frac{\mathbb{P}[t_I + N_{\mathrm{SD}} t_{\mathrm{SD}} \leq x \leq t_I + N_{\mathrm{SD}} t_{\mathrm{SD}} + (\eta' + l - 1) t_{\mathrm{LD}}]}{\mathbb{P}(x > t_I + N_{\mathrm{SD}} t_{\mathrm{SD}})}, & l > 1 \end{cases} \\ &= \begin{cases} 1 - (t_I + N_{\mathrm{SD}} t_{\mathrm{SD}})^{\alpha} \cdot (t_I + N_{\mathrm{SD}} t_{\mathrm{SD}} + \eta' t_{\mathrm{LD}})^{-\alpha}, & l = 1 \\ \left\{ [t_I + N_{\mathrm{SD}} t_{\mathrm{SD}} + (\eta' + l - 2) t_{\mathrm{LD}}]^{-\alpha} - [t_I + N_{\mathrm{SD}} t_{\mathrm{SD}} + (\eta' + l - 1) t_{\mathrm{LD}}]^{-\alpha} \right\} \cdot (t_I + N_{\mathrm{SD}} t_{\mathrm{SD}})^{\alpha}, & l > 1 \end{cases} \end{aligned} \tag{18}$$

$\lfloor\cdot\rfloor$ is the round down function, and $l \in \{1, 2, \ldots, n_m = \lfloor (m - t_I - N_{\mathrm{SD}} t_{\mathrm{SD}}/t_{\mathrm{LD}})\rfloor\}$.

Through the given analysis, we can come to a conclusion that the DTSMP is irreducible and that the mean sojourn time is finite for any state $j \in E$. According to [28, Prop. 8], the limit distribution of DTSMP could be obtained, i.e.,

$$\begin{cases} \pi_1 = \frac{\nu(1)U_1}{\sum_{i\in E}\nu(i)U_i} \\ \pi_2 = \frac{\nu(2)U_2}{\sum_{i\in E}\nu(i)U_i} \\ \pi_3 = \frac{\nu(3)U_3}{\sum_{i\in E}\nu(i)U_i}. \end{cases} \tag{19}$$

C. Evaluation Metrics

Power-Saving Factor: The power-saving factor is introduced to indicate the sleeping time ratio, compared with the overall operation time [3]. It can be given by

$$\mathrm{PS} = \frac{(1-\eta)U_2\nu(2) + (1-\eta')U_3\nu(3)}{\sum_i U_i\nu(i)}, \qquad i \in E. \tag{20}$$

Substituting (7), (13), (14), and (17) into (20), we derive the closed-form equation for the power-saving factor PS.

Expectation of Wake-Up Delay: As mentioned earlier, a packet that arrives in the DRX cycles should be waiting in the buffer until the UE transitions from "S2" (or "S3") to "S1." The remainder of this section focuses on obtaining the expectation of this waiting time, namely, *wake-up delay*. We denote $D_S(l)$ and $D_L(l)$ as delay happens in S2 and S3, respectively. $D_S(l)$ could be calculated as (21), shown at the bottom of the page. Similarly, we could get (22), shown at the bottom of the page, where n_m is similarly defined as (18). Therefore, we can obtain the expected wake-up delay in the form of

$$\begin{aligned} \mathbb{E}(D) &= \mathbb{E}\left[D_S(l)\right] + \mathbb{E}[D_L(l)] \\ &= \pi_2 \sum_{l=1}^{N_{\mathrm{SD}}} f_{21}(lt_{\mathrm{SD}}) D_S(l) + \pi_3 \sum_{l=1}^{n_m} f_{31}(lt_{\mathrm{LD}}) D_L(l). \end{aligned} \tag{23}$$

Substituting (15), (18), (19), (21), and (22) into (23), we can derive the closed-form equation for the expectation of wake-up delay.

IV. Online Estimation of Statistical Parameters

We presented the LTE-DRX analytical model not only to evaluate the power-saving performances but also to seek to provide a practical way to optimize the DRX parameters for the ongoing system. As mentioned earlier, in practical systems, the dynamic DRX mechanism is deployed in a semistatic manner, which gives the eNB a chance to estimate the traffic conditions and power-saving performance of a specific user. Therefore, here, we focus on the following practical issues: 1) which statistical parameter needs to be estimated; 2) how to estimate them from the random samples; 3) how long is sufficient to estimate the parameters without bias.

A. Estimation of Traffic Conditions

Basically, our traffic model is based on the truncated-Pareto distribution. Thus, the traffic condition estimation (TCE) problem could be transformed into a truncated-Pareto distribution parameter estimation problem. In [31], Petersen compared main methods of estimating the parameters of Pareto distribution from which a random sample comes. The compared methods included the *method of moments*, the *maximum-likelihood method*, and the *quantile regression method*. The numerical results show that the *method of moments* horribly behaved in estimating the shape parameter α but was able to estimate location parameter x_m without bias. On the other hand, the *quantile regression method* shows to be the least biased estimator for estimation α, but it generally had greater than 20% for standard errors than the *maximum-likelihood method*. Meanwhile, the *maximum-likelihood method* yielded a smaller mean square error (MSE) than the *quantile regression method* for estimating α. We seek to find an online TCE strategy that could unbiasedly estimate the parameters within numerable random samples. Hence, based on the work in [31], we adopted the *method of moments*

$$\begin{aligned} D_S(l) &= \begin{cases} t_{\mathrm{SD}} - \int_{t_I}^{t_I+\eta t_{\mathrm{SD}}} \frac{\alpha x_m^\alpha}{x^\alpha}\,dx, & l = 1 \\ 2t_{\mathrm{SD}} - \int_{t_I+(l-1)\eta t_{\mathrm{SD}}}^{t_I+l\eta t_{\mathrm{SD}}} \frac{\alpha x_m^\alpha}{x^\alpha}\,dx, & 1 < l < N_{\mathrm{SD}} \end{cases} \\ &= \begin{cases} t_{\mathrm{SD}} - \frac{\alpha x_m^\alpha}{1-\alpha}\left[(t_I+\eta t_{\mathrm{SD}})^{(1-\alpha)} - t_I^{(1-\alpha)}\right], & l = 1 \\ 2t_{\mathrm{SD}} - \frac{\alpha x_m^\alpha}{1-\alpha}\left\{(t_I+l\eta t_{\mathrm{SD}})^{(1-\alpha)} - \left[t_I+(l-1)\eta t_{\mathrm{SD}}\right]^{(1-\alpha)}\right\}, & 1 < l < N_{\mathrm{SD}} \end{cases} \end{aligned} \tag{21}$$

$$\begin{aligned} D_L(l) &= \begin{cases} t_{\mathrm{LD}} - \int_{t_I+N_{\mathrm{SD}}t_{\mathrm{SD}}}^{t_I+N_{\mathrm{SD}}t_{\mathrm{SD}}+\eta' t_{\mathrm{LD}}} \frac{\alpha x_m^\alpha}{x^\alpha}\,dx, & l = 1 \\ 2t_{\mathrm{LD}} - \int_{t_I+N_{\mathrm{SD}}t_{\mathrm{SD}}+(l-1)\eta' t_{\mathrm{SD}}}^{t_I+N_{\mathrm{SD}}t_{\mathrm{SD}}+l\eta' t_{\mathrm{SD}}} \frac{\alpha x_m^\alpha}{x^\alpha}\,dx, & 1 < l < n_m \end{cases} \\ &= \begin{cases} t_{\mathrm{LD}} - \frac{\alpha x_m^\alpha}{1-\alpha}\left[(t_I+N_{\mathrm{SD}}t_{\mathrm{SD}}+\eta' t_{\mathrm{LD}})^{(1-\alpha)} - (t_I+N_{\mathrm{SD}}t_{\mathrm{SD}})^{(1-\alpha)}\right], & l = 1 \\ 2t_{\mathrm{LD}} - \frac{\alpha x_m^\alpha}{1-\alpha}\left\{(t_I+N_{\mathrm{SD}}t_{\mathrm{SD}}+l\eta' t_{\mathrm{SD}})^{(1-\alpha)} - \left[t_I+N_{\mathrm{SD}}t_{\mathrm{SD}}+(l-1)\eta' t_{\mathrm{SD}}\right]^{(1-\alpha)}\right\}, & 1 < l < n_m \end{cases} \end{aligned} \tag{22}$$

to estimate x_m and utilized the *maximum-likelihood method* to estimate α. To get unbiased estimation of α, we go a step further. By transforming the Pareto distribution into exponential distribution, the properties of Gamma distribution moments could be utilized to further improve the maximum-likelihood estimator's (MLE) accuracy. It is particularly necessary to point out that since the random samples could rarely achieve the truncated value,[1] it is impossible to estimate the truncated value during an observation window. Hence, we then assume that the truncated point value m is predefined as a large value. Then, we begin by constructing the MLE for the location parameter.

Suppose we have a random sample of observed packet interarrival times $\{x_1, x_2, \ldots, x_n\}$ obtained by truncated-Pareto distribution. We denote the corresponding joint random variables as $\{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n\}$. Then, the likelihood function, i.e., L, for the Pareto distribution has the following form:

$$L(x_m, \alpha|x) = \prod_{i=1}^{n} \frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}}, \qquad 0 < x_m \leq \min\{x_i\}, \quad \alpha > 0. \tag{24}$$

By analyzing the likelihood function of a series of i.i.d. Pareto random variables, the MLE of the scale parameter can be shown as

$$\hat{x}_m = x_{(1)} = \min\{x_1, x_2, \ldots, x_n\}. \tag{25}$$

This result is convincing, since it is plausible that the finite endpoint of the distribution support should be estimated from the smallest observation in a random sample from the distribution.

It can also be demonstrated that the conditional distribution of the random sample $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$, conditioned on the event that the random variable $\tilde{X}_{(1)} = x$, does not depend on $\hat{x}_m$, which implies that the minimum $\tilde{X}_{(1)}$ is a sufficient statistic for $\hat{x}_m$. Furthermore, the Pareto distribution belongs to the exponential family of distributions, which is complete (will be proved later). This implies that we can construct the minimum variance unbiased estimator. By observing that $\mathbb{E}(X_{(1)}) = (n\alpha x_m/(n\alpha - 1))$, we can then choose the random variable

$$\hat{X}_m = \frac{(n\alpha - 1)X_{(1)}}{n\alpha}. \tag{26}$$

Since this is a function of a complete sufficient statistic, it is a minimum variance unbiased estimator for its mean, which is clearly x_m by construction. For large sample sizes, the statistic $\hat{X}_m$ is asymptotically the same as the minimum $X_{(1)}$; hence, this can be used to remove dependence on α in the estimator of (26).

Next, we examine the MLE for the shape parameter. To find the maximum-likelihood estimate of α, calculus is appropriate. Since L is nonnegative, we can take its logarithm. We do this because it is easier to differentiate $\log L$ than L itself. Logarithms are bijective functions; hence, the value of α that maximizes L also maximizes $\log L$. The brief process is as follows:

$$\begin{aligned}\log L(x_m, \alpha|x) &= \sum_{i=1}^{n} \log\left(\frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}}\right) \\ &= n\log(\alpha) + \alpha n \log(x_m) - (\alpha+1)\sum_{i=1}^{n}\log(x_i).\end{aligned} \tag{27}$$

Therefore

$$\frac{d\log L(\cdot)}{d\alpha} = n/\alpha + n\log(x_m) - \sum_{i=1}^{n}\log(x_i). \tag{28}$$

Setting the derivative equal to zero, a little algebra and an omitted second derivative check confirm that we are maximizing L rather than minimizing L, i.e.,

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n}\left[\log(x_i) - \log(\hat{X}_m)\right]}. \tag{29}$$

Then, we prove that the Pareto distribution belongs to the exponential family as follows. If X is Pareto-distributed with parameters x_m and α, then we define a random variable $Y = \log(X/x_m)$. It is not difficult to show, by constructing the distribution function of Y that is an exponential distribution with parameter α, that this means that a Pareto $P(\alpha, x_m)$ random variable is statistically equivalent to $x_m e^Y$, where $Y \triangleq \mathrm{Exp}(\alpha)$.

Due to the fact that $\log(X_i/\hat{X}_m)$ is exponentially distributed, it can be shown that $T = \sum_{i=1}^{n}[\log(X_i) - \log(\hat{X}_m)]$ has a Gamma distribution with the density function

$$f_T(t) = \frac{\alpha^n}{(n-1)!}t^{n-1}e^{-\alpha t}. \tag{30}$$

As $\hat{\alpha} = n/T$, we get the following:

$$\mathbb{E}(\hat{\alpha}) = \frac{n\alpha}{(n-1)}\int_0^{\infty}\frac{\alpha^{n-1}}{(n-2)!}t^{n-2}e^{-\alpha t}\,dt = \frac{n}{n-1}\alpha \tag{31}$$

$$\begin{aligned}\mathbb{E}(\hat{\alpha}^2) &= \frac{n^2\alpha^2}{(n-1)(n-2)}\int_0^{\infty}\frac{\alpha^{n-2}}{(n-3)!}t^{n-3}e^{-\alpha t}\,dt \\ &= \frac{n^2}{(n-1)(n-2)}\alpha^2.\end{aligned} \tag{32}$$

Hence, the variance of $\hat{\alpha}$ is

$$\mathrm{Var}(\hat{\alpha}) = \frac{n^2\alpha^2}{n-1}\left(\frac{1}{n-2} - \frac{1}{n-1}\right) = \frac{n^2}{(n-1)^2(n-2)}\alpha^2. \tag{33}$$

The MLE $\hat{\alpha}$ of α is biased, as is the estimator

$$\alpha^* = \frac{n-1}{T} = \frac{n-1}{\sum_{i=1}^{n}\left[\log(x_i) - \log(\hat{X}_m)\right]} \tag{34}$$

with T defined earlier. Furthermore

$$\mathrm{Var}(\alpha^*) = \frac{1}{n-2}\alpha^2 < \mathrm{Var}(\hat{\alpha}). \tag{35}$$

[1]The tradeoff between the observation window length and ML's computing complexity should be considered here. As later described, we always tuned the observation window length smaller than the truncated value; thus, it is impossible to observe a whole truncated valued packet interarrival time during an observation window.

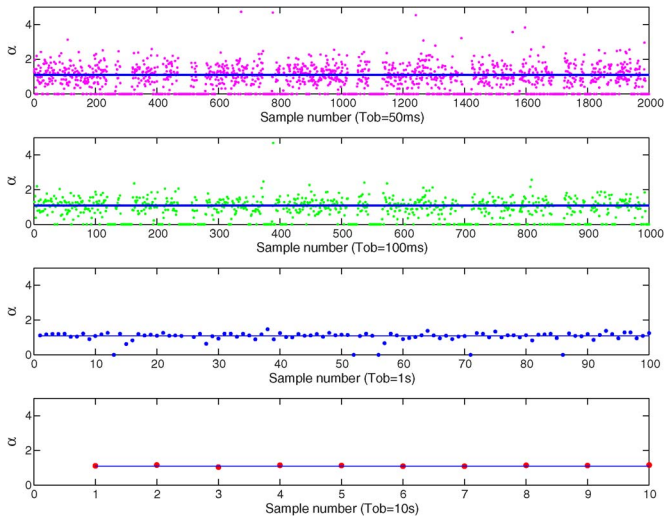

Fig. 4. Estimation of α with different sizes of T_{ob}. The line is an indicator of true value.

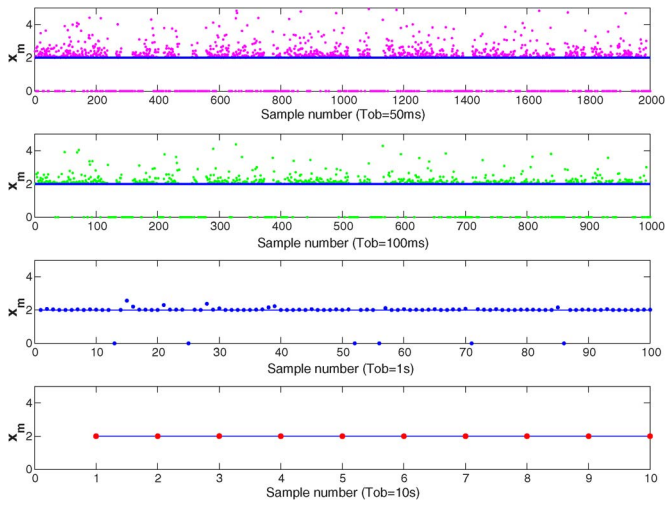

Fig. 5. Estimation of x_m with different sizes of T_{ob}. The line is an indicator of true value.

Thus, α^* is a better estimator of α than $\hat{\alpha}$ and is unbiased for α. By constructing joint conditional distributions, it can be demonstrated that α^* is a complete sufficient statistic for α, and since it is unbiased, it is also the minimum variance unbiased estimator for α (also due to the completeness as before). Hence, (26) can be used to estimate the position parameter (using a large sample approximation), whereas (34) is used with this to estimate the shape parameter.

Next, we study "how long the estimation takes to be sufficient" by numerical methods. First, we generate 1.5×10^5 packet interarrival times by (5). The statistical parameter of the truncated-Pareto distribution is $\alpha = 1.1$, $x_m = 2$ ms, $m = 2$ s, respectively. Then, we generate $(1.5 \times 10^5 + 1)$ packet arrival times, where the first packet arrives at "instant 0," and each packet interarrival time corresponds to one of the obtained samples. Thus, the nth $(n > 1)$ packet arrival time is equal to the summation of $n - 1$ packet interarrival times before. Finally, based on the packets arriving during the different size of the observation window T_{ob}, which is equal to 50 ms, 100 ms, 1 s, and 10 s, the estimation of x_m and α is carried out, respectively.

Each subgraph in Figs. 4 and 5 demonstrates the estimation of α and x_m obtained in each T_{ob}. The total duration of each measurement is 100 s. When no packet arrival happens during T_{ob}, the corresponding estimation value is set to zero. Since the biggest packet interarrival time is equal to the truncated value $m = 2$ s, there should be zero point, as long as T_{ob} is less than 2 s. This case can be seen in both Figs. 4 and 5. These figures also reveal that the estimated quantities converge toward the true values for increasing T_{ob}.

The statistical characters of the estimation results are collected in Table I. The convergence of estimated quantities could also be clearly observed. After comparing the statistical characters, we recommend 1 s as one good choice of T_{ob} for two reasons. First, the bias is acceptable (less than 2%), and the estimation is stable (MSE $< 3\%$). Second, by simulation, we find that the average number of packet arrivals is 87.7067 and 877.0667 for 1 and 10 s, respectively. We should note that the location parameter α here is 1.1. It will be shown later that the number of packet arrivals is increased with the growth of α. It is impractical for the eNB to record thousands of packet arrival times for one specific UE and the ML, for such good many samples will take a lot of computing and storage resources. Therefore, $T_{ob} = 1$ s will bring the eNB sufficient samples, which will be neither too little nor too much.

TABLE I
STATISTICAL CHARACTERS OF ESTIMATION RESULTS FOR PARETO DISTRIBUTION ($x_m = 2$, $\alpha = 1.1$)

Param	T_{ob}	Mean	Bias	MSE
α	50ms	1.1935	0.0935	0.2890
	100ms	1.1223	0.0223	0.1620
	1000ms	1.1188	0.0188	0.0208
	10000ms	1.0989	0.0011	0.0016
x_m	50ms	2.3696	0.3696	0.3676
	100ms	2.2385	0.2385	0.1854
	1000ms	2.0369	0.0369	0.0087
	10000ms	2.0027	0.0027	1.0912e-5

B. Estimation of DRX Operations

The DRX performance has been evaluated by PS and $\mathbb{E}(D)$, but real-time computing (20) and (23) is still quite difficult. The motivation of answering the question of how to estimate the statistical parameters in the expressions of PS and $\mathbb{E}(D)$ directly is strong as the answer will help us find an easier way to obtain the current UE's DRX performance. From (20) and (23), we can see that, apart from α, the other statistical parameters are all associated with DTSMP. Thus, the essential problem is transformed into a DTSMP parameter estimation problem, which is also an important issue in reliable systems and has been studied in [24]. In [24], some useful estimators for DTSMP parameters are provided. In the following, we will first introduce the estimators into our problem, and then, we will demonstrate their performance by numerical evaluations. Note that all the discussions here on DTSMP still follow the same definition in Section III.

Let us assume now that we have an observation of the DTSMP, censored at fixed arbitrary time $M \in \mathbb{N}^*$, which is an

observation of the associated Markov renewal chain $(J_n, S_n)_{n\in\mathbb{N}}$, $(J_0, D_1, \ldots, J_{N(M)-1}, D_{N(M)}, J_{N(M)}, u_M)$, where $u_M = M - S_{N(M)}$ is the censored sojourn time in the last visited state $J_{N(M)}$.

For all states $i, j \in E$, let us introduce

$$N_i(M) = \sum_{n=0}^{M} \mathbb{1}_{\{J_n=i, S_{n+1}\le M\}} \tag{36}$$

where $N_i(M)$ stands for the number of visits to state i of the EMC $(J_n)_{n\in\mathbb{N}}$, up to time M. Thus

$$N_{ij}(M) = \sum_{n=1}^{M} \mathbb{1}_{\{J_{n-1}=i, J_n=j, S_{n+1}\le M\}} \tag{37}$$

where $N_{ij}(M)$ stands for the number of transitions of the EMC $(J_n)_{n\in\mathbb{N}}$ from i to j, up to time M. We defined D_{ik} as the sojourn time of the kth visiting to state i and then defined the empirical estimator of the stationary distribution of the EMC $(J_n)_{n\in\mathbb{N}}$ by

$$\hat{\nu}(i, M) = \frac{N_i(M)}{N(M)}, \qquad i \in E. \tag{38}$$

The estimator for U_i is given by

$$\hat{U}_i(M) = \frac{1}{N_i(M)} \sum_{k=1}^{N_i(M)} D_{ik}. \tag{39}$$

Consequently, an estimator of the mean sojourn time of the DTSMP $\bar{U}$ is

$$\hat{\bar{U}}(M) = \frac{1}{N(M)} \sum_{k=1}^{N(M)} D_k \tag{40}$$

and we get the following estimator of the stationary distribution of the DTSMP:

$$\hat{\pi}_i(M) = \frac{1}{\hat{\bar{U}}(M) N(M)} \sum_{k=1}^{N_i(M)} D_{ik}, \qquad i \in E. \tag{41}$$

According to [24, Lemma 1], for any state $i \in E$ of the DTSMP, the estimators $\hat{\nu}(i, M)$, $\hat{U}_i(M)$, $\hat{\bar{U}}(M)$, and $\hat{\pi}_i(M)$ proposed in (38)–(41) are strongly consistent, as M tends to infinity. Next, we will check their performance and answer the following question: "How long would the estimation take to be sufficient?"

First, we generated a traffic trace that lasted for 1000 s, as we described in the previous section; second, we entered the traffic trace into an LTE-DRX entity, which is written via a MATLAB event-driven simulator (that will be further discussed in Section VI); third, we recorded the $N_i(M)$, $N_{i,j}(M)$, D_{ik}, and D_k and obtained the true value of $\nu(i), U_i, \bar{U}, \pi_i$, $(i, j \in E)$; finally, we verified the estimator's performance at different time scales. Owing to space constraints, we cannot demonstrate all the estimator's performance, but we can show the estimation traces of $\nu(i)$ and U_i in figures. If the $\hat{\nu}(i, M), \hat{U}_i(M)$ values are given, the value of $\hat{\bar{U}}(M)$ and $\hat{\pi}_i(M)$ could be obtained by using algebraic computation. In addition to $T_{ob} = 10$ s, we also

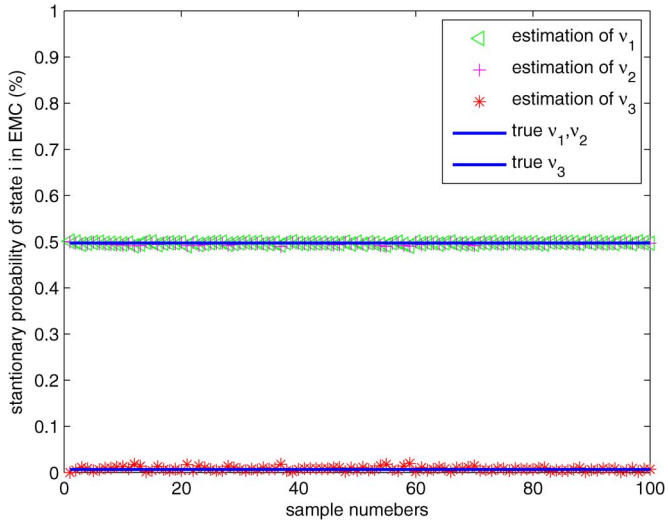

Fig. 6. Estimation of stationary distribution of EMC for $T_{ob} = 10$ s.

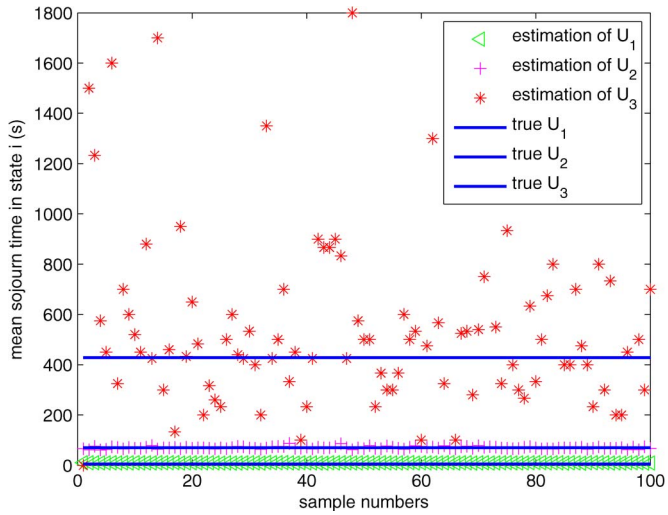

Fig. 7. Estimation of mean sojourn time for $T_{ob} = 10$ s.

directly give the statistical characters of estimation results while T_{ob} is equal to other values.

Figs. 6 and 7 showed that the estimators can effectively estimate most of DTSMP's statistical parameters. However, in Fig. 7, it is obvious that the estimator cannot correctly estimate the mean sojourn time in state "S3." The reason is that the Pareto distribution is kind of a power law probability distribution, which means that there is little chance to see an extremely big packet interarrival time during T_{ob}. By multiplying the large interarrival time and its probability as (39) does, the large fluctuation as shown in Fig. 7 is inevitable. For this reason, we estimate U_3 by our analytical model, which was well coincident with observation (the proof is given in Section VI).

"MAD" in Table III stands for *median absolute deviation*, which is a robust measure of the variability of an univariate sample of quantitative data. From Tables II and III, we can conclude that $T_{ob} = 1$ s is still a good choice. Therefore, the two parts of the statistical parameter estimation algorithm could be simultaneously deployed. Furthermore, to reduce the eNB's burden, we recommend that the UE estimated the DRX performance itself,

TABLE II
STATISTICAL CHARACTERS OF ESTIMATION RESULTS FOR STATIONARY DISTRIBUTION OF EMC ($\nu_1 = \nu_2 = 0.4968$, $\nu_3 = 0.0065$)

Param	T_{ob}	Mean	Bias	MSE
ν_1	500ms	0.5202	0.0238	0.0031
	1000ms	0.5099	0.0131	0.0016
	2000ms	0.5199	0.0151	0.0015
	10000ms	0.4970	2.0e-4	4.3105e-6
ν_2	500ms	0.4809	0.0159	0.0013
	1000ms	0.4878	0.0090	3.5580e-4
	2000ms	0.5199	0.0151	8.3681e-4
	10000ms	0.4955	0.0013	7.2626e-5
ν_3	500ms	0.1126	0.1061	0.0232
	1000ms	0.0405	0.0340	0.0035
	2000ms	0.0249	0.0184	0.0016
	10000ms	0.0077	0.0012	1.8743e-5

TABLE III
STATISTICAL CHARACTERS OF ESTIMATION RESULTS FOR MEAN SOJOURN TIME ($U_1 = 9.7722$, $U_2 = 69.6624$)

Parameter	T_{ob}	Mean	Bias	MAD
U_1	500ms	9.8570	0.0848	1.9991
	1000ms	9.8887	0.1165	1.5187
	2000ms	9.8328	0.0606	1.5868
	10000ms	9.8667	0.0945	0.4444
U_2	500ms	85.8975	16.2351	26.3857
	1000ms	78.3620	8.6996	16.7216
	2000ms	80.5486	10.8862	19.4432
	10000ms	69.6624	4.9096e-5	3.2887

and informed the eNB at the first *on_duration* after the "estimation phase" via the physical uplink control channel (PUCCH). Then, eNB should enter the "optimization phase" and optimize specific UE's DRX parameters based on the strategy discussed in the following section.

V. ONLINE POWER-SAVING STRATEGY

Here, we develop an OPSS, which aims to optimize the DRX configuration based on statistics estimation and the actual requirements. In general, this strategy is divided into an "estimation phase" and an "optimization phase." The "estimation phase" had been introduced in detail in the previous section. Thus, here, we will first discuss the algorithm adopted in the "optimization phase" and then recite the whole procedure of OPSS.

The optimization algorithm should consider maximum delay $t_{\max}$ constraint of the application and minimum power saving $PS_{\rm th}$ imposed by the UE, as to decide a balance tradeoff between power saving and delay. In practice, there are two common scenarios while deploying the DRX optimization algorithm. Hence, in accordance, we introduced two parameters γ_d and γ_{ps} to quantify their difference as in [10].

- Scenario 1: $\gamma_d > \gamma_{ps}$, which indicates that the latency performance is more important than power saving. In this case, first, we should find the feasible ranges for values of the DRX parameters that could guarantee the specified delay constraint $t_{\max}$. Then, the parameters that could maximize power saving will be selected. This scenario is defined for real-time applications such as video streaming.
- Scenario 2: $\gamma_{ps} > \gamma_d$, which indicates that the power-saving performance is more important than latency. In this case, feasible ranges of values for DRX parameters, which are satisfied with the specified UE's power-saving constraint $PS_{\rm th}$, are identified. The parameters should be tuned to minimize the delay within the calculated feasible ranges. This scenario is defined for the UE in the "power save mode," which is activated when the UE has a stringent battery power constraint or for nonreal-time applications, such as those for social networking applications.

Calculating PS and $\mathbb{E}(D)$ for every DRX parameter combination is time consuming, making it nearly impossible to compute them in real time. Hence, we utilized the tradeoff between both computation and storage, which meant offline computations of the DRX performance (e.g., PS and $\mathbb{E}(D)$) based on every possibility traffic condition and DRX parameter configuration pair, and then, it stores them into a table called Ω. Every entry of Ω should include the values of $\{t_I, t_{\rm SD}, N_{\rm SD}, t_{\rm LD}, t_{\rm ON}, \alpha, x_m, PS, \mathbb{E}(D)\}$. An entry is denoted as κ in this work. Note that, here, we include all the configurable parameters in the entries, which may be much fewer in practical systems [8], [10], [12]. Finally, we summarized the main procedures of OPSS in **Algorithm 1** (as shown in the Appendix).

It is worth noting that in our optimization algorithm, we compared the numerically calculated optimal performance with the current DRX performance. The reason is that while the traffic is stable, the DRX parameter configurations that have been optimized in the last cycle may still be the optimum configuration at present. Hence, if there is no need to change the DRX parameters, the eNB will not send the DRX updated configuration information to the UE for the sake of increasing spectrum efficiency. The method of quickly finding $\kappa_{\max}$ from Ω is beyond the scope of this paper, which will be demonstrated in our future works. The searching method adopted in this paper is based on the brute-force research algorithm.

VI. PERFORMANCE EVALUATION

Simulations are conducted to evaluate the proposed analytical model and the OPSS. All the simulations are implemented via a MATLAB event-driven simulator, whereas a single eNB/UE pair with DL self-similar traffic is considered in the simulation environment. We implemented the DRX mechanism following the 3GPP Release 10 specification [5], which included all the required procedures and functions of both the eNB and UE. For ease of description, we named the entire implemented DRX mechanism as "the DRX entity." Furthermore, our proposed OPSS was also implemented in the testbed. Specifically, we implemented the TCE and online DRX parameter configuration optimization in eNB and the DRX operation estimation mechanism in UE, respectively. The required parameter negotiation and update functions are established on the basis of a discrete event simulation method. Each simulation run is executed with a traffic trace generated by the method described in Section IV-A. Moreover, each presented simulation result is the average of 100 simulation runs.

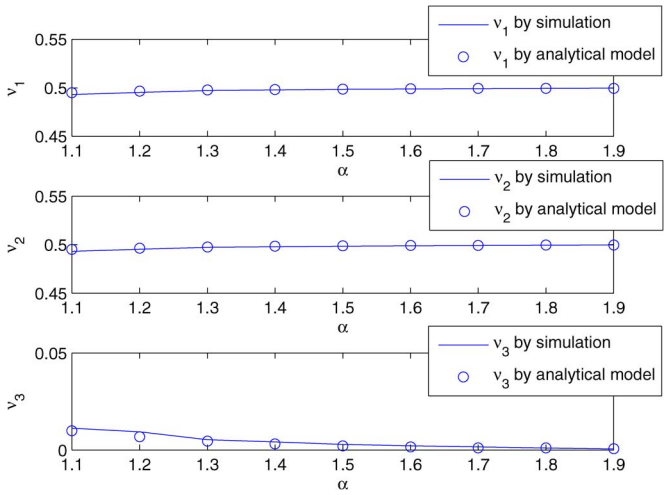

Fig. 8. Validate the analytical model of $\nu_i, (i \in E)$.

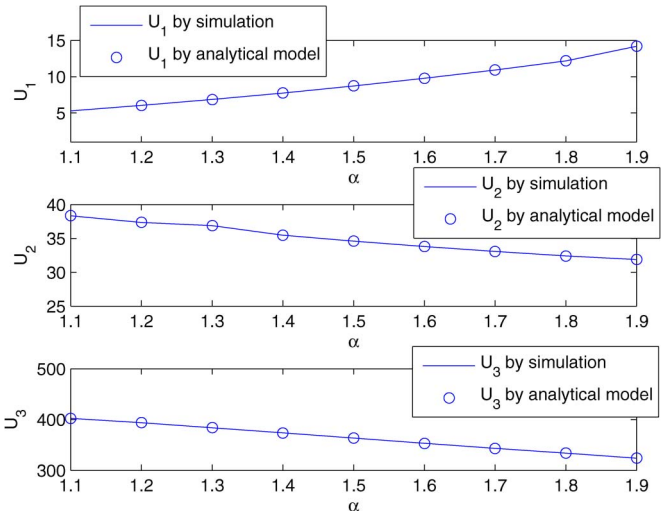

Fig. 9. Validate the analytical model of $U_i, (i \in E)$.

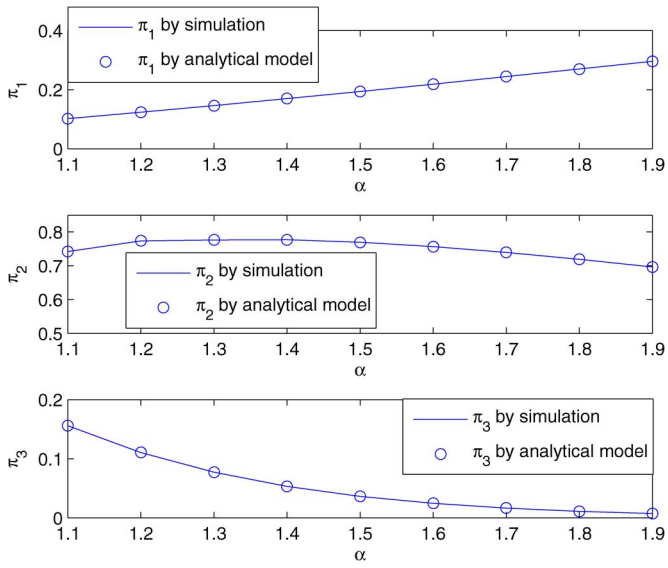

Fig. 10. Validate the analytical model of $\pi_i, (i \in E)$.

A. *Evaluation of the Analytical Models*

In our previous work [22], we have proven that the proposed DRX analytical model could correctly evaluate the PS and $\mathbb{E}(D)$ under various traffic conditions. Moreover, our model is more suitable to multimedia traffic than existing Poisson-based DRX models. Here, we go a step further, verifying not only the accuracy of our model for PS and $\mathbb{E}(D)$ but the accuracy of other relevant statistics as well. As shown later, by doing so, we could study the impact of various traffic conditions on DRX mechanisms precisely.

For the results here, the parameters used in the simulations are: $t_I = 5$ ms, $t_{\rm SD} = 20$ ms, $t_{\rm LD} = 100$ ms, $t_{\rm ON} = 2$ ms, $N_{\rm SD} = 10$, $x_m = 2$ ms, and $m = 2000$ ms. We generated the traffic trace by gradually increasing α with the step of 0.1. Then, we entered the traffic traces into the DRX entity one by one, and the simulations are deployed for each trace. After each simulation, the statistics of DRX operations are collected. Based on these statistics, we could calculate the 1) stationary distribution of EMC, the 2) mean sojourn time in each state, the 3) limit distribution of DTSMP, 4) PS, and 5) $\mathbb{E}(D)$ by the definitions described in Section III. We also calculated the abovelisted numerical characteristics by using the proposed theoretical models. The comparison between the simulation results and analytical results is shown in Figs. 9–13. The well-matched results demonstrating the numerical equations are correctly derived.

With the help of these insight numerical characteristics, we can find out what happens when the traffic condition changes and while the UE deploys a fixed DRX configuration. Fig. 8 shows the accuracy of the analytical model of the stationary distribution at each state in the EMC. It is not surprising to observe the increase of the steady-state probability of "S1" and "S2" and the decrease of the stationary probability of "S3." With the increase of α, the packet interarrival times are gathered in a small-range near zero by (5), which implied that there are much more packet arrivals before t_I and $N_{\rm SD}$ expired. This led to the results that the UE always stayed in the **Active period** and **Light sleep period**. Hence, the probability of a packet interarrival time being larger than $N_{\rm SD}t_{\rm SD} + t_I$ is highly reduced, which led to the decrease of stationary probability of the **Deep sleep period**. In Fig. 9, we can clearly observe that, with the increase of α, the mean sojourn time is increased by 53% for the **Active period** and decreased by 21% and 25% for the **Light sleep period** and **Deep sleep period**, respectively. According to the definition of the limit distribution of DTSMP shown in (19), the variation trend of limit distribution is decided by the trend of the EMC's stationary probability and mean sojourn time in each state. Thus, Fig. 10 revealed the comprehensive situation of the DRX operations with the increasing traffic. The limit steady-state probability of the **Light sleep period** and **Deep sleep period** reduced by 10% and 96%, whereas α increased from 1.1 to 1.9. It suggests that the fixed DRX parameter configurations could not obtain good power-saving performance for varying traffic conditions, which could be directly seen by the degradation of PS in Fig. 11. Since we do not consider the buffer and service delay in this work, $\mathbb{E}(D)$ only reflected the delay caused by the DRX mechanism. As expected, while deploying the fixed DRX parameter configuration, the DRX mechanism was increasingly ineffective with the increase of

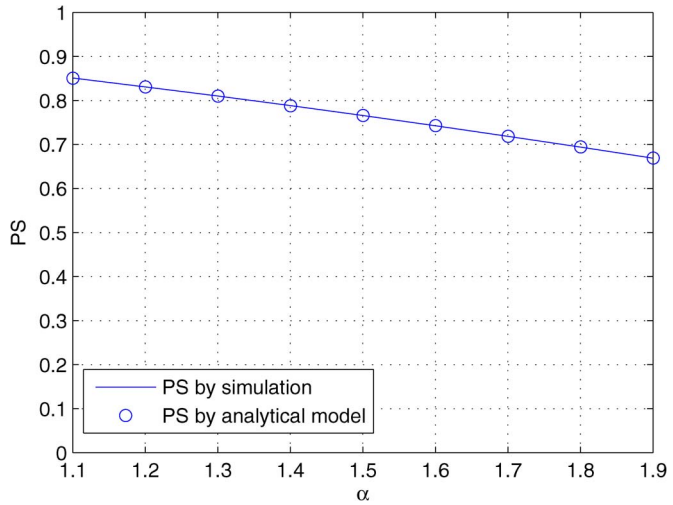

Fig. 11. Validate the analytical model of PS.

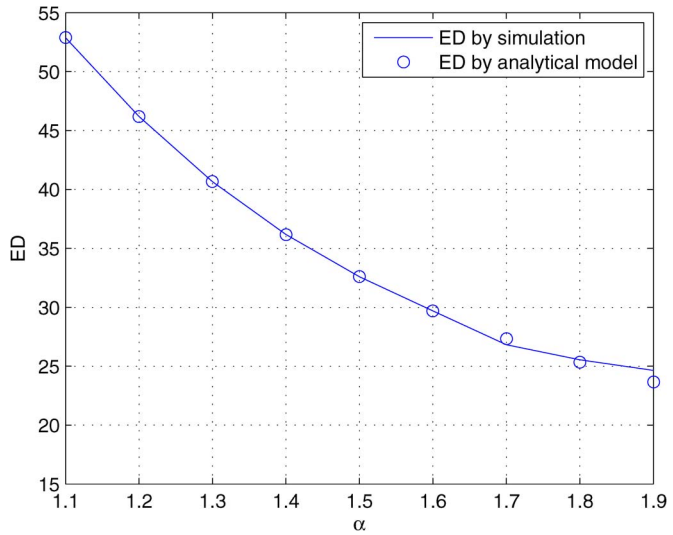

Fig. 12. Validate the analytical model of $\mathbb{E}(D)$.

traffic; therefore, the decrease of $\mathbb{E}(D)$ took place, as shown in Fig. 12. We concluded that to obtain a good performance for varying traffic, a tradeoff between PS and $\mathbb{E}(D)$ must be made.

B. Evaluation of the OPSS

To evaluate the performance of our strategy, we compared the OPSS with three fixed DRX cases and measured PS and $\mathbb{E}(D)$ for various α. The results are reported in Fig. 13. Specifically, the fixed DRX parameter sets for cases 1, 2, and 3 are summarized in Table IV. The selectable parameters are defined as follows: $t_I \in \{1, 2, \ldots, 20\}$, $t_{\mathrm{SD}} \in \{10, 15, \ldots, 50\}$, $t_{\mathrm{LD}} \in \{100, 150, \ldots, 1000\}$, and $N_{\mathrm{SD}} \in \{5, 10, 15, 20\}$. In addition, the η for all cases is equal to 0.1. The traffic trace is generated following the parameters of $x_m = 2$ and $m = 2000$. We assumed that $\gamma_d \leq \gamma_{ps}$ while $\alpha \in [1.1, 1.5]$, which indicated that the OPSS should select the DRX parameters that satisfied $\mathrm{PS}_{\mathrm{th}}$ and simultaneously minimize the delay. We also assumed that $\gamma_{ps} < \gamma_d$ while $\alpha \in (1.5, 1.9]$, which indicated that the OPSS should select the DRX parameters that satisfied $t_{\max}$ and

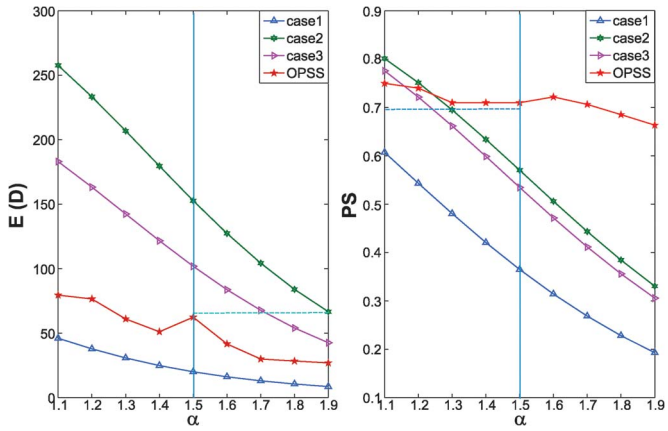

Fig. 13. Evaluation of OPSS.

TABLE IV
FIXED DRX PARAMETER SETS FOR CASES 1, 2, AND 3

case	t_I	t_{SD}	t_{LD}	N_{SD}
1	15	20	100	15
2	10	10	200	5
3	10	10	150	5

maximized PS as well. Particularly, the $\mathrm{PS}_{\mathrm{th}}$ is set as 0.7, and $t_{\max}$ is set as 75 ms in the simulation.

First, we will discuss the situation of $\alpha \in [1.1, 1.5]$ (the left half of each subgraph). In cases 1 and 2, although PS was higher than $\mathrm{PS}_{\mathrm{th}}$, while $\alpha < 1.3$, the mean packet delay was very high. On the other hand, case 3 could obtain good delay performance, but PS was much lower than $\mathrm{PS}_{\mathrm{th}}$. It can be seen that the performance of OPSS could satisfy $\mathrm{PS}_{\mathrm{th}}$ with feasible DRX parameters. Meanwhile, the OPSS could provide dynamic adjustment of DRX parameters under different traffic loads, which could maintain the mean packet delay at a low level.

Next, we will discuss the situation of $\alpha \in [1.6, 1, 9]$ (the right half of each subgraph). It is impressive that the OPSS not only could satisfy the delay constraint but obtained a good power-saving performance as well, which was much better than the fixed parameter strategies. It is worth noting that the advantage of the OPSS came from the potentials of a large number of DRX parameter combinations. It also showed that, by taking all the DRX parameters into consideration, the tradeoff between PS and mean packet delay is not as simple, as shown in [12] and [15]. There could be an optimal DRX parameter configuration under a specific traffic condition, which could achieve the best power-saving (or delay) performance, and have good delay (or power-saving) performance simultaneously. We will leave that for our future work to explain how to efficiently find the optimal DRX parameter configuration from the feasible parameter set.

VII. CONCLUSION

In this paper, we have studied optimizing LTE DRX operations under self-similar traffic. Based on the self-similar traffic model and our previously proposed DRX analytical model, we designed an OPSS. The OPSS could unbiasedly estimate the traffic condition and DRX operations within 1 s and optimize the DRX parameter configurations. Therefore, the OPSS could enhance the DRX operations with a better power-saving

performance while still meeting the packet delay constraint. Extensive MATLAB simulations showed that the OPSS could outperform the conventional LTE DRX mechanism in terms of both energy conservation and packet delay.

APPENDIX

Algorithm 1 OPSS

Estimition phase:

During the observing window T_{ob}:

- eNB records the samples of packet interarrival times $\{x_1, x_2, \ldots, x_n\}$ for the specific UE.
- UE records N_i and N_{ij} $(i, j \in E)$.

After T_{ob}:

- UE estimates $\hat{\nu}_i, (i \in E), \hat{U}_i, (i \in E - \{3\})$ by (38) and (39), respectively, and then sent them to eNB in the first *on_duration*.
- Based on the updates from UE, the eNB estimated the $U_3, \bar{U}, \pi_i$ by (17), (40), and (41), respectively. Meanwhile, eNB also estimated $\hat{X}_m$ and α^* by (26) and (34), respectively. Then, eNB obtained $\mathrm{PS}_{\mathrm{now}}$ and $\mathbb{E}(D)_{\mathrm{now}}$ by (20) and (23), respectively, at the end of this phase. We denote the current DRX parameter set as κ_{now}

Optimization phase:

Update $\gamma_d, \gamma_{\mathrm{ps}}$

if $\gamma_d > \gamma_{\mathrm{ps}}$ **then**

- Update $t_{\max}$;
- Look up the table Ω, and find out the feasible DRX configuration set ω, in which all entries κ could satisfy: $\mathbb{E}(D) < t_{\max}$.
- Find out the largest PS in the feasible set, denote as $\mathrm{PS}_{\max}$, and denote the corresponding DRX parameter set as $\kappa_{\max}$. Moreover, denote the optimized DRX parameter set as κ_{opt}.

if $\mathrm{PS}_{\mathrm{now}} \geq \mathrm{PS}_{\max}$ **then**
if $\mathbb{E}(D) \leq t_{\max}$ **then**
$\kappa_{\mathrm{opt}} \leftarrow \kappa_{\mathrm{now}}$
else
$\kappa_{\mathrm{opt}} \leftarrow \kappa_{\max}$
end if
else
$\kappa_{\mathrm{opt}} \leftarrow \kappa_{\max}$
end if
else

- Update $\mathrm{PS}_{\mathrm{th}}$
- Look up the table Ω, and find out the feasible set ω which satisfies: $\mathrm{PS} > \mathrm{PS}_{\mathrm{th}}$
- Find out the smallest $\mathbb{E}(D)$ in the feasible set, denote as $\mathbb{E}(D)_{\min}$, and denote the corresponding DRX parameter set as $\kappa_{\max}$. Moreover, denote the optimized DRX parameter set as κ_{opt}.

if $\mathrm{PS}_{\mathrm{now}} \geq \mathrm{PS}_{\mathrm{th}}$ **then**
if $\mathbb{E}(D) \leq \mathbb{E}(D)_{\min}$ **then**
$\kappa_{\mathrm{opt}} \leftarrow \kappa_{\mathrm{now}}$
else
$\kappa_{\mathrm{opt}} \leftarrow \kappa_{\max}$
end if
else
$\kappa_{\mathrm{opt}} \leftarrow \kappa_{\max}$
end if
end if

REFERENCES

[1] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36–43, May 2014.

[2] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interference for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16-2004, 2004.

[3] S.-R. Yang, S.-Y. Yan, and H.-N. Hung, "Modeling UMTS power saving with bursty packet data traffic," *IEEE Trans. Mobile Comput.*, vol. 6, no. 12, pp. 1398–1409, Dec. 2007.

[4] *More Battery Life with LTE Connected DRX*. [Online]. Available: http://www.nokiasiemensnetworks.com

[5] *Medium Access Control (MAC) Protocol Specification*, 3GPP TS 36.321 Std. 10.2.0 (Rel. 10), 2011.

[6] "LTE RAN enhancements for diverse data application," Third-Generation Partnership Project (3GPP), Sophia-Antipolis, France, 3GPP TR 36.822 Tech. Rep. 11.0.0 (Rel. 11), 2012.

[7] "DRX parameters in LTE," Nokia, Espoo, Finland, Tech. Rep. R2-071285, 2007.

[8] "On the need for flexible DRX," Nokia, Espoo, Finland, Tech. Rep. R2-071286, 2007.

[9] M. Polignano *et al.*, "Power savings and QoS impact for VoIP application with DRX/DTX feature in LTE," in *Proc. IEEE 73rd VTC—Spring*, 2011, pp. 1–5.

[10] S. C. Jha, A. T. Koç, and R. Vannithamby, "Optimization of discontinuous reception (DRX) for mobile Internet applications over LTE," in *Proc. IEEE VTC—Fall*, 2012, pp. 1–5.

[11] L. Zhou *et al.*, "Performance analysis of power saving mechanism with adjustable DRX cycles in 3GPP LTE," in *Proc. IEEE 68th VTC—Fall*, 2008, pp. 1–5.

[12] S. Fowler, R. S. Bhamber, and A. Mellouk, "Analysis of adjustable and fixed DRX mechanism for power saving in LTE/LTE-advanced," in *Proc. IEEE ICC*, 2012, pp. 1964–1969.

[13] J. Wigard, T. Kolding, L. Dalsgaard, and C. Coletti, "On the user performance of LTE UE power savings schemes with discontinuous reception in LTE," in *Proc. IEEE ICC Workshops*, 2009, pp. 1–5.

[14] S. Alouf, V. Mancuso, and N. C. Fofack, "Analysis of power saving and its impact on web traffic in cellular networks with continuous connectivity," *Pervasive Mobile Comput.*, vol. 8, no. 5, pp. 646–661, Oct. 2012.

[15] 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.

[16] Y.-P. Yu and K.-T. Feng, "Traffic-based DRX cycles adjustment scheme for 3GPP LTE systems," in *Proc. IEEE 75th VTC—Spring*, 2012, pp. 1–5.

[17] C.-H. Hsu, K.-T. Feng, and C.-J. Chang, "Statistical control approach for sleep-mode operations in IEEE 802.16 m systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4453–4466, Nov. 2010.

[18] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, Jun. 1995.

[19] Z. Sahinoglu and S. Tekinay, "On multimedia networks: Self-similar traffic and network performance," *IEEE Commun. Mag.*, vol. 37, no. 1, pp. 48–52, Jan. 1999.

[20] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[21] B. Mandelbrot, "Self-similar error clusters in communication systems and the concept of conditional stationarity," *IEEE Trans. Commun. Technol.*, vol. COM-13, no. 1, pp. 71–90, Mar. 1965.

[22] K. Wang, X. Li, and H. Ji, "Modeling and optimizing the LTE discontinuous reception mechanism under self-similar traffic," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1238–1241, Jul. 2013.

[23] S. Jin, X. Chen, D. Qiao, and S. Choi, "Adaptive sleep mode management in IEEE 802.16 m wireless metropolitan area networks," *Comput. Netw.*, vol. 55, no. 16, pp. 3774–3783, Nov. 2011.

[24] V. Barbu, J. Bulla, and A. Maruotti, "Estimation of the stationary distribution of a semi-Markov chain," *J. Reliab. Stat. Studies*, vol. 5, pp. 15–26, Apr. 2012.

[25] "Feasibility study for OFDM for UTRAN enhancement," Third-Generation Partnership Project (3GPP), Sophia-Antipolis, France, 3GPP TR 25.892 Tech. Rep. 2.0.0 (Rel. 6), 2004.
[26] H. Tanizaki, *Computational Methods in Statistics and Econometrics*. Boca Raton, FL, USA: CRC, 2004, vol. 172.
[27] "Universal Mobile Telecommunications System (UMTS): Selection procedures for the choice of radio transmission technologies of the UMTS," ETSI, Sophia-Antipolis, France, ETSI 30.03 Tech. Rep. 3.2.0, 1998.
[28] V. Barbu, M. Boussemart, and N. Limnios, "Discrete-time semi-Markov model for reliability and survival analysis," *Commun. Stat.–Theory Methods*, vol. 33, no. 11, pp. 2833–2868, Jan. 2004.
[29] C. Bontu and E. Illidge, "DRX mechanism for power saving in LTE," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 48–55, Jun. 2009.
[30] R. G. Gallager, *Discrete Stochastic Processes*, vol. 101. Boston, MA, USA: Kluwer, 1996.
[31] J. L. Petersen, "Estimating the parameters of a pareto distribution," Univ. Montana, Missoula, MT, USA Tech. Rep., 2000.

Ke Wang received the Ph.D. degree in communication and information systems from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014.

He is currently a Lecturer with the School of Information and Communication Engineering, BUPT. His research interests include scheduling algorithms and energy-efficient communications.

Dr. Wang has served as a Reviewer for several journals and conference proceedings, including the IEEE COMMUNICATION LETTERS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Elsevier Computer Communications*, and the 2013 IEEE Global Communications Conference (Globecom). He also served as a Technical Program Committee Member for Globecom 2014 and the 2014 International Conference on Communications. He received the postgraduate innovation fund from the Rohde and Schwarz Beijing University of Posts and Telecommunications in 2013.

Xi Li received the B.E. and Ph.D. degrees in communication and information systems from Beijing University of Posts and Telecommunications, Beijing, China, in 2005 and 2010, respectively.

She is an Associate Professor with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. Her current research interests include broadband wireless communications and heterogeneous networks.

Dr. Li has served as the Chair of the special track on cognitive testbed at the 2011 International ICST Conference on Communications and Networking in China and a Technical Program Committee Member of the track on MAC layer at the 2012/2013 IEEE Wireless Communications and Networking Conference and on MAC and cross-layer design at the 2012 IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications.

Hong Ji (SM'09) received the B.S. degree in communications engineering and the M.S. and Ph.D. degrees in information and communications engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1989, 1992, and 2002, respectively.

From June to December 2006, she was a Visiting Scholar with the University of British Columbia, Vancouver, BC, Canada. She is currently a Professor with BUPT. She is also engaged in national science research projects, including the Hi-Tech Research and Development Program of China (863 Program) and the National Natural Science Foundation of China. Her research interests include heterogeneous networks, peer-to-peer protocols, cognitive radio networks, relay networks, LTE/5G, and cooperative communication.

Dr. Ji serves on the Editorial Board of several journals, including the Wiley *International Journal of Wireless Communications and Mobile Computing* and the Wiley *International Journal of Communication Systems*. She has also served on the Technical Program Committee (TPC) of numerous conferences, e.g., as a TPC member of the 2010/2011/2012/2013 IEEE International Conference on Communications, the 2010/2011/2012/2013 IEEE Global Communications Conference, the 2012/2014 IEEE Wireless Communications and Networking Conference, the 2010/2011 IEEE INFOCOM Workshop, and the 2012 IEEE International Conference on Computer, Information, and Telecommunication Systems, as a TPC Cochair of the 2011 CHINACOM Workshop, and as a Program Cochair of the 2012/2013/2014 International Conference on Wireless Information Networks and Systems.

Xiaojiang (James) Du (SM'03) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 2002 and 2003, respectively, all in electrical engineering.

He is currently an Associate Professor with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA, USA. Between August 2004 and July 2009, he was an Assistant Professor with the Department of Computer Science, North Dakota State University, Fargo, ND, USA, where he received the Excellence in Research Award in May 2009. He has published over 120 journal and conference papers in his areas of interest. His research interests include security, cloud computing, wireless networks, and computer networks and systems.

Dr. Du has received more than $3M in research grants from the National Science Foundation (NSF), Army Research Office, Air Force Research Lab, NASA, the Commonwealth of Pennsylvania, and Amazon. He serves on the Editorial Board of four international journals. He will serve as the Lead Chair of the Communication and Information Security Symposium of the 2015 IEEE International Conference on Communications and a Cochair of the Mobile and Wireless Networks Track of the 2015 IEEE Wireless Communications and Networking Conference. He was the Chair of the Computer and Network Security Symposium of the 2006–2010 IEEE/ACM International Wireless Communication and Mobile Computing Conference. He is or has been a Technical Program Committee (TPC) Member of several premier ACM/IEEE conferences, such as INFOCOM (2007–2015), IM, NOMS, ICC, GLOBECOM, WCNC, BroadNet, and IPCCC. He is a Life Member of the ACM.