Q-PASTE: A Cross-Layer Power Saving Solution for Wireless Data Transmission

Yang Song, Bogdan Ciubotaru, and Gabriel-Miro Muntean
Performance Engineering Laboratory, DCU, Ireland
Email: yang.song2@mail.dcu.ie, {bogdan, munteang}@eeng.dcu.ie

Abstract—This paper introduces a novel quality-oriented cross-layer energy-efficient data communication solution (Q-PASTE) for mobile devices, with two components at different network layers. First component, the Packet/Application manager (PAT) is deployed at the application layer of both service gateway and client host. The gateway level PAT shapes traffic into bursts to reduce the wireless transceiver’s duty cycle. The client-side PAT monitors each active session and informs the Medium Access Control (MAC) layer about their traffic-related behavior. The second component, the Slow Start Exponential and Linear Algorithm (STELA), deployed at MAC layer, adaptively adjusts the sleep/wake-up behavior of mobile device wireless interfaces in order to reduce energy consumption while also maintaining high Quality of Service (QoS) levels. Both mathematical analysis and simulation-based modeling and testing have been performed to evaluate Q-PASTE performance. Comparative results have shown significant energy saving achieved when using Q-PASTE, without compromising content delivery performance.

I. INTRODUCTION

Within the last few decades, the fast development of wireless technologies has determined an unprecedented growth in the popularity of both mobile devices and rich media applications. User experience depends much on the content, context, and also the quality of the communication service provided (i.e. bandwidth, connection reliability etc.). However limited battery lifetime may severely impact user satisfaction.

This paper presents a novel quality-oriented cross-layer energy-efficient solution for wireless communications (Q-PASTE), which increases power conservation at mobile devices, while still providing high Quality of Service (QoS) levels. The proposed solution consists of the Packet/Application manager (PAT) deployed at application-layer, and the Slow Start Exponential and Linear Algorithm (STELA) deployed at Medium Access Control (MAC) layer. Previous works of the authors [1][2] achieve energy efficiency through MAC layer adaptation. In this work, PAT and STELA work together as a cross-layer solution to shape the data traffic into bursts and adapt the behavior of the Wireless Network Interface Card (WNIC) in order to reduce unnecessary state switching and provide high QoS.

The remainder of the paper is organized as follows: Section II briefly describes some state-of-the-art energy efficient MAC solutions and latest traffic shaping technologies designed for wireless enabled mobile devices. Section III introduces and details the proposed algorithm—Q-PASTE. Section IV analyzes theoretically the performance of Q-PASTE, while section V presents Q-PASTE simulation results and compares them with those of other existing solutions such as the ones employed by IEEE 802.11 [3] and IEEE 802.16 [4]. Finally, the paper is concluded and a brief outlook on future work is provided.

II. RELATED WORKS

Different cross-layer solutions have been proposed to achieve energy efficiency or high QoS [5][6]. In this paper, we propose a cross-layer solution which employs the cooperation between MAC layer and application layer for balanced performance.

A. Energy-Efficient MAC

The MAC protocol mainly controls the behavior of the WNIC (i.e. the sleeping schedule) and therefore impacts the overall energy consumption of the transceiver. TRAMA [7] assigns time slots dynamically according to real time information on network topology and traffic load in order to improve efficient channel utilisation. LMAC [8] divides each time slot into a traffic control section and a data section to conserve power. IEEE 802.11 [3] has a built-in Power Saving Mode (PSM) where the Access Point (AP) works as a central coordinator which buffers data for sleeping hosts and notifies the attached hosts if any buffered data exists by beaconing on a regular basis. In [9], one radio interface is allocated for data and control messages transmission and the other one for signalling. Studies such as [10] take advantage of both schedule-based schemes and contention-based strategies. Z-MAC [10] uses a time slot allocation scheme in which each mobile node can contend for idle node’s slots.

B. Adaptive Traffic-Shaping

An intuitive way to conserve power for the wireless interface is to prolong the sleeping period and reduce the number of state switching. Solutions such as [11] use proxies between client and server to enable traffic shaping. In [12], the traffic shaper uses the UDP transport protocol and the data releasing algorithm depends on the estimated available bandwidth and the dynamic client buffer size. PASP [13] introduces proxies at both server and client sides. Server side proxy schedules packets into bursts and informs the client about the schedule of following data transmissions. PSM-throttling [14] presents a client-centric traffic shaping solution for TCP-based applications. It exploits the unused network bandwidth to reshape
TCP data into bursts and allow accurate prediction of the arriving time of packets at the client side.

III. Q-PASTE

Q-PASTE, is a quality-oriented cross-layer solution for wireless data delivery, which increases the energy saving, while maintaining the delivery performance for wireless hosts at high levels. Q-PASTE consists of an application traffic manager (PAT) and a sleeping window scheduler (STELA), as shown in Figure. 1.

The design of Q-PASTE is based on several assumptions. First, network traffic is assumed to have a bursty nature [15]. Burstiness refers to large amount of packets being transmitted within a very short time, followed by a long time of inactivity. Second, data traffic is assumed to be predictable based on historical data [16]. Third, the solution assumes an infrastructure-based network, where the gateway can control the packet scheduling. Moreover, the small gaps between data packets are grouped into longer intervals during which the WNIC can be powered off. Although each of the assumptions have been widely studied, Q-PASTE is the first novel solution which exploits them in an effective way to provide the balance between energy efficiency and high QoS levels.

A. PAT

PAT is employed at the application layer of both the gateway and the client device. At the client side, it keeps track of all active application sessions and informs the MAC layer module, STELA, through a cross-layer information flow whenever an application session ends or a new session starts. At the gateway side, PAT collects real time information of the number of packets ready to be transmitted to the clients and does not allow data releasing until the packets can be shaped into a burst. Although data scheduling can also be fulfilled at the server/client side, the packet shaping function of PAT is implemented as part of gateway functions in our energy efficient solution due to several reasons. First, client side shaping requires extra cost in terms of both hardware and software complexity, as burst should be formed before being received by the WNIC and thus an extra receiver unit should be installed. Second, if continuous data is manipulated at the server side to form bursts, the client node might be communicating with several servers at the same time, and therefore needs to wake up for each burst from each server. Moreover, as a burst is released as soon as the buffer size reaches the threshold, extra delay is introduced when packets from individual servers are considered separately as they will wait longer before being delivered. On the contrary, a service gateway, similar with media gateways which already exist on the market, can act as a centralized controller and gather data from all the servers for the client. This approach is the most efficient in terms of both energy saving and QoS level.

B. STELA

The adaptive sleep/wake-up scheduler, STELA, is employed at the MAC layer of the client host to maximize energy conservation and concerns downlink traffic. In the case of uplink traffic, the WNIC is switched on automatically once data reaches the MAC layer, bypassing STELA. STELA’s algorithm consists of three phases, as shown in Figure. 2. The initiation of the phases is self-adjusted according to the received data. The threshold adjusting phase is employed for optimal behavior of STELA.

1) Exponential Increase Phase: When STELA is restarted by the event of establishment or termination of an application session, the sleeping schedule is set to one beacon interval. The WNIC remains active if there is any packet arriving during channel sampling, or will be switched off immediately otherwise. In the context of bursty traffic, which is generated by PAT, the absence of buffered packets at the AP signals the end of one burst. Therefore the size of the sleeping window, $W_s$, is doubled when the WNIC samples the channel without detecting any incoming packet during the AP’s beaconing period. The phase is terminated once $W_s$ reaches the threshold value $W_{thre}$. This phase is restarted if any data is detected during channel sampling, which indicates the arrival of another data burst.

The exponential increase phase achieves high energy efficiency by enabling long sleeping intervals of WNIC, and at the same time does not compromise user experience due to several reasons. First, packets are transmitted from the gateway in bursts so there will not be any individual packets that would be delayed due to WNIC inactivity. Second, a traffic-based predictive threshold value $W_{thre}$ is set to prevent the aggressive growing of the sleeping window.
2) Linear Increase Phase: The linear increase phase is initiated after the exponential increase phase. During this phase the size of the sleeping window is increased by one beacon interval each time the WNIC samples the channel and detects no incoming data during beaconing period. The novelty of this phase lies in the fact that \( W_s \) grows at a moderate pace when the wireless host expects data to arrive very soon. When the \( W_s \) reaches \( W_{thre} \), the next burst is expected to arrive soon according to historical data arrival pattern, and therefore the wireless interface should wake up frequently enough in order not to miss data reception. Allowing slow growth of \( W_s \) further conserves power for wireless host and at the same time does little harm to the delivery performance.

3) Threshold Adjusting Phase: The threshold value \( W_{thre} \) directly impacts the performance of STELA as it determines the transition between the exponential increase and linear increase phases. A large \( W_{thre} \) value would lead to unacceptable increase in packet delays, while a small value would result in waste of energy. The threshold adjusting phase is triggered when the cross-layer information flow, as depicted in Figure 1, notifies STELA that an application session starts or terminates. Studies such as [16] demonstrate that data arrival patterns are highly regular, and therefore the first two data bursts are monitored and used to configure the optimal \( W_{thre} \) value.

The default value of \( W_{thre} \) is set to one beacon interval while \( W_s \) follows the STELA schedule. Every time the value of \( W_s \) is modified during either the exponential increase phase or the linear increase phase, \( W_s \) is compared with the value of \( W_{thre} \). If \( W_s \) has at least twice the value of \( W_{thre} \), then \( W_{thre} \) is doubled. This process continues until the first packet of a new burst is received. The interval between the two bursts is recorded as \( I_{sb} \).

\( W_{thre} \) is smaller than the size of the sleeping window \( W_s \) when the burst is detected, which guarantees early termination of the exponential increase phase and starting of the linear increase phase. Data arrival pattern fluctuates and consequently the conservative configuration of \( W_{thre} \) enables in time powering of the WNIC before data packets arrive.

IV. PERFORMANCE ANALYSIS

This section analyzes theoretically the performance of Q-PASTE in terms of energy saving and packet delay.

A. Energy Consumption

The WNIC could be in one of the four states: transmitting data (Tx), receiving (Rx), sleeping (Sl) or switching between states (Sw). For each state, the corresponding energy consumption \( E_m \) is calculated as the time \( T_m \) spent by the WNIC in that particular state, multiplied by the energy consumption per time unit \( U_m \). Therefore the total energy used by WNIC, \( E_t \) is the sum of the energy amounts spent in the four states \( M=\{\text{Tx, Rx, Sl, Sw}\} \) and is calculated based on equation (1).

\[
E_t = \sum_{m \in M} E_m = \sum_{m \in M} T_m * U_m
\]

When using the Q-PASTE power saving mechanism, the packet transmitting and receiving processes remain the same, while the new sleep/wake-up state transition schedule leads to different sleeping intervals and number of state switching processes. Due to the significantly lower energy spent during sleep states and the high energy cost of state switches, the Q-PASTE sleeping schedule should increase the sleep time \( T_{sl} \) and decrease switching time \( T_{sw} \) in order to result in energy saving.

Q-PASTE, involves a quick growth of the sleeping window which means the sleeping interval increases and the number of state switches decreases when no packet is predicted to arrive. Moreover, data packets are grouped into bundles which allow longer WNIC sleep intervals. Therefore, total energy consumption \( E_t \) is significantly reduced.

B. Delay Analysis

As part of network QoS, packet delay refers to the time interval elapsed from the moment a packet is sent to the time the packet successfully arrives at the destination. Packet delay \( D_t \) in wireless networks is measured as the sum of the following parts:

- Sender-side delay: \( D_{tr} \) refers to the time it takes to put all packets’ bits to the wireless medium. \( D_{tr} \) depends on the packet length and channel bandwidth.
- Access point (service gateway)-side delay: In infrastructure based networks, packets are first received and processed by the access point (service gateway) before being forwarded to the receiver. Therefore the gateway incurs processing delay \( D_{pr} \). Moreover, if multiple packets are buffered, queuing delay \( D_{qu} \) is introduced.
- Receiver-side delay: WNIC sleeping delay \( D_{sl} \).
- Propagation delay: Propagation delay \( D_{pr} \) refers to the time it takes for a packet to be transferred over a medium. It is the distance between the server and client divided by the propagation speed.

When using the same network architecture and configuration, transmission, processing and propagation delays remain the same. Queuing delay depends on the size of the burst set in the traffic shaper and can be controlled during configuration. Consequently the main variable of the total delay is \( D_{sl} \) which depends on the duty cycle of the radio transceiver. The value of \( D_{sl} \) for each packet is dependent on the packets position in a packet burst and the regularity of data flow. In our analysis, we focus on \( D_{ex} \), which is the extra delay introduced by Q-PASTE, as shown in equation (2).

\[
D_{ex} = D_{sl} - I_{ac} \tag{2}
\]

\( I_{ac} \) is the time taken by a packet to arrive at the mobile host if no power saving is on and is fixed for all compared schemes.

Two scenarios are considered, each having a different effect on the end-to-end packet delay. The first scenario involves a packet being the first packet of a burst and the arrival pattern of each burst follows the exact same pattern. Under these circumstances, the radio transceiver wakes up when the sleeping window reaches the threshold value \( W_{thre} \) which is
smaller than the recorded sleeping interval between bursts. Consequently the first packet of a burst could be detected in time after been processed by the gateway. Therefore, $D_{sl}$ is smaller than the interval between consecutive data bursts $I_{ob}$. This scenario is the ideal situation and has low probability. The second scenario involves a packet being the first of a burst and the bursts arrival pattern fluctuates either with early or late burst arrival. In this case, the extra delay introduced by Q-PASTE can be calculated based on equation (6).

$$D_{sl} = (2^1 + 2^2 + \ldots + 2^N) \times I_{bc} \tag{3}$$

This can be simplified as presented in equation (4).

$$D_{sl} = \sum_{n=1}^{N} 2^n \times I_{bc} \tag{4}$$

As the radio transceiver is switched on when the sleeping window reaches $W_{thre}$, the delay is computed between the times of the last burst and when $W_s$ equals $2^{N-1}$, giving:

$$I_{ac} = \sum_{n=1}^{N-1} 2^n \times I_{bc} \tag{5}$$

Therefore the extra delay introduced by Q-PASTE can be calculated based on equation (6).

$$D_{ex} = D_{sl} - I_{ac} = \sum_{n=1}^{N} 2^n \times I_{bc} - \sum_{n=1}^{N-1} 2^n \times I_{bc} = 2^N \times I_{bc} \tag{6}$$

However, as the binary exponential increase phase is not finished before the packet arrives, the sleeping window is smaller than $W_{thre}$, which means the condition in (7) is valid.

$$D_{ex} < W_{thre} \times I_{bc} \tag{7}$$

Case 2.2, as shown in Figure. 4 represents the scenario where the burst arrives later than in the first case, but earlier than the observed interval $I_{ob}$. When $W_s$ reaches $W_{thre}$, the WNIC samples the radio channel and switches back to sleeping mode, as no packet arrives. Then, the linear increase phase is initiated and $W_s$ is incremented by one and becomes $W_{thre} + 1$. The burst is ready for reception when the radio is switched on the next time, and therefore only $W_{thre}$ beacon periods are added to the overall packet delay.

$$D_{sl} = I_{ob} \tag{8}$$

$$D_{ex} = D_{sl} - I_{ac} = I_{ob} - I_{ac} \tag{9}$$

As we have the inequality from equation (10):

$$I_{ob} - W_{thre} \times I_{bc} < I_{ac} < I_{ob} \tag{10}$$

then:

$$I_{ob} - I_{ob} < D_{ex} < I_{ob} - I_{ob} + W_{thre} \times I_{bc} \tag{11}$$

Equation (11) leads to equation (12):

$$0 < D_{ex} < W_{thre} \times I_{bc} \tag{12}$$

Figure. 5 demonstrates the last scenario, i.e. Case 2.3, where the burst gets to the gateway later than the interval observed during the first round. The radio transceiver wakes up when $W_s$ reaches $W_{thre}$ and receives no packet, then sleeps for $W_{thre} + 1$ beacon intervals and samples the radio channel until the delayed burst is ready for transmission. We assume the size of the sleeping window is $W_{thre} + N$ when the burst is detected and the value of $N$ can be determined as soon as the condition in equation (13) is met.

$$D_{sl} - I_{ob} = (W_{thre} + 1 + W_{thre} + 2 + \ldots + W_{thre} + N) \times I_{bc} \tag{13}$$

This can be simplified as presented in equation (14).

$$D_{sl} - I_{ob} = \sum_{n=1}^{N} (W_{thre} + n) \times I_{bc} \tag{14}$$

Next the delay is calculated based on equation (15).

$$D_{sl} = I_{ob} + \sum_{n=1}^{N} (W_{thre} + n) \times I_{bc} \tag{15}$$
The burst arrives at the client side when sleeping window equals $W_{thre} + N$. However, in order to analyze the worst case scenario in terms of delay, we assume the burst is ready for transmission immediately after the previous wakeup, i.e. when the sleeping window is equal to $W_{thre} + N - 1$. This leads to equation (16).

$$I_{ac} = I_{ob} + \sum_{n=1}^{N-1} (W_{thre} + n) \cdot I_{bc}$$

(16)

$D_{ex}$ can be calculated based on equation (17).

$$D_{ex} = D_{sl} - I_{ac}$$

$$= I_{ob} + \sum_{n=1}^{N} (W_{thre} + n) \cdot I_{bc}$$

$$- I_{ob} - \sum_{n=1}^{N-1} (W_{thre} + n) \cdot I_{bc}$$

$$= (W_{thre} + N) \cdot I_{bc}$$

(17)

Although $N$ depends on the inter-arrival pattern of bursts, the extra delay is a linear function of the beacon interval which exhibits good performance compared with other energy efficient solutions. Above all, it has been proved that for packets that are not burst starting packets, which is true in most cases, the end-to-end delay is very short. Meanwhile, for those beginning packets of each burst, the extra delay introduced by Q-PASTE should be smaller than the observed interval between bursts in the first round of traffic or be a linear function of the beacon interval, which lasts a few milliseconds in most cases. Experimental results also prove that Q-PASTE guarantees both short delay and high energy efficiency.

V. SIMULATION-BASED EVALUATION

We conduct simulation experiments, using Network Simulator NS-3 [17], to evaluate the performance of our solution. Energy consumption and QoS level of Q-PASTE are computed and compared with those when the energy efficient solutions of IEEE 802.11 and IEEE 802.16 are used. The power saving mode in IEEE 802.11 employs a static sleep/wake-up schedule, while IEEE 802.16 consists of three power saving classes. The first one is the similar to IEEE 802.11, the second class lets the sleeping window size grow exponentially to achieve maximum energy saving, and the third class allows for a one-time sleeping window and is typically used for multicast or management traffic instead of data traffic. The binary exponential increase algorithm used in IEEE 802.16 is the most energy efficient and therefore is evaluated in these simulations. We evaluate the performance of IEEE 802.16 with the maximum sleeping window of 2 beacon intervals and 16 beacon intervals which represent the maximum energy efficiency and the best performance, respectively.

The simulation scenario involves two servers transmitting packets to a wireless host via the service gateway which functions as an access point. All traffic follows an on/off pattern with different intervals between packets tested. Packets are stored at the gateway and forwarded once the wireless host is awake and enough packets are shaped into a burst. We show the comparison results of average energy consumption per packet and average packet delay. The common simulation parameters [18] [19] are listed in Table I.

It can be seen from both Figure. 6 and Figure. 7 that Q-PASTE consumes much less energy than the other two schemes while maintains high level of QoS. For example, it saves 16% more energy than IEEE 802.16, while introducing three times less delay when data is generated with 1s on/2s off pattern at 1.0 Mbps. Testing results are affected by several factors:

Traffic pattern. The traffic pattern is affected by the data pattern generated at the server and the maximum number of packets that are buffered by PAT before data could be released to the $MaxPck$. At the server side, three on/off patterns are individually tested to validate the performance of Q-PASTE. It can be seen from Figure. 6(a) that the longer the off period, the more energy is saved by Q-PASTE when compared with the other two schemes. This is due to the fact that Q-PASTE is capable of adapting the sleeping schedule according to the monitored traffic and enables less wake-up of WNIC and thus better energy efficiency in the situation of long intervals between packet bursts. The observed results also prove that Q-PASTE is suitable for other application types and traffic patterns. For example even more energy can be saved for web traffic where the interval between packets could be much longer than the on/off regular traffic pattern. At the service gateway, $MaxPck$ has direct impact on the performance of Q-PASTE as the larger the value is, the longer it takes before data is forwarded and the less switching on time is required by WNIC. Meanwhile, the threshold adjusting algorithm of Q-

\begin{table}[h]
\centering
\caption{SIMULATION PARAMETERS}
\begin{tabular}{|l|l|}
\hline
Parameter & Value \\
\hline
Simulation duration & 200 s \\
Beacon interval & 102.4 ms \\
Packet size & 512 bytes \\
Transmitting energy & 1.5 W \\
Receiving energy & 0.75 W \\
Sleeping energy & 0.01W \\
Switching energy & 0.75 W \\
Switching duration & 2 ms \\
\hline
\end{tabular}
\end{table}

Fig. 5. Illustration of case 2.3, data burst arrives late
PASTE is capable of predicting the future traffic which benefits in the delivery performance, while the energy efficiency is achieved.

Data rate. We study the effect of varying the data rate of each server from 0.5 Mbps to 1.5 Mbps, shown in Figure. 7. The traffic pattern is set to 1s on/1s off. As data rate increases, the energy consumed per each packet decreases, as shown in Figure. 7(a). This is due to the fact that the more packets are transmitted during one time unit, the higher is the possibility that the WNIC has data to receive after being switched on, and therefore less energy is wasted on unnecessary waking up. Figure. 7(b) shows increase in delay for most cases as the queuing delay $D_{qu}$ grows with the increase in the number of packets. However, Q-PASTE-40 has a higher delay as it takes longer time to accumulate 40 packets before the data is released when data rate is low.

Maximum sleeping window size $MaxWin$. $MaxWin$ is applied to IEEE 802.16 only, and refers to the maximum size of the sleeping window. The smaller $MaxWin$ is, the more frequent WNIC is switched on and the more power is consumed. Therefore IEEE 802.16-16, which denotes $MaxWin$ configured as 16 beacon intervals, is much more energy efficient that IEEE 802.16-2, but at the same time affects seriously QoS.

VI. CONCLUSION

This paper introduces Q-PASTE, a cross-layer solution for wireless data delivery which is energy efficient and at the same time maintains high QoS levels. Q-PASTE employs a traffic shaper which forms data bursts at the service gateway and introduces a self-adaptive duty cycle which monitors real time traffic and predicts future data arrival patterns. The performance of Q-PASTE in terms of both energy efficiency and delivery performance is analyzed and evaluated. Experimental results show that, from power conservation perspective, Q-PASTE significantly outperforms two other well known schemes while providing a balanced delivery performance.

Future work includes a more extensive testing under complex network scenarios with multiple mobile stations. Additionally, the traffic shaper of Q-PASTE will be tailored to provide different QoS classes to various application types.

REFERENCES


