Energy-Aware QoS Adjustment of Multimedia Tasks with Uncertain Execution Time⋆

Wan Yeon Lee[1], Heejo Lee[2], and Hyogon Kim[2]

[1] Hallym University, Chunchon 200-702, South Korea
[2] Korea University, Seoul 136-701, South Korea
wanlee@hallym.ac.kr, {heejo,hyogon}@korea.ac.kr

Abstract. In order to make the best use of available energy budget, we propose a QoS adjustment method which maximizes the total QoS-provisioning of multimedia tasks with uncertain execution time. This method utilizes the probability distribution of task's execution time to determine an instant QoS level. Our experiments show that the proposed method gives 52% more QoS-provisioning than the conventional method using a constant QoS level derived from the worst-case time.

1 Introduction

A necessary feature for the mass computation of multimedia applications on wireless electronic devices is acceptable battery lifetime. If critical computation suddenly stops before its completion due to the shortage of available energy, it may result in great loss. Generally, we can reduce the energy consumption rate of battery and thus extend the battery's lifetime by decreasing the computation amount in a running application. In most cases, however, more computation is necessary to provide better quality of service(QoS). As a result, there is a demand to control both the computation amount of an application and its battery lifetime. Another issue to be considered is the uncertain execution time of multimedia applications. Their execution time heavily depends on their input data, but the information of input data is not available before starting execution.

In this paper, we propose an energy-aware QoS adjustment of multimedia tasks with uncertain execution time, which guarantees their worst-case execution time and maximizes the total QoS-provisioning gained from performing their computation in limited-energy environments. The proposed method assigns the highest QoS level to the earliest processing part of a task and decreases the QoS level gradually as the task progresses its execution. Even though the later running parts provide lower QoS level when being executed, they are rarely performed due to their lower probability. Statistically, this can support more QoS-provisioning than assigning a constant QoS level derived from the assumption that the task is always executed for the worst-case execution time.

Graceful degradation [1] or stopping the unimportant task [2] can allow mission-critical tasks to run for a longer period of time when available energy is low. These methods considered the energy management of tasks with fixed execution time but not that with uncertain execution time. Lee *et al.* [3] addressed a similar method to maximize QoS-provisioning while using the limited amount of energy. They proposed a general approach, however, there is no relationship between QoS and energy consumption.

2   Preliminaries

The notion of QoS in multimedia tasks includes various characteristics such as resolution, delay, jitter, loss rate, etc. Among several properties of QoS, the amount of data computed by a task is referred to as *fidelity*. For example, the display size of a video player or the amount of lossy compression applied to a video stream can be the fidelity of the video streaming task. Experimental results [1,4] showed that the energy consumption of a multimedia task is proportional to its fidelity such as the image size being processed, the transmission rate on wireless networks, and its running time.
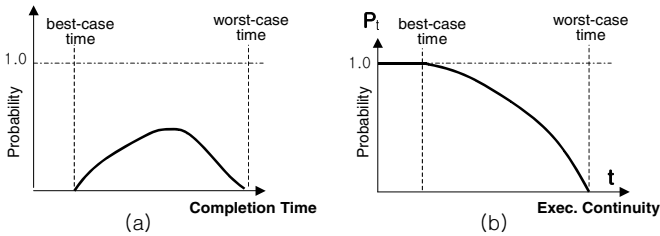

Fig. 1. Probability distribution of task's execution times

The systems can be designed for scheduling multimedia tasks with their worst-case execution times, resulting in significant waste of energy. Approaches based on the average case are likely to suffer from lacking resource, particularly in the worst case. More accurate models are based on the probability distribution of execution times. Figure 1(a) shows an example of probability distribution of task's execution times and Figure 1(b) shows the tail distribution of its cumulative probability distribution, denoted as P_t at a time t. P_t is the probability that the task continues its execution for at least time t.

3   Proposed Method

In this paper, we study how to maximize the fidelity benefit of a single multimedia task with uncertain execution time on a limited-energy system. It is assumed

that the other tasks running on the system have nearly fixed execution times and their operations are stable in most cases. Then, we can formulate the total energy consumption amount of all tasks as follows:

$$\int_0^{T_w} m \cdot \mathcal{F}(t) \cdot dt + b \cdot T_w$$

where T_w is the worst-case execution time, m and b are an application-specific coefficient associated with the multimedia task and a constant coefficient associated with background tasks respectively, and $\mathcal{F}(t)$ is the instantaneous fidelity of the multimedia task at a time t. When E_{max} denotes the amount of available energy budget, $\int_0^{T_w} m \cdot \mathcal{F}(t) \cdot dt + b \cdot T_w \leq E_{max}$. For simplicity, we define B_{max} as $\int_0^{T_w} \mathcal{F}(t) \cdot dt \leq \frac{E_{max} - T_w \cdot b}{m} = B_{max}$.

As a benefit measure of fidelity, we define *Perceptional Resolution* of a task as the resolution of 2-dimensional images computed by the task. User perception against 2-dimensional images is proportional to the square root of their sizes. The problem dealt with in this paper is to maximize the total Perceptional Resolution of a task during its execution, subject to the constraint that the amount of total energy consumption during its execution is no larger than the given energy budget. The problem can be formulated as follows:

$$Maximize \int_0^{T_w} \sqrt{\mathcal{F}(t)} \cdot P_t \cdot dt = \int_0^{T_w} (\mathcal{F}(t)/P_t^2)^{1/2} \cdot P_t^2 \cdot dt \qquad (1)$$

subject to

$$\int_0^{T_w} \mathcal{F}(t) \cdot dt = \int_0^{T_w} (\mathcal{F}(t)/P_t^2) \cdot P_t^2 \cdot dt \leq B_{max}. \qquad (2)$$

By Jensen's inequality [5], this maximization occurs when all values of $\mathcal{F}(t)/P_t^2$ are the same. Then Equation (1) has an upper bound as follows:

$$\int_0^{T_w} \sqrt{\mathcal{F}(t)} \cdot P_t \cdot dt \leq \sqrt{C} \cdot \int_0^{T_w} P_t^2 \cdot dt = \sqrt{B_{max}} \cdot \sqrt{\int_0^{T_w} P_t^2 \cdot dt}$$
$$\text{if and only if } \mathcal{F}(t) = C \cdot P_t^2 = \frac{B_{max}}{\int_0^{T_w} P_t^2 \cdot dt} \cdot P_t^2.$$

Since P_t always decreases as t increases, a task decelerates its fidelity as its execution goes on in the optimal schedule. The previous work [1,6] showed that the overhead to dynamically decelerate fidelity is negligible. This approach can be applicable to similar problems which try to maximize or minimize another metric of QoS, instead of Perceptional Resolution.

4   Evaluation

The proposed method determines an instant QoS level based on the distribution of task's execution times while the conventional method determines a constant QoS level based on the worst-case execution time. For evaluation metric, we define *QoS Increment* as $\frac{PS_p - PS_w}{PS_w} \times 100$, where PS_p and PS_w are the total Perceptional Resolutions in the proposed method and in the conventional method,

respectively. We implement the proposed method and the conventional method on a practical multimedia application and compare their performance. In these experiments, we consider the case that a user enjoys watching a live broadcasting of the 2006 Major League Baseball(MLB) on a mobile device with limited energy budget. Figure 2(a) shows the playing time distribution of the former 83 games and the latter 83 games in the 2006 season of the New York Yankees team [7]. Figure 2(b) shows the average performance of the latter 83 games when they utilizes the proposed method based on the distribution information of the former 83 games. The longest playing time of MLB (i.e., 486 minutes) is used for the worst-case playing time. These experiments show that the proposed method provides 52% QoS Increment when T_w is 486 minutes. As the value of T_w is decreased, its performance goes down but the risk to irresistibly stop the broadcasting in the middle of games is also increased.
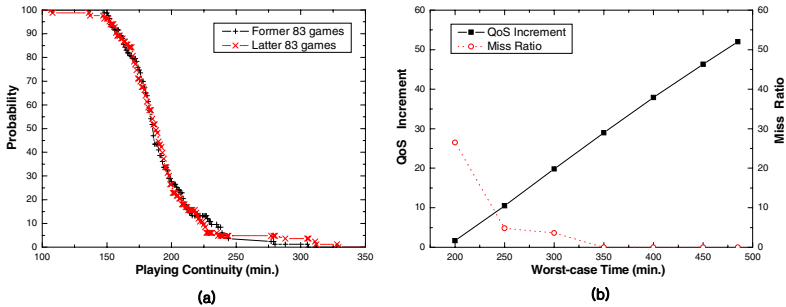

Fig. 2. Experiment results of a multimedia application

References

1. Flinn, J., Satyanarayanan, M.: Managing battery lifetime with energy-aware adaptation. ACM Trans. Comp. Syst. **22**(2) (May 2004) 137–179
2. Tamai, M., Sun, T., Yasumoto, K., Shibata, N., Ito, M.: Energy-aware video streaming with QoS control for portable computing devices. In: ACM NOSSDAV. (2004) 68–73
3. Lee, C., Lehoczky, J., Rajkumar, R., Siewiorek, D.: On quality of service optimization with discrete QoS options. In: IEEE RTAS. (June 1999) 276–286
4. Feeney, L.M., Nilsson, M.: Investigating the energy consumption of a wireless network interface in an ad hoc networking environment. In: IEEE INFOCOM. (April 2001) 1548–1557
5. Krantz, S., Kress, S., Kress, R.: Jenen's Inequality. Birkhauser (1999)
6. Yuan, W., Nahrstedt, K.: Energy-efficient soft real-time CPU scheduling for mobile multmedia systems. In: ACM SOSP. (August 2003) 149–163
7. ESPN: MLB scoreboard. http://sports-ak.espn.go.com/mlb/scoreboard