Energy-Efficient Sleep Scheduling with QoS Considerations in 3GPP LTE-Advanced Networks

Jia-Ming Liang*, Jen-Jee Chen†, Hung-Hsin Cheng*, and Yu-Chee Tseng*
*Department of Computer Science
National Chiao-Tung University, Hsin-Chu 30010, Taiwan
†Department of Electrical Engineering
National University of Tainan, Tainan 70005, Taiwan
Email: {jmliang, hhcheng, yctseng}@cs.nctu.edu.tw, jjchen@mail.nutn.edu.tw

***Abstract*—With the design of data communications in mind, 3GPP LTE-Advanced is probably the most promising technology for next generation mobile communications. For mobile applications, continuous communications at the *user equipments (UEs)* over a long period of time, imposing stringent requirements on power saving. To manage power consumption, 3GPP LTE-Advanced has defined the *Discontinuous Reception (DRX)* mechanism to allow UEs to turn off their radio interfaces and go to sleep in various patterns. Existing literature has paid much attention to evaluate the performance of DRX; however, how to tune DRX parameters to optimize energy cost is still left open. This paper addresses the optimization problem of the DRX mechanism, by asking how to minimize the wake-up periods of the UEs while guarantee their QoS, especially on the aspects of traffic bit-rate, packet delay, and packet loss rate in mobile applications. Efficient schemes to optimize DRX parameters and schedule UEs' packets at the *evolved Node B (eNB)* are proposed. The key idea of these schemes is to analyze and balance the impacts between QoS parameters and DRX configurations. Simulation results show that our scheme can fully satisfy QoS requirements of the UEs while save considerable energy, compared to the existing schemes.**

***Index Terms*—3GPP LTE-A, DRX, power saving, quality of service, sleep scheduling.**

I. Introduction

The *3GPP Long Term Evolution-Advanced (LTE-A)* is the most promising technology for next generation mobile communications. To support various mobile applications, the LTE-A standard has defined multiple *quality-of-service (QoS)* classes for different traffic characteristics on the aspects of traffic bit-rate, tolerable packet delay, and packet loss rate [1]–[3]. On the other hand, since the *user equipments (UEs)* need to continuously communicate over a long period of time, they impose the most stringent requirements on power saving. To save power of UEs, the LTE-A standard has defined the *Discontinuous Reception (DRX)* mechanism to allow UEs to turn off their radio interfaces and go to sleep when the *evolved Node B (eNB)* has no data to delivery to the UEs. The key property of the DRX mechanism is to cooperate the eNB with UEs and make UEs awake periodically to detect the incoming data from the eNB. Especially, each UE adopts a specific "timer" to prolong its wake-up period. Thus, some data posing unexpected delay can be received by the UE without loss beyond the periodic wake-up period.

In the literature, the performance analyses of the DRX mechanism in the LTE-A networks are conducted in [4]–[6]. They show that enabling DRX can significantly save UEs' energy. For DRX configuring, the work [7] proposes a scheme to adaptively adjust DRX cycle length for UEs to improve energy efficiency. In [8], the authors consult UEs' *channel quality identifier (CQI)* and adjust the DRX timer to improve the system utility. However, both [7] and [8] are regardless of the QoS requirements in terms of traffic bit-rate, packet delay, and packet loss rate, which are the mandatory QoS features in the LTE-A networks. As described above, none of work has addressed the DRX optimization problem under the consideration of QoS.

In this paper, we address the DRX optimization problem, which considers UEs' QoS satisfaction and energy conservation. The objective of the problem is to minimize UEs' wake-up periods to save their energy while satisfy the QoS requirements in terms of traffic bit-rate, packet delay, and packet loss rate. We propose an efficient scheme and a packet scheduling method to solve the problem. The key idea of these schemes is to analyze and balance the impacts between QoS parameters and DRX configurations, and try to minimize the wake-up cost without violating UEs' QoS requirements. To the best of our knowledge, this is the first paper to address the DRX optimization problem which considers both energy conservation and QoS guarantee for the UEs in the LTE-A networks. Simulation is conducted to show that our scheme can fully guarantee UEs' QoS requirements while save considerable energy consumption.

The rest of this paper is organized as follows. Preliminaries are given in Section II. Section III presents our schemes. Simulation results are given in Section IV. Conclusions are drawn in Section V.

II. Preliminaries

A. QoS in LTE-A

The LTE-A network supports two types of flows: *guaranteed-bit-rate (GBR)* and *non-guaranteed-bit-rate (NGBR)*. The GBR flow is used for real-time services, such as conversational voice, video, and gaming. The NGBR flow is used for non-real-time services, such as IMS signaling, TCP-based services (e.g., http, ftp, e-mail, etc.). For each GBR flow, it associates with the QoS parameter in terms of *guaranteed-bit-rate (GBR)*, which is the minimum reserved traffic rate (bits/s) admitted by the eNB. For NGBR flows, they associate with a common QoS parameter: *aggregate-maximum-bit-rate (AMBR)*, which is the amount of traffic

978-1-4673-2480-9/13/$31.00

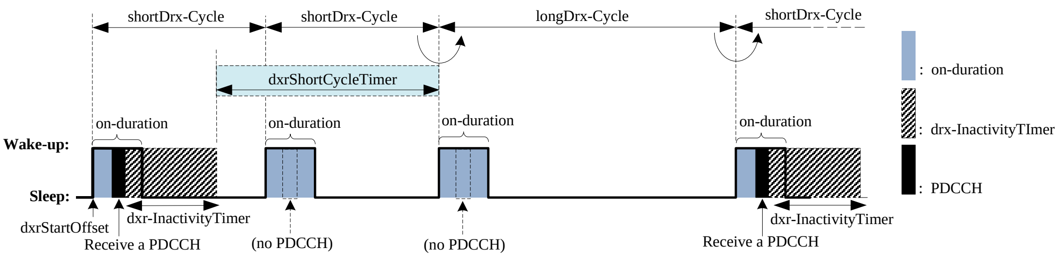

Fig. 1. An overview of the DRX operation.

rate (bits/s) shared by all NGBR flows in a UE. In addition, each flow also associates with a QoS profile, including: *packet delay budget* and *packet loss rate*. The packet delay budget is the maximum waiting time (in ms) that the packet is delivered from the eNB to the UE. The packet loss rate is the probability that a packet is delivered from the eNB but does not be received by the UE. This may happen when the buffered packet waits over its delay budget. Here, we also investigate the impact of *service-request-response time (SRS)* (in ms) [9] for NGBR flows in the LTE-A networks. The SRS is the maximum waiting time for the service request of the applications to be delivered from the UE to the eNB. Usually, the SRS time is lager than the packet delay budget.

B. DRX in LTE-A

The DRX mechanism is managed by the *Radio Resource Control (RRC)*. The DRX configurations are UE-specific, which are determined by the eNB. When DRX is enabled, the UE performs wake-up and sleep operations in a cycle periodically, as shown in Fig. 1. The basic duration of wake-up and sleep operations is a subframe, which is with the length of 1 ms. When the DRX mechanism is activated, there are six parameters to be specified for each UE. The parameters include 1) shortDRX-Cycle, 2) on-duration, 3) drxStartOffset, 4) drx-InactivityTimer, 5) longDRX-Cycle, and 6) drxShortCycleTimer. The shortDRX-Cycle and longDRX-Cycle are the basic operation periods (in subframes) that the UE performs wake-up and sleep operations. Usually, the length of longDRX-Cycle is multiple of the length of shortDRX-Cycle for advanced power conservation. The on-duration is a period (in subframes) in a cycle that the UE has to awake. During the wake-up period, the UE will monitor whether or not there is a *Physical Downlink Control Channel (PDCCH)* delivered from the eNB to indicate the downlink transmission for the UE. The drxStartOffset is the subframe where the first on-duration of the UE starts. The drx-InactivityTimer is used for extending the wake-up period of the UE. Specifically, the UE starts the drx-InactivityTimer and extends the wake-up period when the UE monitors the PDCCH delivered from the eNB. Before drx-InactivityTimer expires, if the UE monitors a new PDCCH from the eNB, the drx-InactivityTimer resets and restarts to count down again. Once the drx-InactivityTimer expires, the UE will start shortCycleTimer and go to sleep while turning off the radio interface to save energy. During the UE's sleep period, all the data for the UE will be buffered in the eNB until the next on-duration comes. If no PDCCH is monitored by the UE during several shortDRX-Cycles, the shortCycleTimer will expire. Once the shortCycleTimer expires, the shortDRX-Cycle ends and the longDRX-Cycle follows. During the longDRX-Cycle, the UE behaves similarly as it staying in the shortDRX-Cycle. Once the UE monitors the PDCCH, it terminates the longDRX-Cycle and starts the shortDRX-Cycle again.

Note that the DRX configurations, QoS requirements, and energy consumption of UEs are dependent with each other. For example, the UE with a larger shortDRX-Cycle, on-duration, and drx-InactivityTimer can have a higher traffic bit-rate, lower packet delay, and lower packet loss rate but also incur higher energy consumption due to longer wake-up periods. Contrarily, the UE with a smaller shortDRX-Cycle, on-duration, and drx-InactivityTimer will decrease the traffic bit-rate, increase the packet delay and packet loss rate but it can improve UEs' energy conservation. Therefore, how to optimize DRX parameters to ensure UEs' QoS while minimize their energy consumption is a difficult problem.

C. Problem Definition

We consider the downlink transmissions in the LTE-A network under Time Division Duplex (TDD) mode. The eNB serves N UEs in a point-to-multipoint manner, where each UE_i,$i = 1..N$, has admitted F_i^G GBR flows and F_i^N NGBR flows, and each GBR flow_j has a guaranteed-bit-rate R_j^G (bits/s) and all NGBR flows in UE_i share an aggregate-maximum-bit-rate R_i^N (bits/s). For each flow_j(including GBR and NGBR flows), it has the QoS profile in terms of *packet delay budget* D_j (ms) and allowable *packet loss rate* P_j^{loss}. The packet size of the flows may vary over time. We assume that the packet size ranges from Q_j^{min} and Q_j^{max} (bits/packet). The inter-arrival time of the packets of flow_j is Z_j ms. In addition, each NGBR flow_j is with a service request-response time S_j (ms) based on the mobile applications, which is larger than the packet delay budget, i.e., ($S_j \gg D_j$). Note that this paper assumes that the packets of a flow are created from a remote source in the Internet. Because the packet has to traverse through the Internet and the core network where the UE locates in before arriving the eNB, the packet may pose some unexpected delay [10]. We model this delay by the probability mass function $P_{i,j}(t)$, which is the delay probability density of the packet belong to flow_j of UE_i with t ms delay, $t = 1, 2, ...$ On the other hand, the wireless resource in the paper is represented as "frames". Each frame is further divided into several subframes. Note that the subframe duration is 1 ms. In each subframe, the basic allocation unit for a UE is a "*resource block (RB)*". Suppose that there are Ω RBs in a subframe. Note that the UE with a higher channel quality can receive more data bits in a RB. Let C_i (bits/RB) be UE_i's transmission rate which may vary over time and is measured during the wake-up period of UE_i. Without loss generality, we assume that C_i ranges from C_i^{min} to C_i^{max} (bits/RB). Our problem asks how to schedule the resource and optimize the DRX parameters for each UE$_i$, including on-duration (O_i), drxStartOffset (L_i), shortDRX-Cycle (T_i^S), longDRX-Cycle (T_i^L), drx-InactivityTimer (Γ_i^I), and drxShortCycleTimer (Γ_i^L) such that the QoS requirements of UE$_i$ (i.e., R_j^G, R_i^N, D_j, P_j^{loss}) can be met while the wake-up periods of all UEs can be minimized.

III. The Proposed Scheme

In this section, we present our scheme to determine DRX parameters (i.e., T_i^S,T_i^L,O_i, L_i, Γ_i^I, Γ_i^L) for each UE$_i$. Once the parameters are determined, they will be sent to UEs and make UEs behave accordingly. In addition, a packet scheduling is also proposed for the eNB to cooperate with UEs. Specifically, our scheme maintains three key properties to reduce wake-up periods of UEs. First, we make all UE$_i$'s DRX cycle be an integer multiple of others'. Thus, it can reduce UEs' unnecessary wake-up periods incurred by resource competition (this will be clear later on). Second, we also optimize the drx-InactivityTimer and help UEs to catch the packets posing unexpected delay to meet their delay budgets. Third, it tries to make UEs go to "deep" sleep without violating their service response time. As the results, the UEs can significantly reduce their wake-up periods and further save energy. The details of the scheme are described as follows.

A. Determination of DRX Cycle Lengths (T_i^S, T_i^L)

To decide T_i^S of each UE$_i, i = 1..N$, we first find the "strictest delay budget" of each UE$_i$ (denoted as D_i^{min}) defined by $D_i^{min} = \min_j\{D_j|flow_j \in UE_i\}$. Without loss of generality, let $D_1^{min} \leq D_2^{min} \leq .. \leq D_N^{min}$. Set $T_1^S = D_1^{min}/2$ and determine $T_i^S, i = 2..N$, as follows.

$$T_i^S = \left\lfloor \frac{D_i^{min}}{T_{i-1}^S} \right\rfloor \times T_{i-1}^S. \quad (1)$$

We can see that Eq. (1) implies $T_i^S \leq D_j$ for all flow$_j$ in UE$_i$ because $T_i^S \leq D_i^{min} \leq D_j$. Since the flows' arriving packets can always be served in UE$_i$'s cycle T_i^S, this guarantees the packets to meet their delay budgets. We also note that Eq. (1) ensures T_i^S to be an integer multiple of T_{i-1}^S for $i = 2..N$. This can help UEs to interleave their wake-up periods and avoid resource competition that may increase the unnecessary wake-up periods. We also note that T_1^S is the basic cycle and the allocation results will repeat after T_N^S/T_1^S cycles due to the cyclic properties (this will be clear later on).

To decide T_i^L of each UE$_i, i = 1..N$, we find the "strictest service request-response time" from the NGBR flows in UE$_i$ (denoted as S_i^{min}), which is defined by $S_i^{min} = \min\{S_j|flow_j \in UE_i\}$. Since the size of longDRX-Cycle T_i^L of UE$_i$ should be an integer multiple of the size of its shortDRX-Cycle T_i^S while T_i^L has to be less than or equal to S_i^{min}, we set $T_i^L, i = 1..N$, as follows.

$$T_i^L = \left\lfloor \frac{S_i^{min}}{T_i^S} \right\rfloor \times T_i^S. \quad (2)$$

Note that Eq. (2) implies $T_i^L \leq S_i^{min} \leq S_j$ for all NGBR flow$_j$ in UE$_i$. Therefore, once a service request arrives in the long cycle T_i^L, it can guarantee the request to be served within $T_i^L \leq S_j$. Therefore, the service response time of all NGBR flows in UE$_i$ can be met.

B. Determination of DRX On-duration and Timers ($O_i, \Gamma_i^I, \Gamma_i^L$)

To determine the on-duration O_i of each UE$_i, i = 1..N$, we first calculate the sum of the maximum packet sizes of the flows in UE$_i$, whose delay budget is equal to its shortDRX-Cycle length T_i^S, i.e., $\sum_{D_j=T_i^S,\forall flow_j \in UE_i} Q_j^{max}$. Then, O_i is set as follows:

$$O_i = \max\left\{ \left\lceil \frac{\sum_{D_j=T_i^S,\forall flow_j \in UE_i} Q_j^{max}}{C_i^{min} \times \Omega} \right\rceil, 1 \right\}. \quad (3)$$

We can see that by reserving O_i subframes as UE$_i$'s on-duration, the most "urgent" packet of flow$_j$ can be served during the short DRX cycle. Here, the most urgent packet indicates the packet with the delay budget equal to T_i^S and it arrives at the beginning of the shortDRX-Cycle. So, such packet needs to be received by UE$_i$ before the cycle ends or it will be dropped. Note that Eq. (3) also implies that UE$_i$ uses the least number of necessary wake-up subframes by reserving necessary resource for urgent packets only, which can reduce the periodic wake-up periods of UEs.

For determining the drx-InactivityTimer Γ_i^I of each UE$_i, i = 1..N$, we first model the *expected packet loss rate*, denoted by $\mathbb{E}_{i,j}(\cdot)$, for flow$_j$ in UE$_i$ by making use of its packet delay probability $P_{i,j}(t)$. Then, a temporal InactivityTimer for each flow$_j$ is chosen to satisfy the flow's packet loss rate. Finally, the best drx-InactivityTimer is determined for UE$_i$ to meet all its flows' packet loss rate. The detail of the procedure is described as follows.

- Let M_j be the number of packets of flow$_j$ that should arrive during D_j ms (thus, $M_j = \lfloor \frac{D_j}{Z_j} \rfloor$). Each packet $m, m = 1..M_j$, may pose delay t_m ms, $t_m = 1,..,D_j$[1] because of traversing through Internet and the core network. Let $\hat{t}_m$ be the subframe number[2] that packet m arrives at the eNB, which can be presented as follows: $\hat{t}_m = t_m + (m-1) \times Z_j + T_{offset}$, where T_{offset} is the expected arrival subframe number of the first packet of flow$_j$ after UE$_i$'s first on-duration ends and Z_j is the expected packet inter-arrival time of flow$_j$. Then, the function of expected packet loss rate $\mathbb{E}_{i,j}(\cdot)$ with the temporal InactivityTimer $\hat{\Gamma}_i^I$ can be expressed as follows.

$$\mathbb{E}_{i,j}(\hat{\Gamma}_j^I, [t_1, t_2, .., t_m, .., t_{M_j}], D_j, T_i^S) = \sum_{t_m=1..D_j,\forall m} Loss(\hat{\Gamma}_j^I, [t_1, t_2, .., t_m, .., t_{M_j}], D_j) \times Prob([t_1, t_2, .., t_m, .., t_{M_j}]),$$

where $Prob(\cdot)$ and $Loss(\cdot)$ are the probability and the packet loss ratio function of certain packet delay distribution $[t_1, t_2, .., t_m, .., t_{M_j}]$, i.e.,

$$Prob([t_1, t_2, .., t_m, .., t_{M_j}]) = \prod_{m=1..M_j} P_{i,j}(t_m), \quad (4)$$

and

$$Loss\left(\hat{\Gamma}_j^I, [t_1, t_2, .., t_m, .., t_{M_j}], D_j, T_i^S\right) = \frac{M_j - \left(\sum_{m=1..M_j}(\phi_m + \eta_m)\right)}{M_j}, \quad (5)$$

[1] We think that the packet is "loss" if it poses a delay over D_j ms.

[2] Without loss of generality, we redefine the first subframe as the subframe number 1 after UE$_i$'s first on-duration ends.

where

$$\phi_m = \begin{cases} 1, & \text{if } X_m \le D_j \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

and

$$\eta_m = \begin{cases} 1, & \text{if } X_m > D_j \text{ and } Y_m < \hat{\Gamma}^I_j \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

In Eq. (5), the denominator is the total number of packet arrivals during the delay budget and the numerator is the number of packets that fail to be received by UE$_i$ due to the expiration of InactivityTimer $\hat{\Gamma}^I_j$. In addition, ϕ_m (in Eq. (5) and Eq. (6)) is an indicator that returns 1 if the arrival of packet m can be received by UE$_i$'s on-duration of the cycle; otherwise, returns 0. Term η_m (in Eq. (5) and Eq. (7)) is also an indicator that returns 1 if the arrival of packet m can be received by the InactivityTimer; otherwise, returns 0. Note that X_m (in Eq. (6) and Eq. (7)) is used to evaluate the waiting time if the arrival of packet m is received until the next on-duration of UE$_i$, i.e., $X_m = \left\lceil \frac{\hat{t}_m}{T^S_i} \right\rceil \times T^S_i - T_{offset} - (m-1) \times Z_j$. Y_m in Eq. (7) is interval between the arrival of packet m and packet y, where packet y satisfies $\hat{t}_y \le \hat{t}_m$ and is successfully received by the on-duration or InactivityTimer of UE$_i$, i.e.,

$$Y_m = \begin{cases} \hat{t}_m - (\lceil \frac{\hat{t}_y}{T^S_i} \rceil \times T^S_i), & \text{if } \phi_y = 1 \\ \hat{t}_m - \hat{t}_y, & \text{if } \eta_y = 1. \end{cases} \quad (8)$$

Thus, it implies that once packet m arrives at the subframe near by that of packet y received successfully by the on-duration or the InactivityTimer, the packet can also be received by UE$_i$ through the extended InactivityTimer triggered by packet y.

- Then, we choose a temporal InactivityTimer $\hat{\Gamma}^{I*}_j$ to meet the required packet loss rate of flow$_j$, i.e., $\hat{\Gamma}^{I*}_j = \min \left\{ \hat{\Gamma}^I_j | \mathbb{E}_{i,j}(\hat{\Gamma}^I_j) \le P^{loss}_{i,j}), \hat{\Gamma}^I_j = 0, 1.. \right\}$.
  Note that a shorter InactivityTimer $\hat{\Gamma}^{I*}_j$ can potentially reduce the wake-up period of the UE when the required packet loss rates are the same.
- Finally, the best InactivityTimer Γ^{I*}_i is chosen for UE$_i$, which satisfies the packet loss rate of all the flows in UE$_i$, i.e.,$\Gamma^{I*}_i = \max \left\{ \hat{\Gamma}^{I*}_j | flow_j \in UE_i \right\}$.

For determining Γ^L_i of each UE$_i, i = 1..N$, we set $\Gamma^L_i = \lceil \frac{S^{max}_i}{T^S_i} \rceil \times T^S_i - (f \% T^S_i)$, where f is the period (in subframes) from the start subframe of the recent shortDRX-Cycle to the subframe that the drxShortCycleTimer is triggered. Here, based on the behavior of mobile applications, if no packet arrives over the maximal service request-time, i.e., $S^{max}_i = \max\{S_j | flow_j \in UE_i\}$, it has a higher probability that there will be no incoming packet arriving. This feature can be used for the UE to go to deep sleep for further energy conservation.

C. Determination of DRX Start-offset (L_i)

We decide $L_i, i = 1..N$ in three steps:

- First, we define the 'crowded' degree for $1..\frac{T^S_N}{T^S_1}$ cycles, denoted as $\mathbb{C}_k, k = 1..\frac{T^S_N}{T^S_1}$, i.e., $\mathbb{C}_k = \sum_{i=1..N} \{T^S_i | \text{if UE}_i\text{'s } L_i \text{ or } O_i \text{ is allocated in cycle } k\}$.

TABLE I
TRAFFIC ADOPTED IN THE SIMULATION [13].

Flow Type	Applications	Traffic Bit-Rate	Packet Delay Budget	Packet Loss Rate
GBR	Voice (G.711)	64 Kbps	100 ms	10^{-2}
GBR	Video (H.264)	33 Mbps	150 ms	10^{-2}
NGBR	HTTP/FTP	169 Kbps	300 ms	10^{-6}

- Second, we recursively choose the cycle k with the smaller $\mathbb{C}_k$ for UE$_i$ in an ascending order of i. Then, we allocate L_i and O_i to the cycle k and the corresponding cycles, respectively, and update $\mathbb{C}_k$.
- Third, we proportionally distribute UEs' L_i based on the length of their $(O_i + \Gamma^I_i)$ in each cycle to prevent the resource competition causing by UEs' extending InactivityTimer.

D. Packet Scheduling in the eNB

For each UE$_i$, we design two "virtual" buffers in the eNB to collect its packets created before and after its on-duration of the short cycle. Then, the eNB only allocates the buffered data which is created before the UE's on-duration in the short cycle because those packets suffer a longer delay. Let $G_{i,j}$ and A_i be the maximum size of data to be received per short cycle for GBR flow$_j$ and all non-GBR flows of UE$_i$, respectively. We set $G_{i,j} = T^S_i \cdot R^G_j$ and $A_i = T^S_i \cdot R^N_i$. Thus, the eNB keeps the allocated GBR and non-GBR data no larger than $G_{i,j}$ and A_i, respectively, when allocating UE$_i$'s data. Once the UE's allocated data reaches the size of $G_{i,j}$ and A_i, the eNB will not deliver data to the UE during its short cycle and thus the UE will go to sleep and save energy. Note that when allocating data, the eNB allocates the "urgent data" to the UEs first. The urgent data means the data that will be dropped at the current subframe if the UE does not received the packet (due to the delay budget). Then, the remaining buffered data will be allocated to the UEs in a proportional manner.

IV. PERFORMANCE EVALUATION

In the section, we develop a simulator in JAVA language to verify the effectiveness of the proposed scheme. The system parameters are listed below. The frame duration is 10 ms. The channel bandwidth is 10 MHz. Thus, we have $\Omega = 100$ RBs in each subframe. The GBR and NGBR flows adopted in the simulation are shown in Table I. The packet delay probability is modeled by the normal distribution with mean 0 and standard deviation $\sigma = 10$ ms [11]. The channel quality for each UE is generated according to [12]. The number of UEs ranges from 10 to 70.

We compare our scheme against the *Counter-Driven DRX (CDD)* scheme [7] and the *Multiple-Threshold DRX (MTD)* scheme [8]. The **CDD** scheme can adjust the length of the DRX cycle for each UE based on the predefined counters and thresholds. The **MTD** scheme designs multiple lengths and thresholds for drx-InactivityTimer and transmission rate to adjust UEs' drx-InactivityTimers. Note that the duration of each experiment result is at least 6000 subframes.

A. Packet Loss Rate

We first investigate the effects of number of UEs on packet loss rate of all schemes. In Fig. 2(a), we can see that the

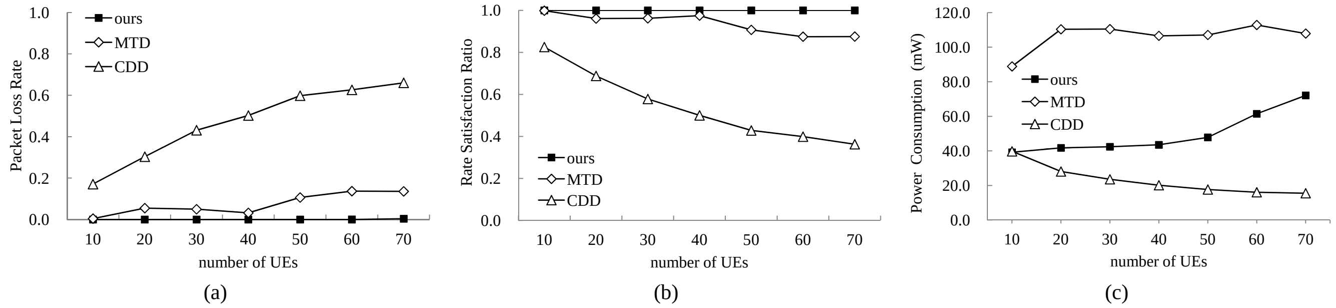

Fig. 2. Comparison on the packet loss rate, rate satisfaction, and power consumption of all schemes under $10 \sim 70$ UEs.

MTD scheme and the **CDD** scheme incur higher packet loss rates because they neglect to optimize the drx-InactivityTimer according to UEs' QoS parameters. Contrarily, our scheme has the minimal packet loss rate (i.e., $\leq 3 \times 10^{-3}$) even if the network is under saturated (i.e., 70 UEs). This is because our scheme can optimize the drx-InactivityTimer for each UE to prolong wake-up period and catch more packets approaching delay budgets.

B. Rate Satisfaction Ratio

Next, we investigate the rate satisfaction ratio of GBR and NGBR flows of UEs, which is defined by the amount of *satisfied* bit-rate to the total amount of admitted bit-rate of UEs. When the satisfaction ratio is 1, it means that the scheme can successfully satisfy the required bit-rate of flows in UEs. Fig. 2(b) show the rate satisfaction ratio of all GBR and NGBR flows for all schemes under different number of UEs. As can be seen, when the network is getting saturated, our scheme can still have the highest satisfaction ratio (i.e., "1.0"). This is because our scheme can well determine the UEs' on-duration, shortDRX-cycle, and drx-InactivityTimer to reduce the packet loss rate and potentially increase the rate satisfaction of all flows.

C. Power Consumption

Finally, we investigate the effects of numbers of UEs on power consumption. Note that the power consumption of the UEs is modeled according to [14]. As shown in Fig. 2(c), the **MTD** scheme has higher power consumption because it enforces to increase the drx-InactivityTimer of UEs with a lower transmission rate even if the network is under non-saturated. This hurts the performance on power saving. On the contrary, the **CDD** scheme has a lowest power consumption but it neglects the QoS satisfaction of UEs. We note that our scheme can not only guarantee the QoS of UEs (as shown in Fig. 2(a) and Fig. 2(b)) but also incur less energy consumption (as shown in Fig. 2(c)). This is because our scheme can optimize DRX configurations by well balancing between UEs' energy consumption and QoS requirements.

V. Conclusions

In this paper, we have addressed the DRX optimization problem with the consideration of both QoS requirements and energy conservation of the UEs. We proposed an energy-efficient scheme and a packet scheduling method to solve the problem. By balancing the impacts between QoS parameters and DRX configurations, simulation results have verified that our scheme can fully guarantee the UEs' QoS requirements while save considerable energy of the UEs.

Acknowledgment

This research is co-sponsored by MoE ATU Plan, by NSC grants 100-2218-E-024-001-MY3, 98-2219-E-009-019, 98-2219-E-009-005, and 99-2218-E-009-005, by AS-102-TP-A06 of Academia Sinica, by ITRI, Taiwan, and by D-Link. This work was also conducted under the "AMI Enhancement Project" of III, which is subsidized by MoEA, ROC.

References

[1] 3GPP TS 23.401, "General Packet Radio Service (GPRS) enhancements for E-UTRAN access," v11.1.0 (Release 11), March 2012.

[2] 3GPP TS 36.213, "Policy and charging control architecture," v11.5.0 (Release 11), March 2012.

[3] 3GPP TS 36.413, "E-UTRAN; S1 Application Protocol (S1AP)," v10.3.0 (Release 10), March 2012.

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

[5] K. Aho, T. Henttonen, J. Puttonen, L. Dalsgaard, and T. Ristaniemi, "User equipment energy efficiency versus LTE network performance," *International Journal on Advances in Telecommunications*, vol. 3, no. 4, pp. 140–151, 2010.

[6] S. Jin and D. Qiao, "Numerical analysis of the power saving in 3GPP LTE Advanced wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1779–1785, 2012.

[7] E. Liu, J. Zhang, and W. Ren, "Adaptive DRX scheme for beyond 3G mobile handsets," *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–5, 2011.

[8] S. Gao, H. Tian, J. Zhu, and L. Chen, "A more power-efficient adaptive discontinuous reception mechanism in LTE," *IEEE Vehicular Technology Conference (VTC)*, pp. 1–5, 2011.

[9] A. Michlmayr, F. Rosenberg, P. Leitner, and S. Dustdar, "End-to-end support for QoS-aware service selection, binding, and mediation in VRESCo," *IEEE Transactions on Services Computing*, vol. 3, no. 3, pp. 193–205, 2010.

[10] A. Pathak, H. Pucha, Y. Zhang, Y. C. Hu, and Z. M. Mao, "A measurement study of Internet delay asymmetry," *ACM International Conference on Passive and Active Network Measurement*, pp. 182–191, 2008.

[11] M. J. Karam and F. A. Tobagi, "Analysis of the delay and jitter of voice traffic over the internet," *IEEE INFOCOM*, pp. 824–833, 2001.

[12] 3GPP TS 36.213, "E-UTRA; Physical layer procedures," v10.6.0 (Release 10), June 2012.

[13] R. Mehmooda, R. Alturkia, and S. Zeadally, "Multimedia applications over metropolitan area networks (MANs)," *Journal of Network and Computer Applications*, vol. 34, no. 5, pp. 1518–1529, 2010.

[14] 3GPP TSG RAN WG2 meeting 57bis, R2-071285, "DRX parameters in LTE," March 2007.