On the Performance Analysis of Energy-Efficient Upstream Scheduling for Hybrid Fiber-Coaxial Networks with Channel Bonding

Ping Lu, Yabo Yuan, Zhongzhen Yang, and Zuqing Zhu, *Senior Member, IEEE*

***Abstract*—We develop a novel QoS-aware energy-efficient upstream traffic scheduling algorithm for channel-bonding cable modems (CMs) in hybrid fiber-coaxial (HFC) networks. Based on their QoS requirements, the algorithm sends incoming packets to different priority queues and controls channel-bonding transmitters (TXs) to forward them. Based on a vacation M/M/n queuing model, we analyze the algorithm's performance with a three-dimensional Markov chain and derive the analytical expressions of packet delay and power consumption.**

***Index Terms*—Hybrid fiber-coaxial (HFC) networks, DOCSIS 3.0, energy-efficient scheduling, vacation queuing model.**

I. INTRODUCTION

CHANNEL bonding is a major technology improvement in the hybrid fiber-coaxial (HFC) networks that comply with the Data over Cable Service Interface Specifications (DOCSIS) 3.0 [1], the newest industry standard for cable equipment vendors. By grooming the capacities of multiple upstream (US) and downstream (DS) channels, a channel-bonding cable modem (CM) can achieve a bi-directional throughput at 100 Mbps or higher. However, since the CMs have to equip and turn on multiple high-speed transceivers, channel-bonding also increases the power consumption in HFC networks dramatically, especially on the CM side [2]. On the other hand, HFC networks are typically engineered to meet anticipated peak traffic demands leading to significant energy waste during periods of low-traffic. Previously, we have reported several traffic-aware algorithms for achieving energy-saving in HFC networks that support DOCSIS 3.0 [3, 4]. However, those works did not consider traffic scheduling on the channel level, and the theoretical analysis was based on an over-simplified M/M/1 model. In this paper, we extend our previous work by designing a QoS-aware energy-efficient US scheduling algorithm. Based on a vacation M/M/n queuing model, we analyze the algorithm's performance with a three-dimensional Markov chain, and investigate the transitions with a quasi-birth-and-death (QBD) process to derive the analytical expressions of packet delay and power consumption.

The rest of the paper is organized as follows. Section II discusses the proposed algorithm. The theoretical analysis is presented in Section III. Section IV discusses the performance evaluations. Finally, Section V summarizes the paper.

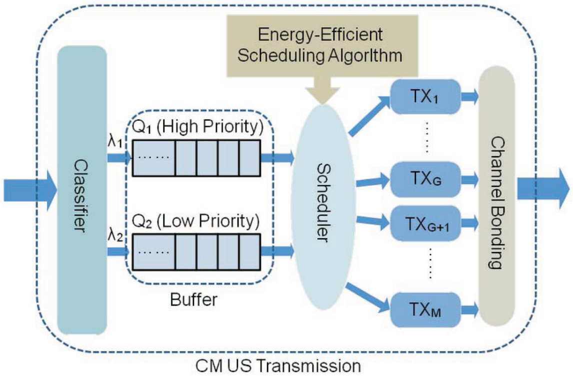


Fig. 1. System model for QoS-aware US transmission on a channel-bonding CM.

II. ENERGY-EFFICIENT UPSTREAM SCHEDULING

Fig. 1 shows the system model of the QoS-aware US transmission on a channel-bonding CM. DOCSIS 3.0 defines a traffic classifier to differentiate incoming packets based on their service flows [1]. We assume that there are two traffic priorities, corresponding to the best-effort (low-priority) and delay-sensitive (high-priority) traffics. After the classifier, the packets are buffered in the two priority queues, denoted as Q_1 and Q_2 for high- and low-priorities, respectively. For the two queues, packets arrive according to the Poisson processes with rates λ_1, λ_2 (packets/time-unit). We assume that the total capacity of Q_1 and Q_2 is finite as N, in terms of number of packets. All packets that arrive afterwards will be dropped, no matter which priority they belong to. Let $L_q(t)$ denote the total number of pending packets in Q_1 and Q_2, at any time instance t. To model channel bonding, we consider M TXs for each CM, denoted as $TX_1, TX_2, \ldots, TX_M$. We assume that the service time per packet on a TX follows the negative exponential distribution with an average of $\frac{1}{\mu}$ time-units. The algorithm keeps the first G TXs active all the time, but may turn the rest $M-G$ TXs off when the traffic load is low. For a TX, the length of each sleeping period (*i.e.* a vacation) follows an exponential distribution with an average of S time-units.

Definition We define three operation modes for a TX: 1) Working, as it is actively on, with average power consumption P_w, 2) Sleeping, as it is not active, with average power consumption P_s, 3) Setting-up, as it is in transition from sleeping to working. According to DOCSIS 3.0 [1], the setting-up time of an RF transceiver should be relatively short, normally in the range of 5 - 10 μsec. We therefore ignore the duration of the setting-up mode in following analysis, but denote its total energy consumption (*i.e.* transition overhead) as E_s.

Manuscript received January 21, 2013. The associate editor coordinating the review of this letter and approving it for publication was J. Wang.

P. Lu, Y. Yuan, and Z. Zhu are with the School of Information Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, P. R. China (e-mail: zqzhu@ieee.org).

Z. Yang is with the National Network New Media Engineering Research Center, Institute of Acoustics, Chinese Academy of Science and the University of Chinese Academy of Science, Beijing, P. R. China (e-mail: yangzz@dsp.ac.cn).

Digital Object Identifier 10.1109/LCOMM.2013.040213.130165

Definition Let T denote the TX's turn-on threshold. When a sleeping period of any TX ends, the scheduling algorithm examines Q_1 and Q_2 to get $L_q(t)$. Specifically, if $L_q(t) \geq T$, the TX goes to the setting-up mode and then the working mode for transmitting data, otherwise, the TX stays in the sleeping mode for another sleeping period.

III. THEORETICAL ANALYSIS

A. *Infinitesimal Generator Matrix* $\mathfrak{Q}$

We model the operation of the scheduling algorithm with a three dimensional Markov chain $(L_q(t), J(t), K(t)), t > 0$. Here, $J(t)$ is the number of high priority packets in Q_1 at t, and $K(t)$ is the number of working TXs at t. Hence, the state space is (i, j, k), where $i \in [0, N]$, $j \in [0, i]$, and $k \in [G, M]$. Both packet-arrival/departure and TX's operation mode change impact the state of the process. As the state space of the three dimensional Markov chain can be divided into levels, we model it with a quasi-birth-and-death (QBD) process [5]. In a QBD process, transitions are allowed only to the neighboring levels or within the same level [5]. We therefore obtain an infinitesimal generator matrix $\mathfrak{Q}$ as follows:

$$\mathfrak{Q} = \begin{bmatrix} B_0 & C_0 & & & \\ A_1 & B_1 & C_1 & & \\ & \ddots & \ddots & \ddots & \\ & & A_{N-1} & B_{N-1} & C_{N-1} \\ & & & A_N & B_N \end{bmatrix}, \tag{1}$$

where sub-matrix A_i encodes the backward transition rates from level i to $i-1$ for $i \geq 1$, sub-matrix B_i is for local transition rates within level i for $i \geq 0$, and sub-matrix C_i is for forward transition rates from level i to $i+1$ for $i \geq 0$. The analytical expressions of A_i, B_i and C_i can be obtained by analyzing the algorithm's operation, and due to the space limitation, we do not list them here.

B. *Matrix Geometric Solution*

Definition Let $\boldsymbol{\pi}$ denote the steady probabilities of the QBD process,

$$\boldsymbol{\pi} = (\pi_0, \boldsymbol{\pi_1}, \ldots, \boldsymbol{\pi_i}, \ldots, \boldsymbol{\pi_N}). \tag{2}$$

Then, each probability sub-matrix $\boldsymbol{\pi_i}$ for $i \in [1, N]$ can be expressed as

$$\boldsymbol{\pi_i} = \begin{cases} \left[\pi_{(i,0,G)} \cdots \pi_{(i,i,G)} \right], & i \in [1, G] \\ \left[\pi_{(i,0,G)} \cdots \pi_{(i,0,i)} \cdots \pi_{(i,i,G)} \cdots \pi_{(i,i,i)} \right], & i \in (G, M] \\ \left[\pi_{(i,0,G)} \cdots \pi_{(i,0,M)} \cdots \pi_{(i,i,G)} \cdots \pi_{(i,i,M)} \right], & i \in (M, N] \end{cases}$$

Lemma 1. *The steady probability of each state* (i, j, k) *in the three dimensional Markov chain* $(L_q(t), J(t), K(t)), t > 0$ *satisfies the following equation:*

$$\boldsymbol{\pi_i} = \pi_0 * \prod_{k=1}^{i} R_i, \quad i \in [1, N], \tag{3}$$

where

$$R_i = \begin{cases} -C_{i-1}(B_i + R_{i+1}A_{i+1})^{-1}, & i \in [1, N) \\ -C_{N-1}B_N^{-1}, & i = N \end{cases}. \tag{4}$$

Proof: It is known that a QBD process satisfies [5]

$$\boldsymbol{\pi}\mathfrak{Q} = \mathbf{0}.$$

Hence, we have

$$\begin{cases} \pi_0 B_0 + \boldsymbol{\pi_1} A_1 = 0 \\ \boldsymbol{\pi_{i-1}} C_{i-1} + \boldsymbol{\pi_i} B_i + \boldsymbol{\pi_{i+1}} A_{i+1} = 0, \ i \in [1, N) \\ \boldsymbol{\pi_{N-1}} C_{N-1} + \boldsymbol{\pi_N} B_N = 0 \end{cases}. \tag{5}$$

We obtain $R_N = -C_{N-1}B_N^{-1}$ from Eq. (4), and after combining it with Eq. (5), we prove

$$\boldsymbol{\pi_N} = \boldsymbol{\pi_{N-1}} R_N. \tag{6}$$

Then, with Eqs. (5) and (6), we can prove $\boldsymbol{\pi_{N-1}} = \boldsymbol{\pi_{N-2}} R_{N-1}$. By repeating these procedures, we prove

$$\boldsymbol{\pi_i} = \boldsymbol{\pi_{i-1}} R_i, \quad i \in [1, N], \tag{7}$$

and hence prove *Lemma* 1. ■

With the sub-matrices in Eq. (1), we can obtain $\{R_i, \ i = 0, \ldots, N\}$. For each steady probability $\boldsymbol{\pi_i}$, we also have

$$\sum_{i=0}^{N} \boldsymbol{\pi_i} = \pi_0 + \sum_{i=1}^{N} \sum_{j} \sum_{k} \pi_{(i,j,k)} = 1. \tag{8}$$

Finally, we can calculate $\boldsymbol{\pi_i}$ with *Lemma* 1 and Eq. (8).

C. *Performance Metrics*

In this subsection, we derive the analytical expressions of two performance metrics to evaluate the proposed algorithm.

1) Average Packet Delays: Let L denote the average number of total packets in Q_1 and Q_2 over the operation time. With Eqs. (1) and (2), we derive L as

$$L = \sum_{i=1}^{G} \sum_{j=0}^{i} i * \pi_{(i,j,G)} + \sum_{i=G+1}^{M} \sum_{j=0}^{i} \sum_{k=G}^{i} i * \pi_{(i,j,k)} + \sum_{i=M+1}^{N} \sum_{j=0}^{i} \sum_{k=G}^{M} i * \pi_{(i,j,k)}.$$

Let L_1 denote the average number of packets in Q_1,

$$L_1 = \sum_{i=1}^{G} \sum_{j=1}^{i} j * \pi_{(i,j,G)} + \sum_{i=G+1}^{M} \sum_{j=1}^{i} \sum_{k=G}^{i} j * \pi_{(i,j,k)} + \sum_{i=M+1}^{N} \sum_{j=1}^{i} \sum_{k=G}^{M} j * \pi_{(i,j,k)}.$$

And L_2, the average number of packets in Q_2, is

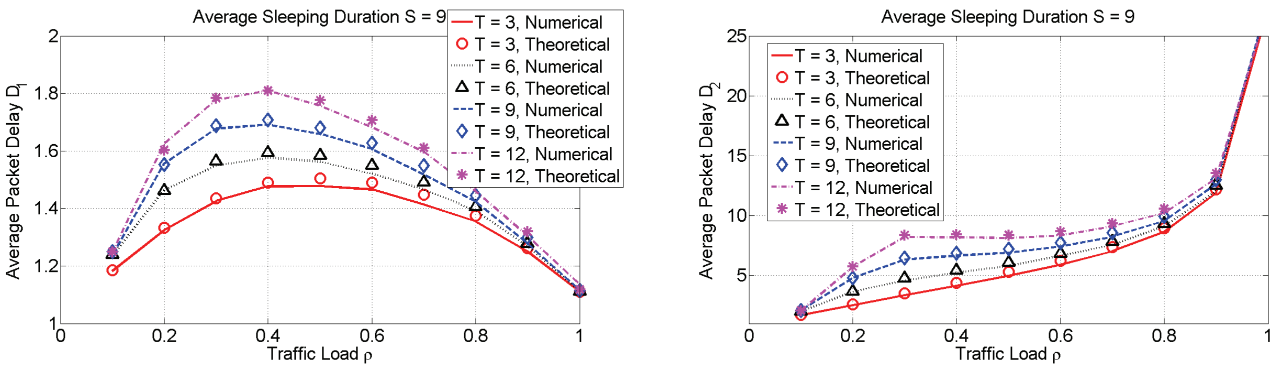
$$L_2 = L - L_1$$

The effective arrival rates of Q_1 and Q_2 are

$$\begin{aligned} \lambda_1^e &= \lambda_1 * [1 - \sum_{j=0}^{N} \sum_{k=G}^{M} \pi_{(N,j,k)}] \\ \lambda_2^e &= \lambda_2 * [1 - \sum_{j=0}^{N} \sum_{k=G}^{M} \pi_{(N,j,k)}] \end{aligned} \tag{9}$$

According to the Little's Law, the average delays of the packets in Q_1 and Q_2 can be obtained as

$$D_1 = \frac{L_1}{\lambda_1^e}, \quad D_2 = \frac{L_2}{\lambda_2^e} \tag{10}$$



(a) Average delay of high-priority packets, D_1

(b) Average delay of low-priority packets, D_2

Fig. 2. Impacts of T (i.e., TXs' turn-on threshold) on average packet delays, with $S = 9$.

Definition Let ρ denote the traffic load, i.e., the ratio of the total effective arrival rate to the total service rate,

$$\rho = \frac{\lambda_1^e + \lambda_2^e}{M * \mu}. \tag{11}$$

2) Average Power Consumption: To calculate the frequency of setting-up modes N_s, i.e., the average number of setting-mode modes that the TXs experience in a time-unit, we define $\theta = 1/S$. Then, N_s can be obtained as

$$N_s = \begin{cases} \theta * \sum_{i=T}^{N} \sum_{j=0}^{i} \sum_{k=G}^{M} \pi_{(i,j,k)}(M-k), & T > M \\ \theta * \left\{ \begin{array}{l} \sum_{i=T}^{M} \sum_{j=0}^{i} \sum_{k=G}^{i} \pi_{(i,j,k)}(M-k) + \\ \sum_{i=M+1}^{N} \sum_{j=0}^{i} \sum_{k=G}^{M} \pi_{(i,j,k)}(M-k) \end{array} \right\}, & T \leq M \end{cases} \tag{12}$$

The average power consumption of the CM is the power consumption of TXs' operation modes averaged over the corresponding steady state probability. Since the average number of working TXs in the system, N_w, can be calculated as

$$N_w = G * \sum_{i=0}^{G} \sum_{j=0}^{i} \pi_{(i,j,G)} + \sum_{i=G+1}^{M} \sum_{j=0}^{i} \sum_{k=G}^{i} k * \pi_{(i,j,k)} + \sum_{i=M+1}^{N} \sum_{j=0}^{i} \sum_{k=G}^{M} k * \pi_{(i,j,k)}, \tag{13}$$

we get the average power consumption as

$$\overline{P} = P_w * N_w + P_s * (M - N_w) + E_s * N_s. \tag{14}$$

IV. PERFORMANCE EVALUATIONS

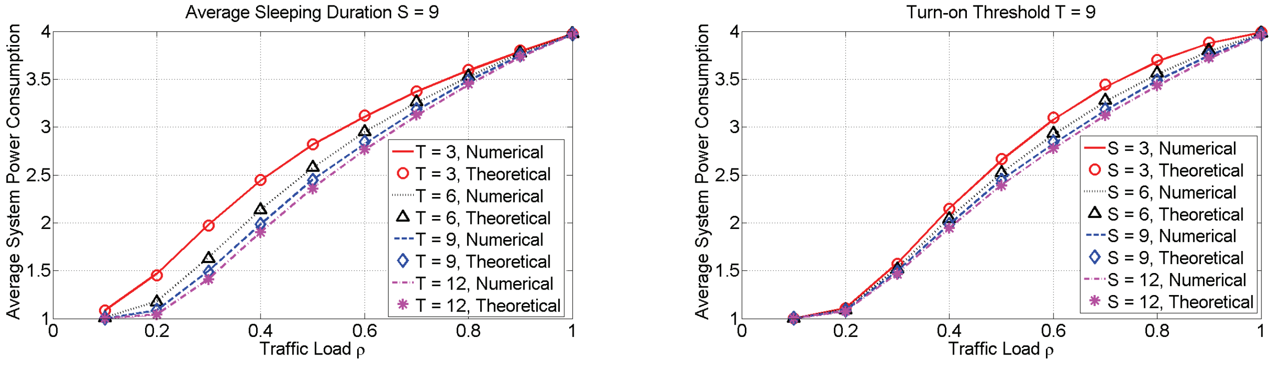
In order to verify the theoretical analysis, we design numerical simulations that use the Monte Carlo method with 6000 time-units. Table I shows the simulation parameters. We set $M = 4$, since to comply with the DOCSIS 3.0 standard, most of the channel-bonding CMs in commercial networks have four TXs. We pick the power consumption values from the realistic data-sheet [6], and normalize them accordingly. The theoretical and numerical results are also utilized to evaluate the performance impacts of the algorithm's key parameters, including the TX's turn-on threshold T and the average sleeping period S.

TABLE I
SIMULATION PARAMETERS

N, System buffer size in numbers of packets	100
M, Number of TXs in the system	4
G, Number of TXs that are normally on	1
μ, Service rate of a TX	1 time-unit/packet
ρ, Traffic load	0 - 1
P_w, TX's power consumption in working mode	1 power-unit
P_s, TX's power consumption in sleeping mode	0 power-unit
E_s, TX's energy consumption in each setting-up mode	5 energy-unit
Number of time-units in a simulation	6000
Number of simulations for statistical accuracy	50

A. TX's Turn-On Threshold T

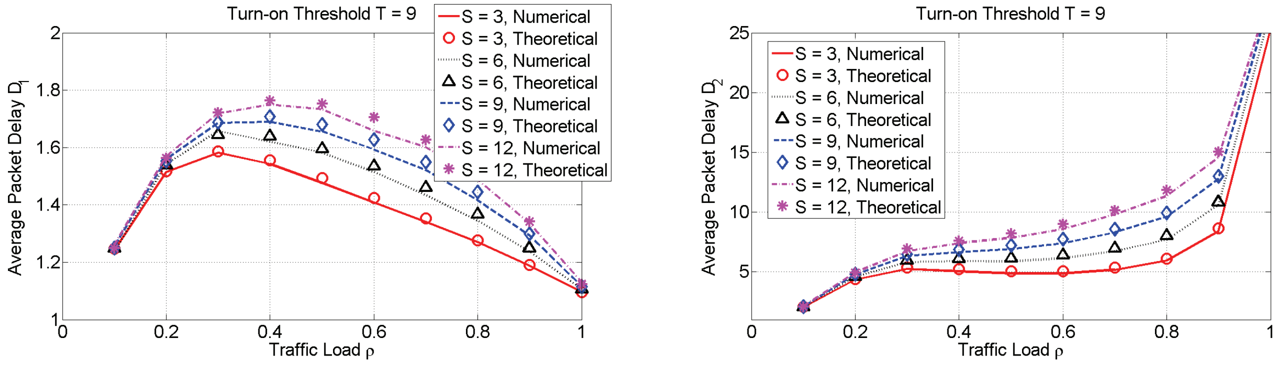
For simplicity, we assume $\lambda_1 = \lambda_2$. We first fix $S = 9$ and investigate the impacts of T on the average packet delays D_1 and D_2, and Fig. 2 shows results. It can be seen that theoretical results match well with the numerical ones. In Fig. 2(a), we observe that D_1 is at a low level ($D_1 \leq 2$), even when the traffic load $\rho \to 1$. When we change T from 3 to 12, the increase on D_1 is within 20% at the same ρ. Hence, the proposed algorithm can satisfy the QoS requirement of the high-priority delay-sensitive packets in Q_1. It is also worth noting that D_1 increases with ρ first for $\rho \in [0, 0.4]$, and then decreases with it. This is because that the algorithm always tries to schedule the high-priority packets in Q_1 first. For $\rho \in [0, 0.4]$, the total service rate from the working TXs is limited as there are still TXs in the sleeping mode. But when $\rho \to 1$, almost all TXs are working and their total service rate is larger than the arrival rate of packets in Q_1, and hence D_1 can decrease when $\rho \to 1$. In Fig. 2(b), we notice that T has a significant impact on D_2. When changing T from 3 to 12, D_2 can be doubled at the same ρ. We also observe that $D_2 \to \infty$ when $\rho \to 1$. Fig. 3(a) shows the impact of T on the average power consumption $\overline{P}$. It can be seen that $\overline{P}$ scales almost linearly with ρ, and $\overline{P}$ decreases with T. Compared to the operation where TXs are always active, the scheduling algorithm can achieve significant power-saving (up to 75%).



(a) Impact of T with $S = 9$.

(b) Impact of S with $T = 9$.

Fig. 3. Average power consumption $\overline{P}$ versus traffic load ρ.



(a) Average delay of high-priority packets, D_1

(b) Average delay of low-priority packets, D_2

Fig. 4. Impacts of S (i.e., average sleeping period) on average packet delays, with $T = 9$.

B. Average Sleeping Period S

We then investigate the impacts of S on $\overline{P}$, D_1 and D_2. Fig. 3(b) shows the impact of S on $\overline{P}$, and as expected, when we increase S, $\overline{P}$ can be reduced. Fig. 4 illustrates the results on average packet delays. We also observe that the impact of S on D_1 is not very significant (as in Fig. 4(a)) and it can significantly affects D_2 (as in Fig. 4(b)). It is interesting to notice that D_1 and D_2 almost stay unchanged for different S when $\rho \in [0, 0.2]$. This is because that for $\rho \in [0, 0.2]$, there is on average only one active TX at any given time.

V. CONCLUSION

We developed a QoS-aware energy-efficient US traffic scheduling algorithm for channel-bonding CMs in HFC networks. Based on a vacation M/M/n queuing model, we analyzed the algorithm's performance with a three-dimensional Markov chain and derived the analytical expressions of packet delay and power consumption. Both theoretical and numerical results indicated that the proposed algorithm achieved effective energy-saving compared to the case where the TXs of a CM are always active, and simultaneously maintained the delay of high-priority packets at a low level.

ACKNOWLEDGMENTS

This work was supported in part by the Fundamental Research Funds for the Central Universities (WK2100060006), the NCET Project (NCET-11-0884), the SRFDP Project (20123402120014), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA06010302).

REFERENCES

[1] Data Over Cable Service Interface Specifications, DOCSIS 3.0, Physical Layer Specification, CM-SP-PHYv3.0-110-111117, Cable Labs (http://www.cablelabs.com), Nov. 2011.

[2] J. Baliga, R. Ayre, K. Hinton, R. Tucker, "Energy consumption in wired and wireless access networks," *IEEE Commun. Mag.*, vol. 49, pp. 70–77, Jun. 2011.

[3] Z. Zhu, "A novel energy-aware design to build green broadband cable access networks," *IEEE Commun. Lett.*, vol. 15, pp. 887–889, Aug. 2011.

[4] Z. Zhu, "Design of energy-saving algorithms for hybrid fiber coaxial networks based on DOCSIS 3.0 standard," *J. Opt. Commun. Netw.*, vol. 4, pp. 449–456, Jun. 2012.

[5] M. Netus, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Johns Hopkins University Press, 1981, pp. 81–95.

[6] F. Gatta, *et al.*, "An embedded 65 nm CMOS baseband IQ 48 MHz-1 GHz dual tuner for DOCSIS 3.0," *IEEE J. Solid-State Circuits*, vol. 44, pp. 3511–3525, Dec. 2009.