Base Station Sleeping and Power Control for Bursty Traffic in Cellular Networks

Jian Wu
Tsinghua University
Beijing, China
wujian09@mails.tsinghua.edu.cn

Yanan Bao
UC Davis
California, USA
ynbao@ucdavis.edu

Guowang Miao
KTH Royal Institute of Technology
Stockholm, Sweden
guowang@kth.se

Zhisheng Niu
Tsinghua University
Beijing, China
niuzhs@tsinghua.edu.cn

***Abstract*—In this paper, we study sleeping and power control of a single-cell cellular network with bursty traffic. The base station (BS) sleeps whenever the system is empty, and wakes up when *N* users are assembled. The service capacity of the BS in the active mode is controlled through its transmitting power. The total power consumption and average delay for bursty traffic that follows the Interrupted Poisson Process (IPP) are analyzed. We discuss when the BS should sleep and the impact of traffic burstiness on it. The impact of the sleeping threshold and the transmitting power on the system performance is also investigated. The numerical results show that given the average traffic load, the more bursty the traffic is, the less the total power is consumed, while the delay performance of the more bursty traffic is better only under certain circumstances.**

I. Introduction

The exponential growth of mobile data traffic has triggered a vast expansion of network infrastructure, resulting in dramatically increasing network energy consumption [1]. Energy-efficient designs are urgently needed from both environmental and economic aspects. In cellular networks BSs consume nearly 60-80% of the total energy [2]. The total power consumption of a BS consists of both the circuit and transmitting power consumption. The circuit power is independent of the transmitting power, and is consumed because of signal processing, battery backup, as well as site cooling. The transmitting power is for reliable data transmission and is mainly consumed by amplifiers, feeder losses and so on [3] [4]. Therefore efforts to reduce both the circuit and transmitting power consumption should be made.

BS sleeping has been proposed recently to realize substantial reduction of energy consumption [5]- [7]. Besides, transmitting power adaptation to match traffic load requirement in the active mode is also an effective way to save energy because of its impact on the operational power consumption of amplifiers and so on [3] [8].

In our previous work we have started the investigation and assumed Poisson traffic arrivals to simplify the analysis [9]. In practice, bursty traffic is much more common, and this motivates the work in this paper where we model the user arrivals using the Interrupted Poisson Process (IPP), which represents typical bursty data traffic [10] [11]. We focus on the N based BS sleeping policy, which turns the BS off when the system is empty and turns it on when N users assemble. This policy has been proposed in literature for queueing analysis [12] [13]. In this paper, first we derive the closed-form total power consumption and average delay for the IPP traffic with N based BS sleeping and adjustable service capacity. Second, when the BS should sleep and the impact of traffic burstiness on it are explored. Moreover, the influences from the sleeping threshold and the transmitting power on system performance are provided.

The rest of this paper is organized as follows: Sec. II describes the system model. Sec. III gives the analysis of the energy and delay performance. The impact of system parameters are investigated in Sec. IV. Numerical results are provided in Sec. V, and Sec. VI concludes the paper.

This work is sponsored in part by the National Basic Research Program of China (2012CB316001), the Nature Science Foundation of China (61201191, 60925002, 61021001), and Hitachi Ltd.

II. System model

A. Traffic model

We consider the downlink of a single BS where users arrive according to an IPP with parameters (λ, α, β). The on and off periods of the IPP are both exponentially distributed with the average length α^{-1} and β^{-1} respectively. In the on period, users arrive according to a Poisson process with the arrival rate λ. The average arrival rate of the IPP traffic is $\hat{\lambda} = \frac{\lambda\beta}{\alpha+\beta}$. Each user requests an exponentially distributed random amount of best-effort data service when it arrives, e.g., file download with average file size l, and the user leaves the system after the service is done.

B. BS power consumption model

We assume the BS has the *active* and *sleep* modes, with the power consumption P_{BS} as follows [3]:

$$P_{BS} = \begin{cases} P_o + \Delta_P P_t, & \textit{active} \text{ mode}, \\ P_{sleep}, & \textit{sleep} \text{ mode}. \end{cases} \tag{1}$$

P_o and P_{sleep} are the circuit power consumption in the active and sleep modes respectively, and Δ_P is the slope of the load-dependent power consumption, where the transmitting power P_t adapts to the system traffic load. It is also assumed that there is a fixed switching energy cost E_s for each mode transition.

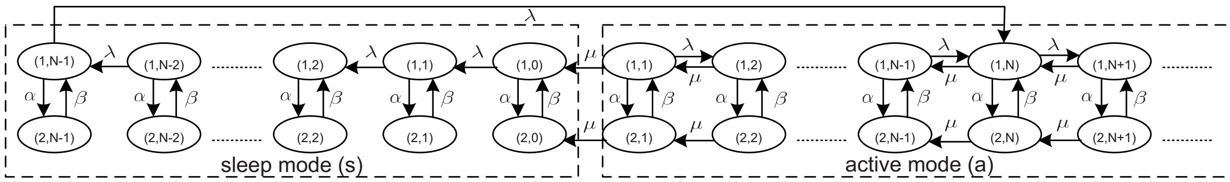


Fig. 1. State transition diagram of the extended IPP/M/1 queueing model.

C. BS sleeping and power control

The BS uses the N based sleeping policy: it goes to sleep when the system is empty and returns to active mode once N users assemble in the system. In the active mode, the transmitting power P_t of the BS is adapted to match the traffic load. Assume that the BS service capacity is x bits per second, which is equally shared by all users being served. So the user departure rate is $\mu = x/l$. The relationship between the service rate x and the transmitting power P_t is

$$x = B\log_2(1+\gamma P_t), \quad P_t \in [0, P_t^{max}] \tag{2}$$

where $\gamma = \frac{\eta g}{N_0 B}$, g[1] represents the channel gain, B is the bandwidth, N_0 denotes the noise density, and η is a constant related to bit error rate (BER) requirement when adaptive modulation and coding is used [14].

The control variables are the sleeping threshold N and the transmitting power P_t. The delay performance is the time between the user arrival and the instant the service is done.

III. The IPP/M/1 queueing model with N based sleeping and power control

In this section, given the sleeping threshold $N (N \geq 1)$ and the transmitting power P_t in the active mode, we analyze the total power consumption and average delay performance for IPP traffic. Since the user departure rate μ is a function of P_t, our sleeping and power control can be modeled using an extended IPP/M/1 queueing model with N based sleeping and adjustable service rate.

A. The extended IPP/M/1 queueing model

The state transition diagram of the queueing model is shown in Fig. 1. The total state space is divided into the active and sleep mode state sets. In each set, the state space is defined as (i, j), where $i = 1$ $(i = 2)$ represents the on (off) period of the IPP traffic, and j counts the number of users in the system. $p^a_{i,j}, (i \in \{1, 2\}, j > 0)$ is the probability that the BS is in the active mode with state (i, j), and $p^s_{i,j}, (i \in \{1, 2\}, 0 \leq j < N)$ is the probability in the sleep mode with state (i, j).

[1]We first study the basic case: users experience homogeneous channels with gain g. The heterogeneous channel conditions will be discussed in future work.

The global balance equations are given as follows:

$$(\lambda+\alpha)p^s_{1,m} = \lambda p^s_{1,m-1} + \beta p^s_{2,m},\ (1 \leq m \leq N{-}1) \tag{3}$$
$$\beta p^s_{2,m} = \alpha p^s_{1,m},\ (1 \leq m \leq N-1) \tag{4}$$
$$(\lambda+\alpha)p^s_{1,0} = \mu p^a_{1,1} + \beta p^s_{2,0}, \tag{5}$$
$$\beta p^s_{2,0} = \mu p^a_{2,1} + \alpha p^s_{1,0}, \tag{6}$$
$$(\lambda+\mu+\alpha)p^a_{1,1} = \mu p^a_{1,2} + \beta p^a_{2,1}, \tag{7}$$
$$(\lambda+\mu+\alpha)p^a_{1,m} = \mu p^a_{1,m+1} + \lambda p^a_{1,m-1} + \beta p^a_{2,m}, (m \geq 2, m \neq N) \tag{8}$$
$$(\lambda+\mu+\alpha)p^a_{1,N} = \mu p^a_{1,N+1} + \lambda(p^a_{1,N-1} + p^s_{1,N-1}) + \beta p^a_{2,N}, \tag{9}$$
$$(\mu+\beta)p^a_{2,m} = \mu p^a_{2,m+1} + \alpha p^a_{1,m},\ (m \geq 1). \tag{10}$$

After some algebraic operations, we obtain the following equations, which are actually the local balance equations.

$$p^s_{1,m} = p^s_{1,m-1},\ (1 \leq m \leq N-1) \tag{11}$$
$$p^s_{2,m} = \frac{\alpha}{\beta} p^s_{1,m},\ (1 \leq m \leq N-1) \tag{12}$$
$$\lambda p^s_{1,0} = \mu(p^a_{1,1} + p^a_{2,1}), \tag{13}$$
$$\lambda(p^s_{1,m} + p^a_{1,m}) = \mu(p^a_{1,m+1} + p^a_{2,m+1}),\ (1 \leq m \leq N-1) \tag{14}$$
$$\lambda p^a_{1,m} = \mu(p^a_{1,m+1} + p^a_{2,m+1}),\ (m \geq N). \tag{15}$$

Define p_1 and p_2 to be the probability that the system is in on and off periods respectively. It is intuitive to know that $p_1 = \frac{\beta}{\alpha+\beta}$ and $p_2 = \frac{\alpha}{\alpha+\beta}$. Summing up Eq.(13)-(15) over all m and substituting Eq.(11)(12) into it, we obtain

$$\lambda p_1 = \mu\big[(p_1 - N p^s_{1,0}) + \big(p_2 - p^s_{2,0} - \frac{\alpha}{\beta}(N-1)p^s_{1,0}\big)\big]. \tag{16}$$

The left-hand side of Eq. (16) is the average user arrival rate, while the right-hand side is the average number of users served by the BS per unit time. Based on this, the probability Pr^s that the BS is in the sleep mode is

$$Pr^s = N p^s_{1,0} + p^s_{2,0} + \frac{\alpha}{\beta}(N-1)p^s_{1,0} = \frac{\mu(\alpha+\beta) - \lambda\beta}{\mu(\alpha+\beta)}, \tag{17}$$

which is independent of the sleeping threshold N.

B. The generation function analysis

To derive the delay and total power consumption, the generation function is analyzed in this section. The generation function $G(z)$ of the system is $G(z) = G_1(z) + G_2(z) = (\sum_{m=0}^{N-1} z^m p^s_{1,m} + \sum_{m=1}^{\infty} z^m p^a_{1,m}) + (\sum_{m=0}^{N-1} z^m p^s_{2,m} + \sum_{m=1}^{\infty} z^m p^a_{2,m}), |z| \leq 1$. As shown in Appendix A, it is transformed into

$$
\begin{aligned}
G(z) = \frac{1}{g(z)}\Big\{ & p^s_{1,0}\big[(\alpha z + \frac{\alpha}{\beta}(\alpha z{+}\mu z{-}\lambda z^2{-}\mu{+}\lambda z)) \sum_{n=1}^{N-1} z^n \\
& + (\alpha z + \beta z + \mu z - \mu) \sum_{n=0}^{N-1} z^n\big] \\
& + p^s_{2,0}\big[\alpha z{+}\beta z{+}\mu z{-}\lambda z^2{-}\mu{+}\lambda z\big]\Big\},
\end{aligned} \tag{18}
$$

where

$$
g(z) = -\lambda(1+\frac{\beta}{\mu})z^2 + (\lambda+\mu+\alpha+\beta)z - \mu. \tag{19}
$$

The polynomial $g(z)$ has a unique root z_0 in the open interval $(0,1)$.

$$
z_0 = \frac{(\lambda{+}\mu{+}\alpha{+}\beta) - \sqrt{(\lambda{+}\mu{+}\alpha{+}\beta)^2 - 4\lambda\mu(1{+}\beta/\mu)}}{2\lambda(1{+}\frac{\beta}{\mu})}. \tag{20}
$$

Making use of $g(z_0)G(z_0) = 0$ in Eq. (18) and $g(z_0) = 0$, we arrive at

$$
p^s_{2,0}\beta + p^s_{1,0}[\frac{z_0 - z_0^N}{1-z_0}\alpha + \frac{1-z_0^N}{1-z_0}(\mu + \beta - \frac{\mu}{z_0})] = 0. \tag{21}
$$

Combining Eq. (21) and Eq. (17), we have

$$
p^s_{1,0} {=} \frac{\mu(\alpha{+}\beta) - \lambda\beta}{\mu(\alpha{+}\beta)} \frac{1}{N(1{+}\frac{\alpha}{\beta}) - [(\frac{\lambda}{\beta}{+}\frac{\lambda}{\mu})z_0 - \frac{\lambda}{\beta}]\frac{1-z_0^N}{1-z_0}}, \tag{22}
$$

$$
p^s_{2,0} {=} \frac{\mu(\alpha{+}\beta) - \lambda\beta}{\mu(\alpha{+}\beta)} \frac{\frac{\alpha}{\beta} - [(\frac{\lambda}{\beta}{+}\frac{\lambda}{\mu})z_0 - \frac{\lambda}{\beta}]\frac{1-z_0^N}{1-z_0}}{N(1{+}\frac{\alpha}{\beta}) - [(\frac{\lambda}{\beta}{+}\frac{\lambda}{\mu})z_0 - \frac{\lambda}{\beta}]\frac{1-z_0^N}{1-z_0}}. \tag{23}
$$

C. The average delay

The average number of users in the system $L_{(N,P_t)} = \sum_{m=0}^{N-1} m(p^s_{1,m}{+}p^s_{2,m}) {+} \sum_{m=1}^{\infty} m(p^a_{1,m}{+}p^a_{2,m})$ can be directly derived from the generation function as follows.

$$
L_{(N,P_t)} = \left.\frac{\mathrm{d}G(z)}{\mathrm{d}z}\right|_{z=1} = \frac{\mathrm{d}}{\mathrm{d}z} \left.\frac{g(z)G_1(z) + g(z)G_2(z)}{g(z)}\right|_{z=1}.
$$

Substituting $g(z)$ and Eqs. (40)(41) in Appendix A into it, we get the average number of users $L_{(N,P_t)}$ in the system.

$$
\begin{aligned}
L_{(N,P_t)} = & \frac{\lambda\beta(\lambda + \alpha + \beta - \mu)}{(\alpha+\beta)(\mu\alpha + \mu\beta - \lambda\beta)} + \frac{\mu p^s_{1,0}}{\mu\alpha + \mu\beta - \lambda\beta}[N\lambda \\
& + \frac{N(N{-}1)}{2}(\alpha{+}\beta)(1{+}\frac{\alpha}{\beta})].
\end{aligned} \tag{24}
$$

Using the Little's Law, the average delay is

$$
\begin{aligned}
D_{(N,P_t)} = & \frac{L_{(N,P_t)}}{\hat{\lambda}} = \frac{\lambda + \alpha + \beta - \mu}{\mu\alpha + \mu\beta - \lambda\beta} \\
& + \frac{1}{\lambda\beta} \frac{N\lambda + \frac{N(N-1)}{2}(\alpha + \beta)(1 + \frac{\alpha}{\beta})}{N(1{+}\frac{\alpha}{\beta}) - [(\frac{\lambda}{\beta} + \frac{\lambda}{\mu})z_0 - \frac{\lambda}{\beta}]\frac{1-z_0^N}{1-z_0}}.
\end{aligned} \tag{25}
$$

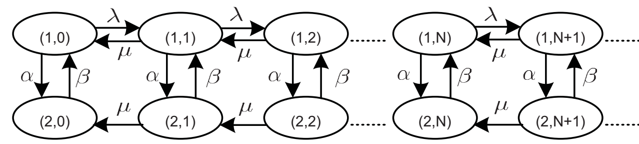


Fig. 2. State transition diagram of the IPP/M/1 queueing model.

D. Total Power Consumption

The total power consumption $P_{(N,P_t)}$ is composed of three parts as shown in Eq.(26).

$$
P_{(N,P_t)} = (1{-}Pr^s)(P_o{+}\Delta_P P_t){+}Pr^s P_{sleep}{+}E_s F_m. \tag{26}
$$

The first two parts are the average power consumption in the active and sleep modes respectively, and the last term $E_s F_m$ is the mode switching cost. The mode transition frequency F_m, defined as the number of mode transitions between active and sleep modes per unit time, is $2\lambda p^s_{1,N-1}$. This is because the BS will be turned on when there is a new user request arrival in state $(1, N-1)$ of the sleep mode. As a result, the total power consumption is

$$
\begin{aligned}
P_{(N,P_t)} {=} & (1{-}\frac{\lambda\beta}{\mu(\alpha{+}\beta)})[P_{sleep} {+} \frac{2E_s\lambda}{N(1{+}\frac{\alpha}{\beta}) - [(\frac{\lambda}{\beta}{+}\frac{\lambda}{\mu})z_0 - \frac{\lambda}{\beta}]\frac{1-z_0^N}{1-z_0}}] \\
& + \frac{\lambda\beta}{\mu(\alpha{+}\beta)}(P_o{+}\Delta_P P_t), \quad N \geq 1.
\end{aligned} \tag{27}
$$

E. Special case: The IPP/M/1 queueing model with power control only

In this section we consider the special case that there is no sleeping control, and only the transmitting power can be adapted to match the traffic load. This is modeled using the IPP/M/1 queueing model with adjustable service rate, and the state space is given in Fig. 2. Using the same analysis method, the generation function $G_{pm}(z)$ is obtained as

$$
\begin{aligned}
G_{pm}(z) = & \frac{1}{g(z)}\big[(\alpha + \beta)(1 - \frac{\lambda\beta}{\mu(\alpha + \beta)})z - p_{1,0}\mu(1 - z) \\
& + p_{2,0}(1 - z)(\lambda z - \mu)\big],
\end{aligned} \tag{28}
$$

with

$$
p_{1,0} = \frac{\mu(\alpha + \beta) - \lambda\beta}{\mu(\alpha + \beta)} \frac{\beta z_0}{(1 - z_0)\mu}, \tag{29}
$$

$$
p_{2,0} = \frac{\mu(\alpha + \beta) - \lambda\beta}{\mu(\alpha + \beta)} \frac{\alpha z_0}{(1 - z_0)(\mu - \lambda z_0)}, \tag{30}
$$

where $g(z)$ and z_0 are the same as those in Eq. (19) and Eq. (20) respectively. Similarly, its total power consumption $P_{(P_t)}$ and average delay $D_{(P_t)}$ are derived through the generation function as follows.

$$
P_{(P_t)} = P_o + \frac{\lambda\beta}{\mu(\alpha + \beta)}\Delta_P P_t, \tag{31}
$$

$$
D_{(P_t)} = \frac{\beta(\alpha + \beta) + \alpha\mu}{(\mu\alpha + \mu\beta - \lambda\beta)\beta} - \frac{\alpha}{\beta} \frac{1}{\alpha + \beta - \frac{\lambda\beta}{\mu} z_0}. \tag{32}
$$

The detail is omitted for the space limitation. Note that there is $D_{(1,P_t)} = D_{(P_t)}$ for the delay performance.

IV. Performance impact of system parameters

A. The traffic burstiness

The burstiness of the IPP traffic is reflected through the variance coefficient C^2 [15], which is given by

$$C^2 = 1 + \frac{2\lambda\alpha}{(\alpha+\beta)^2}. \tag{33}$$

With $\alpha = k\beta$ we have $\hat{\lambda} = \frac{\lambda}{1+k}$ and $C^2 = 1 + \frac{2\lambda k}{(1+k)^2\beta}$. The average arrival rate is independent of β and only relates to λ and k. Given λ and k, the burstiness is only affected by β, and the smaller β is, the more bursty the traffic will be. As a result, in the following of this paper, given λ and k, we investigate the impact of the traffic burstiness by varying β without influencing the average traffic load at the same time.

B. When should the BS sleep?

We compare the joint sleeping and power control with the case that only power control is available to find when it is energy-efficient for the BS to sleep. The following proposition can be obtained and the detailed proof is omitted due to limited space.

Proposition 1. *For the IPP traffic with parameters $(\lambda, k\beta, \beta)$, given the transmitting power P_t, it is energy-efficient to use the N based sleeping control when*

$$N + f(N,\lambda,k,\beta) > \frac{2\lambda E_s}{(1+k)(P_o - P_{sleep})}, \tag{34}$$

where $f(N,\lambda,k,\beta) = -\frac{1}{1+k}[(\frac{\lambda}{\beta}+\frac{\lambda}{\mu})z_0 - \frac{\lambda}{\beta}]\frac{1-z_0^N}{1-z_0}$ *and* $z_0 = \frac{(\lambda+\mu+k\beta+\beta)-\sqrt{(\lambda+\mu+k\beta+\beta)^2-4\lambda(\mu+\beta)}}{2\lambda(1+\beta/\mu)}$ *with the properties:*

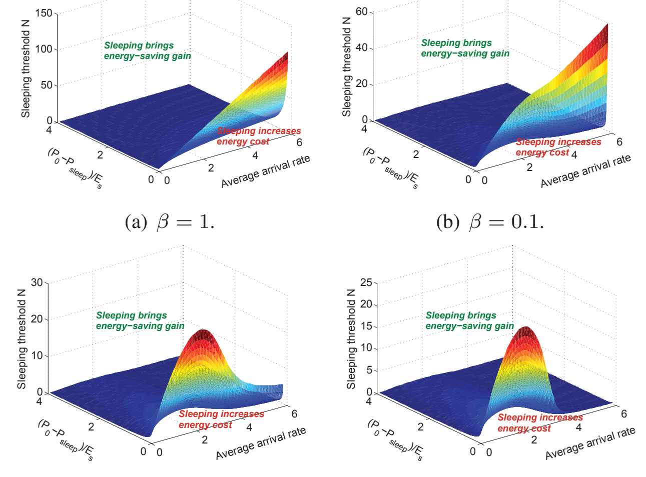
$$f(N,\lambda,k,\beta) > 0;\ \ \frac{\partial f(N,\lambda,k,\beta)}{\partial \beta} < 0;\ \ \frac{\partial f(N,\lambda,k,\beta)}{\partial N} > 0. \tag{35}$$

Remark: First, the condition is depicted in Fig. 3 in 3-D form with different β. The x-axis is average traffic arrival rate and the y-axis is the parameter $\frac{P_o - P_{sleep}}{E_s}$ related to the energy consumption model. Above the surface, BS sleeping saves energy. However, below the surface, BS sleeping wastes energy because of the extra mode switching cost. Comparing this condition with that for the Poisson traffic [9], the IPP traffic always has a wider adaptation range of N.

Second, the properties in (35) indicate that given the average traffic load, the more bursty the traffic is, the smaller N can be to use BS sleeping. From Fig. 3(a) to Fig. 3(d), the four surfaces are lowered as β decreases from 1 to 0.01 when they are plotted together. In other words, as the bustiness of the traffic increases from Fig. 3(a) to Fig. 3(d), the region above the surface in which sleeping brings energy-saving gain expands.

V. Numerical Results

In this section we assume the system bandwidth B = 10MHz, the maximum transmitting power $P_t^{max} = 10$W, and the path loss model $g = 36.7\lg d + 33.05$ (dB), where we set $d = 100$m. The noise power density $N_0 = -174$dBm/Hz, and $\eta = -1.5/\ln(5\varepsilon) = 0.283$ corresponds to the BER requirement of $\varepsilon = 10^{-3}$ [14]. We take the micro BS energy consumption parameters P_o = 100W, Δ_P = 7, P_{sleep} = 30W and set E_s = 25J [3]. For the IPP traffic, $\alpha = \beta$ ($k = 1$), and the average file size l = 2MB. Note that when we make comparisons between the IPP and Poisson traffic, it is assumed that they have the same average user arrival rate.



(a) $\beta = 1$. (b) $\beta = 0.1$.

(c) $\beta = 0.05$. (d) $\beta = 0.01$.

Fig. 3. Whether it is energy efficient to incorporate the N based sleeping control with different β. P_t = 10W, l = 2MB, k = 1.

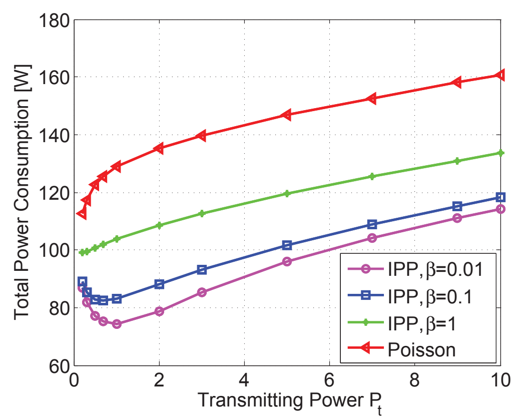


Fig. 4. The total power with varying P_t. $\lambda = 5$, $N = 1$.

Fig. 4 shows the impact of transmitting power on the total power consumption under different traffic burstiness. First, we can see that the more bursty the traffic is, the less the total power will be consumed. Actually, given the control pair (N, P_t) and the average traffic load, it can be proved that $\frac{\partial P_{(N,P_t)}}{\partial \beta} > 0$. Second, for the IPP traffic, the relationship between the total power consumption and P_t is not always monotonic and it greatly depends on the traffic parameters. Sometimes there exists the energy-optimal transmitting power that minimizes the total power consumption. This is similar to

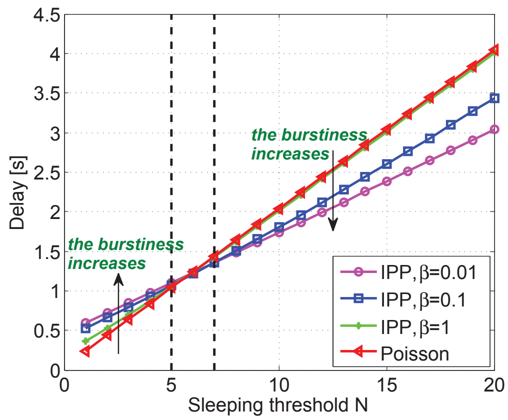


Fig. 5. The delay performance with varying N. $\lambda = 5$, $P_t = 10$W.

that for the Poisson traffic in [9].

Fig. 5 demonstrates the impact of traffic burstiness on the delay performance. It can be observed that there exists a transition area, and on its right where N is relatively large, the delay performance is better as the burstiness increases. The situation is opposite on its left side. Actually, as the traffic load increases, the transition area gradually moves to the right.

VI. Conclusion

In this paper we have analyzed the N based BS sleeping and power control using an extended IPP/M/1 queueing model. The total power consumption and average delay are derived analytically using the generation function method. We explore when the BS should sleep and focus on the impact of traffic burstiness on the system performance. Given the average traffic load, it is found that, first, the region in which sleeping brings energy-saving gain expands as the traffic burstiness increases; second, the more bursty the traffic is, the less the total power is consumed; third, the delay performance of the more bursty traffic is better only under certain circumstances.

Appendix A
The generation function

We rewrite Eqs. (3)(5) as follows.

$$(\lambda+\mu+\alpha)p^s_{1,m} = \lambda p^s_{1,m-1} + \beta p^s_{2,m} + \mu p^s_{1,m}, \qquad (36)$$

$$(\lambda+\mu+\alpha)p^s_{1,0} = \mu p^a_{1,1} + \beta p^s_{2,0} + \mu p^s_{1,0}. \qquad (37)$$

For Eqs. (36)(37) and Eqs. (7)-(9), we multiply each of them by $z^m (m = 0, 1, \cdots)$ appropriately and sum over all m. This process results in

$$(\lambda+\mu+\alpha)G_1(z) = \beta G_2(z) + \lambda z G_1(z) + \mu \sum_{n=0}^{N-1} z^n p^s_{1,0} + \frac{\mu}{z}\Big(G_1(z) - \sum_{n=0}^{N-1} z^n p^s_{1,0}\Big). \qquad (38)$$

Similarly, using Eqs. (4)(6)(10), we have

$$(\mu+\beta)G_2(z) = \alpha G_1(z) + \mu\Big(p^s_{2,0} + \frac{\alpha}{\beta}\sum_{n=1}^{N-1} z^n p^s_{1,0}\Big) + \frac{\mu}{z}\Big(G_2(z) - p^s_{2,0} - \frac{\alpha}{\beta}\sum_{n=1}^{N-1} z^n p^s_{1,0}\Big). \qquad (39)$$

With a polynomial $g(z) = -\lambda(1+\frac{\beta}{\mu})z^2 + (\lambda+\mu+\alpha+\beta)z - \mu$, utilizing Eq. (38)(39), we arrive at

$$g(z)G_1(z) = p^s_{1,0}\Big[\alpha z \sum_{n=1}^{N-1} z^n + (\beta z + \mu z - \mu)\sum_{n=0}^{N-1} z^n\Big] + \beta p^s_{2,0} z, \qquad (40)$$

$$g(z)G_2(z) = (\alpha z + \mu z - \lambda z^2 - \mu + \lambda z)\Big(p^s_{2,0} + p^s_{1,0}\frac{\alpha}{\beta}\sum_{n=1}^{N-1} z^n\Big) + \alpha p^s_{1,0} z \sum_{n=0}^{N-1} z^n. \qquad (41)$$

References

[1] G. Fettweis and E. Zimmermann, "ICT energy consumption-trends and challenges," in *Proc. 2008 International Symp. Wireless Personal Multimedia Commun.*, pp. 1-4.

[2] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," *IEEE ICC GreenComm. Workshop*, Jun. 2009.

[3] G. Auer, V. Giannini, C. Desset, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Trans. Wireless Commun.*, vol. 18, pp. 40-49, Oct. 2011.

[4] A. J Fehske, F. Richter, and G. P Fettweis, "Energy Efficiency Improvements through Micro Sites in Cellular Mobile Radio Networks," *2nd International Workshop on Green Communications, GLOBECOM 2009*, pp. 1-5, Dec. 2009.

[5] Z. Niu, "TANGO: Traffic-Aware Network Planning and Green Operation," *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 25-29, Oct. 2011.

[6] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," *Mobicom*, Sep. 2009.

[7] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE J. Select. Areas Commun.*, vol. 29, no. 8, Sep. 2011.

[8] J. Kwak, K. Son, Y. Yi, and S. Chong, "Greening Effect of Spatio-Temporal Power Sharing Policies in Cellular Networks with Energy Constraints," *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4405-4415, Dec. 2012.

[9] J. Wu, S. Zhou, and Z. Niu, "Traffic-Aware Base Station Sleeping Control and Power Matching for Energy Delay Tradeoffs in Green Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4196-4209, Aug. 2013.

[10] H. Tamura, Y. Yahiro, Y. Fukuda, K. Kawahara and Y. Oie, "Performance analysis of energy saving scheme with extra active period for LAN switches," in *Proc. IEEE Global Telecommunications Conf.*, pp. 198-203, Nov. 2007.

[11] U. Yechiali and P. Naor, "Queueing problems with heterogeneous arrivals and service," *Operations Res.*, vol. 19, pp. 722-734, 1971.

[12] M. Yadin, P. Naor, "Queueing systems with a removable service station," Op. Research, 1963.

[13] D. P. Heyman, "Optimal operating policies for M/G/1 queueing systems," *Operations Research.*, vol. 16, pp. 362-382, Mar. 1968.

[14] X. Qiu, and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Transacations on Communications*, vol. 47. no. 6, pp. 884-895, Jun. 1999.

[15] T. L. Saaty, *Elements of queueing theory*. New York: McGraw-Hill, 1961.