

Delay Analysis of Timer-Based Frame Coalescing in Energy Efficient Ethernet

Nail Akar, *Member, IEEE*

Abstract—IEEE 802.3az, also known as Energy Efficient Ethernet (EEE), aims at reducing the energy consumption of an Ethernet link by placing it in sleep mode when the link is idle. Frame coalescing mechanism proposed for EEE is an effective means to increase the average idle time of the link, thus reducing the overhead stemming from sleep/wake transitions, but at the expense of increased frame delays. Therefore, it is imperative to quantify the energy-delay trade-off while employing frame coalescing. As opposed to existing delay models that focus only on the average delays, a simple but exact queuing model is introduced for timer-based frame coalescing to find the delay distribution when the frame arrival process is Poisson and frame lengths are generally distributed. An expression for average saving in power consumption is also provided.

Index Terms—Energy efficient Ethernet, frame coalescing, queuing model, frame delays.

I. INTRODUCTION

IEEE 802.3az Energy Efficient Ethernet (EEE) is a recent standard with the goal of reducing power consumption of Ethernet physical interfaces during periods of low link utilization via the use of the Low Power Idle (LPI) mechanism [1]. In the LPI mechanism illustrated in Fig. 1, Ethernet frames are transmitted in the Active (A) state and when no further frames are available, a potential exists to enter the Low power (L) state during which no frames will be transmitted. However, a transition from A to L requires an additional Sleep (S) state with a duration of T_S seconds. Moreover, in the L state, short refresh signals of duration T_r are sent with a period of T_q seconds with the goal of keeping the link alive and aligning the receiver with the instantaneous link condition. When frames become available for transmission, a further Wake (W) state is required of duration T_W seconds for the link to wake up and eventually enter the A state again.

When the link is in the L state, a decision is to be made on when to exit that state. In the frame transmission algorithm described in [2] and [3], the W state is entered immediately when a new Ethernet frame arrives. Let T_A and T_L denote the random variables representing the duration of the A and L states, respectively. Power efficiency of EEE, as will be shown later, is governed by the ratio $R = \frac{E[T_L]}{E[T_C]}$ where the cycle time T_C is defined as $T_C = T_L + T_A + T_W + T_S$. With the frame transmission algorithm in place, the link would spend a considerable proportion of time in the S and W states,

Manuscript received January 31, 2013. The associate editor coordinating the review of this letter and approving it for publication was G. Reali.

N. Akar is with the Electrical and Electronics Engineering Department, Bilkent University, Bilkent 06800, Ankara, Turkey (e-mail: akar@ee.bilkent.edu.tr).

This work is supported in part by the Science and Research Council of Turkey (Tübitak) under project grant EEEAG-111E106.

Digital Object Identifier 10.1109/LCOMM.2013.051313.130245

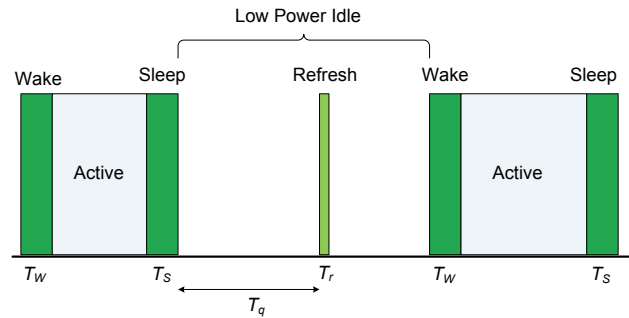


Fig. 1. The operation of an EEE interface supporting the LPI mechanism.

limiting the power efficiency. On the upside, the additional frame delays incurred by EEE are minimal with the frame transmission algorithm.

The burst transmission algorithm, also known as frame coalescing, has been proposed as a means of increasing the ratio R by increasing $E[T_A]$ and $E[T_L]$ relative to T_W and T_S [1],[4]. For this purpose, as opposed to exiting the L state immediately upon a single new frame arrival, a number of frames are allowed to accumulate (or coalesce) at a coalescing buffer when in the L state. As a result, improved power efficiency is shown to be possible but at the expense of increased frame delays [5]. Three types of coalescers may be defined based on when a decision is made to exit the L state. In a timer-based coalescer, upon the arrival of the first frame (say frame f) following departure from the A state (this occurs in either the S or L states), a timer is started which expires after $T - T_W$ seconds. When the timer expires, the link enters the W state. Let us assume that the frame f arrives just at the beginning of the sleep state S. After a duration $T - T_W$, the system should transition to the W state. For this transition to be possible, the S state should be over by then, i.e., $T - T_W \geq T_S$. Therefore, throughout the letter, we assume $T \geq T_W + T_S$ which ensures that the service of the frame f will start exactly T seconds after its arrival. In a frame-based coalescer, N_{max} frames are allowed to accumulate upon the arrival of the first frame after which the W state is entered. Similarly, by taking into account the frame sizes in the coalescing buffer, B_{max} bytes are allowed to accumulate in size-based coalescers during the L state before transitioning to the W state. Hybrid coalescers use a combination of above. For example, the mechanisms proposed in [1] and [5] employ a coalescer in which the W state is entered when either a timer expires or a certain number of frames are accumulated in the buffer. Timer-based coalescers control the ratio R directly and they require the adjustment of one single parameter T . The current letter's scope is limited to the analysis of the delay

distribution of timer-based coalescers only, and more general coalescers including hybrid ones, are left for future research.

A rigorous analysis of the frame transmission algorithm is presented in [3]. In this work, an $M/G/1$ queuing model is proposed and validated to accurately predict the time that an EEE link spends in each of the four possible states so as to characterize the gain in power consumption. A frame-based coalescer is analytically studied in [6] for both energy and delay but with deterministic interarrival and service times. A simulation study for Poisson and bursty traffic accompanies the analytical model of [6]. A power-only model is given for a hybrid frame-coalescer for Poisson frame arrivals and generally distributed frame lengths in [7]. A more general approximative analytical model for a hybrid coalescer has recently been proposed in [5] for both energy and delay with independent and generally distributed frame interarrival times and generally distributed frame lengths. Despite the generality of the model, the focus is on the average frame delay and not on its distribution. However, delay sensitive applications have Quality of Service (QoS) requirements not only in terms of average delay but also in terms of the tail of the delay distribution. In the current letter, we provide a queuing model to find the frame delay distribution exactly when a timer-based coalescer is used for the special case of Poisson arrivals and generally distributed frame lengths.

The remainder is organized as follows. The queuing model of the timer-based coalescer is presented in Section 2. Section 3 presents expressions on average power saving relative to ordinary Ethernet. Numerical examples are provided in Section 4. The final section concludes.

II. QUEUING MODEL

The frame arrival process is Poisson with rate λ . Let A_n denote the inter-arrival time between the frames n and $n+1$. The Probability Density Function (PDF) for $A_n, n \geq 0$ is denoted by $f_A(x) = \lambda e^{-\lambda x}, x \geq 0$. Let B_n denote the service time requirement of the frames with PDF $f_B(x)$ and mean $E[B] = 1/\mu$. Let $g_A(s) = E[e^{-sA}] = \frac{\lambda}{s+\lambda}$ and $g_B(s) = E[e^{-sB}]$. Note that $g_A(s)$ and $g_B(s)$ are the Laplace transforms of $f_A(x)$ and $f_B(x)$, respectively. Incoming frames are assumed to join a coalescing buffer of FIFO type with load denoted by $\rho = \lambda/\mu < 1$, i.e., the queue is stable. Illustration of the sample path of the queuing system associated with timer-based frame coalescing is presented in Fig. 2 in which the unfinished work is plotted as a function of time for an example scenario. When the buffer becomes empty, the link transitions to the L state via the S state. The first frame arrival in the S or L states (frames with duration B_1 or B_7 in Fig. 2) triggers the timer which expires after $T - T_W$ seconds at which the link transitions to the W state. Following timer expiration, the link stays in the W state for T_W seconds. The buffer is not drained until the transmission of this first frame starts which occurs T seconds after its arrival. The buffer is then served until it is completely drained. This process then repeats. Let W_n denote the delay (or waiting time) of the n th frame in the coalescing buffer. From Fig. 2, it is clear that the following

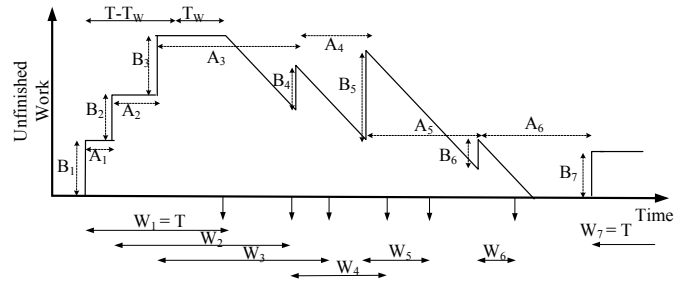


Fig. 2. Illustration of the sample path of the queuing system associated with timer-based frame coalescing.

Lindley equation holds:

$$W_{n+1} = \begin{cases} W_n + B_n - A_n, & \text{if } W_n + B_n - A_n > 0 \\ T, & \text{if } W_n + B_n - A_n \leq 0 \end{cases} \quad (1)$$

In this letter, we study the steady-state distribution of $W \stackrel{d}{=} \lim_{n \rightarrow \infty} W_n$, which is known to exist since the queue is stable. Here, the notation $\stackrel{d}{=}$ refers to equality in distribution. Let $g_W(s) = E[e^{-sW}]$. Based on the results of [8] on Lindley equations, there exists a constant $k > 0$ such that the following algebraic equation holds for $g_W(s)$:

$$g_W(s) = g_W(s)g_B(s)g_A(-s) - \frac{k}{\lambda - s} + \frac{k}{\lambda}e^{-sT} \quad (2)$$

To explain, the random variable $H = W + B - A$ has a two-sided PDF. However, W is right-sided and is obtained by taking out the Laplace transform of the left side of the PDF of H and making up for it by placing a probability mass at T . The identity in (2) is then the transform domain counterpart of the equation (1) as $n \rightarrow \infty$. Solving for k from $g_W(0) = 1$ in (2), we obtain

$$g_W(s) = g_{W_{M/G/1}}(s)g_C(s), \quad (3)$$

where

$$g_{W_{M/G/1}}(s) = \frac{s(1-\rho)}{g_B(s)\lambda - \lambda + s} \quad (4)$$

is the transform of the waiting time $W_{M/G/1}$ in the corresponding $M/G/1$ queue [9], and the second multiplicative term $g_C(s)$ arises due to timer-based frame coalescing:

$$g_C(s) = \frac{1}{1+T\lambda}e^{-sT} + \frac{T\lambda}{1+T\lambda} \frac{1-e^{-sT}}{sT}. \quad (5)$$

The random variable C for which $g_C(s) = E[e^{-sC}]$ is a deterministic quantity T with probability $\alpha = \frac{1}{1+T\lambda}$ and is a uniform random variable in the interval $(0, T)$ with probability $1 - \alpha$. Consequently, the equation (3) provides a complete characterization of the queue waiting time for an Ethernet link employing timer-based frame coalescing. Another way of interpreting the equation (3) is that W turns out to be the sum of two independent random variables $W_{M/G/1}$ and C , the latter characterizing the additional delay stemming from frame coalescing. For the provisioning of EEE links, it is imperative to calculate the Complementary Cumulative

Distribution Function (CCDF) $P(W > t)$ for a given $t > 0$. Using (3), one can easily write for $t \geq T$ the following identity

$$P(W > t) = \alpha P(W_{M/G/1} > t - T) + \frac{(1 - \alpha)}{T} \int_0^T P(W_{M/G/1} > t - \tau) d\tau. \quad (6)$$

If $t < T$, then $P(W > t)$ equals

$$\alpha + (1 - \alpha) \frac{T - t}{T} + \frac{(1 - \alpha)}{T} \int_0^t P(W_{M/G/1} > t - \tau) d\tau. \quad (7)$$

Note that $P(W_{M/G/1} > t)$ can be numerically obtained either by Laplace transform inversion as in [10] or using matrix analytical techniques as in [8], the latter for service time distributions with rational Laplace transforms only. On the other hand, the mean waiting time $E[W] =$

$$E[W_{M/G/1}] + E[C] = \frac{\lambda E[B^2]}{2(1 - \rho)} + \frac{T(2 + T\lambda)}{2(1 + T\lambda)}, \quad (8)$$

where the first term stems from the Pollaczek-Khinchin formula (PK formula) for the mean waiting time in an $M/G/1$ queue [9]. When B_n is exponentially distributed, i.e., $f_B(x) = \mu e^{-\mu x}$, $x \geq 0$, $P(W_{M/G/1} > t) = \rho e^{-\mu(1-\rho)t}$, $t > 0$ [9], and a closed-form expression can be obtained for $P(W > t) =$

$$\begin{cases} \frac{\alpha \rho}{1 - \rho} e^{-\mu(1-\rho)(t-T)} - \frac{\alpha \rho^2}{1 - \rho} e^{-\mu(1-\rho)t}, & t \geq T, \\ \frac{\alpha(1-\rho+\rho^2)}{1 - \rho} + \frac{(1-\alpha)(T-t)}{T} - \frac{\alpha \rho^2}{1 - \rho} e^{-\mu(1-\rho)t}, & t < T. \end{cases} \quad (9)$$

III. ANALYSIS OF POWER CONSUMPTION

In this section, expressions will be provided for the average power consumption of EEE links employing timer-based frame coalescing on the basis of the results provided in [1],[3],[5],[7]. The first step is to find the mean low power state length $E[T_L]$ given T , λ , and ρ . Consider the first frame arriving once a transition from the A state to S state takes place. There are two possibilities; either the first frame arrives while in the S state which occurs with probability $1 - e^{-\lambda T_S}$ or during the L state with probability $e^{-\lambda T_S}$. In the first case, $T_L = T - T_W - D$ where D is uniformly distributed in $(0, T_S)$. In the second case, $T_L = T - T_W + U$ where U is the time needed until one arrival with $E[U] = 1/\lambda$. Consequently, we have

$$E[T_L] = T - T_W + \frac{e^{-\lambda T_S}}{\lambda} - \frac{(1 - e^{-\lambda T_S})T_S}{2}. \quad (10)$$

Since no transmissions are possible in the S , L , and W states, we have $\frac{E[T_L] + T_S + T_W}{E[T_C]} = 1 - \rho$, from which we show that

$$R = \frac{E[T_L]}{E[T_C]} = \frac{E[T_L](1 - \rho)}{E[T_L] + T_S + T_W}. \quad (11)$$

Let P_A , P_L , P_S , and P_W denote the power consumption in the A , L , S , and W states, respectively. Note that P_A yields the average power consumption in a conventional Ethernet interface which is always in the A state. Moreover, it is generally true that $P_A = P_S = P_W$ [7]. Also let P_{EEE} denote the average power consumption in an EEE interface employing timer-based frame coalescing. The average percentage power

saving S in using EEE with timer-based frame coalescing relative to ordinary Ethernet can then be written as follows:

$$S = \frac{P_A - P_{EEE}}{P_A} \times 100 = R \left(1 - \frac{P_L}{P_A}\right) \times 100 \quad (12)$$

The above identity provides a closed-form expression for the reduction in power consumption through the use of the identities (10) and (11).

IV. NUMERICAL EXAMPLE

We first consider a 10 Gbps Ethernet interface and address the problem of finding the optimum timer parameter T and the subsequent power saving by the proposed $M/G/1$ analysis method. We use the parameters used in [1] and [7] based on estimates provided by various manufacturers during the standardization process of EEE. In particular, the parameters T_S , T_W , and $\frac{P_L}{P_A}$ are set to $2.88 \mu s$, $4.48 \mu s$ and 0.1 , respectively. Frame arrival process is assumed to be Poisson and frame service times are assumed to be exponentially distributed due to the existence of the closed form expression (9) for the waiting time CCDF in this particular case. The mean frame length is set to 759.82 Bytes based on a recent measurement of the first 5×10^6 frames taken on the first working day of 2013 at 14:00 by MAWI (Measurement and Analysis on the WIDE Internet) at samplepoint-F (a trans-Pacific link) [11]. Consequently, the mean service time $\mu^{-1} = 0.607856 \mu s$. In this letter, the delay constraint is given in terms of $P(W > W_0) < p_0$ for two QoS parameters W_0 and p_0 . We fix $p_0 = 10^{-3}$ and for various values W_0 and load parameter ρ , we find the maximum value of $T \geq T_S + T_W$ denoted by T^* that satisfies the QoS requirement $P(W > W_0) < p_0$. If there is no such T , we set T^* to zero in which case we say frame coalescing is turned off and the corresponding average power saving becomes zero. We also find the average percentage power saving S^* when the timer parameter T is set to T^* . The results are illustrated in Fig. 3 for varying load ρ . We have the following observations: The optimum value of the timer parameter T^* depends on the load ρ and decreases as ρ increases. For low values of ρ , T^* is slightly less than W_0 since the waiting time is then dominated by frame coalescing and not by the $M/G/1$ component. Beyond a certain value of ρ with fixed W_0 , frame coalescing can not be turned on, i.e., $T^* = 0$. For example, for the case of $W_0 = 100 \mu s$, $p_0 = 10^{-3}$ and when $\rho > 0.95826$, it is already true that $P(W_{M/G/1} > W_0) > p_0$ and the QoS requirement can not be met even without frame coalescing. When $0.95647 < \rho \leq 0.95826$, $P(W_{M/G/1} > W_0) < p_0$ but a timer value $T \geq T_S + T_W$ can not be found such that $P(W > W_0) < p_0$. Therefore, when $\rho > 0.95647$ and under the above-mentioned delay constraint, timer-based frame coalescing can not be turned on and power saving can not be achieved. Moreover, the optimum power saving appears to saturate when $W_0 > 200 \mu s$. This leads us to the conclusion that setting the timer parameter T to a value beyond $200 \mu s$ will have little effect on power saving while it would be detrimental in terms of delay. This observation is in line with identity (11).

It is also interesting to compare the theoretical results of this paper with simulation results obtained by using real traffic patterns which are known to be non-Poisson. For this

TABLE I

THE ADDITIONAL MEAN DELAY AND ADDITIONAL DELAY VARIANCE INCURRED BY TIMER-BASED FRAME COALESCING FOR THE MAWI TRACE COMPARED WITH $E[C]$ (IN SEC.) AND $Var(C)$ (IN SEC^2), RESPECTIVELY, FOR TWO VALUES OF ρ AND T , AND FOR THREE VALUES OF THE LINK SPEED.

ρ	T	100 Mbps		1 Gbps		10 Gbps	
		$E[C]$ ($Var(C)$)	Simulation	$E[C]$ ($Var(C)$)	Simulation	$E[C]$ ($Var(C)$)	Simulation
0.3	200 μs	$1.5033 \cdot 10^{-4}$ ($4.1557 \cdot 10^{-9}$)	$1.7737 \cdot 10^{-4}$ ($9.7768 \cdot 10^{-10}$)	$1.0920 \cdot 10^{-4}$ ($3.8620 \cdot 10^{-9}$)	$1.1670 \cdot 10^{-4}$ ($4.0360 \cdot 10^{-9}$)	$1.0100 \cdot 10^{-4}$ ($3.3992 \cdot 10^{-9}$)	$1.0170 \cdot 10^{-4}$ ($3.4261 \cdot 10^{-9}$)
	1 ms	$5.8424 \cdot 10^{-4}$ ($1.0432 \cdot 10^{-7}$)	$5.8052 \cdot 10^{-4}$ ($1.3596 \cdot 10^{-7}$)	$5.0993 \cdot 10^{-4}$ ($8.6545 \cdot 10^{-8}$)	$5.1842 \cdot 10^{-4}$ ($8.8308 \cdot 10^{-8}$)	$5.0101 \cdot 10^{-4}$ ($8.3669 \cdot 10^{-8}$)	$5.0125 \cdot 10^{-4}$ ($8.3401 \cdot 10^{-8}$)
0.6	200 μs	$1.3362 \cdot 10^{-4}$ ($4.4444 \cdot 10^{-9}$)	$1.8896 \cdot 10^{-4}$ ($2.7364 \cdot 10^{-10}$)	$1.0482 \cdot 10^{-4}$ ($3.6315 \cdot 10^{-9}$)	$1.0768 \cdot 10^{-4}$ ($3.8716 \cdot 10^{-9}$)	$1.0050 \cdot 10^{-4}$ ($3.3667 \cdot 10^{-9}$)	$1.0100 \cdot 10^{-4}$ ($2.9019 \cdot 10^{-9}$)
	1 ms	$5.4599 \cdot 10^{-4}$ ($9.6549 \cdot 10^{-8}$)	$5.6833 \cdot 10^{-4}$ ($1.0561 \cdot 10^{-7}$)	$5.0501 \cdot 10^{-4}$ ($8.4980 \cdot 10^{-8}$)	$5.0923 \cdot 10^{-4}$ ($8.6065 \cdot 10^{-8}$)	$5.0051 \cdot 10^{-4}$ ($8.3502 \cdot 10^{-8}$)	$4.9927 \cdot 10^{-4}$ ($8.2328 \cdot 10^{-8}$)

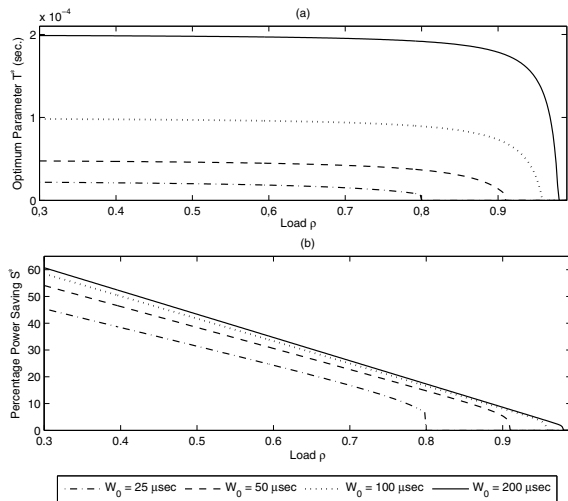


Fig. 3. (a) T^* , the maximum value of the timer parameter T which satisfies the QoS requirement $P(W > W_0) < 10^{-3}$ as a function of ρ for various values of W_0 . (b) Average percentage power saving S^* when the timer parameter T is set to T^* .

purpose, we compare the expected value and the variance of the additional delay incurred by frame coalescing using a real traffic trace, with theoretical results. Note that the additional delay for Poisson arrivals is characterized with the random variable C with mean $E[C] = \frac{T(2+T\lambda)}{2(1+T\lambda)}$ and variance $Var(C) = \frac{T^2(1+2\alpha-3\alpha^2)}{12}$. We used the same MAWI trace mentioned above but scaled the interarrival times to obtain a traffic stream with a given load parameter ρ . We then simulated the Lindley equation (1) for 5×10^6 frame arrivals with MATLAB; first without frame coalescing and then with coalescing. We then compared the additional mean delay and additional delay variance incurred by frame coalescing using simulations with $E[C]$ and $Var(C)$, respectively, to see if the theoretical results may also be used for real traffic traces. Our findings are tabulated in Table IV for two values of T , two values of ρ for which frame coalescing would make sense, and for three different link speeds, namely 100 Mbps, 1 Gbps, and 10 Gbps EEE links. We observe that $E[C]$ and $Var(C)$ provide acceptable approximations to the expected

value and variance of the additional delay incurred by frame coalescing with the accuracy of the approximation improving with increasing link speed and/or timer parameter T .

V. CONCLUSIONS

An exact queuing model for timer-based frame coalescing in EEE links is presented for Poisson Ethernet frame arrivals and generally distributed frame lengths. Future work will consist of more general frame coalescers including frame-based, size-based, and hybrid coalescers, and for more general traffic patterns. A link carrying both commercial voice and best-effort traffic and its power management is another research direction.

REFERENCES

- [1] K. Christensen, P. Reviriego, B. Nordman, M. Bennett, M. Mostowfi, and J. Maestro, "IEEE 802.3az: the road to Energy Efficient Ethernet," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 50–56, Nov. 2010.
- [2] P. Reviriego, J. Hernandez, D. Larrabeti, and J. Maestro, "Performance evaluation of Energy Efficient Ethernet," *IEEE Commun. Lett.*, vol. 13, no. 9, pp. 697–699, Sep. 2009.
- [3] M. Marsan, A. Anta, V. Mancuso, B. Rengarajan, P. Vasallo, and G. Rizzo, "A simple analytical model for Energy Efficient Ethernet," *IEEE Commun. Lett.*, vol. 15, no. 7, pp. 773–775, July 2011.
- [4] P. Reviriego, J. Maestro, J. Hernandez, and D. Larrabeti, "Burst transmission for Energy-Efficient Ethernet," *IEEE Internet Computing*, vol. 14, no. 4, pp. 50–57, July-Aug. 2010.
- [5] 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.
- [6] M. Mostowfi and K. Christensen, "An energy-delay model for a packet coalescer," in *Proc. 2012 IEEE Southeastcon*, pp. 1–6.
- [7] 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.
- [8] N. Akar and K. Sohraby, "System-theoretical algorithmic solution to waiting times in semi-Markov queues," *Performance Evaluation*, vol. 66, no. 11, pp. 587–606, 2009.
- [9] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd edition. Wiley, 1998.
- [10] J. Abate and W. Whitt, "The Fourier-series method for inverting transforms of probability distributions," *Queueing Syst. Theory Appl.*, vol. 10, no. 1-2, pp. 5–88, Jan. 1992.
- [11] K. Cho, K. Mitsuya, and A. Kato, "Traffic data repository at the WIDE project," in *Proc. USENIX 2000 Annual Technical Conference: FREENIX Track*, pp. 263-270. Traces available at <http://mawi.wide.ad.jp/mawi>.