IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 27, 2020, accepted May 29, 2020, date of publication June 3, 2020, date of current version June 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2999602

Power Efficiency and Delay Tradeoff of 100G Energy Efficient Ethernet Protocol

XIAODAN PAN[1], TONG YE[1], (Member, IEEE), AND TONY T. LEE[2]

[1]State Key Laboratory of Advanced Optical Communication System and Networks, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

[2]School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China

Corresponding author: Tong Ye (yetong@sjtu.edu.cn)

This work was supported by the National Nature Science Fund of China under Grant 61571288, Grant 61671286, and Grant 61433009.

ABSTRACT This paper investigates the dual-mode power-saving strategy designed for the 100G Energy Efficient Ethernet. The process of this strategy is a sequence of cycles, where each cycle is an interval elapsed between two consecutive instants when the buffer of the Ethernet interface becomes empty. In each cycle, the interface first enters the fast-wake mode to perform the conditional sleep-mode selection of the strategy. If the number of arrivals during the fast-wake mode reaches a threshold, the interface directly wakes up; otherwise, it proceeds to the deep-sleep mode. The sequence of cycles switches between two types: deep-sleep cycles with deep-sleep mode and light-sleep cycles without deep-sleep mode. We analyze the dual-mode strategy based on the condition of the number of arrivals during the fast-wake mode. We first derive the weights of conditions according to the feature of the conditional sleep-mode operation, and then calculate the unconditioned performance measure of the dual-mode strategy based on the weighted average of that of these two kinds of cycles. Finally, we obtain the close-form expressions of the power efficiency and the mean delay, based on which we provide a set of parameter selection rules. We show that the dual-mode strategy with these rules can select suitable sleep modes according to the instantaneous traffic rate, and thus perform well under bursty input traffic.

INDEX TERMS 100G energy efficient Ethernet, dual-mode strategy, performance tradeoff, threshold selections.

I. INTRODUCTION

In recent years, the demands for high-speed Ethernet have been driven by ever-increasing Internet use and emerging bandwidth-hungry applications, such as 4K/8K video [1] and high-performance computing [2]. 100G Ethernet is gradually displacing the 10G Ethernet and will become mainstream very soon [3]. However, with high data speed, energy consumption per Ethernet device exponentially increases [4]. Meanwhile, the link utilization of typical Ethernet interfaces is only about 5%-30% [4]–[6]. Therefore, a lot of energy is wasted if the Ethernet interfaces run at full power all the time.

In this context, IEEE 802.3az [7], the first Energy Efficient Ethernet (EEE) standard, was published in 2010. This standard defines a Low Power Idle (LPI) mode for 100M/1G/10G Ethernet interfaces. When the buffer is emptied, the Ethernet interface turns off most of its components, and enters the LPI mode through a Sleep operation. In the LPI mode, the interface consumes only $\sim$10% of the full power. The interface uses two thresholds, counter N and timer τ, to control the duration of the LPI mode. Once the number of arrived frames accumulated in the buffer reaches N or the waiting time of the first frame exceeds τ, the interface terminates the LPI mode via a Wakeup operation and starts to transmit frames. This sleep strategy is referred to as single-mode strategy in this paper.

However, the single-mode strategy cannot be directly applied to 100G Ethernet interfaces because the duration of the Wakeup operation for the LPI mode is excessively long [8], [9]. It would take a 100G interface 5.5μs to wake up from the LPI mode [10], [11]. In addition, the data rate of a 100G link is ten times faster than that of a 10G link. As a result, up to 45 frames of 1500 bytes can be accumulated in the transmission queue of a 100G interface during the Wakeup operation, which may increase the queue length in the buffer remarkably. As [11] points out,

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed Mohamed Ahmed Almradi.

this may require a large buffer to accommodate incoming frames.

Therefore, the IEEE 802.3bj standard published in 2014 [10] introduces another power-saving mode, called fast-wake (FW) mode, to 100G Ethernet, and refers to the LPI mode as the deep-sleep (DS) mode. In the FW mode, the interface powers off only a few components, and thus consumes up to 70% of the full power [8]. Even so, it only requires a Wakeup time as short as 0.34μs, which can avoid the sharp increase of the queue length. To distinguish these two kinds of Wakeup, we refer to the Wakeup of the FW mode as FW Wakeup, and that of the DS mode as DS Wakeup. The simulation results of [11]–[14] showed that the combination of the DS mode and the FW mode can help the 100G EEE interface to save more power than the DS mode only.

To use the complementary advantages of the FW mode and the DS mode, a dual-mode strategy was proposed for 100G EEE [11], [15] recently. The dual-mode strategy introduces two additional FW thresholds, FW counter N_f and FW timer T_{FW}. The 100G interface with the dual-mode strategy first enters the FW mode when the buffer becomes empty. The FW mode at most lasts for a duration of T_{FW}. If the traffic rate is high such that the number of arrivals reaches N_f before the FW mode ends, the interface directly wakes up. In this way, the interface can avoid a sharp increase of the queue length. On the contrary, if the traffic rate is low such that fewer than N_f frames arrive before the FW mode terminates, the interface enters the DS mode to reduce the energy consumption. Similar to that of the LPI mode in the single-mode strategy, the duration of the DS mode is controlled by thresholds N and τ, which are called the DS counter threshold and DS timer threshold, respectively.

The dual-mode strategy is essentially a kind of conditional sleeping strategy. Given the parameters T_{FW} and N_f, the usage of the DS mode is conditional on the number of arrivals during the FW mode, i.e., controlled by the traffic rate. When the traffic rate is low, the number of frames that arrive before the FW mode ends will be smaller than N_f with high probability, and the system will tend to enter the DS mode; otherwise, the system will almost not enter the DS mode. Therefore, the FW mode with thresholds T_{FW} and N_f is actually a kind of conditional sleep-mode operation, which can help the system to adaptively select the sleep mode according to the traffic rate. As a comparison, the single-mode strategy is a special case of the dual-mode strategy, since it lets the interface enter the DS mode with probability 1 once the buffer becomes empty regardless the traffic condition.

Clearly, FW thresholds T_{FW} and N_f play a key role in carrying out the conditional sleep-mode operation. If they are not properly selected, the dual-mode strategy may lead to a hostile queue length or a poor power efficiency, which is defined as the percentage of energy that the strategy can save. For instance, if N_f is small while T_{FW} is large, the system will almost not enter the DS mode even if the traffic rate is low, thereby resulting in low power efficiency. Thus, it is necessary to find a proper selection rule for these parameters, based on the understanding of system performance.

However, the feature of the dual-mode strategy imposes a big challenge on the performance analysis. Similar to the single-mode strategy considered in [16], the dual-model strategy can be modelled as a queueing system with vacation times governed by the frame arrival process and two sets of thresholds, the FW thresholds T_{FW} and N_f, and the DS thresholds τ and N. In addition, the dual-mode strategy has a notable feature that it can use the FW thresholds to determine whether to enter the DS mode at the end of the FW mode according to the instantaneous traffic rate. This conditional sleep-mode operation introduces a complicated dependency between the number of arrivals during the vacation time and that during the FW mode. As a result, it is hard to derive the distribution of the number of arrivals during the vacation time, which however is the key to analyze the queueing system with vacation times governed by the arrival process [16]. Thus, previous analysis techniques for the single-mode strategy, such as the model in [16], cannot be directly applied to the dual-mode strategy.

A. RELATED WORKS

Most of the previous works focused on the analysis of single-mode strategies [16]–[23]. The models in Ref. [17]–[23] failed to obtain a unified closed-form formula for the mean delay in the general case. References [17], [18] investigated the simplest cases, where either counter threshold N [17] or timer threshold τ was used [18] to control the duration of the LPI mode. References [19]–[22] studied the general strategy where these two thresholds were employed at the same time. To simplify the analysis, these works analyzed the system respectively in the low-traffic rate region and the high-traffic rate region, and thus can only obtain the piecewise expressions of the mean delay and the power efficiency. To derive the unified expression over the entire region of the traffic rate, Ref. [23] developed an analytical model, starting with the derivation of vacation time distribution. However, the model in [23] is too complex to characterize the behavior of the single-mode strategy. Reference [16] pointed out the failure was incurred by the fact that the single-mode strategy is a kind of queueing system with vacation times controlled by the frame arrival process. Starting with the distribution of the number of arrivals during the vacation time, Ref. [16] derived a generalized P-K formula for the mean delay. Even so, we show in Appendix A that, if the model of [16] is directly applied to the dual-mode strategy, it will become complicated and intractable. This is because the model in [16] was not designed to deal with the complicated dependency between the number of arrivals during the vacation time and that during the FW mode.

Currently, a few works [15], [24], [25] applied the analytical models developed for single-mode strategies to study the dual-mode strategy. Based on the model in [26] and [27], Ref. [24] and [25] analyzed the dual-mode strategy with $N_f = N = 1$, where the interface wakes up immediately as soon as a frame arrives. However, such a strategy has quite low power efficiency and is not suitable for practical applications [11], [12], since the vacation time is short and a large amount of energy is wasted by Wakeup operations. Reference [15] extended the model presented in [21], [22] to study the case where $N > N_f > 1$, which can only derive the power efficiency. Similarly, Ref. [15], [24], [25] did not notice the feature of the dual-mode strategy, and thus cannot find the closed-form mean delay for the dual-mode strategy in the general case. This indicates that simply extending the models of the single-mode strategy for the dual-mode strategy is not feasible. As a result, Ref. [24] specifically mentioned that it was quite difficult to model the dual-mode strategy.

B. OUR WORK AND CONTRIBUTIONS

In this paper, we develop a new approach to study the dual-mode strategy. Our goal is to derive the closed-form solutions of the mean delay and the power efficiency, such that we can propose proper parameter selection rules for the dual-mode strategy.

We consider the dual-mode strategy as a regenerative process, in which the cycle is defined as the interval between two consecutive instants when the buffer becomes empty. We show that under the conditional sleep-mode operation governed by thresholds T_{FW} and N_f, the cycle switches between two types: 1) DS cycle that contains the DS mode; 2) LS (light-sleep) cycle that does not. In particular, such an operation determines the probabilities of the DS cycle and the LS cycle, according to the traffic rate. When the traffic rate is low, the DS cycle occurs with high probability. When the traffic rate is high, the LS cycle appears with high probability.

Motivated by this observation, we propose an analytical model based on the concept of conditional sleep-mode, which treats the performance of the dual-mode strategy as the weighted average of that of the DS cycle and the LS cycle. We first derive the weights, which describe the function of the conditional sleep-mode operation governed by thresholds T_{FW} and N_f. We then analyze the conditional expectation of the system performance given that the DS cycle or the LS cycle occurs. Given that the system is almost in the DS cycle when the traffic rate is low, or the LS cycle when the traffic rate is high, the derivation of conditional performance measures can substantially be simplified and then be handled by the technique in [16]. We finally obtain the closed-form formulas for the mean delay and the power efficiency, which are accurate enough to analyze the dual-mode strategy.

Based on these results, we propose four threshold selection rules for the dual-mode strategy, such that it can achieve a high power efficiency under a given delay requirement. We further show how these rules can be applied in practice. Our simulation results demonstrate that our rules are conservative for the dual-mode strategy under the bursty input traffic. This implies that our rules can provide worst-case performance guarantee though the burstiness of Ethernet traffic could change over time.

In summary, our contributions are listed as follows:

- We successfully model the conditional sleep-mode operation of the dual-mode strategy.
- We derive the close-form formula for the mean delay, mean queue length and power efficiency, which are quite accurate for the dual-mode strategy.
- We propose a set of threshold selection rules for the dual-mode strategy, which can help the interface to perform well under bursty input traffic.

We organize the rest of this paper as follows. In Section II, we introduce the process of the dual-mode strategy, and show that the major function of thresholds T_{FW} and N_f is to perform the conditional sleep-mode operation. In Section III, we propose an analytical model to derive the mean delay, mean queue length and power efficiency. In section IV, we explore the impact of thresholds T_{FW}, N_f, τ, and N on performance tradeoffs, based on which we provide four threshold selection rules. We further evaluate the performance of the dual-mode strategy with the proposed rules under bursty input traffic. Section V concludes this paper.

Besides, for ease of reading, the main notations used in this paper are listed as follows:

- Parameters of the 100G EEE protocol

T_{BtF}	Duration of the transition period busy-to-FW
T_{FtD}	Duration of the transition period FW-to-DS
T_{FtB}	Duration of the FW Wakeup
T_{DtB}	Duration of the DS Wakeup
T_{FW}	FW timer threshold
N_f	FW counter threshold
N	DS counter threshold
τ	DS timer threshold
φ_h	Power consumption in the busy state, transition, and Wakeup periods
φ_f	Power consumption in the FW mode, $\varphi_f = 0.7\varphi_h$
φ_d	Power consumption in the DS mode, $\varphi_d = 0.1\varphi_h$

- System parameters

λ	Frame arrival rate
μ	Frame service rate
$\overline{X}(\overline{X^2})$	first moment (second moment) of frame transmission time
η	power efficiency
D	Mean frame delay
L	Mean queue length

- Variables in the model

$p(q)$	Probability that the interface enters the DS cycle (LS cycle)
$\widehat{V}_d(\widehat{V}_f)$	Mean vacation time of DS single-mode strategy (FW single-mode strategy)
$\widehat{C}_d(\widehat{C}_f)$	Mean cycle time of DS single-mode strategy (FW single-mode strategy)
$\widehat{\eta}_d(\widehat{\eta}_f)$	Power efficiency of DS single-mode strategy (FW single-mode strategy)
$\widehat{D}_d(\widehat{D}_f)$	Mean delay of DS single-mode strategy (FW single-mode strategy)

Other symbols are defined when required.

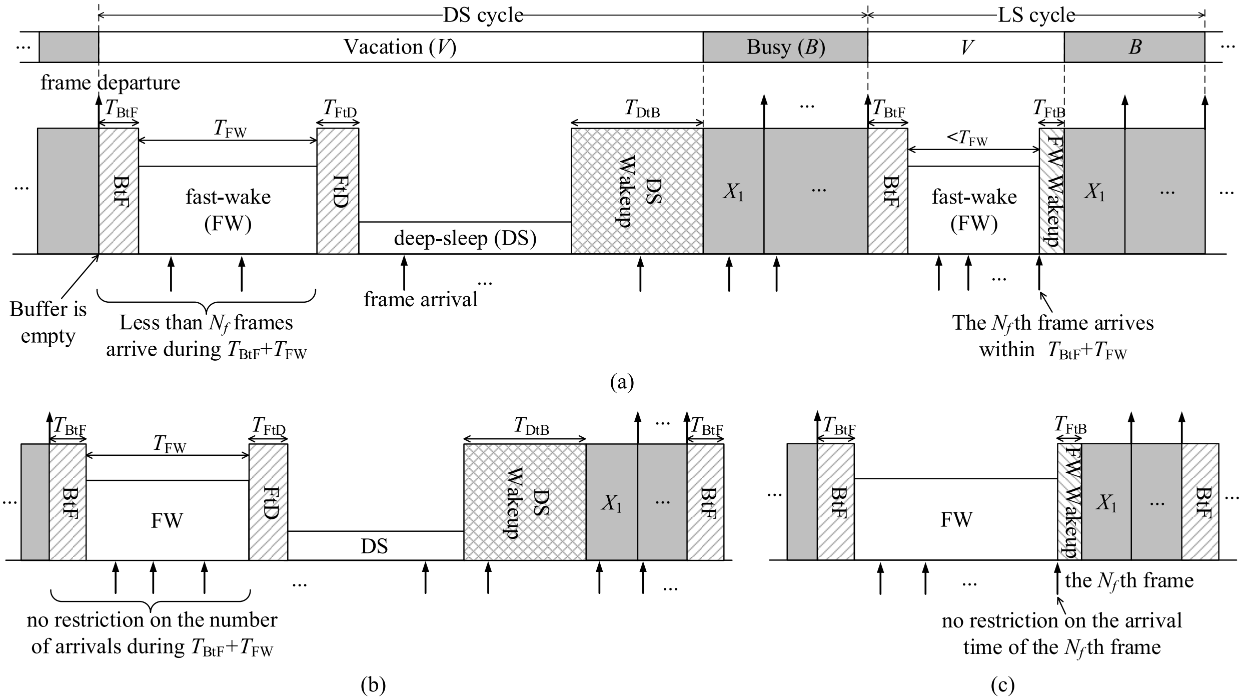


FIGURE 1. **(a) Dual-mode power-saving strategy and two special cases: (b) DS single-mode, and (c) FW single-mode.**

II. PRELIMINARIES

A. DUAL-MODE POWER-SAVING STRATEGY

Fig. 1(a) illustrates the process of the dual-mode strategy. Each time the buffer of the interface becomes empty, a vacation period, denoted by V, begins. The interface first enters the FW mode via a 0.9-μs transition period, named busy-to-FW (BtF) and denoted by T_{BtF}, during which a few components are powered off. In the FW mode, the power consumption is $\varphi_f = 0.7\varphi_h$, where φ_h is the power consumption when the interface is busy with frame transmission. The FW mode at most lasts for a duration of T_{FW}, during which the interface counts the number of arrivals. If less than N_f frames arrive before the FW mode ends, the interface proceeds to the DS mode after a 1.0-μs transition period [11], [15], called FW-to-DS (FtD) and denoted by T_{FtD}. In the DS mode, the interface turns off most of its components and thus its power consumption, denoted by φ_d, is only $0.1\varphi_h$. The length of the DS mode is regulated by two thresholds, DS counter N and DS timer τ. As soon as the queue length reaches N or the waiting time of the first frame reaches τ, the interface terminates the DS mode and wakes up via a 5.5-μs DS Wakeup. We denote the duration of DS Wakeup as T_{DtB}. On the contrary, if N_f frames arrive before the FW mode ends, the interface interrupts the FW mode immediately, and wakes up via a 0.34-μs FW Wakeup. We denote the duration of FW Wakeup as T_{FtB}. After the Wakeup, the interface starts a busy period, denoted by B, and transmits frames until the buffer becomes empty.

Clearly, the dual-mode strategy is a regenerative process. In this process, the system renews and starts a new cycle each time the buffer becomes empty. These cycles can be divided into two types. As Fig. 1(a) shows, a cycle is called an LS cycle if the interface wakes up directly from the FW mode, and called a DS cycle if the interface rouses up from the DS mode. The FW mode helps the interface to choose the type of the cycle, according to the number of arrivals before the FW mode ends. If the number reaches N_f, the interface wakes up directly from the FW mode (i.e., it selects the LS cycle); otherwise, it enters the DS mode at the end of the FW mode (i.e., it selects the DS cycle). From this viewpoint, the FW mode at the beginning of each cycle can be regarded as a sleep-mode selector, which determines the type of cycle to be adopted. With such an operation, the cycle switches back and forth between the DS cycle and the LS cycle, as Fig. 2(a) illustrates.

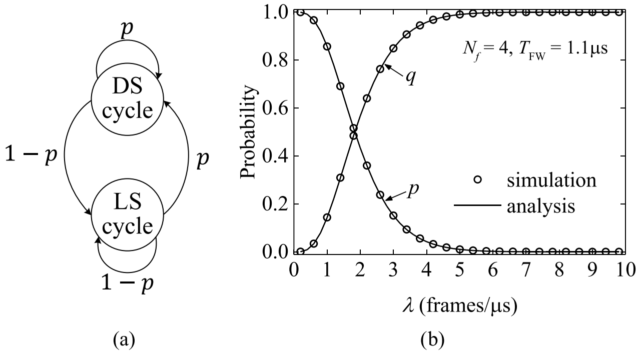


FIGURE 2. **Regenerative process of dual-mode strategy: (a) cycle-type transitions and (b) transition probabilities.**

Benefiting from the conditional sleep-mode operation, the best advantage of the dual-mode strategy is that it can adaptively select the DS cycle and the LS cycle, according to the traffic rate. In particular, such an operation with parameters T_{FW} and N_f determines the probabilities of the DS cycle and the LS cycle, when the traffic rate is given. For example, let's consider the case where the frame arrival process is a Poisson process with rate λ. The probability that the interface enters the DS cycle, denoted by p, is the probability that less than N_f frames arrive before the FW mode ends, which can be given as follows:

$$p = \sum_{k=0}^{N_f-1} e^{-\lambda(T_{\mathrm{BtF}}+T_{\mathrm{FW}})} \frac{[\lambda(T_{\mathrm{BtF}}+T_{\mathrm{FW}})]^k}{k!}. \tag{1}$$

Accordingly, the probability that the interface enters the LS cycle, denoted by q, is given by:

$$q = 1 - p. \tag{2}$$

Fig. 2(b) plots p and q changing with the traffic rate λ, where $N_f = 4$ and $T_{\mathrm{FW}} = 1.1\mu$s. When λ is small and the link is almost idle, the interface enters the DS cycle with probability $p \to 1$ such that it can save relatively more power. With the increase of λ, p decreases while q increases. When λ is large, say $\lambda > 4.0$ frames/μs, and the link is busy, the interface chooses the LS cycle with probability $q \to 1$, such that it can avoid an excessively long queue length.

Hence, the selection of T_{FW} and N_f should be done with care; otherwise, the FW mode cannot fulfill the function of the conditional sleep-mode operation, and may even have a negative impact on system performance. For example, if $N_f \to \infty$, the interface always enters the DS cycle irrespective of the number of arrivals during the FW mode. In this case, the dual-mode strategy reduces to the DS single-mode strategy in Fig. 1(b). The queue length increases rapidly when the traffic rate is high, due to the long DS Wakeup. On the other hand, if $T_{\mathrm{FW}} \to \infty$, the interface never enters the DS cycle, and the dual-mode strategy reduces to the FW single-mode strategy shown in Fig. 1(c). In this case, the power efficiency is poor when the traffic rate is low.

B. ISSUES IN SYSTEM MODELING

In this paper, our goal is to develop an accurate and tractable model to analyze the dual-mode strategy, which will provide the threshold selection rules to optimize system performance. In the modeling, we use the following assumptions:

A1. The frame arrival process is a Poisson process with arrival rate λ.

A2. Frames are transmitted in the first-in/first-out (FIFO) manner.

A3. The frame transmission times $X_1, X_2, \cdots$, are i.i.d. random variables with the first and second moments $\overline{X}$ and $\overline{X^2}$, respectively.

A4. $T_{\mathrm{BtF}} + T_{\mathrm{FW}} + T_{\mathrm{FtD}} < \tau$ and $N_f < N$.

Assumption A4 is justified since it ensures a relatively long duration of the DS mode such that the interface can obtain a high power efficiency when the input traffic rate is low.

Under these assumptions, the dual-mode strategy can be modeled as an M/G/1 queue with vacations governed by two sets of thresholds and the frame arrival process, which is like the 10G EEE studied in [16]. Starting with the derivation of the distribution of the number of arrivals during the vacation, Ref. [16] obtained the unified formulas of the power efficiency and the mean delay of 10G EEE over the entire offered load. However, different from 10G EEE, the vacation of the dual-mode strategy is a two-stage process, and the duration of each stage is controlled by a pair of thresholds, which is quite difficult for analysis. As Appendix A shows, the difficulty is mainly incurred by the fact that thresholds T_{FW} and N_f make the distribution of the number of arrivals during the whole vacation time depend on that of the number of arrivals during the FW mode. Therefore, though the method in [16] can be applied to analyze the dual-mode strategy, the derived formulas are too complex to provide clear physical interpretation.

III. ANALYSIS OF CONDITIONAL SLEEP-MODES

As mentioned before, the analysis of the dual-mode strategy is a significant challenge due to the impact of two sets of thresholds on the vacation process. In this section, we are going to circumvent such difficulty and derive closed-form formulas for the mean delay and the power efficiency, based on the characteristics of the dual-mode strategy.

Recall from Section II that the probabilities of the DS cycle and the LS cycle are controlled by the traffic arrival rate λ, once parameters N_f and T_{FW} are given. In other words, given λ, N_f and T_{FW}, it is possible to obtain the probability that a frame arrives at the DS cycle or the LS cycle, and the fraction of the time that the interface stays in the DS cycle or the LS cycle. In this case, if we can further derive the performances of the DS cycle and the LS cycle separately, we can solve the performance of the dual-mode strategy.

Motivated by this observation, we propose an analytical model based on the concept of conditional sleep-mode. Our idea is to treat the performance of the dual-mode strategy as the weighted average of that of the DS cycle and the LS cycle. We first derive the weights, which describe the function of the conditional sleep-mode operation governed by the two thresholds T_{FW} and N_f. We then analyze the system performance given that it is in the DS cycle or the LS cycle.

A brief profile of our approach is given as follows. Define an indicator

$$\xi = \begin{cases} 0, & \text{if a frame arrives within a DS cycle,} \\ 1, & \text{if a frame arrives within an LS cycle.} \end{cases}$$

and let d_i be the delay of the ith frame. The mean delay of the dual-mode strategy can be expressed as follows:

$$D = E[d_i|\xi = 0]\Pr\{\xi = 0\} + E[d_i|\xi = 1]\Pr\{\xi = 1\}, \tag{3}$$

where $\Pr\{\xi = 0\}$ and $\Pr\{\xi = 1\}$ are respectively the number fractions of the frames that arrive in DS cycles and LS cycles.

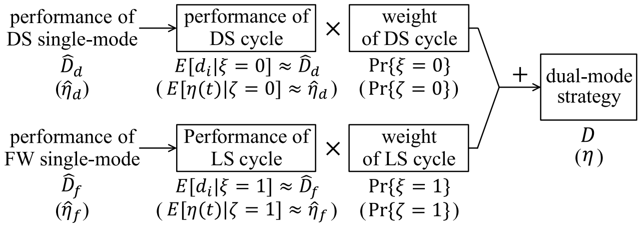


FIGURE 3. **A graphical overview of our model.**

Let $\varphi(t)$ be the power consumption and $\eta(t)=\frac{\varphi_h-\varphi(t)}{\varphi_h}$ be the power efficiency at time t. We also define the indicator

$$\zeta=\begin{cases}0, & \text{if the system is in the DS cycle,}\\ 1, & \text{if the system is in the LS cycle.}\end{cases}$$

Similarly, conditioning on ζ, we have the power efficiency

$$\eta=E[\eta(t)|\zeta=0]\Pr\{\zeta=0\}+E[\eta(t)|\zeta=1]\Pr\{\zeta=1\},\quad (4)$$

where $\Pr\{\zeta=0\}$ and $\Pr\{\zeta=1\}$ are respectively the time fractions of DS cycles and LS cycles.

We derive the weights in (3) and (4) in Section III-A, and solve the conditional expectations of the delay performance and the power efficiency in Section III-B and III-C. We show in Section III-B and III-C that the performance of DS cycle and LS cycle can be described by that of the DS single-mode strategy and the FW single-mode strategy, respectively. Finally, we combine these results according to (3) and (4) in Section III-D, and obtain the mean delay and power efficiency of the dual-mode strategy. Fig. 3 depicts a graphical overview of our model, to facilitate the understanding.

A. CONDITIONAL SLEEP-MODE OPERATION ANALYSIS

Consider a time period $[0,t]$, during which the traffic rate is $\lambda(t)$, and the interface experiences $k_d(t)$ DS cycles and $k_f(t)$ LS cycles. Let $C_d(t)$ and $n_d(t)$ be the mean cycle time of the DS cycle and the mean number of arrivals during a DS cycle, respectively. Similarly, let $C_f(t)$ and $n_f(t)$ be the mean cycle time of the LS cycles and the mean number of arrivals during an LS cycle, respectively. Under condition $\rho=\lambda\overline{X}<1$, the system will eventually reach the steady state. It follows that the probabilities of the DS cycle and the LS cycle are respectively

$$\lim_{t\to\infty}\frac{k_d(t)}{k(t)}=p\quad (5)$$

and

$$\lim_{t\to\infty}\frac{k_f(t)}{k(t)}=q,\quad (6)$$

where p and q are given by (1) and (2). Let n_d, n_f, C_d, and C_f be the limits of $n_d(t)$, $n_f(t)$, $C_d(t)$, and $C_f(t)$ when $t\to\infty$. We thus have the number fractions of the frames that arrive in DS cycles and LS cycles as follows:

$$\Pr\{\xi=0\}=\lim_{t\to\infty}\frac{k_d(t)n_d(t)}{k_d(t)n_d(t)+k_f(t)n_f(t)}=\frac{pn_d}{pn_d+qn_f},\quad (7)$$

$$\Pr\{\xi=1\}=\lim_{t\to\infty}\frac{k_f(t)n_f(t)}{k_d(t)n_d(t)+k_f(t)n_f(t)}=\frac{qn_f}{pn_d+qn_f},\quad (8)$$

and the time fractions of DS cycles and LS cycles as follows:

$$\Pr\{\zeta=0\}=\lim_{t\to\infty}\frac{k_d(t)C_d(t)}{k_d(t)C_d(t)+k_f(t)C_f(t)}=\frac{pC_d}{pC_d+qC_f},\quad (9)$$

$$\Pr\{\zeta=1\}=\lim_{t\to\infty}\frac{k_f(t)C_f(t)}{k_d(t)C_d(t)+k_f(t)C_f(t)}=\frac{qC_f}{pC_d+qC_f}.\quad (10)$$

B. DS CYCLE ANALYSIS

The interface enters the DS cycle if the number of arrivals is less than N_f before the FW mode ends. In a DS cycle, the interface will experience the FW mode and the DS mode in turn. Fig. 1 shows that the DS cycle is quite similar to the cycle of the DS single-mode strategy, except that the number of arrivals during the interval $T_{\mathrm{BtF}}+T_{\mathrm{FW}}$ in the DS cycles must be less than N_f while the cycle of the DS single-mode strategy has no such restriction.

When λ is small, the number of arrivals during $T_{\mathrm{BtF}}+T_{\mathrm{FW}}$ is less than N_f with high probability, and thus the difference between the DS cycle and the cycle of the DS single-mode strategy is negligible. With the increase of λ, the probability that more than N_f frames arrive during $T_{\mathrm{BtF}}+T_{\mathrm{FW}}$ in the DS single-mode strategy increases, and thus the distinction between these two kinds of cycles gradually becomes remarkable.

On the other hand, as Fig. 2(b) shows, when λ is small, the DS cycles predominate in the process of the dual-mode strategy, and thus determine system performance; when λ is large, the DS cycle rarely happens, and thus has little influence on system performance.

Thus, if using the performance of the DS single-mode strategy to describe that of the DS cycle, we can obtain analytical results with reliable accuracy. In the region of small traffic rate, the similarity between the DS cycle and the cycle of the DS single-mode strategy is very high. With the increase of the input traffic rate, though the error increases, the impact of such error on the overall performance decreases since the weight of the DS cycle in the overall performance decreases. That is, the weights can adaptively amend the error. Motivated by this observation, we adopt this method to simplify the analysis of DS cycle.

As Fig. 1(b) illustrates, each time the buffer becomes empty, the interface with the DS single-mode strategy enters the DS mode through a transition of fixed duration $T_s=T_{\mathrm{BtF}}+T_{\mathrm{FW}}+T_{\mathrm{FtD}}$. As soon as N frames arrive at the buffer or the waiting time of the first frame reaches τ, the interface wakes up via the DS Wakeup. This process is similar to that of the 10G EEE studied in [16], and thus we apply the method in [16] to analyze the DS single-mode strategy.

According to [16], the key is to derive the number distribution of the arrivals during a vacation period, which is defined as

$$h_d(n) = \Pr\{n \text{ arrivals during a vacation period of DS single-mode strategy}\}.$$

In Appendix B, we derive $h_d(n)$ based on the event tree that includes all the mutually exclusive events which may occur during the vacation period of the DS single-mode strategy. From the result,we obtain the generating function of $h_d(n)$

$$\begin{aligned} H_d(z) &\triangleq \sum_{n=0}^{\infty} h_d(n) z^n \\ &= \Bigg[e^{-\lambda T_s(1-z)} + \sum_{n=0}^{N-1} e^{-\lambda T_s} \frac{(\lambda T_s)^n}{n!} (z^N - z^n) \\ &\quad - \sum_{n=0}^{N-2} e^{-\lambda \tau} \frac{(\lambda \tau)^n}{n!} (z^N - z^{n+1}) \Bigg] e^{-\lambda T_{\mathrm{DtB}}(1-z)}, \end{aligned}$$

and then the mean number of arrivals during the vacation period

$$\begin{aligned} H_d'(1) = \lambda(T_s + T_{\mathrm{DtB}}) &+ \sum_{n=0}^{N-1} e^{-\lambda T_s} \frac{(\lambda T_s)^n}{n!} (N - n) \\ &- \sum_{n=0}^{N-2} e^{-\lambda \tau} \frac{(\lambda \tau)^n}{n!} (N - n - 1). \end{aligned} \tag{11}$$

According to Little's Law, we have the mean vacation time

$$\widehat{V}_d = \frac{H_d'(1)}{\lambda}. \tag{12}$$

Let $\rho = \lambda \overline{X}$. The mean cycle time is given by

$$\widehat{C}_d = \frac{\widehat{V}_d}{1-\rho} = \frac{H_d'(1)}{\lambda(1-\rho)}. \tag{13}$$

As Fig. 1(b) shows, the interface first enters the FW mode and then the DS mode in each cycle. The duration of the FW mode is T_{FW}, and the mean duration of the DS mode is $\widehat{V}_d - T_s - T_{\mathrm{DtB}}$. Thus, we have the mean power consumption

$$\begin{aligned} \varphi_{\mathrm{DS}} = \big\{ [\widehat{C}_d - T_{\mathrm{FW}} - (\widehat{V}_d - T_s - T_{\mathrm{DtB}})] \cdot \varphi_h + T_{\mathrm{FW}} \cdot \varphi_f \\ + (\widehat{V}_d - T_s - T_{\mathrm{DtB}}) \cdot \varphi_d \big\} / \widehat{C}_d, \end{aligned}$$

from which we derive the power efficiency of the DS single-mode strategy given by (14), as shown at the bottom of this page.

The mean delay can be derived via the P-K formula in [16]:

$$\widehat{D}_d = \frac{\lambda \overline{X^2}}{2(1-\rho)} + \frac{H_d''(1)}{2\lambda H_d'(1)} + \overline{X}, \tag{15}$$

where $H_d''(1)$ is the second moment of $H_d(z)$ and given by (16), as shown at the bottom of this page.

We thus obtain the mean delay of the DS cycle

$$E[d_i | \xi = 0] \approx \widehat{D}_d, \tag{17}$$

and the power efficiency

$$E[\eta(t) | \zeta = 0] \approx \widehat{\eta}_d. \tag{18}$$

C. LS CYCLE ANALYSIS

The interface selects the LS cycle when more than N_f arrivals come before the FW mode ends. Once the LS cycle occurs, the interface will only experience the FW mode. As Fig. 1 shows, the LS cycle is similar to the cycle of the FW single-mode strategy, except that the N_fth frame must arrive within interval $T_{\mathrm{BtF}} + T_{\mathrm{FW}}$ in the LS cycle while the cycle of the FW single-mode strategy has no such constraint.

When λ is large, the difference between the LS cycle and the cycle of the FW single-mode strategy is negligible, because the N_fth frame arrives within interval $T_{\mathrm{BtF}} + T_{\mathrm{FW}}$ of the FW single-mode strategy with high probability. With the decrease of λ, the probability that the N_fth frame arrive during $T_{\mathrm{BtF}} + T_{\mathrm{FW}}$ in the LS single-mode strategy decreases, and thus the distinction between these two kinds of cycles gradually becomes pronounced.

On the other hand, as Fig. 2(b) shows, when λ is large, the LS cycles predominate in the process of the dual-mode

$$\begin{aligned} \widehat{\eta}_d &= \frac{\varphi_h - \varphi_{\mathrm{DS}}}{\varphi_h} \\ &= 1 - \frac{(\widehat{C}_d - \widehat{V}_d + T_{\mathrm{BtF}} + T_{\mathrm{FtD}} + T_{\mathrm{DtB}}) \cdot \varphi_h + T_{\mathrm{FW}} \cdot \varphi_f + (\widehat{V}_d - T_s - T_{\mathrm{DtB}}) \cdot \varphi_d}{\widehat{C}_d \cdot \varphi_h} \\ &= \left[\frac{\varphi_h - \varphi_d}{\varphi_h} - \frac{(T_{\mathrm{BtF}} + T_{\mathrm{FtD}} + T_{\mathrm{DtB}}) \cdot \frac{\varphi_h - \varphi_d}{\varphi_h} + T_{\mathrm{FW}} \cdot \frac{\varphi_f - \varphi_d}{\varphi_h}}{H_d'(1)/\lambda} \right] (1-\rho) \end{aligned} \tag{14}$$

$$\begin{aligned} H_d''(1) = \lambda^2 (T_s + T_{\mathrm{DtB}})^2 &- \sum_{n=0}^{N-1} e^{-\lambda \tau} \frac{(\lambda \tau)^n}{n!} [2\lambda T_{\mathrm{DtB}}(N-n-1) + N(N-1) - n(n+1)] \\ &+ \sum_{n=0}^{N-1} e^{-\lambda T_s} \frac{(\lambda T_s)^n}{n!} [2\lambda T_{\mathrm{DtB}}(N-n) + N(N-1) - n(n-1)] \end{aligned} \tag{16}$$

strategy, and thus determine the system performance; when λ is small, the LS cycle rarely happens, and thus has little influence on system performance.

Thus, following the argument similar to that in Section III-B, we use the performance of the FW single-mode strategy to describe that of the LS cycle.

As Fig. 1(c) illustrates, each time the buffer becomes empty, the interface with the FW single-mode strategy goes into the FW mode through a fixed transition time T_{BtF}. As soon as N_f frames arrive at the buffer, the interface wakes up via the FW Wakeup. Again, we apply the method in [16] to analyze the FW single-mode strategy.

Similarly, based on the event tree that includes all the arrival events which may occur during the vacation period of the FW single-mode strategy, we first derive in Appendix C the distribution of the number of arrivals during a vacation period

$$h_f(n) = \Pr\{n \text{ arrivals during a vacation period of FW single-mode strategy}\}.$$

Accordingly, we have the generating function

$$\begin{aligned} H_f(z) &\triangleq \sum_{n=0}^{\infty} h_f(n) z^n \\ &= \left[\sum_{n=0}^{N_f-1} e^{-\lambda T_{\mathrm{BtF}}} \frac{(\lambda T_{\mathrm{BtF}})^n}{n!} (z^{N_f} - z^n) \right. \\ &\quad \left. + e^{-\lambda T_{\mathrm{BtF}}(1-z)} \right] e^{-\lambda T_{\mathrm{FtB}}(1-z)}, \end{aligned}$$

from which we obtain the mean number of arrivals during a vacation period

$$\begin{aligned} H_f'(1) &= \lambda (T_{\mathrm{BtF}} + T_{\mathrm{FtB}}) \\ &\quad + \sum_{n=0}^{N_f-1} e^{-\lambda T_{\mathrm{BtF}}} \frac{(\lambda T_{\mathrm{BtF}})^n}{n!} (N_f - n). \end{aligned} \tag{19}$$

We then have the mean vacation time

$$\widehat{V}_f = \frac{H_f'(1)}{\lambda}, \tag{20}$$

and the mean cycle time

$$\widehat{C}_f = \frac{\widehat{V}_f}{(1-\rho)} = \frac{H_f'(1)}{\lambda(1-\rho)}. \tag{21}$$

As Fig. 1(c) shows, in each cycle of the FW single-mode strategy, the mean duration of the FW mode is $\widehat{V}_f - T_{\mathrm{BtF}} - T_{\mathrm{FtB}}$. Thus, the mean power consumption is given by

$$\begin{aligned} \varphi_{\mathrm{FW}} = \big\{ & [\widehat{C}_f - (\widehat{V}_f - T_{\mathrm{BtF}} - T_{\mathrm{FtB}})] \cdot \varphi_h \\ & + (\widehat{V}_f - T_{\mathrm{BtF}} - T_{\mathrm{FtB}}) \cdot \varphi_f \big\} / \varphi_h. \end{aligned}$$

Accordingly, we obtain the power efficiency of the FW single-mode strategy as follows:

$$\begin{aligned} \widehat{\eta}_f &= \frac{\varphi_h - \varphi_{\mathrm{FW}}}{\varphi_h} \\ &= \left[1 - \frac{T_{\mathrm{BtF}} + T_{\mathrm{FtB}}}{H_f'(1)/\lambda} \right] \cdot \frac{\varphi_h - \varphi_f}{\varphi_h} \cdot (1-\rho). \end{aligned} \tag{22}$$

In addition, applying the P-K Formula in [16] yields the mean delay

$$\widehat{D}_f = \frac{\lambda \overline{X^2}}{2(1-\rho)} + \frac{H_f''(1)}{2\lambda H_f'(1)} + \overline{X}, \tag{23}$$

where $H_f''(1)$ is the second moment of $H_f(z)$ and is given by

$$\begin{aligned} H_f''(1) &= \lambda^2 (T_{\mathrm{BtF}} + T_{\mathrm{FtB}})^2 \\ &\quad + \sum_{n=0}^{N_f-1} e^{-\lambda T_{\mathrm{BtF}}} \frac{(\lambda T_{\mathrm{BtF}})^n}{n!} \big[2\lambda T_{\mathrm{FtB}} (N_f - n) \\ &\quad + N_f(N_f - 1) - n(n-1) \big]. \end{aligned} \tag{24}$$

Thus, we have the results for the performance of the LS cycle as

$$E[d_i | \xi = 1] \approx \widehat{D}_f, \tag{25}$$

and

$$E[\eta(t) | \zeta = 1] \approx \widehat{\eta}_f. \tag{26}$$

D. PERFORMANCE OF DUAL-MODE STRATEGY

We are now ready to derive the mean delay and the power efficiency of the dual-mode strategy by the unconditional equations (3) and (4). Approximating the mean cycle time of the DS cycle C_d by $\widehat{C}_d$ and that of the LS cycle C_f by $\widehat{C}_f$, we have

$$\Pr\{\zeta = 0\} = \frac{pC_d}{pC_d + qC_f} \approx \frac{p\widehat{C}_d}{p\widehat{C}_d + q\widehat{C}_f}, \tag{27}$$

$$\Pr\{\zeta = 1\} = 1 - \Pr\{\zeta = 0\} \approx \frac{q\widehat{C}_f}{p\widehat{C}_d + q\widehat{C}_f}. \tag{28}$$

Similarly, we approximate n_d (n_f), the mean number of arrivals during a DS cycle (an LS cycle), by $\lambda \widehat{C}_d$ ($\lambda \widehat{C}_f$) according to Little's Law. Then, we have

$$\Pr\{\xi = 0\} = \frac{pn_d}{pn_d + qn_f} \approx \frac{p\lambda\widehat{C}_d}{p\lambda\widehat{C}_d + q\lambda\widehat{C}_f} = \frac{p\widehat{C}_d}{p\widehat{C}_d + q\widehat{C}_f}, \tag{29}$$

$$\Pr\{\xi = 1\} = 1 - \Pr\{\xi = 0\} \approx \frac{q\widehat{C}_f}{p\widehat{C}_d + q\widehat{C}_f}. \tag{30}$$

Substituting (17), (18), (25), (26), and (27) through (30) into (3) and (4), we obtain the mean delay and power efficiency of the dual-mode strategy:

$$D \approx \widehat{D}_d \cdot \frac{p\widehat{C}_d}{p\widehat{C}_d + q\widehat{C}_f} + \widehat{D}_f \cdot \frac{q\widehat{C}_f}{p\widehat{C}_d + q\widehat{C}_f}, \tag{31}$$

$$\eta \approx \widehat{\eta}_d \cdot \frac{p\widehat{C}_d}{p\widehat{C}_d + q\widehat{C}_f} + \widehat{\eta}_f \cdot \frac{q\widehat{C}_f}{p\widehat{C}_d + q\widehat{C}_f}. \tag{32}$$

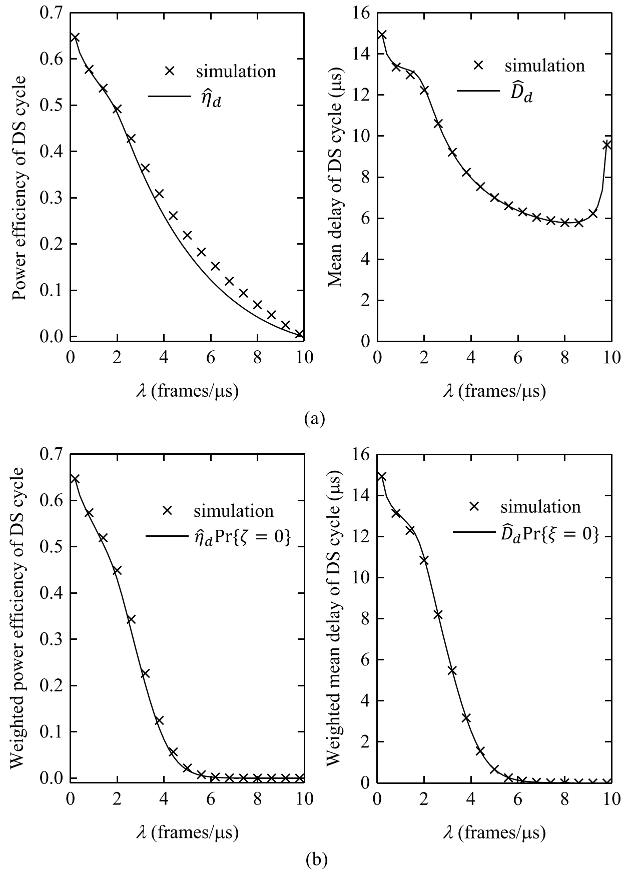


FIGURE 4. Performance of DS cycle: (a) power efficiency and mean delay, (b) weighted power efficiency and weighted mean delay.

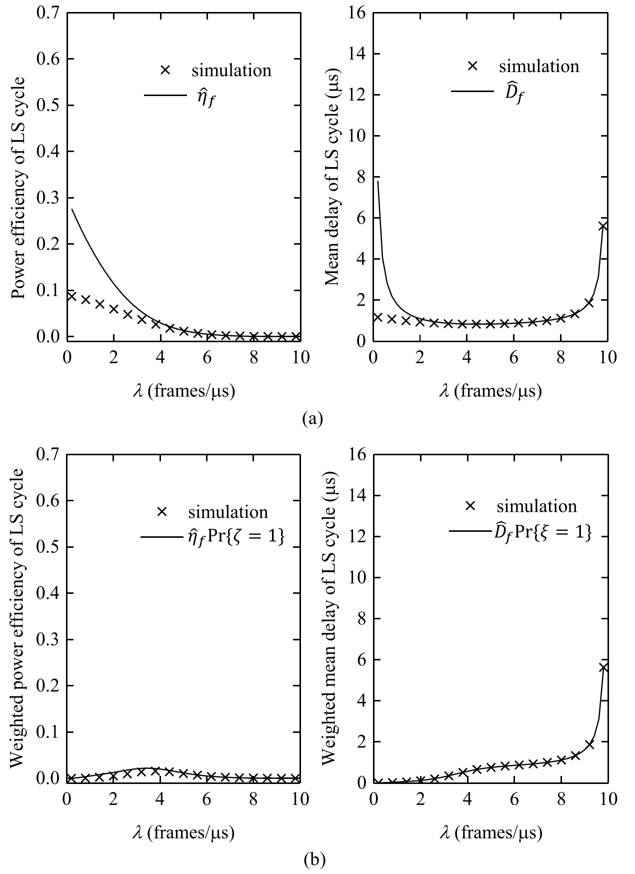


FIGURE 5. Performance of LS cycle: (a) power efficiency and mean delay, (b) weighted power efficiency and weighted mean delay.

As Section I mentions, the mean queue length, denoted by L, is also an important performance metric to study the dual-mode strategy. Following Little's Law, we have the mean queue length:

$$L = \lambda \cdot D. \tag{33}$$

To verify the effectiveness of our modeling methodology, we compare the performances $\widehat{D}_d$ and $\widehat{\eta}_d$ in (17) and (18) and the weighted performances $\widehat{D}_d \Pr\{\xi = 0\}$ and $\widehat{\eta}_d \Pr\{\zeta = 0\}$ to the simulation results of the DS cycle in Fig. 4, where $N_f = 4$, $T_{\mathrm{FW}} = 1.1\mu$s, $N = 41$, and $\tau = 20\mu$s. Fig. 4(a) displays that $\widehat{D}_d$ and $\widehat{\eta}_d$ agree well with the simulated mean delay and the simulated power efficiency of the DS cycle when λ is small, and the difference between $\widehat{\eta}_d$ and the simulated power efficiency becomes remarkable when λ becomes large. However, $\widehat{D}_d \Pr\{\xi = 0\}$ and $\widehat{\eta}_d \Pr\{\zeta = 0\}$ fit the simulated weighted performance very well for all the values of λ, as Fig. 4(b) shows. This clearly indicates that the errors of $\widehat{D}_d$ and $\widehat{\eta}_d$ in (17) and (18) when λ is large can be amended by the weights $\Pr\{\xi = 0\}$ and $\Pr\{\zeta = 0\}$, respectively. Similarly, Fig. 5 confirms that the errors of $\widehat{D}_f$ and $\widehat{\eta}_f$ in (25) and (26) when λ is small can be suppressed by the weights $\Pr\{\xi = 1\}$ and $\Pr\{\zeta = 1\}$. Therefore, as we will verify by simulations in Section IV, the analytical results of (31), (32), and (33) are quite accurate.

IV. THRESHOLD SELECTION RULES

The essence of the EEE protocols is to save power at the expense of delay performance. Similarly, in the dual-mode strategy, such a performance tradeoff can be achieved by tuning thresholds T_{FW}, N_f, N, and τ. In this section, based on the results in Section III, we study the impacts of T_{FW}, N_f, N, and τ on performance tradeoff and seek a way to select these thresholds. We conduct simulations by a discrete-event simulation model [28] coded in C. We consider a 100G EEE interface in the simulation. We assume that the frame size is exponentially distributed with an average of 1250 bytes. Considering the fact that the utilization of a typical Ethernet link ranges from 5% to 30%, we suppose that the steady-state traffic rate is 2.0 frames/μs, which corresponds to a link utilization of 20%.

A. FW MODE VS. DS MODE

The impact of T_{FW} and N_f on the tradeoff between the DS mode and the FW mode is two-fold. As we mention in Section II, for a fixed traffic rate λ, T_{FW} and N_f can make a tradeoff between the mean delay (or the queue length) and the power efficiency by controlling the probabilities of the DS cycle and the LS cycle. On the other hand, the duration of the FW mode increases with T_{FW} and N_f. Intuitively, if the time fraction of the FW mode in the DS cycle is larger, the interface can save less power via the DS cycle when λ is low, but more power via the LS cycle when λ is large. This is another kind of tradeoff between these two modes. In the following, we start our discussion from the first aspect.

The dual-mode strategy is devised to accommodate the high-speed feature of 100G Ethernet [11], [15]. For such high-speed Ethernet, the power saving is mainly carried out when λ is low. The major function of the FW mode is to implement the conditional sleep-mode operation to decide whether the DS mode should be implemented or not.

According to Little's Law, the mean number of arrivals during the BtF and the FW mode is $\lambda(T_{\text{BtF}} + T_{\text{FW}})$. Given that $N_f = \lambda(T_{\text{BtF}} + T_{\text{FW}})$, the probability that fewer than N_f arrivals are accumulated at the end of the FW mode is relatively large if the traffic rate is smaller than $\frac{N_f}{T_{\text{BtF}}+T_{\text{FW}}}$. In this case, the interface implements the DS cycle with large probability, and thus achieves a high power efficiency. On the contrary, when the traffic rate is larger than $\frac{N_f}{T_{\text{BtF}}+T_{\text{FW}}}$, the number of arrivals during $T_{\text{BtF}} + T_{\text{FW}}$ will be larger than N_f with high probability. In this case, the interface almost always enters the LS cycle, and thus keeps its queue length at a low level. We demonstrate this point in Fig. 6, where we set $\lambda = 2.0$ frames$/\mu$s, $N_f = 2$, $T_{\text{FW}} = 0.1\mu$s, $N = 41$, and $\tau = 20\mu$s such that $N_f = \lambda(T_{\text{BtF}} + T_{\text{FW}})$. We see from Fig. 6 that (32) and (33) agree with the simulation results very well. Also, Fig. 6 clearly shows that the dual-mode strategy is able to attain a high power efficiency when $\lambda < \frac{N_f}{T_{\text{BtF}}+T_{\text{FW}}}$, and a bounded queue length when $\lambda > \frac{N_f}{T_{\text{BtF}}+T_{\text{FW}}}$.

As a comparison, Fig. 6 also depicts the performance of the DS single-mode strategy and the FW single-mode strategy, of which the analytical results are plotted according to (17), (18), (25), and (26). We can see from Fig. 6 that the queue length of the DS single-mode strategy grows fast when $\lambda > \frac{N_f}{T_{\text{BtF}}+T_{\text{FW}}}$, while the power efficiency of the FW single-mode strategy is quite low even when $\lambda < \frac{N_f}{T_{\text{BtF}}+T_{\text{FW}}}$. This is attributed to the fact that the FW single-mode strategy and the DS single-mode strategy are both unable to select proper sleep modes according to the traffic rate.

Therefore, the interface with the dual-mode strategy can attain an optimal tradeoff between the DS cycle and the LS cycle in different traffic rate regions if we configure T_{FW} and N_f in accordance with the following rule.

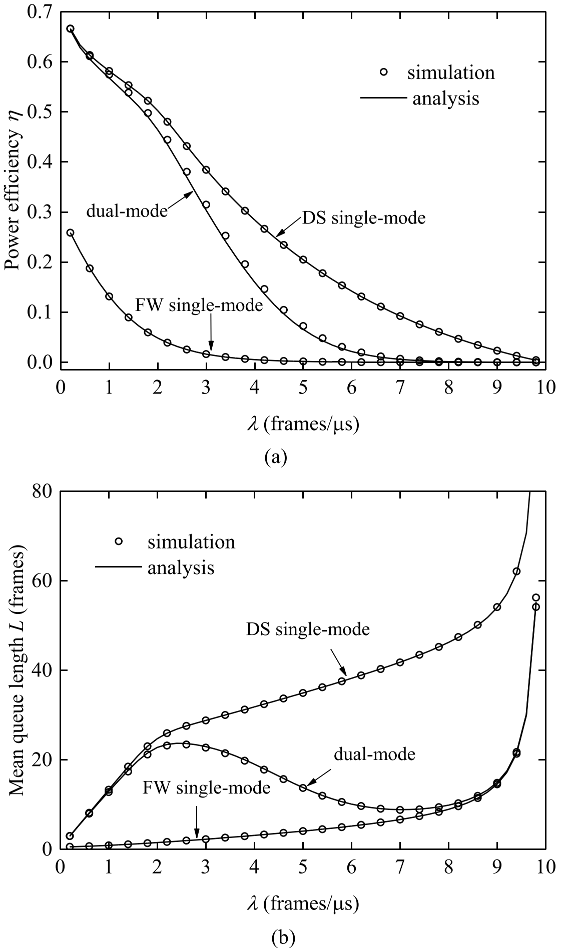


FIGURE 6. Performance evaluation: (a) Power efficiency and (b) mean queue length vs. λ, where $N_f = 2$, $T_{\text{FW}} = 0.1\mu$s, $N = 41$, and $\tau = 20\mu$s.

Rule 1: For a steady-state traffic rate λ^*, the selection of thresholds T_{FW} and N_f should conform to the condition

$$\frac{N_f}{T_{\text{BtF}} + T_{\text{FW}}} = \lambda^*. \tag{34}$$

On the other hand, though relation (34) is given, the values of N_f and T_{FW} will also lead to another kind of performance tradeoff between the FW mode and the DS mode in different traffic rate regions.

When λ is remarkably smaller than λ^*, the number of arrivals before the FW mode ends can hardly reach N_f. Thus, the interface will enter the DS cycle with probability $p \to 1$, which means $\Pr\{\xi = 0\} \to 1$ and $\Pr\{\zeta = 0\} \to 1$. In this case, the system performance is almost determined by that of the DS cycle, that is, (32) and (33) change

$$\eta \approx E[\eta(t)|\zeta = 0] \approx \left\{ \frac{\varphi_h - \varphi_d}{\varphi_h} - \frac{(T_{\text{BtF}} + T_{\text{FtD}} + T_{\text{DtB}})\frac{\varphi_h - \varphi_d}{\varphi_h} + T_{\text{FW}}\frac{\varphi_f - \varphi_d}{\varphi_h}}{H_d'(1)/\lambda} \right\} (1 - \rho) \tag{35}$$

to (35), as shown at the bottom of the previous page, and

$$L \approx \lambda E[d_i|\xi = 0] \approx \frac{\lambda^2\overline{X^2}}{2(1-\rho)} + \frac{H_d''(1)}{2H_d'(1)} + \lambda\overline{X}. \quad (36)$$

Recall that we assume $\tau > T_{\mathrm{BtF}} + T_{\mathrm{FW}} + T_{\mathrm{FtD}}$ and $N > N_f$ in Assumption A4, and the transition time $T_{\mathrm{FtD}} = 0.9\mu$s is very short. Since λ is remarkably smaller than λ^*, the probability that the number of arrivals during time interval $T_{\mathrm{BtF}} + T_{\mathrm{FW}} + T_{\mathrm{FtD}}$ reaches N is negligible, i.e.,

$$e^{-\lambda(T_{\mathrm{BtF}}+T_{\mathrm{FW}}+T_{\mathrm{FtD}})}\frac{[\lambda(T_{\mathrm{BtF}} + T_{\mathrm{FW}} + T_{\mathrm{FtD}})]^n}{n!} \to 0 \quad (37)$$

for $n \geq N$. In other words, the trigger condition of the DS Wakeup can be satisfied only after the FW mode ends. In this case, the queue length builds up during the vacation time is almost independent of N_f and T_{FW}. This point can be confirmed by substituting condition (37) into (36), which yields (38), as shown at the bottom of the this page. On the other hand, the power efficiency η in this case depends on the ratio of the FW mode to the vacation time. Intuitively, if T_{FW} becomes large, such ratio will be large and thus η will decline. We demonstrate this point by substituting (37) into (35), as shown in (39), as shown at the bottom of the this page. The above shows that η in this case decreases with the increase of T_{FW}. In summary, when λ is remarkably smaller than λ^*, large N_f and T_{FW} impair the power efficiency, but have no benefit to the reduction of the mean queue length.

When λ is remarkably larger than $\lambda^* = \frac{N_f}{T_{\mathrm{BtF}}+T_{\mathrm{FW}}}$, the interface enters the LS cycle with probability $q \to 1$, which leads to $\Pr\{\xi = 1\} \to 1$ and $\Pr\{\zeta = 1\} \to 1$. In this case, the system performance is dominated by that of the LS cycle, and the power efficiency changes to:

$$\eta \approx E[\eta(t)|\zeta = 1] \approx \left[1 - \frac{T_{\mathrm{BtF}}+T_{\mathrm{FtB}}}{H_f'(1)/\lambda}\right]\frac{\varphi_h-\varphi_f}{\varphi_h}(1-\rho). \quad (40)$$

Equation (40) shows that the mean vacation time of the LS cycle (i.e., $H_f'(1)/\lambda$ in (40)) and the power efficiency η increase with N_f and T_{FW}. However, the increment of η is quite limited, which can be demonstrated by (40) as follows. Since the mean vacation time of the LS cycle is larger than $T_{\mathrm{BtF}} + T_{\mathrm{FtB}}$, the second term in (40) satisfies

$$1 - \frac{T_{\mathrm{BtF}} + T_{\mathrm{FtB}}}{H_f'(1)/\lambda} < 1.$$

Therefore, the power efficiency of the LS cycle given by (40) is upper bounded by:

$$\eta < \frac{\varphi_h - \varphi_f}{\varphi_h}(1-\rho) = \frac{\varphi_h - 0.7\varphi_h}{\varphi_h}(1-\rho) = 0.3(1-\rho), \quad (41)$$

where $1 - \rho$ is much smaller than 1 when the traffic rate λ is high. On the other hand, when N_f and T_{FW} increase, the number of frames accumulated in the buffer at the end of the FW mode grows, which in turn increases the queue length. Briefly, when λ is remarkably larger than λ^*, increasing N_f and T_{FW} incurs a large queue length, while only leading to a slight improvement in power efficiency.

Fig. 7 plots the power efficiency η and the mean queue length L changing with N_f, where $\lambda^* = 2.0$ frames/μs, and N_f and T_{FW} satisfy Rule 1. In addition, we set $N = 41$ and $\tau = 20\mu$s. We check the system performance changing with N_f and T_{FW} when $\lambda = 0.5$ frames/μs and 7.0 frames/μs. Fig. 7 clearly shows that with the increase of N_f, η decreases linearly while L remains almost unchanged when $\lambda = 0.5$ frames/μs, and η converges from 0 to 0.09 while L first slightly decreases and then grows by four times when $\lambda = 7.0$ frames/μs.

This result suggests that T_{FW} and N_f should be as small as possible, such that the system can save more power when the traffic rate is small, while keeping the queue length short when the traffic rate becomes high. From (34), we have $N_f = \lambda^*(T_{\mathrm{BtF}} + T_{\mathrm{FW}})$. In addition, since N_f is an integer, the smallest possible integer of N_f is $\lceil\lambda^* T_{\mathrm{BtF}}\rceil$, where $\lceil x\rceil$ is the smallest integer larger than x. This yields the second selection rule for T_{FW} and N_f as follows:

Rule 2: For a given steady-state traffic rate λ^*, threshold N_f should be selected as the following formula:

$$N_f = \lceil\lambda^* T_{\mathrm{BtF}}\rceil. \quad (42)$$

According to Rules 1 and 2, we have $T_{\mathrm{FW}} = \lceil\lambda^* T_{\mathrm{BtF}}\rceil/\lambda^* - T_{\mathrm{BtF}}$. For example, if $\lambda^* = 2.0$ frames/μs, Rule 2 will set $N_f = 2$ and $T_{\mathrm{FW}} = 0.1\mu$s.

B. TRADEOFF IN DS CYCLE

Once T_{FW} and N_f are set according to Rules 1 and 2, the dual-mode strategy can adaptively implement the LS cycle and the DS cycle. When the interface enters the DS cycle, the vacation time is controlled by N and τ. As Sections I and II mention, the Wakeup condition of the DS cycle of the dual-mode strategy is quite similar to that of the 10G EEE [16]. In the

$$L \approx \frac{\lambda^2\overline{X^2}}{2(1-\rho)} + \frac{(N+\lambda T_{\mathrm{DtB}})^2 - N - \sum_{n=0}^{N-1} e^{-\lambda\tau}\frac{(\lambda\tau)^n}{n!}[2\lambda T_{\mathrm{DtB}}(N-n-1) + N(N-1) - n(n+1)]}{2\left\{N - \sum_{n=0}^{N-1} e^{-\lambda\tau}\frac{(\lambda\tau)^n}{n!}[N-(n+1)] + \lambda T_{\mathrm{DtB}}\right\}} + \lambda\overline{X} \quad (38)$$

$$\eta \approx \left[\frac{\varphi_h - \varphi_d}{\varphi_h} - \frac{(T_{\mathrm{BtF}} + T_{\mathrm{FtD}} + T_{\mathrm{DtB}})\frac{\varphi_h-\varphi_d}{\varphi_h} + T_{\mathrm{FW}}\frac{\varphi_f-\varphi_d}{\varphi_h}}{\frac{N}{\lambda} - \sum_{n=0}^{N-1} e^{-\lambda\tau}\frac{(\lambda\tau)^n}{n!}\cdot\frac{N-n-1}{\lambda} + T_{\mathrm{DtB}}}\right](1-\rho) \quad (39)$$

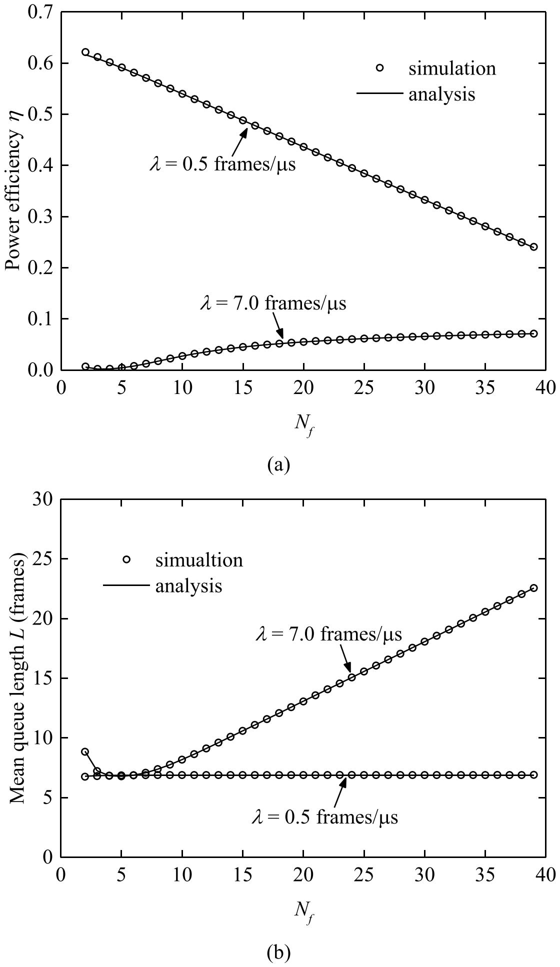


FIGURE 7. (a) power efficiency and (b) mean queue length vs. N_f, where $N_f = \lambda^*(T_{\text{BtF}} + T_{\text{FW}})$, $N = 41$, and $\tau = 20\mu$s.

following, we show that the impact of N and τ on the performance of the dual-mode strategy is also similar to what happens in the 10G EEE [16].

As Section I mentions, a 10G EEE interface uses N and τ to control the duration of the LPI mode. According to Little's Law, there are $\lambda^*\tau$ arrivals during time interval τ on average. Thus, Ref. [16] recommends setting $N - 1 = \lambda^*\tau$ or $(N - 1)/\tau = \lambda^*$, such that N and τ can compensate each other under different traffic rates as follows:

- When $\lambda < \lambda^*$, the waiting time of the first arrival during the vacation tends to reach τ before N frames arrive. In this case, timer τ triggers the DS Wakeup with high probability, and thus the delay of the frame that arrives during the vacation time is bounded.
- When $\lambda > \lambda^*$, N frames tend to arrive before timer τ expires, and the Wakeup operation is almost always triggered by N, such that the queue length at the end of the vacation is bounded.

On the other hand, Ref. [16] also shows that the duration of the LPI mode increases with N and τ. If N and τ are large, the LPI mode is long such that the fraction of energy consumed by the transition operations is small, and thus power efficiency is large. As N and τ go to infinity, the power efficiency converges to a constant related to the traffic load [16]. However, the power efficiency is improved at the expense of delay performance. If N and τ are large, the frames that arrive during the LPI mode have to wait for a long time. Thus, given $(N - 1)\tau = \lambda^*$, Ref. [16] suggests selecting the values of N and τ to meet a preset delay requirement.

To demonstrate that N and τ have a similar effect on the performance of the dual-mode strategy, Fig. 8 plots the power efficiency, the mean delay, and the mean queue length with the following parameter setups:

1) $N_f = 2$, $T_{\text{FW}} = 0.1\mu$s, $N = 41$, and $\tau = 20\mu$s, which corresponds to the $\tau\&N$ policy in Fig. 8;
2) $N_f = 2$, $T_{\text{FW}} = 0.1\mu$s, $N = 41$, and $\tau \to \infty$, which corresponds to the N policy in Fig. 8, namely the interface only uses N to regulate the vacation time of the DS cycle;
3) $N_f = 2$, $T_{\text{FW}} = 0.1\mu$s, $N \to \infty$, and $\tau = 20\mu$s, which corresponds to the τ policy in Fig. 8, namely the interface only uses τ to regulate the vacation time of the DS cycle.

As Fig. 8 shows, when $\lambda < \frac{N-1}{\tau} = 2.0$ frames/μs, the N policy suffers a large delay, since it takes a long time to accumulate N arrivals in the buffer when the frame arrivals are sparse. When $\lambda > \frac{N-1}{\tau}$, the τ policy suffers a serious queue length performance, since the frame arrivals are intensive and the frames accumulate rapidly during the interval τ. In contrast, the $\tau\&N$ policy is able to adaptively adopt N or τ to wake up the interface according to the traffic rate. It uses τ to trigger the DS Wakeup when $\lambda < \frac{N-1}{\tau}$, and uses N to trigger the DS Wakeup when $\lambda > \frac{N-1}{\tau}$. Thus, the $\tau\&N$ policy achieves the minimum of the mean delay and the power efficiency of the N policy and the τ policy, i.e., we have

$$D \approx \min\{D_\tau, D_N\},$$
$$L \approx \lambda \cdot \min\{D_\tau, D_N\},$$

and

$$\eta \approx \min\{\eta_\tau, \eta_N\},$$

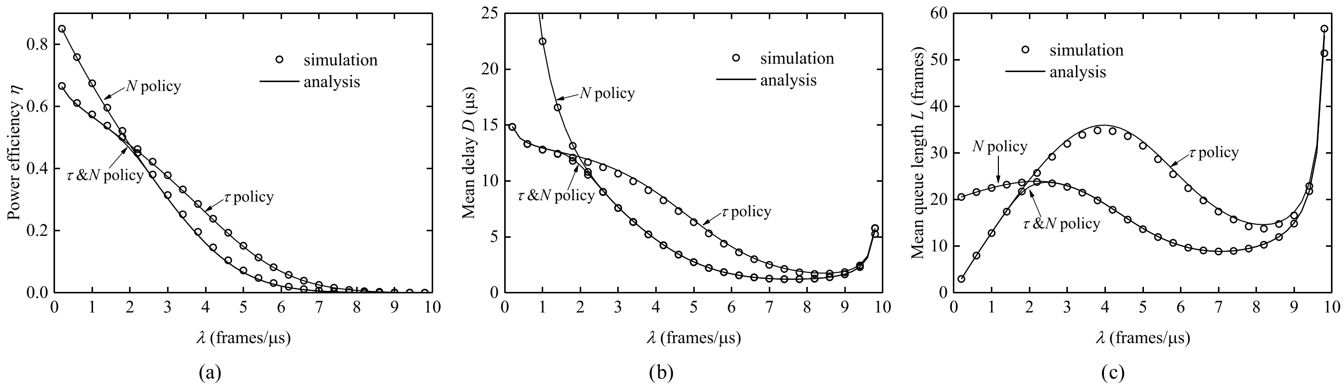
where $D_\tau = \lim_{N\to\infty} D$ and $\eta_\tau = \lim_{N\to\infty} \eta$ are the mean delay and the power efficiency of the τ policy, and $D_N = \lim_{\tau\to\infty} D$ and $\eta_N = \lim_{\tau\to\infty} \eta$ are those of the N policy. In particular, according to the results in Fig. 8, there are

$$D \approx D_\tau \approx D_N \tag{43}$$

and

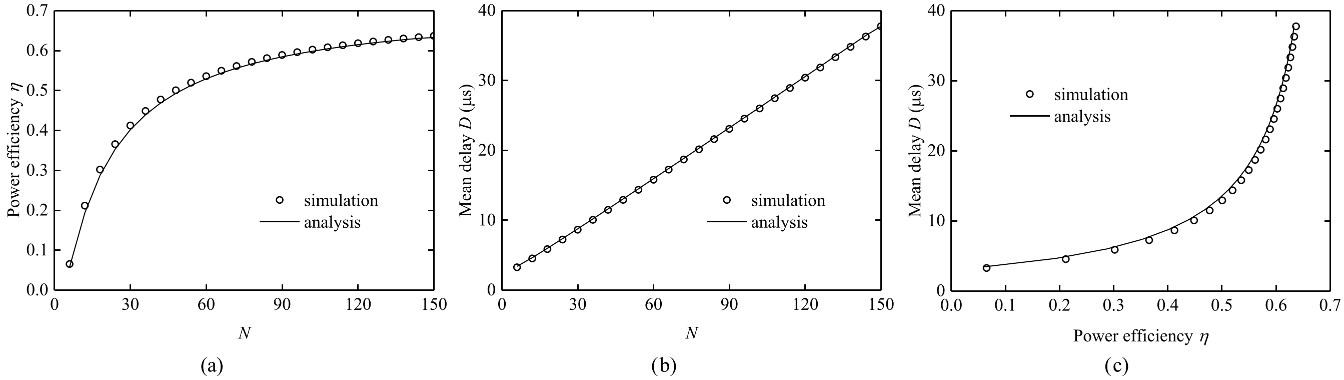
$$\eta \approx \eta_\tau \approx \eta_N \tag{44}$$

when $\lambda = \frac{N-1}{\tau}$. This point is quite similar to that of 10G EEE studied in [16]. Thus, we use the rule similar to EEE 1 presented in [16]:



(a) (b) (c)

FIGURE 8. Performance comparison of three policies: (a) power efficiency, (b) mean delay, and (c) mean queue length, where $T_{\mathrm{FW}} = 0.1\mu s$, $N_f = 2$, and $\frac{N-1}{\tau} = \lambda^* = 2.0$ frames/μs.



(a) (b) (c)

FIGURE 9. Performance evaluation: (a) power efficiency vs. *N*, (b) mean delay vs. *N*, and (c) mean delay vs. power efficiency, where $T_{\mathrm{FW}} = 0.1\mu s$, $N_f = 2$, and $\frac{N-1}{\tau} = \lambda^* = 2.0$ frames/μs.

Rule 3: For a given steady-state traffic rate λ^*, thresholds N and τ should be selected as:

$$\lambda^* = \frac{N-1}{\tau}. \tag{45}$$

We proceed to determine the values of thresholds N and τ on the basis of Rule 3. Fig. 9(a) and 9(b) depict η and D versus N, where $\lambda^* = 2.0$ frames/μs, $T_{\mathrm{FW}} = 0.1\mu$s, $N_f = 2$, and τ varies with N according to (45). Fig. 9(a) shows that η increases with N, and converges to an upper bound when $N \to \infty$. This can be demonstrated by setting limit $N, \tau \to \infty$ on (32). In (32), the mean cycle time of the DS cycle $\widehat{C}_d \to \infty$ as $N \to \infty$, and thus the upper bound of the power efficiency η is

$$\eta^* = \lim_{N,\tau\to\infty} \eta = \lim_{N,\tau\to\infty} \widehat{\eta}_d = \frac{\varphi_h - \varphi_d}{\varphi_h}(1-\rho) = 0.9(1-\rho).$$

In the example Fig. 9(a) gives, the power efficiency converges to $\eta^* = 0.72$ when N and τ approach infinity. On the other hand, as Fig. 9(b) indicates, the mean delay linearly increases with N and τ. Therefore, as Fig. 9(c) plots, there is a tradeoff between D and η with the increase of N and τ. In particular, the power efficiency η can be remarkably improved with a small delay cost when η is small, but η can only be enhanced slightly even if we trade a large delay cost when η is close to its upper bound. Such a tradeoff is also quite similar to the situation in 10G EEE.

Given the steady-state traffic rate λ^* and Rule 3, we assume that the delay requirement is D^*. According to (43), at the traffic rate λ, there is equation (46), as shown at the bottom of this page. We thus have the following rule:

Rule 4: In terms of a given delay requirement D^*, threshold N can be selected via (46).

C. PERFORMANCE UNDER BURSTY TRAFFIC

In practice, the traffic in Ethernet links is bursty by nature. For example, the test data in [29] indicates that the traffic

$$D \approx D_N = \lim_{\tau\to\infty} D = \left\{\frac{\lambda\overline{X^2}}{2(1-\rho)} + \frac{(N+\lambda T_{\mathrm{DtB}})^2 - N}{2\lambda(N+\lambda T_{\mathrm{DtB}})} + \overline{X}\right\} \cdot \frac{p(N+\lambda T_{\mathrm{DtB}})}{p(N+\lambda T_{\mathrm{DtB}}) + qH_f'(1)} + \widehat{D}_f \cdot \frac{qH_f'(1)}{p(N+\lambda T_{\mathrm{DtB}}) + qH_f'(1)} \tag{46}$$

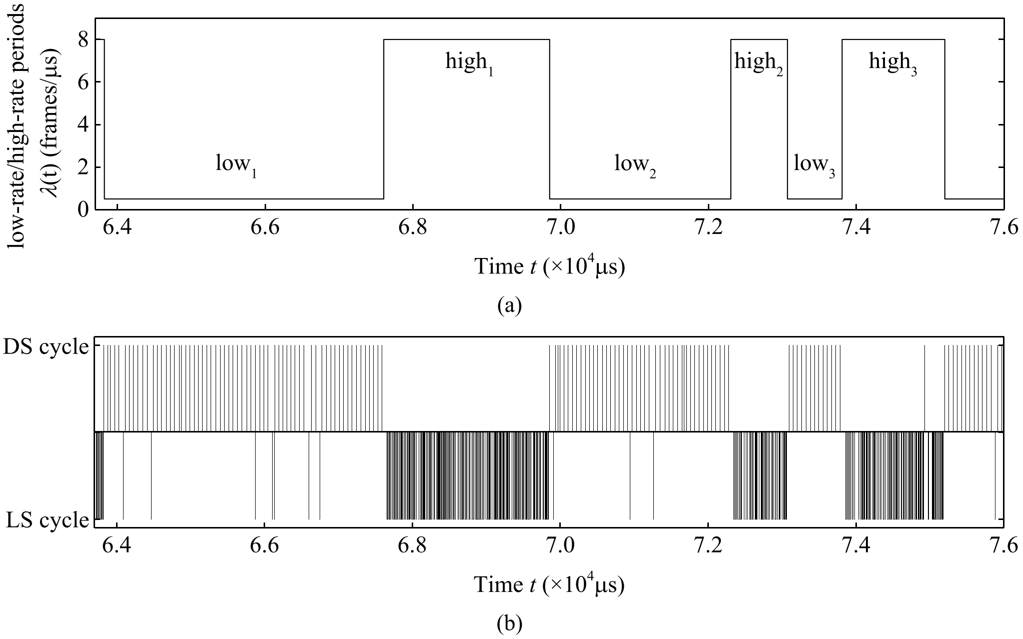


FIGURE 10. Working cycles of the interface: (a) 6 low-rate/high-rate periods in the simulation and (b) cycle records.

TABLE 1. Mean queue length and power efficiency of three strategies during each high-rate/low-rate period in Fig. 10.

performance	strategy	low$_1$	high$_1$	low$_2$	high$_2$	low$_3$	high$_3$	mean value of 3 MMPP cycles
mean queue length (frames)	dual-mode	7.74	11.81	7.65	11.75	7.32	12.65	11.67
	DS single-mode	8.52	46.13	7.82	44.56	7.55	45.70	42.31
	FW single-mode	1.72	9.98	1.72	9.88	1.90	10.34	9.33
power efficiency	dual-mode	0.623	0.007	0.618	0.010	0.613	0.009	0.393
	DS single-mode	0.619	0.051	0.615	0.063	0.581	0.069	0.403
	FW single-mode	0.204	0	0.204	0.001	0.203	0	0.126

rate of a link in commercial data center networks exhibits a time-of-day feature: the traffic rate almost periodically fluctuates around an average value in the long run. In other words, though the instantaneous traffic rate can change fast and irregularly, the statistical average traffic rate varies very slowly [29], [30]. In the following, we further demonstrate via simulation that the dual-mode strategy can perform well under the bursty traffic, as long as the thresholds are set according to the statistical average traffic rate.

As Fig. 6 illustrates, the advantage of the dual-mode strategy is that, given thresholds N_f and T_{FW}, it can adaptively select sleep modes according to the instantaneous input traffic. Consider the case where the input traffic rate λ varies around the steady state traffic rate $\lambda^* = N_f/(T_{\mathrm{BtF}} + T_{\mathrm{FW}}) =$ 2.0 frames/μs. We can see from Fig. 2(b) and Fig. 6 that

(1) if $\lambda < \lambda^*$, the dual-mode strategy will select the DS cycle more frequently than the LS cycle, such that the interface can save more power;

(2) if $\lambda > \lambda^*$, the dual-mode strategy will enter the LS cycle more often than the DS cycle, such that the interface can suppress the queue length.

To demonstrate this benefit clearly, we simulate the dual-mode strategy under the bursty traffic. We assume that the input traffic is a two-state Markov-modulated Poisson Process (MMPP) [31], where the frame arrival process alternates between two kinds of periods: high-rate period and low-rate period. The high-rate period and the low-rate period are exponentially distributed with mean durations $1/\alpha$ and $1/\beta$, respectively. The traffic rate is λ_h during the high-rate period and λ_l during the low-rate period. Thus, the average traffic rate of the two-state MMPP is [32]

$$\lambda = \frac{\lambda_h \cdot \beta + \lambda_l \cdot \alpha}{\alpha + \beta}.$$

The burstiness of the two-state MMPP is defined as

$$b = \frac{1}{\alpha + \beta}.$$

clearly, burstiness b increases with the decrease of α and β.

In reality, the timescales of the high-rate period and the low-rate period range from a few microseconds to several seconds [33] in different applications. Moreover, the low-rate period is usually several times longer than the high-rate period [34]. We thus set $\alpha : \beta = 4 : 1$, $\lambda_h = 8.0$ frames/μs, and $\lambda_l = 0.5$ frames/μs to generate the bursty input traffic with $\lambda^* = 2.0$ frames/μs in our simulation. Assume that the mean delay requirement is $D^* = 12\mu$s. According to Rules 1

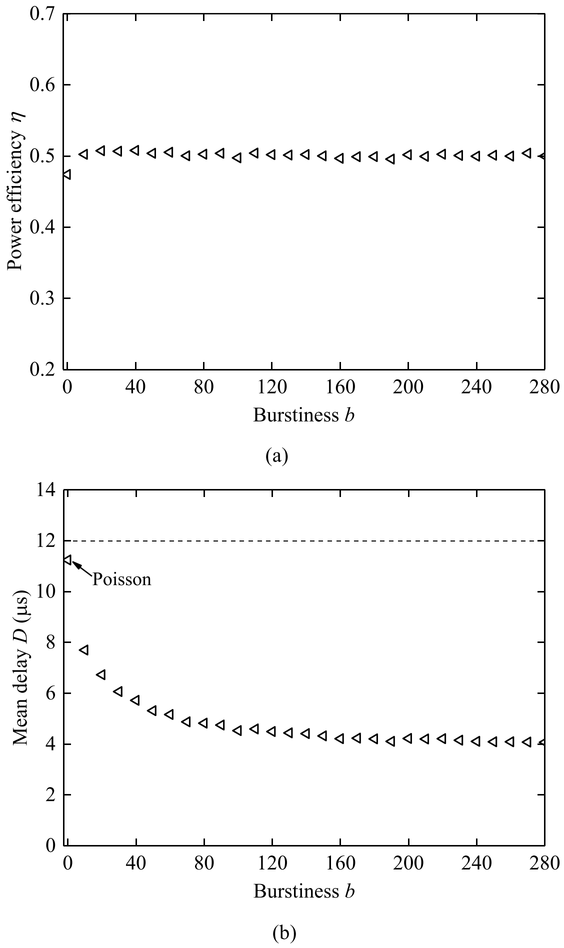


FIGURE 11. **Performance under bursty traffic: (a) power efficiency and (b) mean delay vs. burstiness *b*, where $\alpha : \beta = 4 : 1$, $\lambda_h = 8.0$ frames/μs, and $\lambda_l = 0.5$ frames/μs, and $D^* = 12$ frames/μs.**

through 4, we set $N_f = 2$, $T_{\text{FW}} = 0.1\mu$s, $N = 41$, and $\tau = 20\mu$s for the dual-mode strategy.

We observe the behavior of the dual-mode strategy in three consecutive cycles of the two-state MMPP in Fig. 10(a), where $1/\alpha = 1$ms, $1/\beta = 4$ms. Fig. 10(b) clearly displays that the dual-mode strategy almost selects the DS cycle during the low-rate periods and the LS cycle during the high-rate periods. Consequently, as Table 1 shows, the dual-mode strategy can suppress the queue length during all these cycles, and achieve a high power efficiency when the input traffic rate is low.

As a comparison, Table 1 also provides the simulation results of the DS single-mode strategy with $T_{\text{FW}} = 0.1\mu$s, $N = 41$, $\tau = 20\mu$s, and the FW single-mode strategy with $N_f = 2$, $N = 41$, and $\tau = 20\mu$s. Table 1 clearly shows that, though the DS single-mode strategy can achieve almost the same power efficiency with the dual-mode strategy during the low-rate periods, its queue length is much longer than that of the dual-mode strategy during the high-rate periods. The FW single-mode can always suppress the queue length very well, but its power efficiency is much lower than those of the dual-mode strategy and the DS single-mode strategy. We thus conclude that the dual-mode strategy can make a good performance compromise between the DS single-mode and the FW single-mode strategies under the bursty traffic.

Fig. 11 further plots the power efficiency and the mean delay when burstiness b changes from 0 to 280. From Fig. 11, one can see that the power efficiency almost keeps unchanged, while the mean delay decreases and finally converges to 3.9μs with the increase of b. Note that the two-state MMPP traffic with $b = 0$ reduces to the Poisson traffic. This indicates that our analytical results cannot accurately delineate the performance of the dual-mode strategy under the bursty traffic. Even so, Fig. 11 shows that the mean delay of the dual-mode strategy under bursty input traffic is upper bounded by the analytical result. This implies that our rules proposed based on our analytical results are conservative under the bursty input traffic. Since the instantaneous Ethernet traffic changes irregularly and the burstiness varies over time [30], we conclude from Fig. 11 that our rules can provide worst-case performance guarantee.

Therefore, the dual-mode strategy can be applied in practice as follows. The Ethernet interface monitors the input traffic and calculates the statistical average traffic rate every fixed time period, e.g., one week or one month [35]. According to such collected history information, it estimates the statistical average traffic rate of the next month, and sets the thresholds using Rules 1 through 4 according to the estimated value.

V. CONCLUSION

In this paper, we propose an analytical model based on the concept of conditional sleep-mode to analyze the dual-mode strategy designed for the 100G EEE protocol. The key idea of our approach is to calculate the performance of the dual-mode strategy as the weighted average of that of the DS cycle and the LS cycle. Using this approach, we derive the power efficiency, the mean delay, and the mean queue length. Our simulation results confirm that our analysis is quite accurate. In addition, our analysis shows that the FW thresholds (i.e., N_f and T_{FW}) mainly control the probabilities and the durations of the DS mode and the FW mode, while the DS thresholds (i.e., N and τ) balance the power efficiency and the mean delay of the DS cycle. Based on this observation, we provide four parameter selection rules, which are easy to implement and can help the dual-mode strategy to select the suitable sleep mode according to the fluctuation of traffic rate.

APPENDIX A
DIFFICULTY OF EXACT ANALYSIS OF DUAL-MODE STRATEGY

As Section II mentions, thresholds N_f and T_{FW} in the dual-mode strategy make the distribution of the number of arrivals during a vacation depend on that during the FW mode. If we directly apply the model in [16] to the dual-mode strategy, the analysis will be very complex. In the following, we elaborate on this difficulty.

According to [16], modeling of the dual-mode strategy starts from the distribution of the number of arrivals during

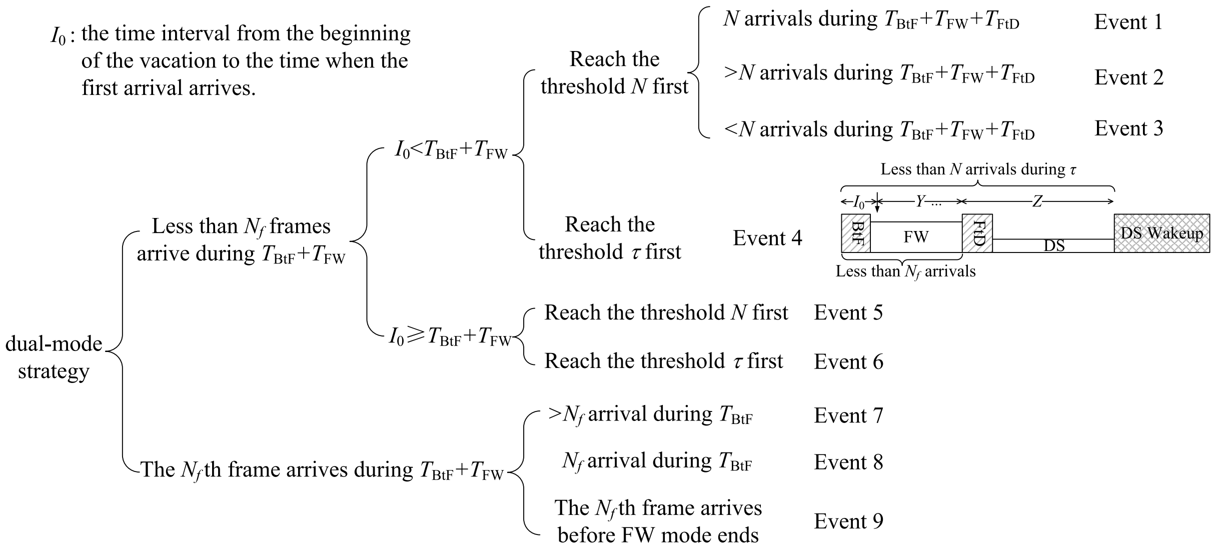

I_0: the time interval from the beginning of the vacation to the time when the first arrival arrives.
dual-mode strategy
Less than N_f frames arrive during $T_{BtF}+T_{FW}$
$I_0<T_{BtF}+T_{FW}$
Reach the threshold N first
N arrivals during $T_{BtF}+T_{FW}+T_{FtD}$
Event 1
$>N$ arrivals during $T_{BtF}+T_{FW}+T_{FtD}$
Event 2
$<N$ arrivals during $T_{BtF}+T_{FW}+T_{FtD}$
Event 3
Reach the threshold τ first
Event 4
Less than N arrivals during τ
I_0
Y
Z
BtF
FW
FtD
DS
DS Wakeup
Less than N_f arrivals
$I_0 \geqslant T_{BtF}+T_{FW}$
Reach the threshold N first
Event 5
Reach the threshold τ first
Event 6
The N_fth frame arrives during $T_{BtF}+T_{FW}$
$>N_f$ arrival during T_{BtF}
Event 7
N_f arrival during T_{BtF}
Event 8
The N_fth frame arrives before FW mode ends
Event 9

FIGURE 12. Arrival events of the dual-mode strategy.

a vacation, based on a tree of all possible events generated by the arrivals during the vacation time, as Fig. 12 shows. The most difficult part for the derivation of such distribution is to calculate the following distribution

$$a(n) = \Pr\{\text{there are } n \text{ arrivals during } V_{SL}\},$$

where V_{SL} is the time interval from the beginning of the vacation to the time that the FW Wakeup or the DS Wakeup is triggered. Such difficulty is mainly incurred by the analysis of Events 3 and 4. To avoid tediousness, we take Event 4 as an example to illustrate the difficulty.

In the following, we are going to derive the conditional probability defined as follows:

$$a(n|\text{Event } 4) \triangleq \Pr\{\text{there are } n \text{ arrivals during } V_{SL} \text{ given that Event 4 happens}\}.$$

To facilitate our discussion, we divide V_{SL} into three parts, as Fig. 12 shows. The first part, denoted by I_0, is the time interval from the beginning of the vacation to the time that the first frame arrives, namely the first arrival interval. It is clear that I_0 is an exponentially distributed random variable as follows

$$f(x) = \lambda e^{-\lambda x}, \quad x < T_{BtF} + T_{FW}. \tag{47}$$

The second part, denoted by Y, is the duration from the first arrival instant to the end of the FW mode, and thus,

$$Y = T_{BtF} + T_{FW} - I_0. \tag{48}$$

The last part, denoted by Z, is the period from the end of the FW mode to the end of the DS mode, and thus,

$$Z = \tau - Y = I_0 + \tau - T_{BtF} - T_{FW}. \tag{49}$$

Let n_y and n_z be the number of arrivals in the intervals Y and Z, respectively. If $N_f \leq n < N$, n_y must be less than $N_f - 1$, to guarantee that the interface enters the DS mode. In this case, we have equation (50), as shown at the bottom of the next page. If $1 \leq n \leq N_f - 1$, $n_y \leq n - 1$ is sufficient to ensure that the interface enters the DS mode. We thus have equation (51), as shown at the bottom of the next page. We can see from (50) that the thresholds T_{FW} and N_f result in intractable convolution and integral when $N_f \leq n < N$.

APPENDIX B
DERIVATION OF $h_d(n)$

Here, we employ the method in [16] to calculate $h_d(n)$, the distribution of the number of arrivals during a vacation time of the DS single-mode strategy. We divide the vacation time into two parts. The first part is the period from the beginning of the vacation to the instant that the DS Wakeup is triggered, and the second part is the duration of the DS Wakeup T_{DtB}. Let $a_d(n)$ be the distribution of the number of arrivals during the first part and $b_d(n)$ be that during the second part. The distribution $h_d(n)$ can be expressed by

$$h_d(n) = a_d(n) * b_d(n). \tag{52}$$

We derive the distribution $a_d(n)$ by analyzing the mutually exclusive events that may occur during the vacation time, as Fig. 13 shows. In particular, we consider the following four cases:

1. $n = 0$.
   The interface will not wake up if no arrival comes during the vacation time, and thus
   $$a_d(0) = 0. \tag{53}$$
2. $n = 1, 2, \cdots, N - 1$.
   As Events 4 and 6 show, when the waiting time of the first arrival frame reaches τ before N frames arrive, the DS mode is ended with less than N frames in the

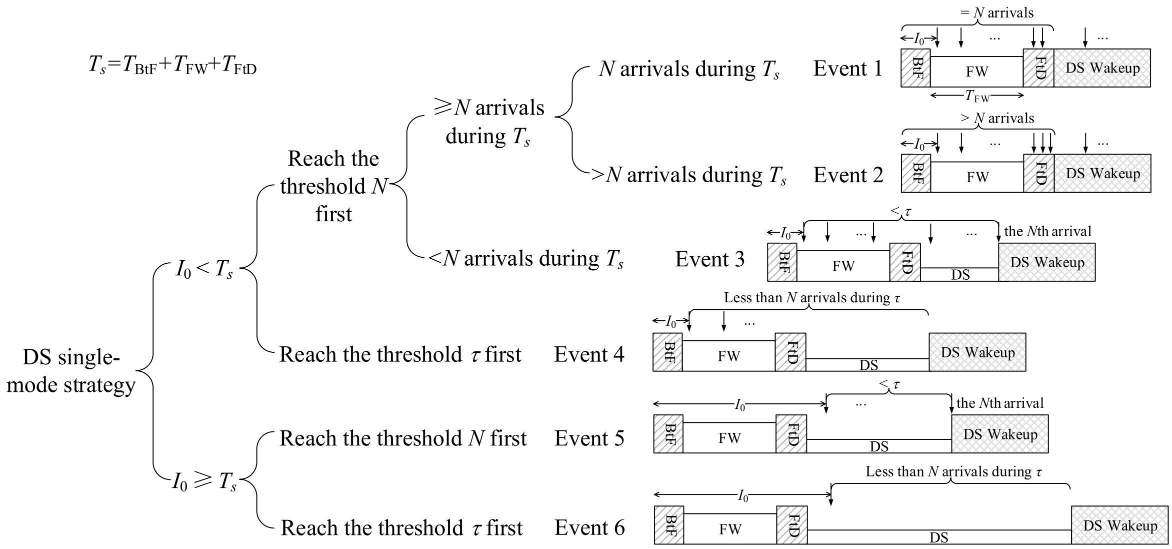

T_s=$T_{\rm BtF}$+$T_{\rm FW}$+$T_{\rm FtD}$
DS single-mode strategy
$I_0 < T_s$
$I_0 \geqslant T_s$
Reach the threshold N first
Reach the threshold τ first
$\geqslant N$ arrivals during T_s
N arrivals during T_s
$>N$ arrivals during T_s
$<N$ arrivals during T_s
Reach the threshold N first
Reach the threshold τ first
Event 1
Event 2
Event 3
Event 4
Event 5
Event 6
= N arrivals
> N arrivals
$<\tau$
the Nth arrival
Less than N arrivals during τ
I_0
$T_{\rm FW}$
BtF
FW
FtD
DS
DS Wakeup

FIGURE 13. Arrival event tree of DS single-mode strategy.

buffer. Therefore, we have

$$
\begin{aligned}
a_d(n) &= \Pr\{n-1 \text{ arrivals in interval } \tau\} \\
&= e^{-\lambda\tau}\frac{(-\lambda\tau)^{n-1}}{(n-1)!}.
\end{aligned} \tag{54}
$$

3. $n = N+1, N+2, \ldots$
This case corresponds to Event 2, where more than N frames arrive during time interval T_s. In this case, the interface wakes up immediately when the FtD terminates. Thus, we have

$$
a_d(n) = \Pr\{n \text{ arrivals in interval } T_s\} = e^{-\lambda T_s}\frac{(\lambda T_s)^n}{n!}. \tag{55}
$$

4. $n = N$.
If N frames arrive before the waiting time of the first arrival reaches time τ, the interface wakes up with exactly N frames in the buffer, as Events 1, 3, and 5 in Fig. 13 illustrate. Since $\sum_{n=0}^{\infty} a_d(n) = 1$, we can

$$
\begin{aligned}
a(n|\text{Event 4}) &= \int_0^{T_{\rm BtF}+T_{\rm FW}} \lambda e^{-\lambda x}\cdot\sum_{k=0}^{N_f-2}\Pr\{n_y = k, n_z = n-k-1|I_0 = x\}dx \\
&= \int_0^{T_{\rm BtF}+T_{\rm FW}} \lambda e^{-\lambda x}\cdot\sum_{k=0}^{N_f-2}\Pr\{n_y = k|I_0 = x\}\Pr\{n_z = n-k-1|I_0 = x\}dx \\
&= \int_0^{T_{\rm BtF}+T_{\rm FW}} \lambda e^{-\lambda x}\cdot\sum_{k=0}^{N_f-2}\Bigg\{e^{-\lambda(T_{\rm BtF}+T_{\rm FW}-x)}\frac{[\lambda(T_{\rm BtF}+T_{\rm FW}-x)]^k}{k!}\cdot \\
&\qquad e^{-\lambda(x+\tau-T_{\rm BtF}-T_{\rm FW})}\frac{[\lambda(x+\tau-T_{\rm BtF}-T_{\rm FW})]^{n-k-1}}{(n-k-1)!}\Bigg\}dx \\
&= \lambda e^{-\lambda(\tau+T_{\rm BtF}-T_{\rm FW})}\sum_{k=0}^{N_f-2}\frac{\lambda^{n-1}}{k!(n-k-1)!}\int_0^{T_{\rm BtF}+T_{\rm FW}} e^{\lambda x}\cdot x^k\cdot(\tau-x)^{n-k-1}dx
\end{aligned} \tag{50}
$$

$$
\begin{aligned}
a(n|\text{Event 4}) &= \int_0^{T_{\rm BtF}+T_{\rm FW}} \lambda e^{-\lambda x}\cdot\sum_{k=0}^{n-1}\Bigg\{e^{-\lambda(T_{\rm BtF}+T_{\rm FW}-x)}\frac{[\lambda(T_{\rm BtF}+T_{\rm FW}-x)]^k}{k!}\cdot \\
&\qquad e^{-\lambda(x+\tau-T_{\rm BtF}-T_{\rm FW})}\frac{[\lambda(x+\tau-T_{\rm BtF}-T_{\rm FW})]^{n-k-1}}{(n-k-1)!}\Bigg\}dx \\
&= \frac{(\lambda\tau)^{n-1}}{(n-1)!}\left[e^{-\lambda\tau}-e^{-\lambda(\tau+T_{\rm BtF}+T_{\rm FW})}\right]
\end{aligned} \tag{51}
$$

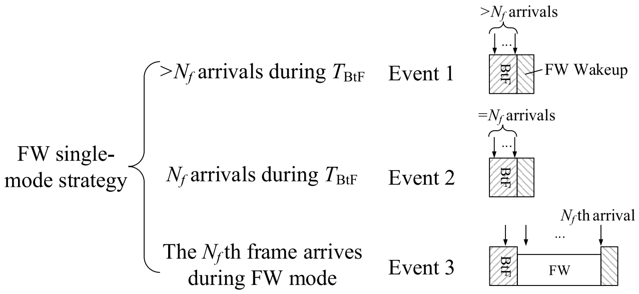

FW single-mode strategy
>N_f arrivals during T_{BtF}
Event 1
>N_f arrivals
...
BtF
FW Wakeup
N_f arrivals during T_{BtF}
Event 2
=N_f arrivals
...
BtF
The N_fth frame arrives during FW mode
Event 3
N_fth arrival
...
BtF
FW

FIGURE 14. **Arrival event tree of FW single-mode strategy.**

obtain $a_d(n)$ for $n = N$ as

$$a_d(N) \ = 1 - \sum_{n=0}^{N-1} a_d(n) - \sum_{n=N+1}^{\infty} a_d(n). \\ = \sum_{n=0}^{N} e^{-\lambda T_s} \frac{(\lambda T_s)^n}{n!} - \sum_{n=1}^{N-1} e^{-\lambda \tau} \frac{(\lambda \tau)^n}{n!} \tag{56}$$

Since the duration of the DS Wakeup T_{DtB} is a constant, $b_d(n)$ is given by

$$b_d(n) = e^{-\lambda T_{\mathrm{DtB}}} \frac{(\lambda T_{\mathrm{DtB}})^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{57}$$

Substituting the expressions of $a_d(n)$ and $b_d(n)$ into (52), we can obtain the distribution $h_d(n)$.

APPENDIX C
DERIVATION OF $h_f(n)$

In this part, we derive $h_f(n)$, the distribution of the number of arrivals during a vacation of the FW single-mode strategy. We divide the vacation time of the FW single-mode strategy into two parts. The first part is from the beginning of the vacation time to the instant when the FW Wakeup is triggered. The second part is the duration of the FW Wakeup T_{FtB}. Let $a_f(n)$ be the distribution of the number of arrivals during the first part and $b_f(n)$ be that during the second part. The distribution $h_f(n)$ can be expressed by

$$h_f(n) = a_f(n) * b_f(n). \tag{58}$$

We first derive the distribution $a_f(n)$ based on the event tree in Fig. 14. We consider the following three cases:

1. $n = 0, 1, \cdots, N_f - 1$.
   The interface does not wake up until N_f frames are accumulated in the buffer. We have

$$a_f(n) = 0. \tag{59}$$

2. $n = N_f + 1, N_f + 2, \ldots$
   If there are more than N_f arrivals during transition time T_{BtF}, as Event 1 in Fig. 14 shows, the interface wakes up immediately when the BtF terminates. Therefore, we have

$$a_f(n) = e^{-\lambda T_{\mathrm{BtF}}} \frac{(\lambda T_{\mathrm{BtF}})^n}{n!}. \tag{60}$$

3. $n = N_f$.
   In Events 2 and 3, the interface wakes up from the FW mode with exactly N_f arrivals in the buffer. In Event 2, exact N_f frames arrive during the BtF, and the interface wakes up immediately when the BtF ends. In Event 3, there are less than N_f arrivals during the BtF, and the N_fth frame arrives and triggers the FW Wakeup after the interface enters the FW mode. Accordingly, we can obtain $a_f(n)$ for $n = N_f$ as follows

$$a_f(N_f) \ = \Pr\{\text{exact } N_f \text{ arrivals during } T_{\mathrm{BtF}}\} \\ + \Pr\{\text{less than } N_f \text{ arrivals during } T_{\mathrm{BtF}}\} \\ = \sum_{n=0}^{N_f} e^{-\lambda T_{\mathrm{BtF}}} \frac{(\lambda T_{\mathrm{BtF}})^n}{n!}. \tag{61}$$

As the duration of the FW Wakeup T_{BtF} is a constant, the distribution of the number of arrivals during T_{BtF} is given by

$$b_f(n) = e^{-\lambda T_{\mathrm{BtF}}} \frac{(\lambda T_{\mathrm{BtF}})^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{62}$$

Finally, substituting the expressions of $a_f(n)$ and $b_f(n)$ into (58) yields the distribution $h_f(n)$.

REFERENCES

[1] M. Kitamura, D. Shirai, K. Kaneko, T. Murooka, T. Sawabe, T. Fujii, and A. Takahara, "Beyond 4K: 8K 60p live video streaming to multiple sites," *Future Gener. Comput. Syst.*, vol. 27, no. 7, pp. 952–959, Jul. 2011.

[2] R. R. Expósito, G. L. Taboada, S. Ramos, J. Touriño, and R. Doallo, "Performance analysis of HPC applications in the cloud," *Future Gener. Comput. Syst.*, vol. 29, no. 1, pp. 218–229, Jan. 2013.

[3] H. Frazier, *Evolution from 10G to 40G & 100G*, Standard 802.3ba, IEEE 802.3 Higher Speed Study Group, May 2007.

[4] H. Barrass, *Energy Efficient Ethernet*, Standard 802.3az, IEEE 802 LAN/MAN Standards Committee, Jul. 2007.

[5] B. Nordman, *EEE Savings Estimates*, Standard 802.3az, IEEE Energy Efficient Ethernet Study Group, May 2007.

[6] D. Meisner, B. T. Gold, and T. F. Wenisch, "PowerNap: Eliminating server idle power," *SIGARCH Comput. Archit. News*, vol. 37, pp. 205–216, Mar. 2009.

[7] *IEEE Standard for Information Technology-Local and Metropolitan Area Networks-Specific Requirements-Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment 5: Media Access Control Parameters, Physical Layers, and Management Parameters for Energy-Efficient Ethernet*, IEEE Standard 802.3az-2010, Oct. 2010, pp. 1–302.

[8] H. Barrass, *Options for EEE in 100G*, Standard 802.3bj, 100 Gb/s Backplane and Copper Cable Task Force, Jan. 2012.

[9] M. Gustlin and H. Barrass, *EEE support for 100 Gb/s*, Standard 802.3bj, 100 Gb/s Backplane and Copper Cable Task Force, Jan. 2012.

[10] *IEEE Standard for Ethernet Amendment 2: Physical Layer Specifications and Management Parameters for 100 Gb/s Operation Over Backplanes and Copper Cables*, IEEE Standard 802.3bj-2014, Sep. 2014, pp. 1–368.

[11] M. Mostowfi, "A simulation study of energy-efficient Ethernet with two modes of low-power operation," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1702–1705, Oct. 2015.

[12] M. Mostowfi, "Packet coalescing for dual-mode energy efficient Ethernet: A simulation study," in *Proc. 8th EAI Int. Conf. Simulation Tools Techn.*, 2015, pp. 335–342.

[13] K. P. Saravanan, P. M. Carpente, and A. Ramirez, "Exploring multiple sleep modes in on/off based energy efficient HPC networks," in *Proc. 33rd IEEE Int. Conf. Comput. Design (ICCD)*, Oct. 2015, pp. 54–61.

[14] K. P. Saravanan and P. M. Carpenter, "PerfBound: Conserving energy with bounded overheads in On/Off-based HPC interconnects," *IEEE Trans. Comput.*, vol. 67, no. 7, pp. 960–974, Jul. 2018.

[15] S. Herrería-Alonso, M. Rodríguez-Pérez, M. Fernández-Veiga, and C. López-García, "Frame coalescing in dual-mode EEE," 2015, *arXiv:1510.03694*. [Online]. Available: http://arxiv.org/abs/1510.03694

[16] X. Pan, T. Ye, T. T. Lee, and W. Hu, "Power efficiency and delay tradeoff of 10GBase-T energy efficient Ethernet protocol," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 2773–2787, Oct. 2017.

[17] M. Mostowfi and K. Christensen, "An energy-delay model for a packet coalescer," in *Proc. IEEE Southeastcon*, Mar. 2012, pp. 1–6.

[18] N. Akar, "Delay analysis of timer-based frame coalescing in energy efficient Ethernet," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1459–1462, Jul. 2013.

[19] S. Herrería-Alonso, M. Rodríguez-Pérez, M. Fernández-Veiga, and C. López-García, "How efficient is energy-efficient Ethernet?" in *Proc. 3rd Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2011, pp. 1–7.

[20] S. Herreria-Alonso, M. Rodriguez-Perez, M. Fernandez-Veiga, and C. Lopez-Garcia, "A power saving model for burst transmission in energy-efficient Ethernet," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 584–586, May 2011.

[21] S. Herrería-Alonso, M. Rodríguez-Pérez, M. Fernández-Veiga, and C. López-García, "Optimal configuration of energy-efficient Ethernet," *Comput. Netw.*, vol. 56, no. 10, pp. 2456–2467, Jul. 2012.

[22] S. Herreria-Alonso, M. Rodriguez-Perez, M. Fernandez-Veiga, and C. Lopez-Garcia, "A GI/G/1 model for 10 Gb/s energy efficient Ethernet links," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3386–3395, Nov. 2012.

[23] K. J. Kim, S. Jin, N. Tian, and B. D. Choi, "Mathematical analysis of burst transmission scheme for IEEE 802.3az energy efficient Ethernet," *Perform. Eval.*, vol. 70, no. 5, pp. 350–363, May 2013.

[24] M. Mostowfi and K. Shafie, "An analytical model for the power consumption of dual-mode EEE," *Electron. Lett.*, vol. 52, no. 15, pp. 1308–1310, Jul. 2016.

[25] M. Mostowfi and K. Shafie, "Average packet delay in dual-mode EEE: An analytical model," *Electron. Lett.*, vol. 52, no. 21, pp. 1759–1761, Oct. 2016.

[26] M. A. Marsan, A. F. Anta, V. Mancuso, B. Rengarajan, P. R. Vasallo, and G. Rizzo, "A simple analytical model for energy efficient Ethernet," *IEEE Commun. Lett.*, vol. 15, no. 7, pp. 773–775, Jul. 2011.

[27] K. T. Marshall, "Bounds for some generalizations of the GI/G/1 Queue," *Oper. Res.*, vol. 16, pp. 841–848, Aug. 1968.

[28] K. Watkins, *Discrete Event Simulation in C*. New York, NY, USA: McGraw-Hill, 1993.

[29] T. Benson, A. Akella, and D. A. Maltz, "Network traffic characteristics of data centers in the wild," in *Proc. 10th Annu. Conf. Internet Meas. (IMC)*, 2010, pp. 267–280.

[30] M. Gupta and S. Singh, "Using low-power modes for energy conservation in Ethernet LANs," in *Proc. IEEE 26th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, May 2007, pp. 2451–2455.

[31] W. Fischer and K. Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook," *Perform. Eval.*, vol. 18, no. 2, pp. 149–171, Sep. 1993.

[32] M. H. van Hoorn and L. P. Seelen, "The SPP/G/1 queue: A single server queue with a switched Poisson process as input process," *Oper.-Res.-Spektrum*, vol. 5, no. 4, pp. 207–218, Dec. 1983.

[33] Y. Zhao, B. Zhang, C. Li, and C. Chen, "ON/OFF traffic shaping in the Internet: Motivation, challenges, and solutions," *IEEE Netw.*, vol. 31, no. 2, pp. 48–57, Mar. 2017.

[34] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," in *Proc. 1st ACM Workshop Res. Enterprise Netw. (WREN)*, 2009, pp. 65–72.

[35] K. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot, "Long-term forecasting of Internet backbone traffic," *IEEE Trans. Neural Netw.*, vol. 16, no. 5, pp. 1110–1124, Sep. 2005.

XIAODAN PAN received the B.S. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2014. She is currently pursuing the Ph.D. degree with the State Key Laboratory of Advanced Optical Communication Systems and Network, Shanghai Jiao Tong University, Shanghai, China. Her major research interest is energy efficient Ethernet.

TONG YE (Member, IEEE) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1998 and 2001, respectively, and the Ph.D. degree in electronics engineering from Shanghai Jiao Tong University, Shanghai, China, in 2005.

He was with The Chinese University of Hong Kong for one and a half year as a Postdoctoral Research Fellow. He is currently an Associate Professor with the State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University. His research interests include the design of optical network architectures, optical network systems and subsystems, and silicon-ring-based optical signal processing.

TONY T. LEE received the B.S.E.E. degree from National Cheng Kung University, Taiwan, and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic Institute, NYU, Brooklyn, NY, USA.

He was with AT&T Bell Laboratories, Holmdel, NJ, USA, from 1977 to 1983. He was with Telcordia Technologies, Morristown, NJ, USA, from 1983 to 1993. From 1991 to 1993, he was a Professor of electrical engineering with the Polytechnic Institute, NYU. From 1993 to 2013, he was a Chair Professor with the Information Engineering Department, The Chinese University of Hong Kong. From 2013 to 2018, he was a Zhiyuan Chair Professor with the Electronics Engineering Department, Shanghai Jiao Tong University, and an Emeritus Professor of information engineering with The Chinese University of Hong Kong. He is currently a Professor with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen).

Dr. Lee is a Fellow of the HKIE. He received many awards, including the 1989 Leonard G. Abraham Prize Paper Award from the IEEE Communication Society, the 1999 Outstanding Paper Award from the IEICE of Japan, and the 1999 National Natural Science Award from China. He has served as an Editor for the IEEE Transactions on Communications and an Area Editor for the *Journal of Communication Network*.