An Extra Power Saving Scheme for Prolonging Lifetime of Mobile Handset in the 4G Mobile Networks

Jenhui Chen[1]*, Woei-Hwa Tarn[2], Jiann-Der Lee[2]

1 Department of Computer Science and Information Engineering, School of Electrical and Computer Engineering, College of Engineering, Chang Gung University, Kweishan, Taoyuan, Taiwan, **2** Department of Electrical Engineering, School of Electrical and Computer Engineering, College of Engineering, Chang Gung University, Kweishan, Taoyuan, Taiwan

Abstract

In the fourth generation or next generation networks, services of non-real-time variable bit rate (NRT-VBR) and best effort (BE) will dominate over 85% of the total traffic in the networks. In this paper, we study the power saving mechanism of NRT-VBR and BE services for mobile handsets (MHs) to prolong their battery lifetime (i.e., the sustained operation duration) in the fourth generation networks. Because the priority of NRT-VBR and BE is lower than that of real-time VBR (RT-VBR) or guaranteed bit rate (GBR) services, we investigate an extended sleep mode for lower priority services (e.g., NRT-VBR and BE) in an MH to conserve the energy. The extended sleep mode is used when the MH wakes up from the sleep mode but it cannot obtain the bandwidth from base station (BS). The proposed mechanism, named *extra power saving scheme* (EPSS), uses the $M/M/k/k$ Markovian queuing model to estimate the extended sleep duration to let MHs conserve their battery energy when the networks traffic is congested. To study the performance of EPSS, an accurate analysis model of energy is presented and validated by taking a series of simulations. Numerical experiments show that EPSS can achieve 43% extra energy conservation at most when downlink resource is saturated. We conclude that the energy of MHs can be conserved further by applying EPSS when the traffic load is saturated. The effect of energy saving becomes more obvious when the portion of NRT-VBR and BE services is greater than that of RT-VBR and GBR services.

Competing Interests: The authors have declared that no competing interests exist.

* Email: jhchen@mail.cgu.edu.tw

Introduction

The fourth generation (4G) personal communication networks provide telecommunication and data access together for nomadic users by integrating telephone and computer networks into a single system [1,2]. These nomadic users access Internet services via mobile handsets (MHs), e.g., smart phones or pads, and these MHs work relying on the battery energy. According to the latest reports of statistics of mobile data traffic [3,4], the non-real-time variable bit rate (NRT-VBR) and best effort (BE) traffic, e.g., data (Internet downloading, web browsing, E-mail, and social networking), file sharing (FTP and P2P), and video (buffered video streaming), dominates around 86.4% of mobile data traffic in early 2013. These reports forecast that mobile data traffic will grow at a compound annual growth rate of 66% from 2012 to 2017, reaching 11.2 exabytes per month by 2017 [3]; video streaming will dominate two-thirds of the traffic, where NRT video (buffered video streaming) dominates three-fourths of the video streaming traffic. However, according to the guideline of related 4G standards [5,6], NRT-VBR service does not have the high priority to acquire the system resource (e.g., bandwidth) for transmission [7]. It means that less system resources are allocated to high proportion of services. This certainly leads a serious competition [8].

On the other hand, because the energy of the MH battery is limited, how to prolong the lifetime of the MH battery to support multiple services has become an valuable issue. Many power saving mechanisms (PSMs) have been broadly studied such as adaptive traffic coalescing (ATC) scheme [9], power saving studies based on quality-of-service (QoS) constraints [10], statistical sleep window control (SSWC) approach [11], fold-and-demultiplex (FD) method for power saving class (PSC) type I and II services [12], and a hybrid mechanism for voice over IP (VoIP) service to improve power saving efficiency [13]. Because PSM uses the method that an MH turns off its radio transceiver in idle time to conserve its energy, too much waiting time will decrease the efficiency of PSM. Those mentioned PSMs, however, do not consider a problem that MHs may not obtain the downlink bandwidth and thus waste their energy due to traffic congestion. Actually, the problem of wasting battery energy on waiting for available bandwidth can be solved by implementing a smarter scheduler in a base station (BS) to order those MHs to enter a pre-calculated *extended sleep duration* for energy conservation. In this paper, we present an MH PSM named *extra power saving scheme* (EPSS) to prolong the lifetime of the MH battery by avoiding those unnecessary and uncertain waiting time for available bandwidth. From the performance evaluation, we find that the proposed mechanism can achieve better power saving efficiency as

compared with the 4G communication protocols when the traffic load is large.

The rest of this paper are organized as follows. A detailed description of the unnecessary energy wasting because MHs wait for available bandwidth in 4G systems and the solution of EPSS is given in Section. The analytical model of the proposed EPSS is presented to show how to obtain the estimated extended time and derive the energy consumption in Section. Section shows the validation of the analytical model and related numerical experiments. Finally, the conclusion is given in Section.

Extra Power Saving Scheme

Problem Description

Latest 4G communication techniques, the IEEE 802.16m [5] and long-term evolution advanced (LTE-A) [6], utilize 100% packet-switched air interface to provide numerous services for MHs. To ensure that the data receiving meet the QoS requirement of variant application services, these two standards provide different QoS classes for different application services which are treated as service flows. Table 1 and Table 2 summarize the priority order of QoS classes of IEEE 802.16m and LTE-A, respectively. As mentioned in the preceding section, around 86.4% of Internet traffic is NRT-VBR, BE, and non-GBR types. These types of traffic are lower priority services as compared with GBR services [4]. Because the NRT-VBR services have the delay-tolerant property [15], they are usually served in a lower priority as compared with real-time (RT) and guaranteed bit rate (GBR) services. This setting will lead to the consequence that NRT-VBR and non-GBR services cannot obtain the bandwidth when they compete with services having higher priority than theirs. This problem will also appear when all services are NRT-VBR or non-GBR and they compete bandwidth with each other because these services are served in a first come, first served (FCFS) basis. Then we can conclude if the downlink packet has higher QoS priority, e.g., rtPS/GBR service, the scheduler will allocate the resources first to support its transmission. In contrast, if the downlink packages are nrtPS/non-GBR services, it may need longer waiting time till the scheduler serves the services with higher QoS priority. Especially when the system is in saturated condition [16].

In order to prolong the battery lifetime of an MH, the PSM turns off an MH radio interface when the MH is in idle duration. The IEEE 802.16m PSM provides three types of PSCs for different service types to conserve the battery energy. The PSC type I is used for NRT and BE traffic, PSC type II is used for unsolicited grant service (UGS) and RT-VBR traffic, and PSC type III is used for multicast and management traffic, respectively. Similarly, the LTE-A PSM uses discontinuous reception and discontinuous transmission (DRX/DTX) mechanism to conserve the battery energy in downlink and uplink connections, respectively. When an MH is in active mode, it can receive the packets from the BS regularly. When an MH is in sleep mode, it turns off the transceiver power and wakes up periodically in a period of time (i.e., listening window). In the listening window, the MH listens the message from the BS to confirm whether the BS has any packets to the MH. If the BS has any packets to the MH, it must inform the MH in the listening window to let the MH keep awake and receive the packet. If the MH does not receive a positive notification from the BS, it will go back to another sleep cycle.

Based on preceding described PSM, when the MH is in sleep mode, the BS cannot communicate with the MH except in the pre-negotiated listening window. If the BS has packets to the MH, the BS has to send a positive notification to inform the MH. Because the positive notification does not include the information of the bandwidth allocation message, the MH must keep awake until it receives the downlink bandwidth allocation information. This leads the MH to take more unnecessary energy consumption to wait, especially the scheduler cannot allocate available bandwidth for the nrtPS/non-GBR transmission. We find that both of IEEE 802.16 and LTE-A do not have effective solution for this problem. According to the aforementioned increasing trend of nrtPS/non-GBR services in Section, this problem will become more series in the future.

Mechanism

As mentioned in Section and Subsection, when an MH receives a positive notification in the listening window, the MH keeps awake to wait for the downlink bandwidth allocation message to receive upcoming packets. To alleviate the energy wasting of MHs in case the downlink bandwidth are not available, EPSS modifies the positive notification message (e.g., the AAI_TRF-IND message used in IEEE 802.16m [5] and the physical downlink control channel (PDCCH) used in LTE-A [6]) to let the BS be able to notify those MHs which cannot obtain the bandwidth immediately to enter the extended sleep mode for battery power conservation. We note that EPSS can be applied to either IEEE 802.16m or LTE-A protocol because both of these two standards use the similar mechanism to perform the PSM for nrtPS or non-GBR connections and have the same problem of energy wasting.

[Figure 1] illustrates the operation of PSC type I in the IEEE 802.16m protocol and the described solution of EPSS. When an MH enters the sleep mode, the MH repeats the sleep cycle with binary exponential increment, i.e., $T_{\min}, 2T_{\min}, 4T_{\min}, \ldots, T_{\max}$. A

Table 1. IEEE 802.16 QoS Classes and Priorities [14].

service	abbrev.	priority	definition	typical applications
Unsolicited Grant Service	UGS	1	Real-time data streams comprising fixed-size data packets issued at periodic intervals	T1/E1 transport
Extended Real-time Polling Service	ertPS	2	Real-time service flows that generate variable-sized data packets on a periodic basis	VoIP
Real-time Polling Service	rtPS	3	Real-time data streams comprising variable-sized data packets that are issued at periodic intervals	MPEG Video
Non-real-time Polling Service	nrtPS	4	Delay-tolerant data streams comprising variable-sized data packets for which a minimum data rate is required	FTP with guaranteed minimum throughput
Best Effort	BE	5	Data streams for which no minimum service level is required and therefore may be handled on a space-available basis	HTTP

doi:10.1371/journal.pone.0103429.t001

Table 2. 3GPP Standardized QCI and Priorities [17].

QCI	Priority	Resource Type	Delay Budget	Example Services
1	2	GBR	100 ms	Conversational Voice
2	4	GBR	150 ms	Conversational Video (Live Streaming)
3	3	GBR	50 ms	Real-Time Gaming
4	5	GBR	300 ms	Non-Conversational Video (Buffered Streaming)
5	1	non-GBR	100 ms	IMS Signaling
6	6	non-GBR	300 ms	Video (Buffered Streaming), TCP-based (e.g. www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7	7	non-GBR	100 ms	Voice, Video, Interactive Gaming
8	8	non-GBR	300 ms	Video (buffered streaming) TCP-based (e.g., www, email, chat, FTP P2P file sharing, progressive video, etc.)
9	9	non-GBR	300 ms	Video (buffered streaming) TCP-based (e.g., www, email, chat, FTP P2P file sharing, progressive video, etc.)

doi:10.1371/journal.pone.0103429.t002

listening window T_L is at the beginning of each sleep cycle (except the first one) to let MHs receive AAI_TRF-IND signaling messages. If there is any downlink packets destined to the MH, BS will use SLPID field and bitmap in AAI_TRF-IND to notify the MH and transmit these packets later. When the MH receive the notification in the listening window, the MH will keep awake to wait the DL-MAP and receive the packets. If the packet transmission time is longer than the default listening window, the BS sends a sleep control header (SCH) to extend the listening window (e.g., the grey interval in T_i) till the end of current sleep cycle. When the MH finishes receiving the downlink traffic, i.e., there is no packets in the buffer of the BS for the MH, the BS will send an SCH to terminate the downlink transmission and the MH will go back to the sleep window till the end of sleep cycle. As was mentioned above, current IEEE 802.16m standard only assume an ideal condition that the MH will obtain the downlink bandwidth immediately after receiving the positive AAI_TRF-IND message. Nevertheless, this is not true that the BS can provide needed bandwidth to the MH in time to receive data. As we described in Section and Subsection, nrtPS/non-GBR services have lower QoS priority for bandwidth allocation than rtPS/GBR, this leads the MH to wait longer for available bandwidth. The worst case may happen when the bandwidth is saturated that the MH keep awake till the end of the sleep cycle even longer (e.g., the grey arrow in $2T_i$). Then, the detail of extra power saving scheme is described as follows.

If the BS use AAI_TRF-IND message to keep the MH awake but the scheduler cannot provide available bandwidth, the BS will send the AAI_SLP-RSP messages to update the sleep parameter of MH. Then the MH will enter an extended mode for a specified time duration T_E. T_E is estimated from the traffic and available bandwidth at that time. In the extended mode, the MH can turn off its transceiver as it is in sleep mode and avoid the energy wasting. The algorithm of EPSS is shown in [Figure 2]. When the MH in the listening window of sleep mode, the BS checks if there is any downlink traffic for it. If there is no traffic for the MH, it will go back to next sleep window. Otherwise, the BS will check whether the bandwidth is available. If the bandwidth is available, then the packet transmission will be performed. On the contrary, the MH will be ordered to enter the extended mode for T_E. When the MH sleeps for T_E, it will directly perform the packet transmission and the scheduler should promote the MH to higher priority. In this way, the MH can decrease the unnecessary waiting time when the system is saturated and achieve load balance of the system.

System Model

The system is modeled as a one BS to n MHs communication model. Because the PSM is used for downlink transmission, only downlink bandwidth is considered. The service queue of downlink for each MH follows the FCFS basis. The BS only maintains n NRT-VBR downlink queue for n MHs (i.e., each MH has only one downlink queue for NRT and BE services in the BS buffer). For simplicity, a simple bandwidth allocation policy is considered. Since the downlink bandwidth resource is limited, we assume that k resource spaces in total at most are allowed to assign among *downlink sessions* in a frame. In this paper, all NRT-VBR/non-GBR/BE services accepted by the call admission control (CAC) mechanism are treated as sessions. In order to increase bandwidth utilization, the number of allowed sessions could be greater than the number of total resource spaces per frame, i.e., $n > k$.

The PSM of EPSS is modeled as the Markov chain model [18] as shown in [Figure 3]. Similar models and assumptions can be found in [19,20], however, it is totally different from EPSS model and has to be redesigned. When an MH enters the sleep mode, the MH repeats the sleep cycle X_i, where

$$X_i = \begin{cases} X_{\min}, & \text{for } i=1 & \text{(1a)} \\ 2^{i-1}X_{\min} & \text{for } 1<i<m & \text{(1b)} \\ X_{\max}, & \text{for } i=m. & \text{(1c)} \end{cases}$$

An MH wakes up periodically to listen AAI_TRF-IND during T_L at the beginning of each sleep cycle X_i, where $X_{\min} \le X_i \le X_{\max}$, and then goes back to sleep window till the end of the sleep cycle. The value of sleep window T_i in different states is

$$T_i = 2^{i-1}X_{\min} - T_L, \text{ for } 1 \le i \le m. \tag{2}$$

In the state transition diagram, all the states are represented as a set $S = \{S_E, S_0, S_1, \ldots, S_m\}$, where S_E represents the state that the MH is informed with a positive AAI_TRF-IND but cannot be
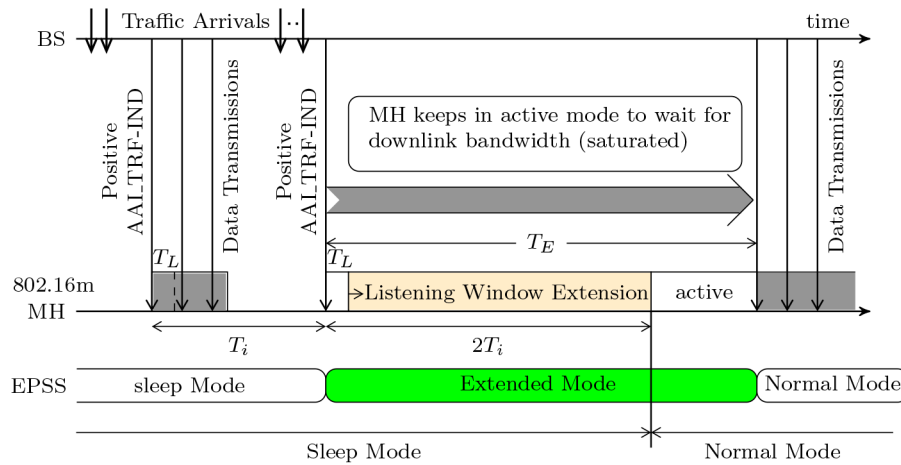
Figure 1. The type I power saving operation in IEEE 802.16m standards.
doi:10.1371/journal.pone.0103429.g001
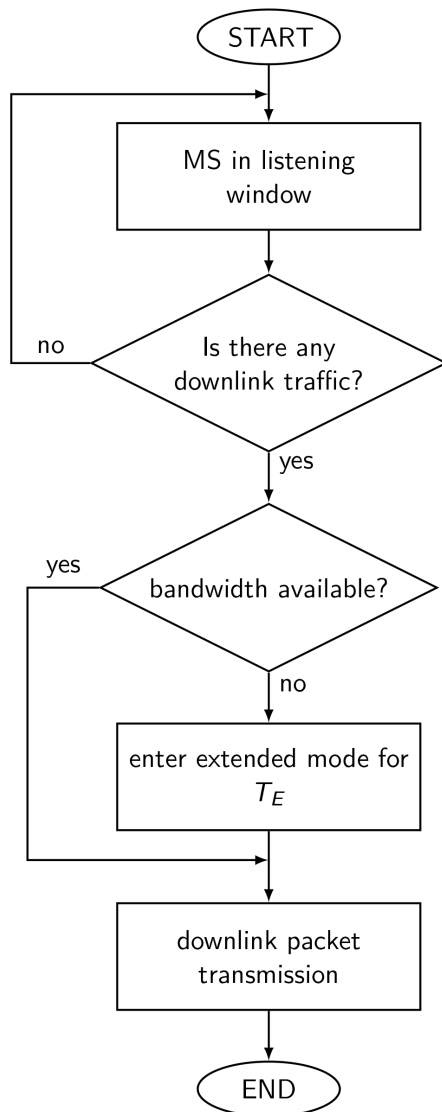
Figure 2. The algorithm of EPSS scheme.
doi:10.1371/journal.pone.0103429.g002

provided resource. In EPSS, the MH will enter the extended sleep mode to avoid unnecessary energy consumption. S_0 represents the normal mode state and the subset $S_S = \{S_1, S_2, \ldots, S_m\}$ represents the sleep mode states. All elements in S_S follow the sequence of $S_1 \prec S_2 \prec \cdots \prec S_m$ in chronological order. An MH sleeps a time duration T_i in the state S_i. After the duration time T_i (T_0 represents the idle time before MH enters the sleep mode and $T_i, 1 \leq i \leq m$, represents the ith sleep window), an MH may transit from S_i to S_{i+1} with the probability p_i or enter the normal state S_0 with the probability $1 - p_i$ depending on the downlink traffic to the MH. Assume the downlink arrivals to an MH are a Poisson process with a mean rate λ [21,22]. Let p_i be the probability that the MH transits from S_i to S_{i+1}, where $S_i \in S_S$, in the sleep mode. Because the probability p_i depends on the probability of no downlink arrivals to the MH during a time interval $X_i = t_{i+1} - t_i$, where t_i represents the time the MH enters S_i, we have

$$p_i = \Pr\{\text{no arrivals in time } X_i\} = e^{-\lambda X_i}. \qquad (3)$$

State Transition Probability of EPSS

In EPSS, the MH will enter S_E if it cannot obtain the resource immediately, i.e., the network load is saturated. Let p_b be the downlink transmission blocking probability (i.e., the transition probability from S_0 to S_E as shown in [Figure 3]). As mentioned in Section, the system has k parallel sessions in a frame and n sessions are permitted to be served in the system. The queueing model of EPSS can be formulated as the *Erlang-B formula* [18] with an $M/M/k/k$ model. According to the $M/M/k/k$ model, p_b can be obtained by

$$p_b = \frac{(n\lambda)^k / (k! \mu^k)}{\sum_{i=0}^{k} (n\lambda)^i / (i! \mu^i)}. \qquad (4)$$

Since the transition probability from S_0 to S_1 is p_0 (see [Figure 3]), the state transition probability from S_0 to S_0 (i.e., an MH stays in S_0) is equal to $1 - p_b - p_0$. In EPSS, since an MH transits from extended mode to normal mode (i.e., S_E to S_0) without any conditions once the extended time is decreased to zero, the state transition probability from S_E to S_0 is equal to 1.
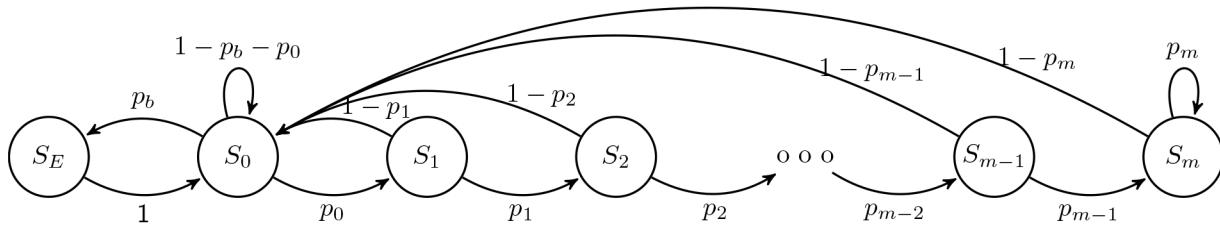
Figure 3. The state transition diagram of EPSS.
doi:10.1371/journal.pone.0103429.g003

According to the *flow balance* property of steady state in Markov chain of EPSS, we have

$$\begin{cases} P(S_1) = P(S_0)p_0, & \text{(5a)} \\ P(S_2) = P(S_1)p_1 = P(S_0)p_1p_0, & \text{(5b)} \\ \quad \vdots \\ P(S_{m-1}) = P(S_{m-2})p_{m-2} = P(S_0)p_{m-2}p_{m-3}\cdots p_1p_0, & \text{(5c)} \\ P(S_m) = \dfrac{p_{m-1}}{1-p_m}P(S_{m-1}) = \dfrac{P(S_0)p_{m-1}p_{m-2}\cdots p_1p_0}{1-p_m}. & \text{(5d)} \end{cases}$$

The sleep mode probability can be obtained by summing up all probabilities of sleep mode states, then according to (5), we can obtain

$$P(S_S) = P(S_1) + P(S_2) + \cdots + P(S_m)$$

$$= P(S_0)\left(p_0 + p_0p_1 + \cdots + p_0p_1\cdots p_{m-2} + \frac{p_0p_1\cdots p_{m-2}p_{m-1}}{1-p_m}\right) \quad (6)$$

Following the steady state property, in state S_0, we have

$$P(S_E) + \sum_{i=1}^{m} P(S_i)(1-p_i) = P(S_0)p_0 + P(S_0)p_b. \quad (7)$$

Solving $P(S_E)$ for (7), we have

$$P(S_E) = P(S_0)(p_0 + p_b) - \sum_{i=1}^{m} P(S_i)(1-p_i)$$

$$= P(S_0)(p_b + p_0) - \left(P(S_0)\sum_{i=1}^{m-1}(1-p_i)\prod_{j=0}^{i-1}p_j + P(S_0)\prod_{j=0}^{m-1}p_j \right)$$

$$= P(S_0)p_b. \quad (8)$$

To solve $P(S_0)$, we use the axiom of probability theory

$$P(S_E) + P(S_0) + P(S_1) + \cdots + P(S_m)$$
$$= P(S_E) + P(S_0) + P(S_S) = 1, \quad (9)$$

Substituting (6) and (8), into (9) leads to

$$1 = P(S_0)\left(p_b + 1 + p_0 + \cdots + p_0p_1\cdots p_{m-2} + \frac{p_0p_1\cdots p_{m-2}p_{m-1}}{1-p_m}\right)$$

$$= P(S_0)\left(p_b + 1 + \sum_{i=1}^{m-1}\prod_{j=0}^{i-1}e^{-\lambda X_j} + \frac{\prod_{i=0}^{m-1}e^{-\lambda X_i}}{1-e^{-\lambda X_m}}\right). \quad (10)$$

Then $P(S_0)$ can be obtain as

$$P(S_0) = \left(p_b + 1 + \sum_{i=1}^{m-1}\prod_{j=0}^{i-1}e^{-\lambda X_j} + \frac{\prod_{i=0}^{m-1}e^{-\lambda X_i}}{1-e^{-\lambda X_m}}\right)^{-1}. \quad (11)$$

Substitute (11) into (8) and (5), we can obtain the steady state probability $P(S_E)$ and $P(S_S)$.

Energy Consumption

The energy consumption can be derived by using preceding obtained probabilities $P(S_S)$, $P(S_E)$, and $P(S_0)$. Let ϖ_S and ϖ_N denote the energy consumption per unit time in sleep mode and normal mode, respectively. Thus, the energy consumption of the EPSS per unit time, denoted as $E[\varpi_E]$, can be obtained by

$$E[\varpi_E] = (P(S_S) + P(S_E))\varpi_S + P(S_0)\varpi_N. \quad (12)$$

Figure 4. The rate transition diagram for the $M/M/k$ queueing mode.
doi:10.1371/journal.pone.0103429.g004
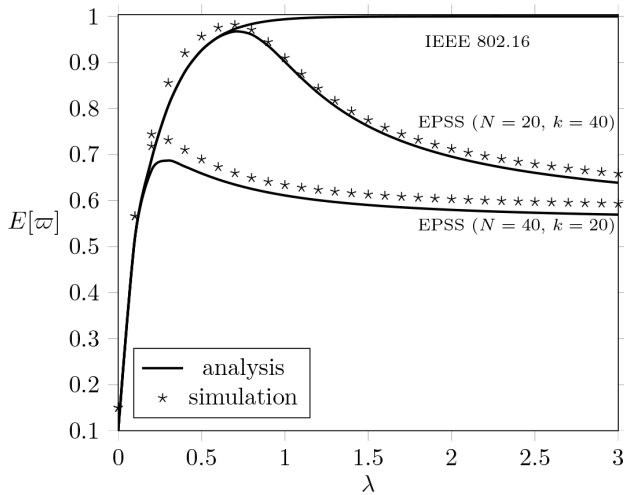
Figure 5. Simulation versus analysis in terms of arrival rate λ when $T_0=5$, $T_{\min}=4$, $T_{\max}=512$, and $\mu=0.5$.
doi:10.1371/journal.pone.0103429.g005

In comparison with EPSS, we use the similar model and obtain the steady state probability of the legacy IEEE 802.16 PSM

$$P_{16}(S_0) = \left(1 + \sum_{i=1}^{m-1}\prod_{j=0}^{i-1}e^{-\lambda X_j} + \frac{\prod_{i=0}^{m-1}e^{-\lambda X_i}}{1-e^{-\lambda X_m}}\right)^{-1}. \quad (13)$$

In a similar way, we can substitute the probability $P(S_S)$ and derive the energy consumption of the legacy IEEE 802.16 PSM per unit time, denoted as $E[\varpi_{16}]$, as

$$E[\varpi_{16}] = P(S_S)\varpi_S + P_{16}(S_0)\varpi_N. \quad (14)$$

Extended Time Estimation

The delay of access network bandwidth is an important metric of performance. In order to estimate T_E, we take the $M/M/k$ Markovian queueing model to derive the estimation function of

T_E. We assume that there are at most k parallel sessions (transmissions) allowed at a frame in the system. [Figure 4] shows the Markov chain model with states $s(i)$, where $i \geq 0$ represents the number of busy channels. When the process is in state $s(i)$, for $0 \leq i < k$, i channels are busy. The effective downlink traffic to an BS at $s(i)$ is $n\lambda$, where n represents the number of MH served by the BS. Suppose the downlink transmission time is exponentially distributed with mean $1/\mu$. Because a busy channel is released with rate μ (i.e., the service rate of a channel), the process moves from $s(i)$ to $s(i-1)$ for $1 \leq i \leq k$ with rate $i\mu$.

According to the process, the steady-state probabilities π_i for $s(i)$ can be expressed as

$$\pi = \begin{cases} \dfrac{(n\lambda)^i}{i!\mu^i}\pi_0, & \text{for } i<k, & (15a) \\[2ex] \dfrac{(n\lambda)^i}{k^{(i-k)}k!\mu^i}\pi_0, & \text{for } i\geq k. & (15b) \end{cases}$$

To solve π_0, we use $\pi_0 + \pi_1 + \cdots + \pi_k + \cdots = 1$ to obtain

$$\pi_0 = \left(\sum_{i=0}^{k-1}\frac{(n\lambda)^i}{i!\mu^i} + \sum_{i=k}^{\infty}\frac{(n\lambda)^i}{k^{i-k}k!\mu^i}\right)^{-1}. \quad (16)$$

Based on the derivation of [18], we can rewrite (16) by substituting $r = n\lambda/\mu$ and $\rho = n\lambda/(k\mu)$ as

$$\pi_0 = \left(\frac{r^k}{k!(1-\rho)} + \sum_{i=0}^{k-1}\frac{r^i}{i!}\right)^{-1}, \quad (17)$$

where r represents the work load rate to a server (i.e., a channel) and ρ means the work load rate to the system. We notice that the reason we use r and ρ to replace $n\lambda/\mu$ and $n\lambda/(k\mu)$ here is to easily understand the status of the traffic load in the whole system.

Let $E[L_q]$ be the expected number of MHs that cannot be served and enters S_E, i.e., $i>k$. Then

$$E[L_q] = \sum_{i=k+1}^{\infty}(i-k)\pi_i = \sum_{i=k+1}^{\infty}(i-k)\frac{r^i}{k^{(i-k)}k!}\pi_0$$
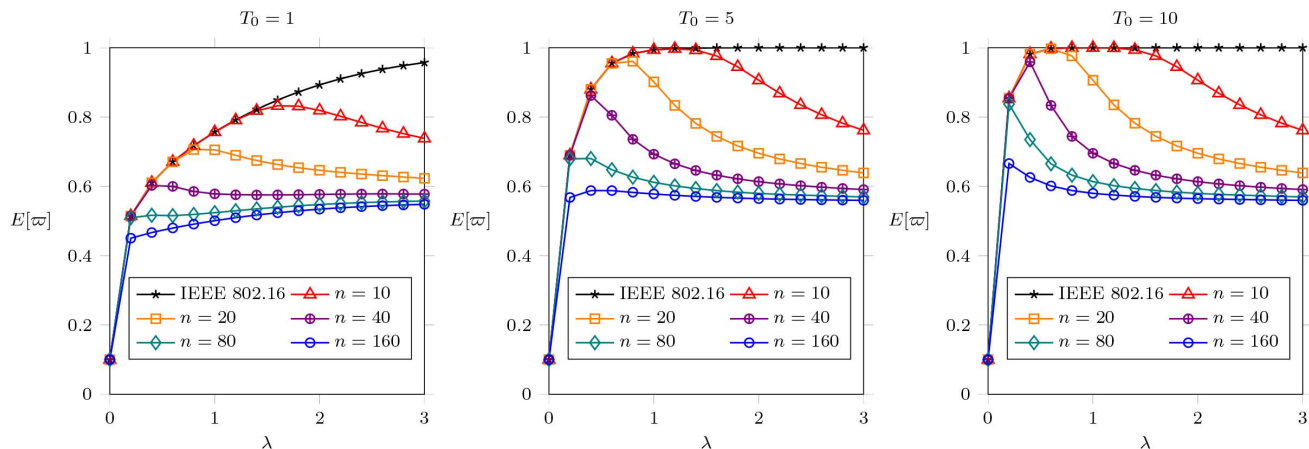
Figure 6. The effect of different T_0 on $E[\varpi]$ when $T_{\min}=4$, $T_{\max}=512$, $T_L=1$, $R=0.1$, $k=40$, and $\mu=0.5$.
doi:10.1371/journal.pone.0103429.g006
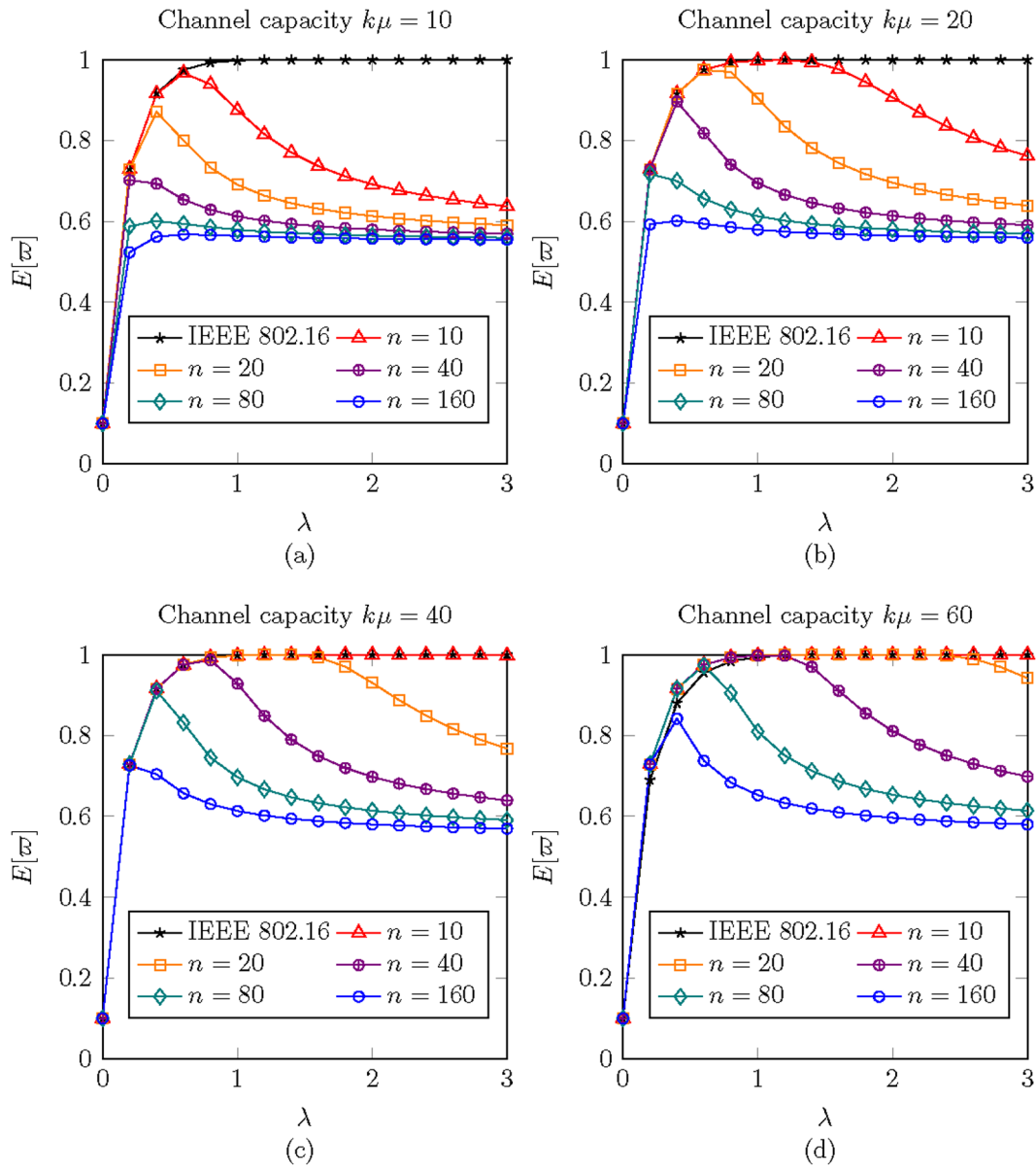
Figure 7. The effect of different $k\mu$ on $E[\varpi]$ when $T_0=6$, $T_{\min}=4$, $T_{\max}=512$, $T_L=1$, $R=0.1$, and $\mu=0.5$.
doi:10.1371/journal.pone.0103429.g007

$$= \frac{r^k \pi_0}{k!} \sum_{j=1}^{\infty} j\rho^j = \frac{r^k \rho \pi_0}{k!} \sum_{j=1}^{\infty} j\rho^{j-1}$$

$$= \frac{r^k \rho \pi_0}{k!} \frac{d}{d\rho} \sum_{j=1}^{\infty} \rho^j$$

$$= \left(\frac{r^k \rho}{k!(1-\rho)^2} \right) \pi_0. \qquad (18)$$

Substituting (17) into (18), we get

$$E[L_q] = \frac{r^k \rho}{k!(1-\rho)^2} \left(\frac{r^k}{k!(1-\rho)} + \sum_{i=0}^{k-1} \frac{r^i}{i!} \right)^{-1}. \qquad (19)$$

Because the downlink transmission arrival rate of nrt-PS to the BS is $n\lambda$, it is easy to use Little's formula to get $E[T_E]$:

$$E[T_E] = \frac{E[L_q]}{n\lambda}. \qquad (20)$$

From (20), the BS can estimate the expected time duration $E[T_E]$ according to the current traffic load to the MH. Note that the computed $E[T_E]$ will be adopted by the scheduler of the BS to
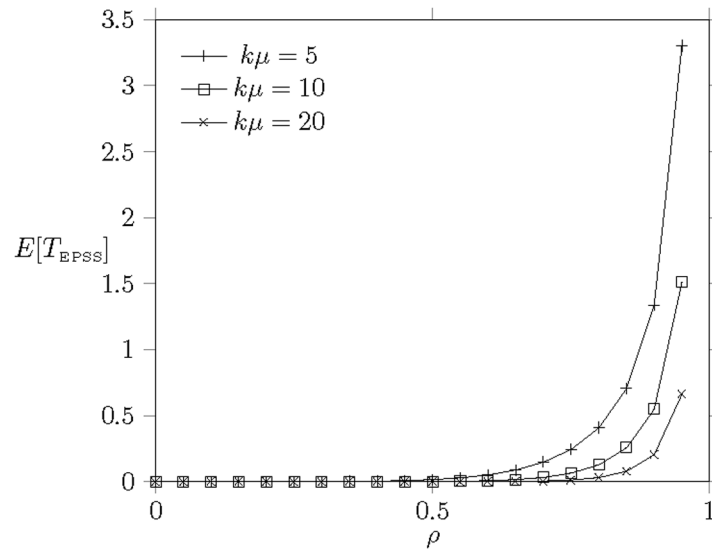
Figure 8. The downlink $E[T_{\mathrm{EPSS}}]$ versus work load ρ when $T_0=6$, $T_{\min}=4$, $T_{\max}=512$, $T_L=1$, $R=0.1$, $\mu=0.5$, and $n=20$.
doi:10.1371/journal.pone.0103429.g008

schedule the needed resource to the MH accordingly after the period of sleeping time $E[T_E]$. Thus, the MH goes to S_E for $E[T_E]$ and comes back to receive the downlink traffic. Let T_{EPSS} denote the total time an MH wakes from the sleep mode and finishes the downlink transmission. Thus the mean downlink transmission delay of each downlink attempt achieved by EPSS will be

$$E[T_{\mathrm{EPSS}}] = \frac{1}{\mu} + E[T_E]$$

$$= \frac{1}{\mu} + \frac{r^k \rho}{n\lambda k!(1-\rho)^2} \left(\frac{r^k}{k!(1-\rho)} + \sum_{i=0}^{k-1} \frac{r^i}{i!} \right)^{-1}. \quad (21)$$

Numerical Experiments

Simulation Model

The simulator is developed by the authors to evaluate the energy consumption of the legacy IEEE 802.16 PSM and EPSS. The simulation model is based on all system assumptions presented in Section. There are n independent MHs which connect with one BS for Internet access. The downlink traffic to an MH is an independent Poisson source with an aggregate mean burst data generation rate of λ packet arrivals/frame and the total traffic to the downlink channel is equal to $n\lambda$. The arrival data consist of consecutive packets arrive at the BS and are scheduled for downlink burst transmission. Once the downlink traffic arrives at the BS, the BS will wake up the MH at the next listening window if the MH is in the sleep mode. The scheduler follows the FCFS basis to schedule the arrival data. In EPSS, we emulate the
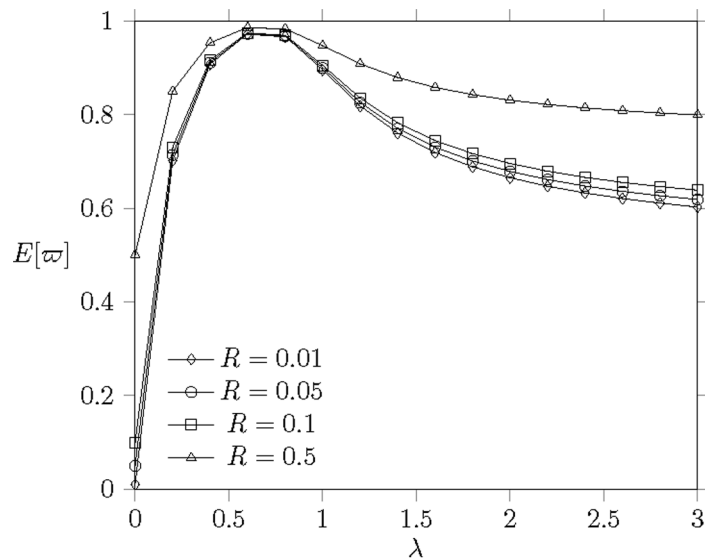
Figure 9. The energy consumption under different R when $T_0=6$, $T_{\min}=4$, $T_{\max}=512$, $T_L=1$, $k=40$, $\mu=0.5$, and $n=20$.
doi:10.1371/journal.pone.0103429.g009

Table 3. Analysis Parameters for IEEE 802.16.

Parameter	Value
simulation time	120 000 frames
bandwidth	56 Mbps
frame size	5 ms
downlink/uplink ratio per frame	5/3
LRU size (16-QAM, coding rate 3/4)	432 bits
minimum resource allocation size per frame	12 LRUs
T_0	5 frames
T_{min}	4 frames
T_{max}	512 frames
mean size of video streaming per request	176 Mbits
mean size of web browsing per request	6 Mbits

doi:10.1371/journal.pone.0103429.t003

behavior of MHs which are served. If an MH does not get enough resource to receive the downlink traffic after it wakes up from sleep mode. The BS will count $E[T_E]$ that is obtained from (20) and order the MH to enter the extended mode. After the duration of $E[T_E]$, the MH will wake up from the extended mode and get a priority queue to be allocated the DL radio resource.

The transmission of arrival in a session is assumed exponentially distributed with mean rate μ packet arrivals/frame. Thus the total downlink channel capacity can be represented as $k\mu$, where k indicates the scheduled number of sessions per frame. When k increases, this implies the number of served session increases, then the allocation resource for each k will decrease. When μ increases, this implies the modulation rate (carried bits per one logic resource unit) increases. In other words, the larger the $k\mu$ is, the larger the capacity of service in a frame is. In the simulation, the MH will enter the normal mode after it goes out the extended mode and

completes the reception of the downlink data. The MH will enter the sleep mode again till it successfully waits a T_0 period without any downlink traffic. The provided work load for the system is set as $\rho = n\lambda/(k\mu)$.

To study the effect of the power saving achieved by EPSS, we investigate the relationship between ϖ_S and ϖ_N. Let

$$R = \frac{\varpi_S}{\varpi_N} \qquad (22)$$

indicate the ratio of energy consumption in the sleep mode to the normal mode. To normalize the energy consumption we let $\varpi_N = 1$. For example, when $R = 0.1$, which means the energy consumption in the sleep mode is $1/10$ of the energy consumption in the normal mode, i.e., $\varpi_S = 0.1$. In the following evaluation and comparison, the normalized energy consumption per frame,
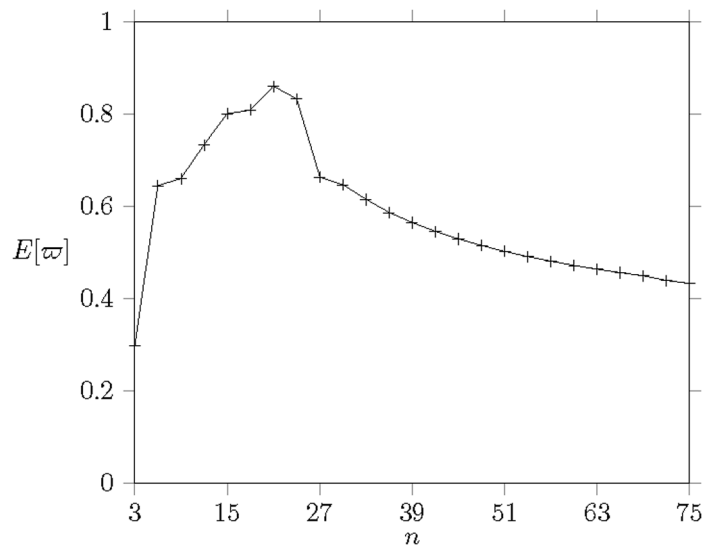
Figure 10. The energy consumption against the number of MSs n in realistic traffic where the ratio of video traffic to web traffic is 2 : 1.
doi:10.1371/journal.pone.0103429.g010
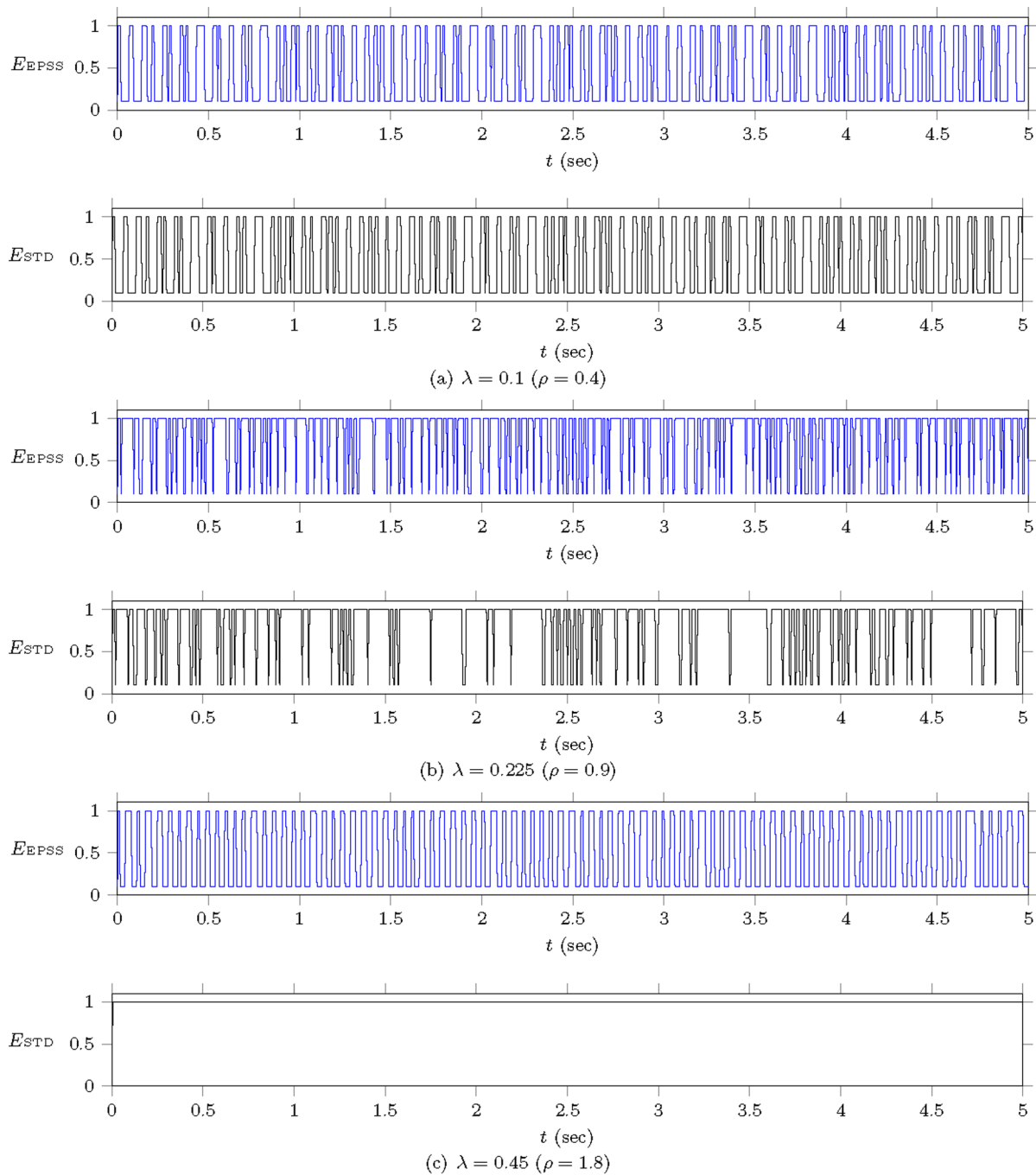
Figure 11. Example of energy consumption for EPSS and IEEE 802.16 in a single MH with $T_0 = 6$**,** $T_{\min} = 4$**,** $T_{\max} = 512$**,** $T_L = 1$**,** $n = 10$**,** $k = 10$**, and** $\mu = 0.25$**. (a) middle traffic load** $\rho = 0.4$**. (b) saturated traffic load** $\rho = 0.9$**. (c) overloaded traffic load** $\rho = 1.8$**.**
doi:10.1371/journal.pone.0103429.g011

denoted by $E[\varpi]$, is used as our performance measurement. When $E[\varpi] = 1$, it implies that the MHs are always in normal mode and do not take advantage of sleep mode.

Other system parameters are set as follows: the initial sleep window $T_{\min} = 4$, the listening window $T_L = 1$, the number of states in sleep mode $m = 8$, i.e., $T_{\max} = 512$, the given idle time duration $T_0 = 6$ to enter the sleep mode, the ratio of energy consumption $R = 0.1$, and the channel capacity $k\mu = 20$. Our analytical model has been validated against the simulation

experiments. We perform the simulation 1200 times to average each plotting value as shown in [Figure 5]. The solid lines are the numerical results obtained from our derived formulae. We can see from [Figure 5] that the simulation results matches the numerical results.

Numerical Examples

In this subsection, we use some numerical examples to illustrate the effects of the number of MHs n, the initial idle time duration

T_0, the downlink channel capacity $k\mu$, the maximum number of sleep mode states m, the energy consumption ratio R, and $T_{\min}$ on output measure $E[\varpi]$ (the normalized energy consumption). Further, we also show the effect of $k\mu$ on the downlink transmission delay $E[T_{\text{EPSS}}]$ caused by EPSS.

Effect of n. [Figure 6] illustrates the output measure $E[\varpi]$ by varying different downlink arrival rate λ when $T_{\min}=4$, $m=8$, $T_L=1$, $R=0.1$, and $k\mu=20$ ($k=40$, $\mu=0.5$) under different $T_0=1$, 5, and 10. [Figure 6] shows that $E[\varpi]$ of IEEE 802.16 PSM increases as λ increases (the larger the downlink arrival rate is, the higher the energy consumption it will be). Taking the subfigure $T_0=5$ as an example, the upper bound of IEEE 802.16 reaches 1 as λ reaches 1.6 ($\rho=n\lambda/(k\mu)=16/20=0.8$). This result implies that the energy consumption achieved by IEEE 802.16 cannot be reduced any more when the work load is approximately saturated.

In contrast, initially $E[\varpi_E]$ increases as λ increases but $E[\varpi_E]$ will decreases when download traffic load approaches saturation, i.e., $\rho \approx 0.8$. Unlike the IEEE 802.16 PSM, EPSS estimates an extended sleep time to let MH enter the extended mode for extra power saving, thus the energy consumption will decrease accordingly when ρ is more than 1. [Figure 6] shows that EPSS can achieve $E[\varpi_E]=0.57$ at most when ρ is very large. This implies that EPSS is capable of conserving the energy consumption as the channel utilization is saturated.

[Figure 6] shows that the energy consumption is proportionally decreasing when n increases. We first note that n is a major factor that determines $E[\varpi]$. The larger the number of n implies the more competition among MHs to obtain the bandwidth is intense. In other words, MHs in highly competitive circumstance, the delay of obtaining the bandwidth for transmission is longer. And thus EPSS enforces MHs to enter the extended mode, which could avoid the MH spending unnecessary idle time to waste the energy.

Effect of T_0. [Figure 6] additionally plots T_0 against $E[\varpi]$, where $T_0=1$, 5, and 10 ($T_0=10$ is a maximum recommended value for practice [23]). This consequence shows the value of T_0 will affect $E[\varpi]$. When T_0 is larger, MHs will obtain lower energy saving efficiency, i.e., $E[\varpi]$ is higher, because an MH should wait a longer time to enter the sleep mode. It makes sense that MHs get quicker response time to receive the downlink transmission if the MH prolongs T_0 to enter the sleep mode. However, this effect will decrease the efficiency of energy conservation.

Effect of $k\mu$. [Figure 7] plots $E[\varpi]$ against the arrival rate λ to each MH. Each subfigure represents different $k\mu$ as $k\mu=10$, 20, 40, and 60, where $\mu=0.5$, i.e., $1/\mu=2$ frames. The parameter $\mu=0.5$ implies each arrival to an MH dominates 2 frames long in average to be served. [Figure 7] shows that, in general, $E[\varpi]$ grows up proportionally with the increase of λ before the offered load rate is saturated ($\rho<1$). First, $E[\varpi_{16}]$ will reach 1 when $\rho \approx 1$. The energy consumption cannot be reduced further because the IEEE 802.16 PSM cannot reflect the offered load. On the contrary, EPSS can reflect the traffic load and enforce the MHs to enter the extended mode for power conservation.

[Figure 7] shows that energy consumption is proportional to $k\mu$. For example, in [Figure 7](a), EPSS will conserve more energy when the channel capacity $k\mu$ is small. In contrast to [Figure 7](c), [Figure 7](d) shows the $E[\varpi_E]$ is the same as $E[\varpi_{16}]$ when n is small and $k\mu$ is large. This result indicates that EPSS exhibits its effect on $E[\varpi]$ as the network is saturated. There are two cases to lead the situation of network saturation. One is $n \leq k$, i.e., the number of MHs n less than or equal to the number of serving channel in a frame. The other one is $n>k$, i.e., the number of MHs n greater than the number of serving channel in a frame. In case one, the system saturation only takes place when each MH

has so much traffic to be received, so it is always in normal mode. In the second case, the resource is less than the request of all MHs. In this case, it leads some MHs idle to wait the BS to allocate resource. It motivates the EPSS to be proposed to reduce the MH idle time and conserve more energy. The effect of EPSS on $E[\varpi]$ will be more obvious when $k\mu$ is small or n is large.

Another affected result, the mean downlink transmission delay $E[T_{\text{EPSS}}]$, caused by $k\mu$ is shown in [Figure 8]. As we can expect that $E[T_{\text{EPSS}}]$ grows rapidly as the network load is saturated, i.e., $\rho>0.8$, because there are no more bandwidth to be allocated. This result explains the phenomenon that $E[\varpi]$ decreases when $\rho>0.8$ (see [Figure 7]) because EPSS enforces the MHs into the extended mode for energy conservation.

Effect of R. [Figure 9] illustrates the effect of energy consumption of sleep mode to normal mode on $E[\varpi]$. The revealed data indicate that the higher ratio of R achieves lower $E[\varpi]$. The improvement is significant when R from 0.5 to 0.1, i.e., ϖ_S is from $1/2$ of ϖ_N to $1/10$ of ϖ_N. However, when R from 0.1 to 0.01 (10 time smaller than 0.1), the improvement of $E[\varpi]$ is getting smaller. The gap of $E[\varpi]$ improvement between the gap between 0.05 and 0.1 and the gap between 0.1 and 0.5 is $0.02007/0.1606=0.125=1/8$. This observation means the improvement from $R=0.5$ to $R=0.1$ is 8 times of the improvement from $R=0.1$ to $R=0.05$. Based on the discovery, we recommend the design of $\varpi_S=0.1\varpi_N$ is an efficient value to achieve good energy conservation.

Realistic Traffic

In the following, we present a realistic traffic study of buffered video streaming and web browsing to investigate the effect of EPSS on energy conservation. The traffic model refers to the statistical results in [3,4]. We have developed a C++ program to emulate the two traffic types. The system parameters used in the simulation follow the IEEE 802.16 standard and are shown in Table 3. The minimum resource allocation size is 12 logic resource units (LRUs) where each LRU can bear 432 bits. Therefore, there are $k=18$ available resource blocks a frame (downlink) deducting the control overheads, e.g., super frame headers, mapping information, and control messages. We assume the ratio of video streaming traffic to web browsing traffic is 2 : 1. Each MH can only make a request to the BS for service. The simulation time is 10 minutes (120 000 frames) long to ensure that simulation results are stable.

The buffered video streaming traffic we consider here is the behavior of people watching the YouTube video clip. The resolution of video clip is 480p (854 × 480 pixels) resolution, which consumes bandwidth about 1000 kbps. The length of video clip is 3 minutes long on average, i.e., each downloading of a video clip will consume up to 176 Mbits bandwidth. The web browsing behaves like the burst traffic model, i.e., a download uses 2 Mbps on average, while having "peaks" bursting up to 2.4 Mbps. Each downloading of a web consumes total bandwidth of 6 to 12 Mbits on average and the user will spend 10 seconds on average to read the web.

To serve the two traffic types, a simple round-robin scheduler is built. Each request (video or web downloading) has at least a resource block quota for service when $k \geq n$. If $k<n$, the scheduling policy among nodes will be based on the round-robin basis. The scheduler serves each video streaming request at most 2000 frames (10 seconds). If a video transmission has been served 10 seconds, the scheduler will release the channel and serve another request.

The simulation results are shown in [Figure 10]. [Figure 10] shows the obtained results are very close to the numerical results.

We notice that the network load is not saturated when $n = 18$ ($\rho = 0.87$) because there are 6 web requests (dominating 1/3 total traffic). As we mentioned that the web browsing behaves like brusty traffic. Therefore, the web browsing traffic only dominates one resource block a frame sustaining 606 frames (≈ 3 seconds) and then releases the channel. As we saw from [Figure 10], EPSS can reflect the network load and enforce the MHs entering the sleep mode to conserve energy if they cannot obtain the bandwidth immediately. The realistic traffic study shows EPSS has the capability of saving more energy when MHs face the saturated condition.

[Figure 11] demonstrates the process of detailed energy consumption resulted by applying the EPSS and legacy IEEE 802.16 PSM. The simulation is performed for 5 seconds (1000 frames). To show the impact of EPSS on energy conservation, the simulation parameters are set as $k = 10$, $n = 10$, $\mu = 0.25$, and $T_0 = 6$. Three different traffic loads, $\lambda = 0.1$ ($\rho = n\lambda/(k\mu) = 0.4$) for a middle load, $\lambda = 0.225$ ($\rho = 0.9$) for a saturated load, and $\lambda = 0.45$ ($\rho = 1.8$) for an overloaded load, are investigated to observe the effectiveness of EPSS on energy conservation. The energy consumption in ϖ_N and ϖ_S is 1 and 0.1. The energy consumption record is sampled at each frame. In the middle load ([Figure 11](a)), the energy consumption of EPSS, denoted by E_{EPSS}, is similar with that achieved by IEEE 802.16, denoted by E_{STD}. It is because that the traffic load is not saturated (i.e., $\rho = 0.4$) and the extended sleep mode is not activated. When the traffic load increases (i.e., $\rho = 0.9$ for saturated condition), as shown in [Figure 11](b), EPSS can enforce the MH to enter the extended mode for extra energy conservation as compared with the IEEE 802.16 PSM. Finally, we use an extreme case (i.e., $\rho = 1.8$ is sometimes taken by call admission control (CAC) for channel utilization improvement) to observe the effectiveness of EPSS. [Figure 11](c) shows that the MH adopts EPSS is able to enter the extended mode to conserve its energy when it cannot

obtain the bandwidth in the active mode. This effect becomes more obviously when the traffic load is overloaded. For instance, visual recognition applications via mobile equipments for location identification are widely used recently [24–26]. These applications will query a great number of image and video and consume the bandwidth. In this case, EPSS has the capability of conserving the MH's energy. As compared with EPSS, MHs with the IEEE 802.16 PSM will wait for the downlink bandwidth and waste the energy all the time.

Conclusion

In this paper, we invent a new power saving scheme, called *extra power saving scheme* (EPSS), to let MHs enter the *extended sleep mode* for extra power saving when the downlink traffic load is saturated. We show the impact of EPSS on energy consumption $E[\varpi]$ is significant when the number of MHs n or the traffic load ρ is large. We comprehensively show how the various parameters of the system, T_0, n, $k\mu$, and R, affect the performance of $E[\varpi]$. The performance analyses (i.e., the expected energy consumption and access delay) are verified by simulation, which includes an investigation of realistic traffic flows. Our invention gives a guideline to design PSM by considering the situation of downlink traffic load. EPSS can be adopted by the downlink scheduler of the BS to conserve the energy consumption of MHs. EPSS is suitable for applications that need a great deal of downloading such as mobile visual search. Moreover, how to combine with scheduling policies to optimize the efficiency of power saving is an important future work for study.

Author Contributions

Conceived and designed the experiments: JC WHT. Performed the experiments: JC WHT. Analyzed the data: JC WHT. Wrote the paper: JC WHT JDL.

References

1. G Americas (2013) 3GPP release 11: understanding the standards for HSPA+ and LTE-advanced enhancements. LTE Whitepaper & Tutorials.
2. Stevens-Navarro E, Lin Y, Wong VWS (2008) An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. IEEE Trans. Veh. Technol57(2): 1243–1254. doi: 10.1109/TVT.2007.907072
3. Cisco (2013) Cisco visual networking index: global mobile data traffic forecast update, 2012–2017. Cisco VNI Mobile Forecast.
4. Palo Alto Networks (2013) An analysis of application usage and related threats–regional findings. The Application Usage and Threat Report. Avaliable: http://www.paloaltonetworks.com.
5. IEEE 802.16 Working Group (2011) IEEE standard for local and metropolitan area networks–part 16: air interface for broadband wireless access systems, amendment 3. IEEE Std. 802.16m.
6. GPP (2013) Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; Stage 2. 3GPP TS 36.300 V10.10.0.
7. Alshamrani A, Shen X, Xie LL (2011) QoS Provisioning for Heterogeneous Services in Cooperative Cognitive Radio Networks. IEEE J. Sel. Areas Commun29(4): 819–830. doi: 10.1109/JSAC.2011.110413
8. Niyato D, Hossain E (2008) A game theoretic analysis of service competition and pricing in heterogeneous wireless access networks. IEEE Trans. Wireless Commun7(12): 5150–5155. doi: 10.1109/T-WC.2008.070609
9. Wang R, Tsai J, Maciocco C, Tai TC, Wu J (2011) Reducing power consumption for mobile platforms via adaptive traffic coalescing. IEEE J. Sel. Areas Commun29(8): 1618–1629. doi: 10.1109/JSAC.2011.110911
10. Cicconetti C, Lenzini L, Mingozzi E, Vallati C (2010) Reducing power consumption with QoS constraints in IEEE 802.16e wireless networks. IEEE Trans. Mobile Comput9(7): 1008–1021. doi: 10.1109/TMC.2010.53
11. Hsu CH, Feng KT, Chang CJ (2010) Statistical control approach for sleep-mode operations in IEEE 802.16m systems. IEEE Trans Veh. Technol59(9): 4453–4466. doi: 10.1109/TVT.2010.2070086
12. Tseng YC, Chen JJ, Yang YC (2011) Managing power saving classes in IEEE 802.16 wireless MANs: a fold-and-demultiplex method. IEEE Trans. Mobile Comput10(9): 1237–1247. doi: 10.1109/TMC.2010.215
13. Choi HH, Lee JR, Cho DH (2007) Hybrid power saving mechanism for VoIP services with silence suppression in IEEE 802.16e systems. IEEE Commun. Lett11(5): 455–457. doi: 10.1109/LCOMM.2007.070035
14. IEEE 802.16 WG (2006) IEEE standard for local and metropolitan area networks–part 16: air interface for fixed and mobile broadband wireless access systems, amendment 2: physical and medium access control layers for combined fixed and mobile operation in licensed bands. IEEE Std 802.16e-2005.
15. Chen J, Tarn WH, Lee JD (2014) A downlink and uplink alignment scheme for power saving in IEEE 802.16 protocol. The Scientific World Journal 2014: 1–11. doi: 10.1155/2014/217973
16. Vu HL, Chan S, Andrew LLH (2010) Performance analysis of best-effort service in saturated IEEE 802.16 networks. IEEE Trans. Veh. Technol59(1): 460–472. doi: 10.1109/TVT.2009.2033191
17. GPP (2013) Technical specification group services and system aspects; policy and charging control architecture; overall description. 3GPP TS 23.203 V12.2.0.
18. Gross D, Shortle JF, Thompson JM, Harris CM (2008) Fundamentals of Queueing Theory, the 4th Edt. Hoboken, NJWiley. doi: 10.1002/9781118625651
19. Jeong DG, Jeon WS (2006) Performance of adaptive sleep period control for wireless communications systems. IEEE Trans. Wireless Commun5(11): 3012–3016. doi: 10.1109/TWC.2006.04772
20. Hwang E, Kim KJ, Son JJ, Choi BD (2010) The power-saving mechanism with periodic traffic indications in the IEEE 802.16e/m. IEEE Trans. Veh. Technol59(1): 319–334. doi: 10.1109/TVT.2009.2032193
21. Wu Y, Min G, Li K, Javadi B (2012) Modeling and analysis of communication networks in multicluster systems under spatio-temporal bursty traffic. IEEE Trans. Parallel Distrib. Syst23(5): 902–912. doi: 10.1109/TPDS.2011.198
22. Azad AP, Alouf S, Altman E, Borkar V, Paschos GS (2011) Optimal control of sleep periods for wireless terminals. IEEE J. Sel. Areas Commun29(8): 1605–1617. doi: 10.1109/JSAC.2011.110910
23. Sanghvi K, Jain PK, Lele A, Das D (2008) Adaptive waiting time threshold estimation algorithm for power saving in sleep mode of IEEE 802.16e. Proc. COMSWARE: 334–340, Bangalore, India. doi: 10.1109/COMSWA.2008.4554436

24. Guan T, He Y, Duan L, Yang J, Gao J, et al. (2014) Efficient bag-of-features generation and compression for on-device mobile visual location recognition. IEEE Multimedia 21(2): 32–41. doi: 10.1109/MMUL.2013.31

25. Ji R, Gao Y, Zhong B, Yao H, Tian Q (2011) Mining flickr landmarks by modeling reconstruction sparsity. ACM Trans. Multimedia Computing, Communications, and Applications 7S(1): 1–22. doi: 10.1145/2037676.2037688

26. Guan T, He Y, Gao J, Yang J, Yu J (2013) On-device mobile visual location recognition by integrating vision and inertial sensors. IEEE Trans. Multimedia 15(7): 1688–1699. doi: 10.1109/TMM.2013.2265674