CrossMark

ORIGINAL PAPER

QoS-aware energy-saving mechanism for hybrid optical-wireless broadband access networks

Ruyan Wang[1] · Alei Liang[1] · Chao Zhou[1] · Dapeng Wu[1] · Hong Zhang[1]

Abstract Hybrid optical-wireless broadband network (HOWBAN) takes full advantage of the high capacity and reliability of the passive optical network and the flexibility, ubiquity of the wireless network. Similar to other access networks, the issue of high energy consumption is a great challenge for HOWBAN. In HOWBAN, optical network units (ONUs) consume a great amount of energy. The sleep of ONUs can greatly improve the energy efficiency of HOWBAN. However, the quality of service (QoS) will be decreased while the packets are waiting in ONUs and optical line terminal. In this paper, we propose a QoS-aware energy-saving mechanism. A dynamic bandwidth allocation mechanism is designed to guarantee the QoS, where different priorities are considered. Meanwhile, we employ different sleep strategies by taking different priorities' tolerant delays into account to prolong the sleep time of ONUs. Then, based on the evaluation of packet delay, the optimal sleep parameter is derived to maximum the energy efficiency. In addition, a load balancing and resource allocation mechanism is adopted in the wireless domain to reduce the delay and congestion caused by ONUs' sleep. Results show that the proposed mechanism can effectively improve the energy efficiency and meet the QoS requirements of packets.

Keywords Passive optical network · QoS-aware · Energy-saving · Wireless mesh network · Resource allocation

✉ Dapeng Wu
wudapengphd@gmail.com

1 Chongqing University of Posts and Telecommunications, Chongqing 400065, China

1 Introduction

Passive optical network (PON) has the advantages of high bandwidth, long distance and high reliability, but it is not flexible and its deployment cost is high [1]. On the contrary, wireless mesh network (WMN) has the advantages of promising flexibility and low cost, but its bandwidth capacity is low [2,3]. HOWBAN combines the advantages of PON and WMN, thus becomes an important solution to the next-generation access networks. Figure 1 presents a HOWBAN architecture that integrates Ethernet PON (EPON) with the WMN. Each PON has a unique optical line terminal (OLT), which is located at the Central Office (CO) and connected to multiple optical network units (ONUs). On the other hand, the WMN is composed of several Mesh Points (MPs) [4]. The MP directly connected to the ONU is called wireless gateway, which completes the packet transmission between the wireless network and the PON. In the upstream direction, the users access the network through the MPs and users' packets are forwarded to the ONUs along wireless links. In the downstream direction, the packets from the Internet are sent to the OLT, and then are broadcasted to the ONUs. Consequently, the packets are routed to the users by MPs [5].

Statistics show that the energy consumption of information and communication technology (ICT) accounts for about 8% of the world's total energy consumption [6,7]. Meanwhile, as an important part of the entire network, the access network consumes about 70% of the ICT energy [8,9]. However, its equipment's utilization rate is less than 15% [10]. With traffic's explosive growth, the huge energy consumption will restrict HOWBAN's large-scale application. The sleep of ONUs can effectively improve the energy efficiency of HOWBAN, but it will lead to the degradation of QoS due to the increase of packets' queuing, and rerouting delay. There-
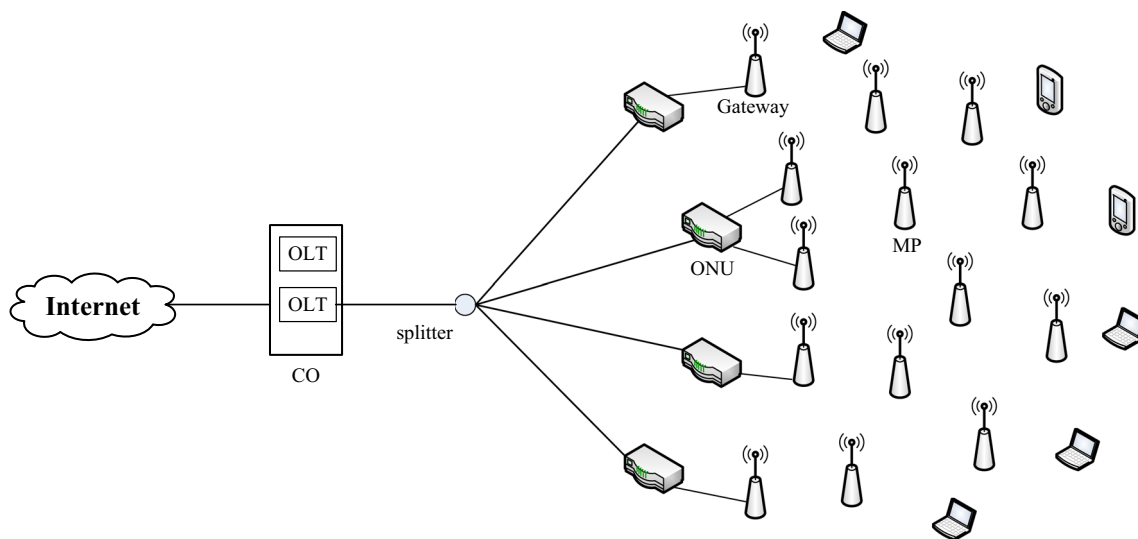
Fig. 1 Architecture of HOWBAN

fore, it is a great challenge to maximum the energy efficiency when different QoS requirements are met.

From the perspective of traffic characteristic, the packets of HOWBAN mainly consist of two types, namely high-priority packets and low-priority packets. Wherein the high-priority packets' delay requirements are higher, thus the long sleep time of ONU will lead to a sharp decline of QoS performance. On the contrary, low-priority packets are not sensitive to delay, so longer sleep time of ONU will improve the energy efficiency and won't degrade the QoS performance. Obviously, the sleep of ONU has different effects on different priorities because of their different delay tolerances. Therefore, it is necessary to design the optimal sleep time to improve the energy efficiency based on different priorities' delay tolerances. In addition, the connection status between the ONU and the wireless gateway changes dynamically due to the ONU's switching between sleep state and active state. In other words, the sleep of ONU will break the connection. Then, the upstream packets cannot reach the ONU and are buffered in the WMN, which will degrade the QoS performance of high-priority packets significantly. Therefore, the connection status and link quality should be fully considered to guarantee high-priority packets' delay requirements by selecting another active ONU to reroute the high-priority packets.

Currently, researchers have proposed several energy-saving mechanisms. Reference [11] proposes an energy-saving mechanism based on dynamic bandwidth allocation, where the OLT allocates uplink slots to the ONU and the ONU sleeps in other time slots. However, the mechanism does not consider packets' different priorities and the delay of high-priority packets is large. Reference [12] proposes a load-aware sleep mechanism. Whenever the ONU load is less than a threshold and the remaining load can be transferred to other

ONUs, the ONU sleeps for a period. However, the mechanism does not consider users' QoS performance. In addition, the transition increases the packet delay. Reference [13] proposes a periodic sleep energy-saving mechanism. The OLT calculates an optimal wake-up time and sends it to the ONU. Then, the ONU sends a confirmation message to the OLT and sleeps. Nevertheless, the energy efficiency is low. Reference [14] extends EEE packet coalescing by considering QoS priorities and improves the QoS support for video delivery by proposing a green QoS differentiated routing strategy. More specifically, packets of all links will be firstly assembled at the node and each time when high-priority packets arrive, the coalescing period will be interrupted and all the waiting packets will be sent out in the next transmitting period. To better guarantee QoS for video traffic, the capacity and delay-aware routing (CaDAR) scheme is adopted to route high-priority traffic flows, whereas a concentration routing strategy is used to reroute low-priority traffic flows. Reference [15] proposes a green mechanism with different classes of service. The OLT assigns time slots to the ONU depending on the class of service and the ONU sleeps in other time slots. However, such mechanism cannot take full advantage of existing connections between other active ONUs and wireless gateways when the ONU is sleeping. Therefore, the energy efficiency is also low.

To solve the above problems, a QoS-aware energy-saving mechanism (QESM) is designed. A dynamic bandwidth allocation mechanism is designed by taking different priorities into account to guarantee the QoS. Then, considering different tolerant delays of different priories, we design different sleep strategies to prolong the sleep time of ONUs. In order to maximum the energy efficiency, the optimal sleep parameter is obtained based on analysis of the packet delay. Meanwhile, a load balancing and resource allocation mechanism

is designed in the wireless domain to avoid the degradation of QoS performance due to the sleep of ONUs. Eventually, the network not only meets the QoS of packets, but also achieves the highest energy efficiency. Our contributions are summarized as follows:

Firstly, we design a dynamic bandwidth allocation mechanism to guarantee different QoS requirements and improve the bandwidth utilization where different priorities are considered.

Secondly, based on tolerant delays and analysis of the packet delay, the optimal sleep time is derived to prolong ONUs sleep time, further improve the energy efficiency.

Lastly, we propose a load balancing and resource allocation mechanism to reduce high-priority packets' delay and ONUs' congestion due to ONUs' sleep.

The remainder of the paper is structured as follows. The proposed mechanism is described in greater detail in Sect. 2. We evaluate the performance of our proposed mechanism, and compare it with some of previous works in Sect. 3. Finally, Sect. 4 concludes the paper.

2 QoS-aware energy-saving mechanism

As known, there are different services in the network, such as video on demand, voice over Internet protocol and e-mail. As a result, the packets in the network have different priorities and QoS requirements. Although researches have proposed many mechanisms to improve the energy efficiency of the HOWBAN, they seldom consider different QoS requirements simultaneously. Therefore, the QoS-aware energy-saving mechanism is proposed to improve the energy efficiency when different QoS requirements are met. The proposed energy-saving mechanism consists of three parts: dynamic bandwidth allocation, determination of the optimal energy efficiency, load balancing and resource allocation.

2.1 Dynamic bandwidth allocation

In EPON, the OLT allocates bandwidth (time slots) to each ONU by polling them. During the granted time slots, ONUs transmit or receive packets, whereas outside the granted slots, ONUs sleep to save energy. However, as mentioned above, long sleep time will increase packet delay and short sleep time will increase the energy consumption due to frequent wake-up. Thereby, to improve energy efficiency, a reasonable bandwidth allocation mechanism is designed to reduce the switching frequency and increase the ONU sleep time.

When ONU i completes data transmission within the granted time slots, it sends a report message to the OLT, which contains the queue statuses of high-priority and low-priority packets. After receiving all ONUs' report messages, the OLT derives the granted slots for all ONUs in the next

polling cycle. The polling cycle T_{cycle} consists of granted slots and guard slots, can be expressed as follows:

$$T_{\text{cycle}} = \sum_{i=1}^{N} (T_{\text{grant}}^i + T_{\text{g}}) \tag{1}$$

where N is the number of ONUs, T_{grant}^i is ONU i's granted slots, and T_{g} is the guard slot.

Long sleep time will increase the waiting delay and degrade QoS performance of high-priority packets due to their sensitivity to delay. On the contrary, low-priority packets can tolerate some delay, but short sleep time will lead to the frequent wake-up of ONU and low energy efficiency. For convenience, we call the ONU that has high-priority packets as HONU, and the ONU that only has low-priority packets as LONU. After receiving all ONUs' report messages, the OLT can judge whether it is a HONU or a LONU. To improve the energy efficiency and meet different priorities' QoS, we allow HONU to sleep within one polling cycle and LONU to sleep for more than one polling cycle. In traditional mechanisms, the ONU needs to exchange messages with the OLT in each polling cycle to determine its sleep and wake-up time in the next polling cycle and monitor whether there are downstream packets to be received. In order to reduce the energy consumption caused by the exchange of messages, the ONU that sleeps continuously doesn't need to exchange messages with the OLT in each polling cycle and the polling cycle T_{cycle} is set to be constant.

In the proposed mechanism, to guarantee high-priority packets' QoS, the OLT allocates bandwidth to high-priority packets firstly, then to low-priority packets. However, when the LONU that has slept for more than one polling cycle wakes, it should be allowed to transmit packets. Therefore, to ensure the QoS of the low-priority packets in the LONU, bandwidth b_L^k is allocated to the k-th LONU. It should be noted that b_L^k can be known by the OLT from LONU's report messages. When allocating bandwidth, if there is no LONU, the OLT will not reserve bandwidth b_L^k. As a result, the average bandwidth allocated to a HONU is obtained as Eq. (2).

$$B_i^{\text{avg}} = \frac{1}{N-n} \times \left[\frac{(T_{\text{cycle}} - N \times T_{\text{g}}) \times R}{8} - \sum_{k=1}^{n} b_L^k \right] \tag{2}$$

where R is the upstream transmission rate between OLT and ONU, n is the number of LONU.

Let B_i^H and B_i^L be the i-th HONU's request bandwidth for high-priority queue and low-priority queue, respectively. To ensure the fairness of each HONU and avoid that other HONUs cannot get the bandwidth, the granted bandwidth for HONU i's high-priority queue is as follows:

$$G_i^{HF} = \begin{cases} B_i^{\text{avg}}, & \text{if } B_i^H \geq B_i^{\text{avg}} \\ B_i^H, & \text{if } B_i^H < B_i^{\text{avg}} \end{cases} \tag{3}$$

In general, some HONUs' bandwidth requests are higher than B_i^{avg}, whereas some HONUs' bandwidth requests are lower than B_i^{avg} due to the uneven distribution of users and traffic. Therefore, there may be remaining bandwidth B_{offset}, can be calculated as:

$$B_{\text{offset}} = \sum_{i=1}^{N} \left(B_i^{\text{avg}} - B_i^H - B_i^L \right) \cdot \delta \left(B_i^{\text{avg}} - B_i^H - B_i^L \right) \tag{4}$$

where $\delta \left(B_i^{\text{avg}} - B_i^H - B_i^L \right)$ is a step function.

In order to improve bandwidth utilization, the remaining bandwidth B_{offset} is proportionally allocated to the high-priority queues which haven't got enough bandwidth as follows:

$$G_i^{HS} = \frac{R_i^H \times B_{\text{offset}}}{\sum_{i=1}^{N} \left(B_i^H - B_i^{\text{avg}} \right) \cdot \delta \left(B_i^H - B_i^{\text{avg}} \right)} \tag{5}$$

where R_i^H is HONU i's lacking bandwidth, it can be expressed as Eq. (6).

$$R_i^H = B_i^H - B_i^{\text{avg}} \tag{6}$$

After the above bandwidth allocations for high-priority queues, HONU i's granted bandwidth for high-priority queue is expressed as Eq. (7).

$$G_i^H = G_i^{HF} + G_i^{HS} \tag{7}$$

Moreover, if there is still remaining bandwidth, and there are bandwidth requests for low-priority packets, we further allocate the remaining bandwidth to them similarly to Eq. (5). Otherwise, the process of bandwidth allocation ends. Let G_i^L be the granted bandwidth for HONU i's low-priority queue, then ONU i's total granted bandwidth is shown in Eq. (8).

$$B_{\text{grant}}^i = \begin{cases} G_i^H + G_i^L, & \text{if ONU}\,i \text{ is a HONU} \\ b_L^i, & \text{else.} \end{cases} \tag{8}$$

At the beginning of each polling cycle, the OLT broadcasts grant messages to all ONUs. Then, ONU i transmits packets within the granted slots and sleeps outside the granted slots. The overhead time spent switching an ONU from sleep state to active state is T_{oh}. Therefore, the HONU i's sleep time T_H^i is obtained as Eq. (9).

$$\begin{aligned} T_H^i &= T_{\text{cycle}} - T_{\text{grant}}^i - T_{\text{oh}} \\ &= T_{\text{cycle}} - \frac{B_{\text{grant}}^i \times 8}{R} - T_{\text{oh}} \end{aligned} \tag{9}$$

The LONU can tolerate some delay, so the sleep time of LONU i is set to be more than one polling cycle as follows:

$$T_L^i = m \cdot T_{\text{cycle}} \tag{10}$$

The sleep time T_L^i has a great effect on the energy efficiency. The longer time the ONU sleeps, the higher energy efficiency can be achieved, but the delay of packets increases. On the contrary, the shorter sleep time signifies a shorter packet delay, but the energy efficiency decreases. Obviously, the tradeoff between the energy efficiency and packet delay is crucial to allow ONUs to sleep as long as possible and meanwhile meet the delay requirements of packets. According to the analysis of delay performance in Sect. 2.2, the optimal sleep parameter m can be obtained.

2.2 Determination of the optimal energy efficiency

As mentioned earlier, the ONU sleep time is dynamically derived based on different priorities and their tolerant delays to achieve the optimal energy efficiency and satisfy users' QoS. However, the sleep parameter m determines the ONU sleep time T_L^i, further affects packets' delay and the energy efficiency. In this paper, the optimal sleep parameter m is derived according to analysis of packets' delay under specific arrival rate λ.

Without loss of generality, low-priority packets are assumed to arrive at a Poisson process with rate λ and the service times have a general distribution. Therefore, the M/G/1 queuing model is adopted to analyze the packet delay.

In the WMN, the propagation delay is negligible because the distance between two adjacent MPs is relatively short. Then, the delay for link u–v d_{uv} consists of transmission delay, slot synchronization delay and queuing delay, it can be obtained as follows [16]:

$$d_{uv} = \frac{1}{\mu C_{uv}} + \frac{1}{2\mu C_{uv}} + \frac{\lambda_{uv}}{\mu C_{uv} \left(\mu C_{uv} - \lambda_{uv} \right)} \tag{11}$$

where $1/\mu$ is the average packet length, C_{uv} is the capacity of link u–v and λ_{uv} is the arrival rate at link u–v.

Based on the delay for a single-hop link, the delay for a wireless path can be derived as follows:

$$D(W) = \sum_{o \in P_k} d_o \tag{12}$$

where P_k is the k-th path and o is a single-hop link on P_k.

The upstream packets which arrive at ONU during the sleep period have to wait until ONU's sleep period ends. Besides, after the sleep time expires, the ONU needs to synchronize and recover clock. Therefore, the average waiting delay during sleep and synchronization periods is obtained as Eq. (13).

$$D[I] = \frac{T_L + T_{\text{oh}}}{2} \tag{13}$$

After the ONU wakes, it transmits packets during the grant slots. The expected waiting time in queue is calculated as Eq. (14) [17].

$$D[A] = \frac{\lambda E\left\{X^2\right\}}{2(1 - \rho)} \tag{14}$$

where X is the service time, $\rho = \lambda/\mu = \lambda\bar{X}$ and $E\left\{X^2\right\}$ is the second moment of service time, which can be expressed as Eq. (15).

$$E\left\{X^2\right\} = X^2 + \sigma^2 \tag{15}$$

where σ^2 denotes the variance of service time and can be obtained as follows:

$$\sigma^2 = \alpha \times (x - T_{\text{avg}})^2 + \beta \times (mT_{\text{cycle}} - T_{\text{avg}})^2 \tag{16}$$

In Eq. (16), T_{avg} denotes the average value of the service time for all types of traffic and is calculated as Eq. (17).

$$T_{\text{avg}} = \frac{mT_{\text{cycle}} + T_{\text{grant}}}{2} \tag{17}$$

In Eq. (16), α is the ratio of sleep slots to total slots and β is the ratio of granted slots to total slots, can be obtained as follows:

$$\alpha = \frac{T_{\text{grant}}}{mT_{\text{cycle}} + T_{\text{grant}}} \tag{18}$$

$$\beta = \frac{mT_{\text{cycle}}}{mT_{\text{cycle}} + T_{\text{grant}}} \tag{19}$$

Based on the above analysis, the total delay of a packet is derived as Eq. (20).

$$D = D[W] + D[I] + D[A] \tag{20}$$

In EPON, ONU sleep time is limited to 50 ms due to the multi-point control protocol. Therefore, we can obtain the constraint as follows:

$$mT_{\text{cycle}} \leq 50 \tag{21}$$

Let D_r be the delay requirement of the low-priority packet, then we can obtain constraint (22).

$$D \leq D_r \tag{22}$$

Based on the above two constraints (21) (22), the maximum m can be derived. Then, the optimal energy efficiency can be achieved. Therefore, the maximum m is the optimal sleep parameter we want.

2.3 Load balancing and resource allocation

In this section, a load balancing and resource allocation mechanism is proposed to reduce high-priority packets' delay and ONUs' congestion.

The path quality has a great effect on the end-to-end delay of wireless path. Good path quality signifies low path delay. In addition, the ONU load affects the arrived packets' waiting delay. As known, high ONU load will result in high waiting delay. Therefore, both path quality and ONU load should be considered to improve packets' delay performances.

We use the airtime metric defined in the hybrid wireless mesh protocol (HWMP) to evaluate the path quality, where channel access overhead O_{ca}, protocol overhead O_{p}, and frame error rate e_{f} are fully considered [18]. The airtime metric of link o is calculated as follows:

$$C_o = \left[O_{\text{ca}} + O_{\text{p}} + \frac{B_{\text{t}}}{r}\right] \cdot \frac{1}{1 - e_{\text{f}}} \tag{23}$$

where B_{t} is the number of bits in a test frame and r is the bit rate.

In HWMP, the airtime metric is the default parameter for path selection. When node A wants to find a path to node B, it will broadcast a Path Request (PREQ) message. Then, the nodes that receive the PREQ message will forward it to other nodes. When a node is forwarding a PREQ message, it will add the airtime metric between it and its predecessor to the Metric field of the PREQ message and forward the PREQ message to its successor. In this way, a node can know the airtime metrics of other links. The destination node will receive some PREQ messages which come from different paths and then select the optimal path based on the airtime metrics recorded in the Metric field of the PREQ message. Eventually, the destination node sends a Path Reply (PREP) message to the source node to establish the forwarding path.

Based on the airtime metric, the path quality is obtained as Eq. (24).

$$P_k = \sum_{o \in P_k} C_o \tag{24}$$

where P_k is the k-th path and o is a single-hop link on P_k.

In the HOWBAN, ONUs periodically broadcast state information to MPs. The MPs directly connected to the ONUs receive the broadcast information firstly, and then forward the information to other MPs which haven't received the information yet. Finally, the MPs can receive the broadcast of an ONU that locate far away. Then, each MP derives the wireless path with the optimal path quality to ONUs based on Eq. (24). In addition, the state information broadcasted by ONUs also includes ONUs' load information. To accurately assess an ONU's load state, the average load of ONU i can

be calculated as Eq. (25).

$$X_i^{(0)}(k) = \frac{NP_k \times AP}{M} \qquad (25)$$

where NP_k is ONU i's total received packets during the k-th slot, AP is packet size, and M is slot length, as well as ONU's broadcast cycle.

ONU i can record the sequence of the previous n slots' average loads as follows:

$$X^{(0)} = \left\{ x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n) \right\} \qquad (26)$$

When selecting the destination ONUs, MPs need to determine ONUs' load states in the next slot according to the previous load records to keep away from the high-load ONUs. Obviously, ONUs need to predict their average loads in the next slot and broadcast them to MPs. The accuracy of the predicted results will directly affect the selection of destination ONUs and then affect transmission delay of upstream packets. Discrete gray prediction model (DGM) develops a mathematical model based on small and incomplete discrete information. Moreover, it has the advantages of requiring less sample data, high prediction accuracy and low computational complexity. Currently, DGM (1, 1) is the most widely used model, which is a sequence-oriented DGM with one variable and the first-order differential equations. Besides, when raw data approximately follows exponential distribution, the prediction accuracy is much higher. We use the DGM (1, 1) to predict ONUs' loads in the next slot to determine whether there is congestion because ONUs' received upstream packets approximately follow the exponential distribution [19]. Based on the DGM (1, 1) established in "Appendix", ONU i's average load in the next slot $x_i^{(0)}(n+1)$ can be obtained. After receiving the broadcast messages from ONUs, which contain ONUs' current average loads and the predicted average loads, MPs calculate the average load of all ONUs in current slot and next slot as follows:

$$X(n) = \frac{1}{N}\sum_{i=1}^{N} x_i^{(0)}(n) \qquad (27)$$

$$X(n+1) = \frac{1}{N}\sum_{i=1}^{N} x_i^{(0)}(n+1) \qquad (28)$$

To accurately determine whether ONU i is congested, we consider ONU i's loads of two consecutive slots. If $x_i^{(0)}(n) > (1+\gamma)X(n)$ and $x_i^{(0)}(n+1) > (1+\gamma)X(n+1)$, ONU i is congested. On the contrary, if $x_i^{(0)}(n) < (1-\gamma)X(n)$ and $x_i^{(0)}(n+1) < (1-\gamma)X(n+1)$, ONU i is at low load. Otherwise, ONU i is steady. In order to relieve high-load ONU's congestion, we need to transfer part of high-load ONUs' traffic to other low-load ONUs. Among the low-load ONUs, the

ONU that is on the path with optimal path quality is chosen as the transferred ONU. The load transfer ratio can be obtained by Eq. (29).

$$p = \frac{x_i^{(0)}(n+1) - X(n+1)}{x_i^{(0)}(n+1)} \qquad (29)$$

where p is the ratio of the traffic to be transferred to the total traffic in ONU i.

To transmit the delay-sensitive high-priority packets, the wireless path with the optimal path quality is chosen and the ONU on the path is called primary ONU. When the load of the primary ONU is high, we reroute the low-priority packets to other ONUs to reduce the primary ONU's load and packets' waiting delay. When the primary ONU sleeps, the wireless path with the suboptimal path quality is chosen and the ONU on the path is called secondary ONU. Then, high-priority packets will be transmitted through the secondary ONU and low-priority packets will be transmitted through the primary ONU after it wakes to avoid the congestion of the secondary ONU. To make it clear, we summarize the main procedures of the QoS-aware energy-saving mechanism and show the pseudo-code in Algorithm 1.

Algorithm 1 QoS-aware Energy-saving Mechanism (QESM)

1: Allocate bandwidth for all ONUs;
2: **for** each ONU i **do**
3:     record the sequence of the previous n slots' average loads;
4:     predict the average loads in the next slot: $x_i^{(0)}(n+1)$;
5:     broadcast messages that contain $x_i^{(0)}(n+1)$ to MPs;
6: **end for**
7: **for** each MP **do**
8:     update messages;
9:     calculate the average loads of all ONUs in current slot and next slot: $X(n), X(n+1)$;
10: **end for**
11: **if** $\left(x_i^{(0)}(n) > (1+\gamma)X(n) \right) \&\& x_i^{(0)}(n+1) > (1+\gamma)X(n+1)$ **then**
12:     calculate the load transfer ratio p;
13:     transfer part of high-load ONUs' traffic to other low-load ONUs;
14: **else**
15:     wait for transfer;
16: **end if**
17: **for** each ONU i **do**
18:     **if** ONU i is a HONU **then**
19:         sleep time is T_H;
20:     **else**
21:         sleep time is T_L;
22:     **end if**
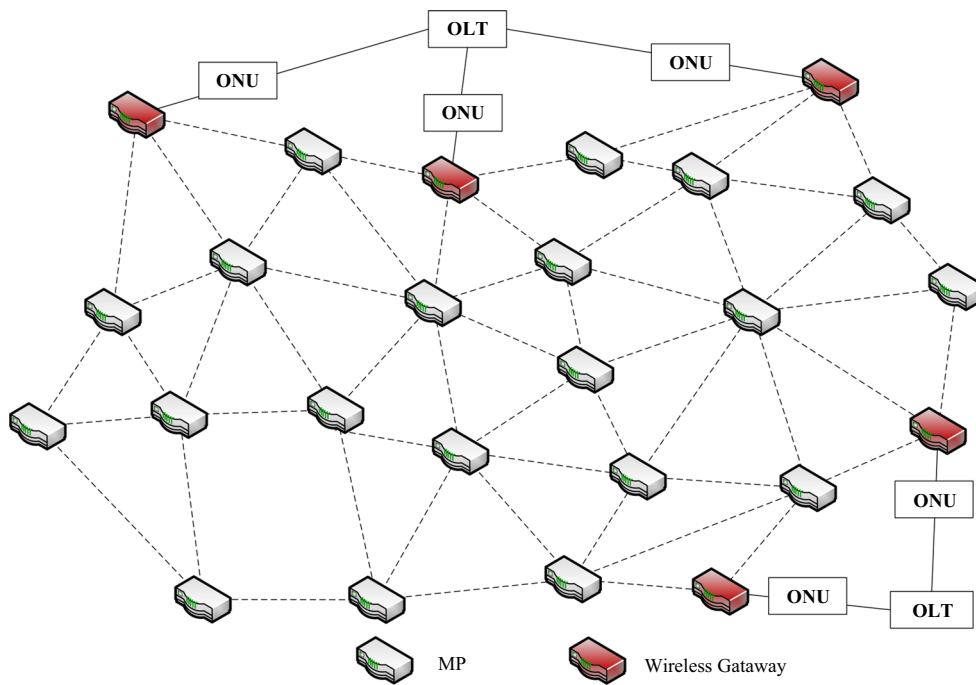23:   sleep and wake based on designed time;
24: **end for**

Fig. 2 Simulation topology

3 Performance evaluation

In this section, we present the performance evaluation of our proposed mechanism by Network Simulator 2. The network topology is shown in Fig. 2 [10], which is composed of 20 MPs, 5 ONUs and 2 OLTs. Then, we compare the proposed mechanism (QESM) with DQAE mechanism [13] and SLAB mechanism [15].

To verify the effectiveness of the proposed mechanism, we compare and analyze the above three mechanisms at different network loads. Performance indicators are as follows:
a. ONUs' total energy consumption (TE)

The value of TE is the sum of total energy consumed by all ONUs under both active and sleep states, can be expressed as:

$$\text{TE} = P_{\text{onu}}^{\text{active}} \left(\sum_{n=1}^{N} T_{\text{active}}^{n} \right) + P_{\text{onu}}^{\text{sleep}} \left(\sum_{n=1}^{N} T_{sl}^{n} \right) \quad (30)$$

where $P_{\text{onu}}^{\text{active}}$ and $P_{\text{onu}}^{\text{sleep}}$ are the power of each ONU with the active state and sleep state, respectively, T_{sl}^{n} and T_{active}^{n} are the sleep duration and active duration of ONU n, respectively.
b. Energy consumption per Erlang (EPE)

EPE is the ratio of total energy consumption to the network load (measured in Erlang).
c. Average path delay from the access MP to the OLT (APD)

APD is the ratio of all packets' path delays to the number of packets, can be calculated as follows:

$$\text{APD} = \frac{\sum_{q=1}^{\text{num}} \left(t_{q,\text{receive}} - t_{q,\text{arrive}} \right)}{\text{num}} \quad (31)$$

where $t_{q,\text{arrive}}$ represents the time when the q-th packet arrives at the access MP, $t_{q,\text{receive}}$ represents the time when the OLT receives the q-th packet, and num is the number of all packets.
d. Sleep duration ratio (SDR)

SDR is the ratio of total sleep duration for each ONU to the total simulation time for all ONUs, can be expressed as:

$$\text{SDR} = \frac{\sum_{n=1}^{N} T_{sl}^{n}}{N \cdot T_{\text{sim}}} \quad (32)$$

where T_{sim} is the simulation time for each ONU.

The simulation parameters are given in Table 1.

The total energy consumption of ONUs directly reflects the energy efficiency of the energy-saving mechanisms. Figure 3 illustrates the total ONU energy consumptions of the three mechanisms at different network loads. Clearly, with the increase of the network load, the total energy consumptions rise gradually. This is because that the ONU needs to forward more and more packets and the sleep time of ONU is relatively shorter. Moreover, the QESM mechanism chooses different ways of sleep according to different priorities. That is, the ONU that has high-priority packets sleeps within one polling cycle, whereas the ONU that only has low-priority packets sleeps for more than one polling cycle. Therefore,

Table 1 Simulation parameters

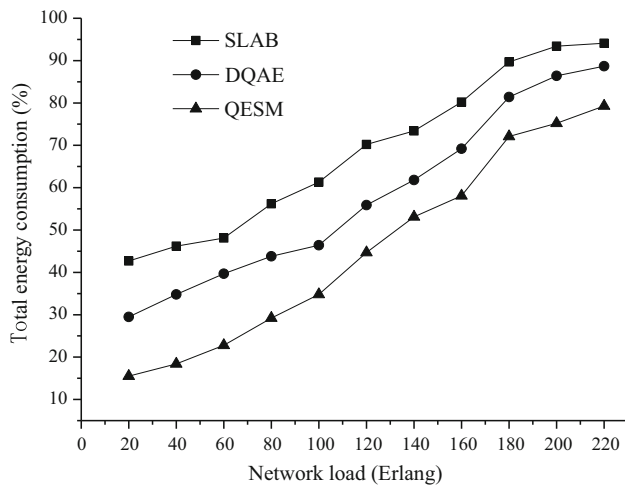| Parameter	Value
Simulation time	1000 s
Scenario size	1600 m × 1600 m
Maximum capacity of a wireless link	54 Mbps
Total capacity of each ONU	100 Mbps
ONU power consumption in active state, sleep state	5.052, 0.75 W
Length of a polling cycle	2 ms
ONU overhead time	0.125 ms
OLT-ONU link capacity	1 Gbps

Fig. 3 Total energy consumption of ONUs

Fig. 4 Energy consumption per unit of Erlang

Fig. 5 Average path delay for high-priority packets

QESM's energy efficiency is higher than those of the other two mechanisms due to the increase of ONUs' sleep time.

Figure 4 shows the energy consumptions per Erlang of the three mechanisms. The results show that the energy consumptions per Erlang are higher when the network load is low. This is because that the ONU needs to wake-up even the number of packets is small when the network load is low. In addition, the performance of QESM is better compared with the other two due to QESM's different responses to different priorities.

The average path delay for high-priority packets is an important indicator of users' QoS performance. Figure 5 shows that the average path delay for high-priority packets rises with the growing network load. The main reason is that the increasing number of packets leads to the congestion in nodes and over links. Then, the queuing delay rises and the sleep of some ONUs further increases the queuing delay. However, a load balancing and resources allocation mechanism is adopted in the QESM, where high-priority packets are forwarded on the path with the optimal path quality and low-load ONU. Therefore, high-priority packets' average path delay in QESM is lower than those in the SLAB and DQAE.
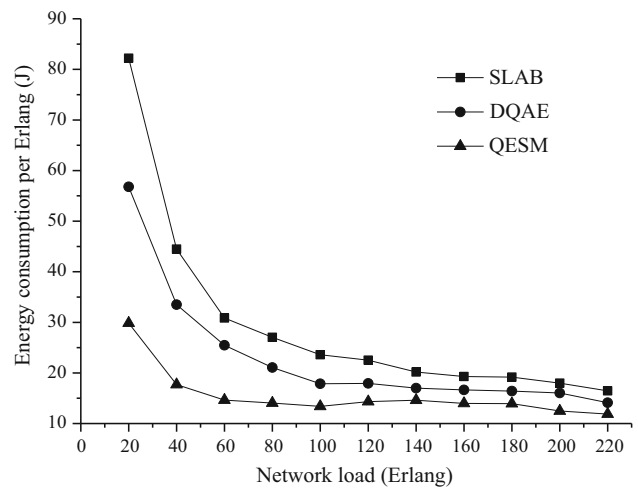
Figure 6 shows the average path delays for low-priority packets at different network loads. As can be seen, the average path delays for low-priority packets rise with the increasing network load due to the increase of queuing delay and transmission delay. The QESM mechanism takes full advantage of low-priority packets' tolerant delay to increase ONUs' sleep time and further improve the energy efficiency. Therefore, when the network load is low, the average path delay for low-priority packets in QESM is a little higher than those in the other two. However, when the network load increases to a certain extent, the average path delay for low-priority packets in QESM is lower than those in the other two due to the load balancing and resources allocation mechanism's adoption.

Figure 7 shows the SDRs of QESM, DQAE and SLAB mechanisms, respectively. As can be seen, the SDRs decrease gradually with the increase of network load due to the decrease of sleep time. The QESM mechanism allows the
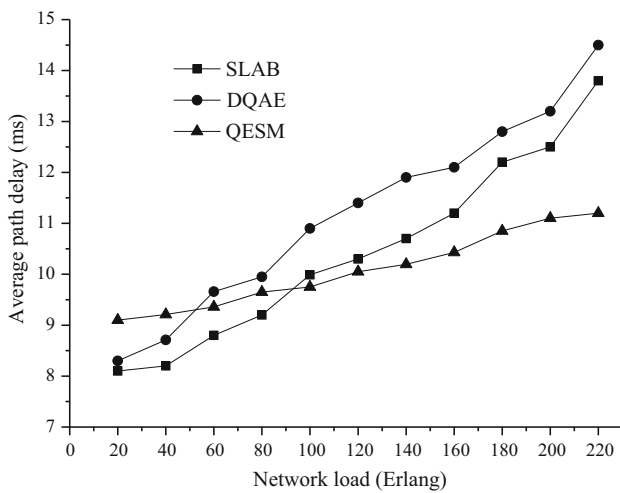
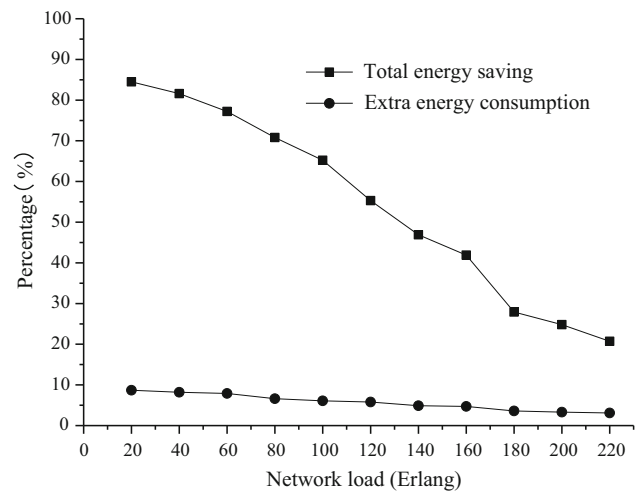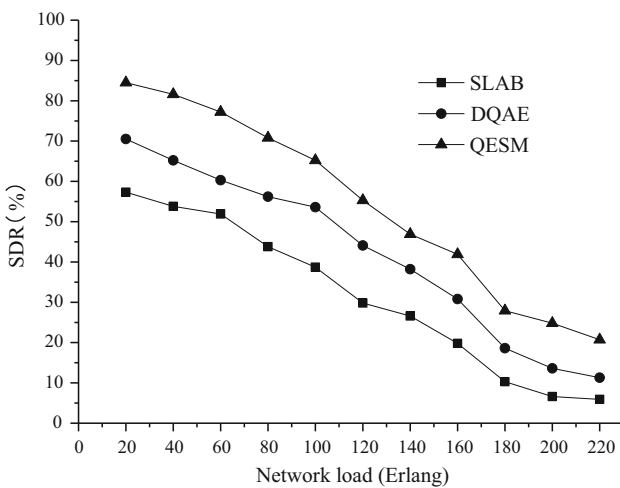**Fig. 6** Average path delay for low-priority packets

Fig. 7 SDR at different network loads

Fig. 8 Extra energy consumption

4 Conclusions

In this paper, a novel QoS-aware energy-saving mechanism is proposed to reduce the energy consumption of HOWBAN. Based on consideration of different priorities, a dynamic bandwidth allocation mechanism is proposed to guarantee the QoS; at the same time, different sleep strategies are designed for different priorities to prolong the sleep time of ONUs. In order to further maximum the energy saving, the optimal sleep parameter is derived based on analysis of the packet delay. In addition, a load balancing and resource allocation mechanism is designed in the wireless domain to avoid the degradation of QoS performance due to the sleep of ONUs. Results show that the energy efficiency can be dramatically improved by our introduced mechanisms.

Appendix

The detailed process of the DGM (1, 1) is as follows:
Step 1 Accumulated generating operation

The sequence of the previous n slots' average loads for ONU i is expressed as follows:

$$X^{(0)} = \left\{ x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n) \right\} \tag{33}$$

Making one accumulated generating operation (1-AGO)

$$X^{(1)} = \left\{ x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n) \right\} \tag{34}$$

ONU that only has low-priority packets to sleep for more than one consecutive polling cycle. Therefore, the sleep time increases due to the decrease of ONUs' overhead time. From the results, we can see that the SDR performance of the QESM mechanism is higher than those of DQAE and SLAB mechanisms by 22.5 and 38.3%, respectively.

In our proposed mechanism, the high-priority packets arriving during ONUs' sleep periods will be transferred to other active ONUs. However, the process of transfer will consume extra energy. As can be seen from Fig. 8, the extra energy consumption decreases with network load's increasing. This is because that the higher is the network load, the smaller is the number of sleeping ONUs. As a result, the smaller is the number of packets that should be transferred. On the whole, the little extra energy consumption doesn't affect the high energy efficiency of our proposed mechanism.

where $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$, $k = 1, 2, \ldots, n$.

1-AGO makes any nonnegative, fluctuant and nonfluctuant column translate to nondecreasing and increasing column, which weakens the randomness and strengthens the regularity.

Step 2 Establishment of DGM (1, 1)

The DGM (1, 1) mode is established as follows:

$$x^{(1)}(k+1) = \beta_1 x^{(1)}(k) + \beta_2 \tag{35}$$

Step 3 Solutions of parameters

Let $\hat{\beta} = [\beta_1, \beta_2]^T$ be the parameter column, and

$$Y = \begin{bmatrix} x^{(1)}(2) \\ x^{(1)}(3) \\ \vdots \\ x^{(1)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} x^{(1)}(1) & 1 \\ x^{(1)}(2) & 1 \\ \vdots & \vdots \\ x^{(1)}(n-1) & 1 \end{bmatrix} \tag{36}$$

Then, we can get

$$\hat{\beta} = \left(B^T B\right)^{-1} B^T Y = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \tag{37}$$

Substituting Y and B into Eq. (37), β_1, β_2 are given by

$$\beta_1 = \frac{\sum_{k=1}^{n-1} x^{(1)}(k+1) \cdot x^{(1)}(k) - \frac{1}{n-1} \cdot \sum_{k=1}^{n-1} x^{(1)}(k+1) \cdot \sum_{k=1}^{n-1} x^{(1)}(k)}{\sum_{k=1}^{n-1} \left[x^{(1)}(k)\right]^2 - \frac{1}{n-1} \left[\sum_{k=1}^{n-1} x^{(1)}(k)\right]^2} \tag{38}$$

$$\beta_2 = \frac{1}{n-1} \cdot \left[\sum_{k=1}^{n-1} x^{(1)}(k+1) - \beta_1 \sum_{k=1}^{n-1} x^{(1)}(k)\right] \tag{39}$$

Step 4 Acquirement of prediction equation

Substituting β_1, β_2 into Eq. (35), then the recurrence function of DGM (1, 1) is obtained as follows

$$\hat{x}^{(1)}(k+1) = \beta_1^k x^{(0)}(1) + \frac{(1-\beta_1^k)\beta_2}{1-\beta_1}, \quad k = 1, 2, \ldots, n-1 \tag{40}$$

The prediction equation is obtained by inverse 1-AGO as

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad k = 1, 2, \ldots, n-1 \tag{41}$$

where $\hat{x}^{(0)}(k+1)$ is the predicted value of ONU i's average load in the next slot.

References

1. Zhou, H., Mao, S., Agrawal, P.: Optical power allocation for adaptive transmissions in wavelength-division multiplexing free space optical networks. Digital Commun. Netw. **1**(3), 171–180 (2015)

2. Peng, M., Li, Y., Jiang, J., Li, J., Wang, C.: Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies. IEEE Wirel. Commun. **21**(6), 126–135 (2014)

3. Luo, C., Guo, S., Guo, S., Yang, L., Min, G., Xie, X.: Green communication in energy renewable wireless mesh networks: routing, rate control, and power allocation. IEEE Trans. Parallel Distrib. Syst. **25**(12), 3211–3220 (2014)

4. Shaddad, R.Q., Mohammad, A.B., Al-Gailani, S.A., Al-Hetar, A.M., Elmagzoub, M.A.: A survey on access technologies for broadband optical and wireless networks. J. Netw. Comput. Appl. **41**(5), 459–472 (2014)

5. Sarigiannidis, A.G., Iloridou, M., Nicopolitidis, P., Papadimitriou, G., Pavlidou, F.N., Sarigiannidis, P.G., Louta, M.D., Vitsas, V.: Architectures and bandwidth allocation schemes for hybrid wireless-optical networks. IEEE Commun. Surv. Tutor. **17**(1), 427–468 (2015)

6. Zhang, Z., Yi, Y., Yang, J., Jing, X.: Energy efficiency based on joint mobile node grouping and data packet fragmentation in short-range communication system. Int. J. Commun. Syst. **27**(4), 534–550 (2014)

7. Wu, D.P., He, J., Wang, H.G., Wang, C.G., Wang, R.Y.: A hierarchical packet forwarding mechanism for energy harvesting wireless sensor networks. IEEE Commun. Mag. **53**(8), 92–98 (2015)

8. Valcarenghi, L., Van, D.P., Raponi, P.G., Castoldi, P., Campelo, D.R., Wong, S., Yen, S., Kazovsky, L.G., Yamashita, S.: Energy efficiency in passive optical networks: where, when, and how. IEEE Netw. **26**(6), 61–68 (2012)

9. Li, Y., Liao, C., Wang, Y., Wang, C.: Energy-efficient optimal relay selection in cooperative cellular networks based on double auction. IEEE Trans. Wirel. Commun. **14**(8), 4093–4104 (2015)

10. Chowdhury, P., Tornatore, M., Sarkar, S., Mukherjee, B.: Building a green wireless-optical broadband access network (WOBAN). J. Lightw. Technol. **28**(16), 2219–2229 (2010)

11. Dhaini, A.R., Ho, P.H., Shen, G.: Toward green next-generation passive optical networks. IEEE Commun. Mag. **49**(11), 94–101 (2011)

12. Kantarci, B., Mouftah, H.T.: Towards energy-efficient hybrid fiber-wireless access networks. In: Proceedings of ICTON, pp. 1–4 (2011)

13. Schütz, G., Correia, N.: Design of QoS-aware energy-efficient fiber–wireless access networks. J. Opt. Commun. Netw. **4**(8), 586–594 (2012)

14. Liu, X., Ghazisaidi, N., Ivanescu, L., Kang, R.: On the tradeoff between energy saving and QoS support for video delivery in EEE-based FiWi networks using real-world traffic traces. J. Lightw. Technol. **29**(18), 2670–2676 (2011)

15. Shi, L., Mukherjee, B., Lee, S.S.: Energy-efficient PON with sleep-mode ONU: progress, challenges, and solutions. IEEE Netw. **26**(2), 36–41 (2012)

16. Sarkar, S., Yen, H.H., Dixit, S., Mukherjee, B.: A novel delay-aware routing algorithm (DARA) for a hybrid wireless-optical broadband access network (WOBAN). IEEE Netw. **22**(3), 20–28 (2008)

17. Bertsekas, D.P., Gallager, R.G., Humblet, P.: Data Networks. Prentice-Hall, Upper Saddle River (1992)

18. Hiertz, G.R., Max, S., Zhao, R., Denteneer, D., Berlemann, L.: Principles of IEEE 802.11s. In: Proceedings of International Conference on Computer Communications and Networks, pp. 1002–1007 (2007)

19. Hiertz, G.R., Denteneer, D., Max, S., Taori, R., Cardona, J., Berlemann, L., Walke, B.: IEEE 802.11s: the WLAN mesh standard. IEEE Wirel. Commun. **17**(1), 104–111 (2010)

Ruyan Wang received his Ph.D. degree in 2007 from the University of Electronic and Science Technology of China (UESTC) and the M.S. degree from Chongqing University of Posts and Telecommunications (CQUPT), China, in 1997. From Dec. 2002, he is a professor in Special Research Centre for Optical Internet and Wireless Information Networks at CQUPT. His research interests include: optical burst switching networks, network performance analysis, and multimedia information processing.

Alei Liang is working for his M.S. degree in Chongqing University of Posts and Telecommunications. His research interest includes hybrid optical-wireless broadband access networks and green computing technologies.

Chao Zhou received his M.S. degree in 2016, from Chongqing University of Posts and Telecommunications (CQUPT), China. He has several years of research experience in QoS control of high speed networks and network performance analysis.

Dapeng Wu received his M.S. degree of communication and information system in June 2006, Chongqing University of Posts and Telecommunications, the Ph.D. degree from Beijing University of Posts and Telecommunications in 2009. His research interests include next generation network, IP QoS architecture, network reliability, performance evaluation in communication systems.

Hong Zhang is a Ph.D. student in Chongqing University of Posts and Telecommunications. He received the M.S. degree in Chongqing University of Posts and Telecommunications in 2013. His research includes optical communication, survivability technology and next generation passive optical network.