QoE Based Random Sleep-Awake Scheduling in Heterogeneous Cellular Networks

Abbas Farrokhi, Ozgur Ercetin
Faculty of Engineering and Natural Sciences,
Sabanci University
Istanbul, TURKEY
Email: afarrokhi@sabanciuniv.edu, oercetin@sabanciuniv.edu

***Abstract*—In this paper, we investigate an optimal resource on-off switching framework that minimizes the energy consumption of a heterogeneous cellular network. Specifically, our goal is to minimize the energy consumption of the cellular network while satisfying the quality of user experience (QoE, e.g., buffer starvation probability). For an ON/OFF bursty arrival process, we introduce recursive equations to obtain the buffer starvation probability of a mobile device (MD) for streaming services. The MD is in the coverage area of a femtocell base station (FBS) which is implemented at the cell edge of a macrocell base station (MBS), and when its buffer gets empty, the media player of the MD restarts the service after a certain amount of packets are prefetched (this event is known as start-up delay in the literature). Numerical simulations illustrate how our system significantly reduces the overall energy consumption of the network while guaranteeing a target starvation probability in comparison to the case where the MD is covered only by one MBS.**

***Index Terms*—Energy efficiency, heterogeneous cellular network, sleep/awake strategy, buffer starvation, start-up delay, quality of experience.**

I. Introduction

With the extensive popularization of smart mobile terminals and the resulting increasing mobile internet traffic, the conventional homogeneous cellular networks which consist of MBSs have faced a great challenge to meet this overwhelming demand of network capacity. In order to address this issue and provide better coverage, heterogeneous networks have been introduced in the LTE-Advanced standardization [1], [2], [3]. A heterogeneous network uses a mixture of macrocells and small cells such as microcells, picocells, and femtocells. Moreover, meeting this overwhelming traffic demand led to a significant increase in the power consumption and the operating cost of a cellular network [4]. The rapid increase in energy cost and CO_2 emissions has made the network operators realize the importance of designing their networks in an energy efficient manner. The energy consumption of the cellular networks mostly comes from the BSs, which consume about 60% to 80% of that of the whole network [5]. And since the energy consumption of a BS mainly comes from the cooling, controller, baseband signal processor and other circuits (in literature it is known as the fixed power consumption of a BS), rather than the transmit power which consumes only 3.1% [6], turning BSs into sleep mode whenever possible is a promising strategy to reduce the energy consumption. Because of the high fluctuations in traffic demand over space and time in cellular networks [7], some BSs could be switched off when the traffic load in their coverage area is low, and the users in sleeping cells can be served by neighboring active BSs [8]. Nevertheless, applying sleep/active strategy and turning some BSs into sleep mode may deteriorate the Quality of Service (QoS). Therefore, in order to make a tradeoff between QoS and energy efficiency of cellular networks, researchers have been investigating different active/sleep schedules while guaranteeing acceptable QoS such as delay [5], coverage performance [9], blocking probability [10]. In this paper, we focus on the QoE of a user, where it is the starvation probability of user buffer. The probability of buffer starvation, as an important performance measure, has various applications in different fields, such as video streaming services. The event of starvation happens when the buffer gets empty, and after each such event, the media player of the MD resumes the service when there is a certain amount of packets accumulated in the buffer (prefetching). Therefore, the media streaming service is under the influence of two factors which are the prefetching process and the starvation event. In fact, as the prefetching process gets shorter the starvation event occurs with a higher probability, and a longer prefetching process results in a larger start-up (initial buffering) delay.

In this paper, we introduce recursive equations, based on the approach in [11], to obtain the starvation probability of a buffer for an ON/OFF bursty arrivals and in a time-slotted queuing system. Unlike [12] where the authors obtain the buffer starvation probability of a user that could be served only through a single source, we evaluate the starvation probability while the MD is within the coverage area of two BSs, namely MBS and FBS, i.e., the MD may be served by either of these BSs depending on which one is in active mode. Note that, analyzing the aggregated active/sleep period length distribution analytically has been an unsolved challenging problem in the literature. In [13], the authors use Monte Carlo simulations to investigate the characteristics of the OFF-period length distribution in an aggregated ON/OFF process. In our work, using a three state Markov chain and applying the first step analysis, we investigate analytically the aggregated active/sleep period length distribution for the first time in the literature.

This work was supported in part by a grant from Argela Technologies, Turkey.

The rest of the paper is organized as follows: In Section II, we describe the system model under consideration. In Section III, we present the calculation of buffer starvation probability with an ON/OFF bursty arrival. In Section IV we write down the optimization problem. In Section V, we validate our analysis via simulation results. Lastly, our conclusions are given in Section VI.

II. SYSTEM DESCRIPTION

A. Network Model in a Discrete-Time System

We consider a heterogeneous cellular network consisting of two base stations where a FBS is implemented within the coverage area of a MBS. Our main goal is to optimize the energy consumption of this heterogeneous cellular network while satisfying user QoE, which in this work is guaranteeing a target buffer starvation probability for streaming services. To this end, we consider a single media file with finite size N. The media content is pre-stored in the media server (e.g., video on demand (VoD) service). After a request by the MD, the server (either MBS or FBS) segments the file into packets and transfers them to the MD. In order to correctly model the packet arrivals to the media player of the MD, we should consider several important points. Firstly, we consider an ON/OFF bursty traffic model where the sources (BSs) may stay for relatively long durations in ON and OFF modes, and the packet arrival occurs only when a BS is in ON mode. Secondly, we divide the time into small slots with duration h, and denote by ρ_m, ρ_f the probability of packet arrival from MBS and FBS to the media player of MD in a time slot, respectively. Assuming that in a continuous-time scenario the packet arrival from MBS and FBS is modeled according to poisson processes with rates λ_m and λ_f, respectively, ρ_m and ρ_f can be defined as

$$\rho_m = (\lambda_m h)e^{-\lambda_m h}, \tag{1}$$

$$\rho_f = (\lambda_f h)e^{-\lambda_f h}. \tag{2}$$

Thirdly, we denote by ψ_m the probability that MBS is active, given that the *system* is in ON mode, and by ψ_f the probability that FBS is active, given that the *system* is in ON mode. Note that the *system* is active whenever either MBS or FBS is in ON mode. To obtain the probabilities ψ_m, ψ_f, we use the Markov chain of our system model as shown in Fig. 1. The state space of this Markov chain is $\{(s_m(k), s_f(k)) \ : s_i(k) \in (\text{ON}, \text{OFF}) \ ; \ i = m, f\}$, where $s_m(k)$, $s_f(k)$ denote the state of MBS and FBS at time k, and t_{ij} denotes the transition probability from state i to state j. In Fig. 1, the states 0, 1, 2 denote the state space (OFF,OFF), (OFF,ON), and (ON,OFF), respectively. In this model, we consider the users that are in the coverage area of FBS which is implemented within the MBS's area, and thus only one base station is needed to be active for serving these users.

We denote by π_1, π_2 the steady state probabilities of states 1 and 2, i.e., the proportion of time that each FBS and MBS is

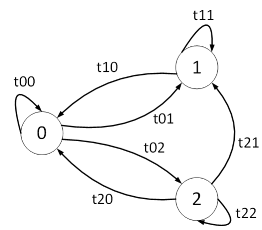

Fig. 1: Markov Chain

in active mode at the steady state. Therefore, we obtain ψ_m and ψ_f as follows.

$$\psi_m = \frac{\pi_2}{\pi_2 + \pi_1}, \tag{3}$$

$$\psi_f = \frac{\pi_1}{\pi_2 + \pi_1}. \tag{4}$$

Using equations (1)-(4), we obtain the probability of packet arrival to the media player of a MD during a time slot h as follows.

$$\varsigma = \psi_m \rho_m + \psi_f \rho_f. \tag{5}$$

The probability ς denotes the probability of packet arrival from either MBS or FBS to the MD's buffer during a time-slot h. We model the arrival process as a bernoulli process with this success probability ς. In addition, we assume that at the buffer of a MD packet departure follows an exponential distribution with rate μ. Using this assumption we obtain the probability of packet departure, denoted by ω, in a time-slot h as follows.

$$\omega = 1 - e^{-\mu h}. \tag{6}$$

Considering ω as the probability that a packet completes its service during a small time slot h, we model the service process at the media player of the MD as a bernoulli process with probability ω.

B. Base Stations Active/Sleep Schedules

The active/sleep period durations of BSs are modeled as four independent and identically distributed (i.i.d.) random variables. More specifically, we model the active period durations of MBS and FBS according to an exponential distribution with rates α_m and α_f, respectively. The sleep period durations of the BSs are modeled as exponential distributions with rates β_m, β_f for macrocell and femtocell, respectively. Recall that we are considering those users that are under the coverage area of femtocell base station, so the users could be served by FBS or MBS depending on which one is active. However, we assume that the arrival rate from a FBS is more than that a MBS provides to the MDs.

C. Energy Consumption Model

The expected energy consumption of this cellular network is given by $E_{total} = E_f + \gamma E_m$, where E_m, E_f denote the expected energy consumptions of MBS and FBS, respectively. In this model, we assume that femtocell's energy consumption is $1/\gamma$ of that a MBS consumes per unit time. Note that E_f and E_m

are proportional to the time that each FBS and MBS spends in active mode. Hence to obtain the values of E_f and E_m using the Markov chain shown in Fig. 1, we obtain the steady state probabilities of states 1 and 2, respectively.

III. Probability of Buffer Starvation for an ON/OFF Bursty Traffic

In this section, we define a recursive approach to obtain the buffer starvation probability of a MD that is in the coverage area of the FBS, and FBS is implemented within the coverage of the MBS, i.e. the mobile device could be served by both BSs. We denote by $P_i(n)$ the probability of starvation for a file of n packets, given that there are i packets in the buffer of the MD upon arrival of the first packet of this file. In our system, we aim to obtain the starvation probability in downloading a file of size N while x packets of this file (x packets out of N packets) are prefetched before the service begins. Therefore, the starvation probability in our system model corresponds to $P_i(n)$ with $i = x-1$ and $n = N-x+1$. To compute $P_i(n)$, we introduce recursive equations. To this end, we define a quantity $Q_i^{ON}(k)$, $0 \le i \le N-1$, $0 \le k \le i$, which is the probability that k packets out of i leave the MD's buffer upon an arrival at the ON state, i.e., there is no packet arrival when the system is in OFF mode (both BSs are switched off). To apply the recursive equations, we start from the case $n = 1$.

$$P_i(1) = 0, \quad \forall i \ge 1. \tag{7}$$

When the file size is 1 and the only packet observes a non-empty queue, the probability of starvation is zero. If i is zero, i.e. upon arrival we find the buffer empty, the starvation occurs for sure, thus yielding

$$P_0(n) = 1, \quad n = 1, ..., N. \tag{8}$$

For $n \ge 2$, we have the following recursive equation:

$$P_i(n) = \sum_{k=0}^{i+1} Q_{i+1}^{ON}(k) P_{i+1-k}(n-1), \quad 0 \le i \le N-1. \tag{9}$$

According to (9), when the first packet of the file arrives and finds i packets in the system, the starvation does not happen. However, the starvation might happen in the service of remaining $n-1$ packets. Upon the arrival of the next packet, k packets out of $i+1$ leave the system with probability $Q_{i+1}^{ON}(k)$. Since the total number of packets is N, the starvation probability must satisfy $P_i(n) = 0$ for $i+n > N$. In order to obtain $P_i(n)$ using (9), we should first obtain the term $Q_i^{ON}(k)$.

A. Calculating $Q_i^{ON}(k)$

Note that $Q_i^{ON}(k)$ is the probability that k packets out of i leave the buffer of the MD during an inter-arrival period. First, we denote the random variable (r.v.) of inter-arrival period by τ, and let $T(z) = \mathsf{E}[z^\tau]$ be its probability generating function. Secondly, we denote by ν the r.v. of the number of packets that leave the MD's buffer during an inter-arrival period, and let $N(z) = \mathsf{E}[z^\nu]$ be its probability generating function. Using the probability generating function $T(z)$, we obtain the probability generating function of the number of bernoulli departures, with a success probability as defined in (6), during the inter-arrival period τ, i.e. we obtain $N(z)$ from $T(z)$. Finally, by evaluating the inverse transform of $N(z)$, we obtain the probability mass function (pmf) of r.v. ν, from which we obtain the term $Q_i^{ON}(k)$.

B. Probability Generating Function of Inter-Arrival Period τ

Considering that an arbitrary packet has been generated by the system, we denote the time period from the instant at which this packet is generated until the point when the system goes to sleep mode, i.e., both MBS and FBS goes to sleep mode, by active period number 1, and the following sleep period by sleep period number 1. Then, we number the subsequent active (sleep) periods by the numbers $2, 3, \ldots$. We define the event ϕ_m as the event in which the next packet arrives during active period number m, (m=$1, 2, \ldots$). The probability of ϕ_m is given as follows.

$$\mathsf{Pr}(\phi_m) = q^{m-1} p, \quad m \ge 1, \tag{10}$$

where p denotes the probability of packet arrival in an active period, and $q = 1-p$. In other words, the probability p denotes the event in which the time duration from the beginning of an active period until the next packet arrival is less than or equal to the duration of that active period. We let r.v. R denote the time duration from the beginning of an active period until the next packet arrival in that active period, and r.v. Y denote the time duration of an active period. According to our system model which is shown in Fig. 1, and using the first step analysis we obtain the probability mass function (pmf) of r.v. Y as follows.

$$T_1(1) = t_{10}, \tag{11}$$
$$T_1(k) = t_{11} T_1(k-1), \tag{12}$$
$$T_2(1) = t_{20}, \tag{13}$$
$$T_2(k) = t_{22} T_2(k-1) + t_{21} T_1(k-1), \tag{14}$$

where $T_i(.)$ denotes the number of steps that it takes to get to state zero, given that we are initially at state i(i=$1, 2$), and t_{ij} denotes the transition probability from state i to state j. To obtain $T_1(k)$ in a closed formula, we rewrite (12) as follows:

$$\begin{aligned} T_1(k) &= t_{11} T_1(k-1) = t_{11}^2 T_1(k-2) = \ldots \\ &= t_{11}^{k-1} T_1(1) = t_{11}^{k-1} t_{10}, \quad k = 1, 2, 3, \ldots. \end{aligned} \tag{15}$$

To obtain $T_2(k)$ in a closed formula, we rewrite (14) as follows.

$$\begin{aligned} T_2(k) &= t_{22} T_2(k-1) + t_{21} T_1(k-1) \\ &= t_{22}[t_{22} T_2(k-2) + t_{21} T_1(k-2)] + t_{21} T_1(k-1) \\ &= t_{22}^2 T_2(k-2) + t_{22} t_{21} T_1(k-2) + t_{21} T_1(k-1) = \ldots \\ &= t_{22}^{k-1} T_2(1) + t_{22}^{k-2} t_{21} T_1(1) + t_{22}^{k-3} t_{21} T_1(2) + \ldots + t_{21} T_1(k-1). \end{aligned} \tag{16}$$

Inserting (13) and (15) in (16) results in:

$$
\begin{aligned}
T_2(k) &= t_{22}^{k-1}t_{20} + t_{22}^{k-2}t_{21}t_{10} + t_{22}^{k-3}t_{21}t_{11}t_{10} + \ldots + t_{21}t_{11}^{k-2}t_{10} \\
&= t_{22}^{k-1}t_{20} + t_{21}t_{10}[t_{22}^{k-2} + t_{22}^{k-3}t_{11} + t_{22}^{k-4}t_{11}^{2} + \ldots + t_{11}^{k-2}] \\
&= t_{22}^{k-1}t_{20} + t_{21}t_{10}t_{22}^{k-2}\sum_{r=2}^{k}\left(\frac{t_{11}}{t_{22}}\right)^{r-2} \\
&= t_{22}^{k-1}t_{20} + t_{21}t_{10}\left(\frac{t_{22}^{k-1} - t_{11}^{k-1}}{t_{22} - t_{11}}\right), \quad k = 1, 2, 3, \ldots
\end{aligned} \tag{17}
$$

Using (15) and (17), we obtain the aggregated active period length distribution as follows.

$$
F_Y(y) = t_{11}^{y-1}t_{10}\psi_f + \left(t_{22}^{y-1}t_{20} + \frac{t_{21}t_{10}(t_{22}^{y-1} - t_{11}^{y-1})}{t_{22} - t_{11}}\right)\psi_m, \tag{18}
$$

where t_{ij} denotes the transition probability from state i to state j. Considering that the packet arrival to the MD's buffer is modeled as a bernoulli process with a success probability defined in (5), we obtain the pmf of r.v. R as follows.

$$
F_R(r) = \varsigma(1-\varsigma)^{r-1}, \quad r = 1, 2, 3, \ldots \tag{19}
$$

Now, by the use of $F_Y(y)$ and $F_R(r)$ we obtain the probability of packet arrival during an active period as follows.

$$
\begin{aligned}
p &= \Pr(R \leq Y) = \sum_{y=1}^{\infty} F_Y(y)\sum_{r=1}^{y} F_R(r) \\
&= \sum_{y=1}^{\infty} F_Y(y)\sum_{r=1}^{y}\varsigma(1-\varsigma)^{r-1} \\
&= \sum_{y=1}^{\infty} F_Y(y)\varsigma[1 + (1-\varsigma) + (1-\varsigma)^2 + \ldots + (1-\varsigma)^{y-1}] \\
&= \sum_{y=1}^{\infty} F_Y(y)\varsigma\frac{1-(1-\varsigma)^y}{\varsigma} = \sum_{y=1}^{\infty} F_Y(y) - \sum_{y=1}^{\infty} F_Y(y)(1-\varsigma)^y \\
&= \sum_{y=1}^{\infty}(t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)t_{11}^{y-1} + (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m t_{22}^{y-1} \\
&\quad - \sum_{y=1}^{\infty}[(t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)t_{11}^{y-1}(1-\varsigma)^y \\
&\quad + (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m t_{22}^{y-1}(1-\varsigma)^y] \\
&= (t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)\frac{1}{1-t_{11}} + (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m\frac{1}{1-t_{22}} \\
&\quad - (t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)\frac{1-\varsigma}{1-(1-\varsigma)t_{11}} \\
&\quad - (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m\frac{1-\varsigma}{1-(1-\varsigma)t_{22}} \\
&= d_1 + d_2
\end{aligned} \tag{20}
$$

where the values of d_1 and d_2 are given as follows.

$$
d_1 = \left(\frac{1}{1-t_{11}} - \frac{1-\varsigma}{1-(1-\varsigma)t_{11}}\right)\left(t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m\right), \tag{21}
$$

$$
d_2 = \left(\frac{1}{1-t_{22}} - \frac{1-\varsigma}{1-(1-\varsigma)t_{22}}\right)\left(t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}}\right)\psi_m. \tag{22}
$$

In Fig. 2, we illustrate the inter-arrival time τ in terms of three subsections, i.e., C_k, S_k, D_k. Note that C_k denotes the time duration of active period number k given that this period ends before the arrival of the next packet. S_k denotes the time duration of sleep period number k. D_k denotes the time duration from the beginning of active period number k until the arrival of the next packet given that this packet has arrived in this active period. Note that C_k, S_k, D_k, $k \geq 1$, are i.i.d. random variables. In Fig. 2, the sleep events point to the time instances at which both BSs are in sleep mode, and thus the system is in sleep mode. The activation events, after a sleep event, point to the time instances at which either FBS or MBS wakes up, i.e. the system is in active mode.

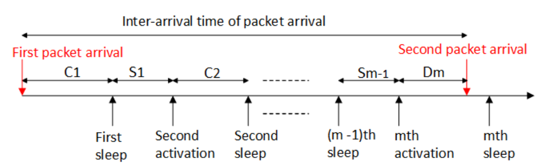

Fig. 2: Illustration of random variables τ, C_k, S_k, and D_k, $k \geq 1$, given that the next packet arrival occurs in active period number m.

Accordingly, using (10) we define the probability generating function of inter-arrival time τ as follows.

$$
\begin{aligned}
T(z) = \mathsf{E}[z^{\tau}] &= \sum_{m=1}^{\infty}\Pr(\phi_m)\mathsf{E}[z^{\tau} \mid \phi_m], \\
&= \sum_{m=1}^{\infty} pq^{m-1}\mathsf{E}[z^{\sum_{k=1}^{m-1}(C_k+S_k)+D_m}], \\
&= W(z)\sum_{m=1}^{\infty} pq^{m-1}(U(z)V(z))^{m-1}, \\
&= W(z)\frac{p}{1-qU(z)V(z)},
\end{aligned} \tag{23}
$$

where $U(z)$, $V(z)$, $W(z)$ denote the probability generating functions of random variables C_k, S_k, and D_k, $k \geq 1$, respectively.

C. Probability Generating Function of Random Variable C_k

To obtain the probability generating function $U(z)$, we first need to obtain the pmf of r.v. C_k. To this end, we first derive the probability $\Pr(C_k > m)$, and then we obtain the cdf of random variable C_k as $Z_{C_k}(m) = 1 - \Pr(C_k > m)$. Finally, from the obtained cdf, we derive the pmf of r.v. C_k

as $F_{C_k}(m) = Z_{C_k}(m) - Z_{C_k}(m-1)$.

$$\begin{aligned}
\Pr(C_k > m) &= \Pr(Y_k > m \mid N_k) = \frac{\Pr(Y_k > m,\ N_k)}{\Pr(N_k)} \\
&= \frac{\Pr(Y_k > m,\ N_{k-1},\ Y_k < R_k)}{\Pr(N_{k-1},\ Y_k < R_k)} \\
&= \frac{\Pr(m < Y_k < R_k)\Pr(N_{k-1})}{\Pr(Y_k < R_k)\Pr(N_{k-1})} \\
&= \frac{\sum_{r=m+2}^{\infty} F_{R_k}(r) \sum_{y=m+1}^{r\text{-}1} F_{Y_k}(y)}{\sum_{r=2}^{\infty} F_{R_k}(r) \sum_{y=1}^{r\text{-}1} F_{Y_k}(y)}.
\end{aligned} \tag{24}$$

In (24), if we set $m = 0$, the numerator and denominator will be the same. Hence, we first obtain the numerator and then set $m = 0$ in the obtained result to get the answer for the denominator.

$$\begin{aligned}
&\sum_{r=m+2}^{\infty} F_{R_k}(r) \sum_{y=m+1}^{r\text{-}1} F_{Y_k}(y) \\
&= \sum_{r=m+2}^{\infty} F_{R_k}(r) \sum_{y=m+1}^{r\text{-}1} [((t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)t_{11}^{y-1} \\
&\quad + (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m t_{22}^{y-1})] \\
&\quad = \sum_{r=m+2}^{\infty} F_{R_k}(r)[\frac{1}{1-t_{11}}(t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)(t_{11}^{m} - t_{11}^{r-1}) \\
&\quad + \frac{1}{1-t_{22}}(t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m(t_{22}^{m} - t_{22}^{r-1})] \\
&= \frac{\varsigma}{1-t_{11}}(t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m) \sum_{r=m+2}^{\infty}(t_{11}^{m} - t_{11}^{r-1}) \\
&\quad + \frac{\varsigma}{1-t_{22}}(t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m \sum_{r=m+2}^{\infty}(t_{22}^{m} - t_{22}^{r-1}) \\
&= c_1 t_{11}^{m}(1-\varsigma)^m + c_2 t_{22}^{m}(1-\varsigma)^m,
\end{aligned} \tag{25}$$

where c_1 and c_2 are given as follows.

$$c_1 = \frac{\varsigma(1-\varsigma)}{1-t_{11}}(\frac{1}{1-(1-\varsigma)} - \frac{t_{11}}{1-(1-\varsigma)t_{11}})(t_{10}\psi_f - \frac{t_{21}t_{10}\psi_m}{t_{22}-t_{11}}), \tag{26}$$

$$c_2 = \frac{\varsigma(1-\varsigma)}{1-t_{22}}(\frac{1}{1-(1-\varsigma)} - \frac{t_{22}}{1-(1-\varsigma)t_{22}})(t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m. \tag{27}$$

By setting $m = 0$ in (25) we obtain the denominator of (24). Hence, we rewrite the (24) as follows.

$$\begin{aligned}
\Pr(C_k > m) &= \frac{\sum_{r=m+2}^{\infty} F_{R_k}(r) \sum_{y=m+1}^{r\text{-}1} F_{Y_k}(y)}{\sum_{r=2}^{\infty} F_{R_k}(r) \sum_{y=1}^{r\text{-}1} F_{Y_k}(y)} \\
&= \frac{c_1 t_{11}^{m}(1-\varsigma)^m + c_2 t_{22}^{m}(1-\varsigma)^m}{c_1 + c_2}
\end{aligned} \tag{28}$$

From (28), we obtain the pmf of random variable C_k as follows.

$$\begin{aligned}
F_{C_k}(m) &= Z_{C_k}(m) - Z_{C_k}(m-1) \\
&= (1 - \Pr(C_k > m)) - (1 - \Pr(C_k > m\text{-}1)) \\
&= \frac{c_1}{c_1+c_2}(1 - t_{11}(1-\varsigma))t_{11}^{m-1}(1-\varsigma)^{m-1} \\
&\quad + \frac{c_2}{c_1+c_2}(1 - t_{22}(1-\varsigma))t_{22}^{m-1}(1-\varsigma)^{m-1}
\end{aligned} \tag{29}$$

Using (29) we obtain the probability generating function $U(z)$ as follows.

$$\begin{aligned}
U(z) = \mathsf{E}[z^C] &= \sum_{r=1}^{\infty} z^r F_{C_k}(r) \\
&= \frac{c_1}{c_1+c_2}(1 - t_{11}(1-\varsigma)) \sum_{r=1}^{\infty} z^r t_{11}^{r-1}(1-\varsigma)^{r-1} \\
&\quad + \frac{c_2}{c_1+c_2}(1 - t_{22}(1-\varsigma)) \sum_{r=1}^{\infty} z^r t_{22}^{r-1}(1-\varsigma)^{r-1} \\
&= \frac{c_1(1 - t_{11}(1-\varsigma))z}{(c_1+c_2)(1 - t_{11}(1-\varsigma)z)} + \frac{c_2(1 - t_{22}(1-\varsigma))z}{(c_1+c_2)(1 - t_{22}(1-\varsigma)z)}.
\end{aligned} \tag{30}$$

D. Probability Generating Function of Random Variable D_k

To obtain the probability generating function $W(z)$, we first need to obtain the pmf of r.v. D_k. To this end, we first derive the probability $\Pr(D_k > m)$, and then we obtain the cdf of random variable D_k as $Z_{D_k}(m) = 1 - \Pr(D_k > m)$. Finally, from the obtained cdf, we will end up to the pmf of r.v. D_k as $F_{D_k}(m) = Z_{D_k}(m) - Z_{D_k}(m-1)$.

$$\begin{aligned}
\Pr(D_k > m) &= \Pr(R_k > m \mid \phi_k) = \Pr(R_k > m \mid R_k \le Y_k, N_{k-1}) \\
&= \frac{\Pr(R_k > m,\ R_k \le Y_k,\ N_{k-1})}{\Pr(R_k \le Y_k,\ N_{k-1})} = \frac{\Pr(m < R_k \le Y_k)\Pr(N_{k-1})}{\Pr(R_k \le Y_k)\Pr(N_{k-1})} \\
&= \frac{\Pr(m < R_k \le Y_k)}{\Pr(R_k \le Y_k)} = \frac{\sum_{y=m+1}^{\infty} F_{Y_k}(y) \sum_{r=m+1}^{y} F_{R_k}(r)}{\sum_{y=1}^{\infty} F_{Y_k}(y) \sum_{r=1}^{y} F_{R_k}(r)},
\end{aligned} \tag{31}$$

where the term N_k denotes the event in which the next packet arrival does not happen in the first k active period. Note that the denominator in (31) is equal to p given in (20). Substituting

(18) and (19) in (31) results in

$$\begin{aligned}
\mathsf{Pr}(D_k > m) &= \frac{\sum_{y=m+1}^{\infty} F_{Y_k}(y) \sum_{r=m+1}^{y} \varsigma(1-\varsigma)^{r-1}}{p} \\
&= \frac{\sum_{y=m+1}^{\infty} F_{Y_k}(y)(1-\varsigma)^m - \sum_{y=m+1}^{\infty} F_{Y_k}(y)(1-\varsigma)^y}{p} \\
&= [\sum_{y=m+1}^{\infty} (t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)(1-\varsigma)^m t_{11}^{y-1} \\
&\quad + (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m (1-\varsigma)^m t_{22}^{y-1} \\
&\quad - \sum_{y=m+1}^{\infty} (t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}(1-\varsigma)^y \psi_m) t_{11}^{y-1}(1-\varsigma)^y \\
&\quad + (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m (1-\varsigma)^y t_{22}^{y-1}(1-\varsigma)^y \ ]/p \\
&= [(t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)\frac{t_{11}^m(1-\varsigma)^m}{1-t_{11}} \\
&\quad + (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m \frac{t_{22}^m(1-\varsigma)^m}{1-t_{22}} \\
&\quad - (t_{10}\psi_f - \frac{t_{21}t_{10}}{t_{22}-t_{11}}\psi_m)\frac{t_{11}^m(1-\varsigma)^{m+1}}{1-(1-\varsigma)t_{11}} \\
&\quad - (t_{20} + \frac{t_{21}t_{10}}{t_{22}-t_{11}})\psi_m \frac{t_{22}^m(1-\varsigma)^{m+1}}{1-(1-\varsigma)t_{22}} \ ]/p \\
&= \frac{d_1 t_{11}^m (1-\varsigma)^m + d_2 t_{22}^m (1-\varsigma)^m}{d_1+d_2}.
\end{aligned} \tag{32}$$

The values of d_1, d_2 are given in (14) and (15), respectively. From (32), we obtain the pmf of random variable D_k as follows

$$\begin{aligned}
F_{D_k}(m) &= Z_{D_k}(m) - Z_{D_k}(m-1) \\
&= (1-\mathsf{Pr}(D_k > m)) - (1-\mathsf{Pr}(D_k > m\text{-}1)) \\
&= \mathsf{Pr}(D_k > m\text{-}1) - \mathsf{Pr}(D_k > m) \\
&= \frac{d_1 t_{11}^{m-1}(1-\varsigma)^{m-1} + d_2 t_{22}^{m-1}(1-\varsigma)^{m-1}}{d_1+d_2} \\
&\quad - \frac{d_1 t_{11}^m (1-\varsigma)^m + d_2 t_{22}^m (1-\varsigma)^m}{d_1+d_2} \\
&= \frac{d_1(1-t_{11}(1-\varsigma))}{d_1+d_2} t_{11}^{m-1}(1-\varsigma)^{m-1} \\
&\quad + \frac{d_2(1-t_{22}(1-\varsigma))}{d_1+d_2} t_{22}^{m-1}(1-\varsigma)^{m-1}.
\end{aligned} \tag{33}$$

Using (33) we obtain the probability generating function $W(z)$ as follows:

$$\begin{aligned}
W(z) &= \mathsf{E}[z^D] = \sum_{r=1}^{\infty} z^r F_{D_k}(r) \\
&= \frac{d_1(1-t_{11}(1-\varsigma))}{d_1+d_2} \sum_{r=1}^{\infty} z^r t_{11}^{r-1}(1-\varsigma)^{r-1} \\
&\quad + \frac{d_2(1-t_{22}(1-\varsigma))}{d_1+d_2} \sum_{r=1}^{\infty} z^r t_{22}^{r-1}(1-\varsigma)^{r-1} \\
&= \frac{d_1(1-t_{11}(1-\varsigma))z}{(d_1+d_2)(1-t_{11}(1-\varsigma)z)} + \frac{d_2(1-t_{22}(1-\varsigma))z}{(d_1+d_2)(1-t_{22}(1-\varsigma)z)}.
\end{aligned} \tag{34}$$

E. Probability Generating Function of Random Variable S_k

Using the Markov chain shown in Fig. 1, we obtain the pmf of sleep period as follows.

$$F_{S_k}(m) = t_{00}^{m-1}(t_{01}+t_{02}). \tag{35}$$

From this pmf we obtain the probability generating function of r.v. S_k as follows.

$$\begin{aligned}
V(z) &= \mathsf{E}[z^S] = \sum_{r=1}^{\infty} z^r F_{S_k}(r) \\
&= (t_{01}+t_{02}) \sum_{r=1}^{\infty} z^r t_{00}^{r-1} = \frac{(t_{01}+t_{02})z}{1-t_{00}z}
\end{aligned} \tag{36}$$

F. Probability Generating Function of Random Variable ν

Recall that r.v. ν denotes the number of packets that leave the buffer of the MD during an inter-arrival period τ and its probability generating function is denoted by $N(z)$. We express the generating function of r.v. ν as follows.

$$\begin{aligned}
N(z) &= \mathsf{E}[z^\nu] = \sum_{t=1}^{\infty} \mathsf{E}[z^\nu | \tau = t] F_\tau(t), \\
&= \sum_{t=1}^{\infty} F_\tau(t) \sum_{k=0}^{t} z^k \mathsf{Pr}(\nu = k | \tau = t), \\
&= \sum_{t=0}^{\infty} F_\tau(t) \sum_{k=0}^{t} z^k \binom{t}{k} \omega^k (1-\omega)^{t-k}, \\
&= \sum_{t=1}^{\infty} F_\tau(t)(1-\omega)^t \sum_{k=0}^{t} \binom{t}{k} (\frac{\omega z}{1-\omega})^k, \\
&= \sum_{t=0}^{\infty} F_\tau(t)(1-\omega)^t (1+\frac{\omega z}{1-\omega})^t = \sum_{t=1}^{\infty} F_\tau(t)(1+\omega(z-1))^t,
\end{aligned} \tag{37}$$

where $F_\tau(t)$ denotes the pmf of inter-arrival period τ, and ω denotes the probability of service completion as defined in (6). On the other hand, the probability generating function of r.v. τ is equal to

$$T(z) = \mathsf{E}[z^\tau] = \sum_{t=1}^{\infty} z^t F_\tau(t). \tag{38}$$

By comparing the equations (37) and (38), we conclude the following expression which results in the probability generating function of r.v. ν

$$\begin{aligned}
N(z) &= T(1+\omega(z-1)), \\
&= \frac{W(1+\omega(z-1))p}{1-qU(1+\omega(z-1))V(1+\omega(z-1))}.
\end{aligned} \tag{39}$$

Let $F_\nu(t)$ denote the pmf of r.v. ν. By evaluating the inverse transform of $N(z)$, we obtain $F_\nu(t)$. Meanwhile, recall that $Q_i^{ON}(k)$ denotes the probability that k packets out of i leave the MD's buffer during the inter-arrival period τ. According to [11], the term $Q_i^{ON}(k)$ is obtained as follows.

$$Q_i^{ON}(k) = F_\nu(k), \quad 0 \le k \le i-1, \tag{40}$$

$$Q_i^{ON}(i) = \sum_{n=i}^{\infty} F_\nu(n). \tag{41}$$

Therefore, inserting (40) and (41) in the recursive equation (9) gives us the probability of starvation for streaming a file with size N, given that there are x packets (start-up delay) accumulated in the buffer before the service begins.

IV. Constrained Optimization Problem

In this section, we use the results from the previous sections to investigate the energy efficiency related optimization problem subject to a QoE constraint in terms of starvation probability. We formulate an optimization problem that minimizes the energy consumption of heterogeneous cellular network while guaranteeing a target buffer starvation probability for a MD as follows.

$$\begin{aligned} \underset{\beta_m,\beta_f}{\text{Minimize}} \quad & E_{total} = E_f + \gamma E_m \\ \text{s.t.} \quad & P_i(n) \leq \varepsilon, \\ & i = x-1, \end{aligned} \tag{42}$$

where E_m, E_f denote the expected value of the energy consumptions of MBS and FBS, respectively, and x denotes the start-up delay. Using the Markov chain shown in Fig. 1, we obtain the values of E_m, E_f as the proportion of time that MBS and FBS are in active mode in the steady state. We also assume that a femtocell's energy consumption is $1/\gamma$ of that a MBS consumes per unit time. Note that β_m, β_f denote the rates at which MBS and FBS go to active mode, respectively. Therefore, E_m, E_f increase with the increase in rates β_m, β_f. On the other hand, the starvation probability, which is given in (9), is a decreasing function of β_m and β_f. In order to solve the above problem, we first find the values of β_m and β_f that satisfy the buffer starvation constraint with equality, and then, we solve the minimization problem considering the β_m^*'s and β_f^*'s.

V. Numerical Results

In this section, we first investigate the energy minimization problem, and then compare the buffer starvation probability of a MD in a heterogeneous and a homogeneous cellular network. In the following, we assume that $\gamma = 10$.

A. Energy Consumption Optimization Subject to a QoE Constraint

Fig. 3 illustrates the minimum amount of energy consumed for streaming a file with size N while guaranteeing a target starvation probability ε which is set to 0.15. The file size in this experiment ranges between 100 and 300 in terms of packets, and the start-up delay x is set to 50 packets. For the heterogeneous network, we let the rates α_m, α_f be 0.1 and 0.15, respectively. The rate β_m varies between 0.01 and 0.11, and the rate β_f varies between 0.05 and 0.15. We set λ_m, λ_f, μ, and time-slot h to 1.5, 1.7, 1, and 10^{-5}, respectively. In the case of homogeneous cellular network with a single MBS, in order to satisfy the QoE constraint (i.e., the buffer starvation probability to be less than or equal to ε), the rate β_m should vary between 0.11 and 0.21. The total energy consumption of the network increases with the increase in the rates at which the BSs go to active mode. Nevertheless, our system model, in which the MD is covered by two BSs, significantly reduces the overall energy consumption of cellular network while guaranteeing a target starvation probability in comparison to the case where the MD is covered only by a MBS as demonstrated in Fig. 3.

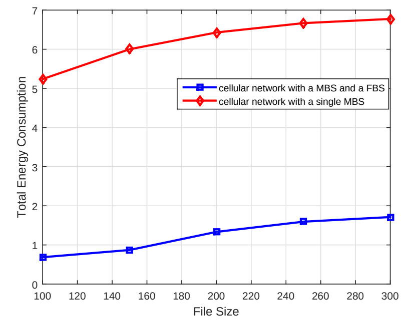

Fig. 3: Total energy consumption of cellular network with initial-buffering delay x= 50, and target starvation probability $\varepsilon = 0.15$.

B. The Probability of Starvation at the Buffer of a Mobile Device with respect to File Size

In Fig. 4, we plot the buffer starvation probability with initial buffering delay $x = 30$. The file size increases from 100 to 600 and $\alpha_m = \beta_m = 0.1$, $\alpha_f = \beta_f = 0.15$. The values of h, λ_m, λ_f are the same as in section *V.A*. The probability of buffer starvation increases with the increase in file size, however, the probability of having a buffer starvation while streaming a file (with the same size) in our system with two BSs is much less than the case where the MD could be served only through a single MBS. Moreover, as the file size increases, the starvation probability in a system with one MBS increases much more than that of our system. The reason is that in a heterogeneous cellular network, the MD could be served by either MBS or FBS, and since the arrival rate from a FBS is usually more than that of a MBS, a starvation event occurs less frequently.

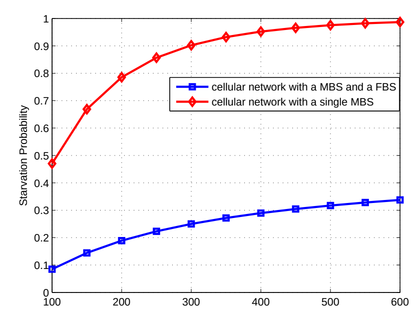

Fig. 4: Buffer starvation probability with initial-buffering delay x= 30.

C. The Probability of Starvation at the Buffer of a Mobile Device with respect to Start-Up Delay

Fig. 5 depicts the impact of start-up delay on the starvation probability. In this set of experiments, N= 600 and the start-up delay varies between 30 and 100 packets. We let α_m, β_m be 0.1, and α_f, β_f be 0.15. λ_m, λ_f, and h are set to 1.5, 1.7, 10^{-5}, respectively. First, for the same file size and the same start-up delay, the starvation probability of a MD in our system model is much less than that of a MD in a system with a single MBS. Second, a slight increase in start-up delay can greatly improve the starvation probability in our system compared to the case where the MD is in a homogeneous cellular network.

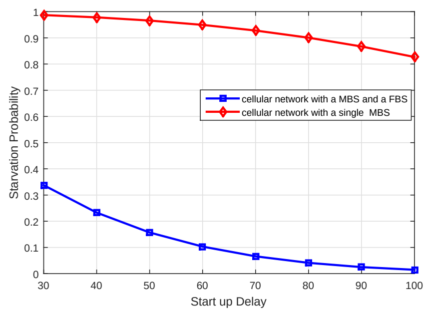

Fig. 5: Buffer starvation probability for streaming a file of size N= 600.

D. Buffer Starvation with respect to Energy Consumptions

In this experiment, we illustrate how the starvation probability is related to the energy consumption of BSs in a heterogeneous cellular network. We set the file size N, and start-up delay x to 300 and 60, respectively. The rate of going to active mode for MBS increases from 0.01 to 0.11, and this rate for FBS ranges between 0.05 and 0.15. The values of h, λ_m, λ_f are the same as in the first part of this section. It is clear that the starvation probability increases with the decrease in MBS's and FBS's energy consumption.

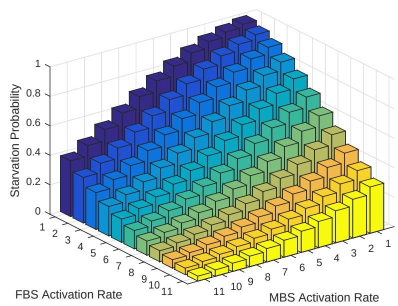

Fig. 6: Buffer starvation probability with N= 300 and x= 60.

VI. Conclusion

In this work, we considered an uncoordinated energy saving mechanism, where MBS and FBS goes in and out of sleep and active modes randomly throughout the system operation. This simple system model is used to demonstrate the efficacy of heterogeneous cellular networks in terms of meeting QoE guarantees of MDs, where QoE is defined as buffer starvation probability of MD. Considering an on/off bursty traffic, we derived the buffer starvation probability of a MD in a system with multiple servers, where the MD could be served by a MBS or FBS depending on which one is in active mode, for the first time. In addition, by the use of a three state Markov chain and applying the first step analysis we investigated the aggregated active/sleep period length distribution analytically. The simulation results reveal that the proposed system model provides significant energy savings compared to a homogeneous cellular network. In conclusion, our proposed framework can be used both for the energy efficient design and operation of different types of base stations in a heterogeneous networks and for improving the mobile devices' quality of experiences. We believe our model can be used as a useful starting point for future studies on interruption analysis in video streaming for mobile devices in a system with multiple servers, and specially in studying the aggregated active/sleep mode duration distribution. Interesting future direction to extend this work include developing analytical approaches towards analyzing the buffer starvation probability of mobile devices in a heterogeneous network with more than one femtocell base station.

References

[1] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 10–21, June 2011.

[2] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 22–30, June 2011.

[3] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo, H. Dhillon, and T. Novlan, "Heterogeneous cellular networks: From theory to practice," *Communications Magazine, IEEE*, vol. 50, no. 6, pp. 54–64, June 2012.

[4] G. Fettweis and E. Zimmermann, "Ict energy consumption-trends and challenges," in *Proceedings of the 11th International Symposium on Wireless Personal Multimedia Communications*, vol. 2, no. 4, 2008, p. 6.

[5] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 8, pp. 1525–1536, September 2011.

[6] H. Karl *et al.*, "An overview of energy-efficiency techniques for mobile communication systems," *Report of AG Mobikom WG7*, 2003.

[7] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 40–49, October 2011.

[8] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 74–79, November 2010.

[9] Y. S. Soh, T. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 5, pp. 840–850, May 2013.

[10] G. Jie, Z. Sheng, and N. Zhisheng, "A dynamic programming approach for base station sleeping in cellular networks," *IEICE Transactions on Communications*, vol. 95, no. 2, pp. 551–562, 2012.

[11] I. Cidon, A. Khamisy, and M. Sidi, "Analysis of packet loss processes in high-speed networks," *Information Theory, IEEE Transactions on*, vol. 39, no. 1, pp. 98–108, 1993.

[12] Y. Xu, E. Altman, R. El-Azouzi, M. Haddad, S. Elayoubi, and T. Jimenez, "Analysis of buffer starvation with application to objective qoe optimization of streaming services," *Multimedia, IEEE Transactions on*, vol. 16, no. 3, pp. 813–827, 2014.

[13] M. Wellens, J. Riihijarvi, and P. Mahonen, "Modelling primary system activity in dynamic spectrum access networks by aggregated on/off-processes," in *Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2009. SECON Workshops '09. 6th Annual IEEE Communications Society Conference on*, June 2009, pp. 1–6.