# Optimal Trade-off Between Energy Efficiency and Average Delay

Edmund M. Yeh, Thierry E. Klein

# Optimal Trade-off Between Energy Efficiency and Average Delay

Edmund M. Yeh
Department of Electrical Engineering
Yale University
New Haven, CT 06520, USA
e-mail: edmund.yeh@yale.edu

Thierry E. Klein
Wireless Research Laboratory
Bell Labs - Lucent Technologies
Murray Hill, NJ, 07974 USA
e-mail: tek@lucent.com

## I. Introduction

In wireless networks consisting of sensors and mobile devices, battery power is a severe resource constraint. Energy-efficient communication techniques are therefore crucial to the long-term survivability of the network. In communication theory, it is well known that for a bandlimited additive white Gaussian noise (AWGN) channel with bandwidth $W$ and one-sided noise spectral level $N_0$, the minimum energy per information bit $E_{b,min}$ achieves its *minimum* value $N_0 \ln 2$ as $W \to \infty$. Interpreting this trade-off between energy efficiency and bandwidth in the more general context of available *degrees of freedom*, it can be verified that, for a *fixed bandwidth* and noise level, the minimum energy per bit decreases as the transmission time per bit increases. Thus, greater energy efficiency is gained by using *low power* and *long transmission delays*. Using this observation, recent work in [1, 2] examines a scheduling problem in which energy efficiency is traded off with the need to transmit a finite number of packets by a *fixed deadline*. The findings in [1, 2] are important for communication scenarios where a group of time-sensitive data packets (such as observations from sensors) must be delivered to its destination (such as a sensor fusion site) by a given deterministic deadline. There are also many situations, however, where time-critical information must reach its destination promptly *on an ongoing basis*. In this paper, we focus on the latter scenario, and investigate the trade-off between energy efficiency and the *average system delay* of bits. We show that minimizing a weighted combination of energy per bit and average delay per bit is equivalent to a convex optimization problem which can be solved exactly using an iterative "off-line" algorithm, where packet arrival times and packet lengths are known beforehand. The structure of the solution turns out to be very different from that of the solution to the deadline problem of [1, 2].

## II. Model

We focus on a single user communicating over a noisy channel. For simplicity of exposition, assume that all packets to be transmitted contain $b$ bits. The results presented here generalize naturally to the case of variable packet lengths. Further, assume that the whole of each packet must be sent at a fixed rate. Let $e(\tau)$ denote the minimum energy necessary to send a packet of $b$ bits subject to some fixed reliability criterion in $\tau$ seconds, at rate $b/\tau$. We refer to $e(\tau)$ as the *energy function.* We make the following assumptions regarding $e(\tau)$: (1) $e(\tau)$ is a positive, differentiable, decreasing, strictly convex function of $\tau$, and (2) $\lim_{\tau\to 0} e(\tau) = \infty$. To see that these assumptions are reasonable for the AWGN channel, consider a few illustrative examples. If we use the Shannon capacity formula, then $e(\tau) = \tau N_0 W(2^{b/(\tau W)} - 1)$, which clearly satisfies assumptions 1 and 2. Operating below Shannon capacity implies that reliable communication is *possible*, but it does not

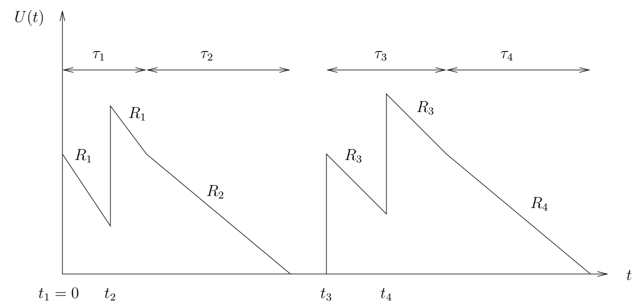


Figure 1: The unfinished work function for four packet arrivals. The packet arrival times are $t_1, \ldots, t_4$ and the packet transmission times are $\tau_1, \ldots, \tau_4$. Each packet contains $b$ bits.

guarantee any given level of error probability for the transmitted packet. To guarantee that each packet be received with error probability no more than a specified level, it is necessary to consider error exponents [3]. For the AWGN channel, it can be shown that the minimum energy to send a packet of $b$ bits in $\tau$ seconds with error probability no more than $P_e$ is $e(\tau) = \tau N_0 W(1+\rho)(2^{c/(\rho\tau W)} - 1)$, where $c = \rho b - \log_2(P_e/4)$, and $\rho \in [0, 1]$. Thus, assumptions 1 and 2 are again satisfied. Finally, using a standard uncoded $M \times M$ QAM system over an AWGN channel, it can be verified that to meet a (packet) error probability of $P_e$, $e(\tau) = \frac{-\tau N_0 W \ln(P_{es}/4)}{3b}(2^{b/(\tau W)} - 1)$, where $P_{es} = 1 - (1 - P_e)^{2\log_2 M/b}$. Thus, assumptions 1 and 2 are quite appropriate for a number of communication schemes.

## III. Optimal Off-line Schedule

Suppose a single user wishes to transmit $n$ packets, each with $b$ bits, over a noisy channel. Assume that the packet arrival times $t_1 = 0 < t_2 < \cdots < t_n$ are *known at time 0.* Let $e(\tau)$ be the energy function. The transmitter must choose the transmission times $(\tau_1, \ldots, \tau_n)$ (and therefore transmission rates $R_i = b/\tau_i$) to minimize a weighted combination of the minimum total energy $\sum_{i=1}^{n} e(\tau_i)$ to send $n$ packets, and the average bit delay. Now the key step is to realize that, because of Little's Law, minimizing the average bit delay is equivalent to minimizing the *area under the unfinished work function* $U(t)$, where $U(t)$ is the total number of untransmitted bits in the system at time $t$. Figure 1 shows an example of the unfinished work function for four packet arrivals. Note that $U(t)$ is a function of the transmission times $(\tau_1, \ldots, \tau_n)$. The off-line optimization problem can thus be stated as follows:

$$\min_{(\tau_1,\ldots,\tau_n)\in[0,\infty)^n} \sum_{i=1}^{n} e(\tau_i) + \lambda \int_0^\infty U(t)dt \qquad (1)$$

where $\lambda > 0$ is a weighting factor.

We now show that the objective function in (1) is in fact *convex* in $\quad = (\tau_1, \ldots, \tau_n)$. From Figure 1, it can be seen that $\int_0^\infty U(t)dt$ is the sum of triangular and rectangular areas. Let $\gamma_i$ be the *waiting time* for the $i$th packet, *i.e.,* the time between arrival time $t_i$ and the instant when the transmitter starts sending packet $i$. We see that when the $i$th packet is waiting for service, it contributes a rectangular area of $b\gamma_i$ to $\int_0^\infty U(t)dt$. When packet $i$ is in service, it contributes a triangular area of $\frac{1}{2}b\tau_i$. Therefore, the objective in (1) now becomes $\sum_{i=1}^n [e(\tau_i) + \lambda b(\frac{\tau_i}{2} + \gamma_i)]$. Next, observe that the waiting times are given by the recursion $\gamma_0 = \gamma_1 = 0$, and for $i = 2, \ldots, n$, $\gamma_i = \max\{0, \gamma_{i-1} + \tau_{i-1} - d_{i-1}\}$, where for $j = 1, \ldots, n-1$, $d_j = t_{j+1} - t_j$ is the inter-arrival time between packet $j$ and $j+1$. Thus, (1) is equivalent to minimizing

$$\sum_{i=1}^{n} \left[ e(\tau_i) + \lambda b \quad \frac{\tau_i}{2} + \max\{0, \gamma_{i-1} + \tau_{i-1} - d_{i-1}\} \quad \right] \qquad (2)$$

over all $(\tau_1, \ldots, \tau_n) \in [0, \infty)^n$. Since the max function is *increasing and convex* in its argument and the energy function $e(\tau)$ is convex in $\tau$, we conclude that the objective in (1)-(2) is convex in , and therefore the solution is given by the Kuhn-Tucker Theorem. Note first that due to assumption 2, the optimal solution $^*$ must satisfy $\tau_i^* > 0\ \forall i$. It is therefore possible to restrict the search for the optimal solution to stationary points. Next, note that the objective function in (2) takes on different forms depending on whether each $\gamma_i^*$ is positive or zero, where $\gamma_i^*$ is the waiting time of the $i$th user corresponding to the optimal solution $^*$. The transmission of packet $i$ is said to *overlap* with those of $k$ $(0 \leq k \leq n-i)$ subsequent transmissions, namely the transmissions of packets $i+1$ to $i+k$, if $\gamma_{i+1}^* > 0, \gamma_{i+2}^* > 0, \ldots, \gamma_{i+k}^* > 0$ and $\gamma_{i+k+1}^* = 0$ (with $\gamma_{n+1}^*$ defined as 0). It turns out that if the transmission of packet $i$ overlaps with $k$ subsequent transmission in the optimal solution, then $\tau_i^* = y_k$, where $y_k$ is the root of the equation $e'(\tau) + \lambda b(2k+1)/2 = 0$. Notice that since $e(\tau)$ is assumed to be strictly convex, the $y_k$'s are strictly decreasing in $k$. Thus, we obtain an important monotonicity result.

**Lemma 1** *Let $^* = (\tau_1^*, \ldots, \tau_n^*)$ be the optimal solution to (1)-(2), and let $^* = (\gamma_1^*, \ldots, \gamma_n^*)$ be the corresponding optimal waiting times. Suppose for some $i, 1 \leq i \leq n$, and some $j, 1 \leq j \leq n-i$, $\gamma_{i+1}^* > 0, \gamma_{i+2}^* > 0, \ldots, \gamma_{i+j}^* > 0$ and $\gamma_{i+j+1}^* = 0$, then $\tau_i^* < \tau_{i+1}^* < \cdots < \tau_{i+j}^*$.*

We now point to two fundamental differences between the optimal solution to the average delay problem in (1)-(2) and the solution to the deadline problem in [1, 2]. First, Lemma 1 says that during any period of overlap in the optimal solution, the transmission times must be *strictly increasing* in the index. This is a reflection of the "slow truck effect": a long transmission time for a packet ahead of many other packets waiting behind contributes to the delay of all waiting packets. This result contrasts sharply with Lemma 1 of [1], where the optimal transmission times in the deadline problem are *decreasing* in the index. Second, our focus on the average *system delay* implies that the solution to (1)-(2) accounts for both the delay cost of waiting *and* the delay cost of transmission. When the "price" $\lambda$ of system delay is sufficiently large, the optimal transmission times $\tau_i^*$ must be short, so that it is entirely possible for the transmitter to be idle at times, having finished sending all packets in queue. This contrasts sharply with the finding in the deadline problem of [1, 2], where packet transmission times are stretched as far as possible (subject to meeting the deadline), so that the transmitter is never idle (before the deadline).

Using the results developed above, we now demonstrate an iterative algorithm which produces the optimal solution $^*$ to problem (1)-(2) by successively checking for transmission overlaps and iteratively updating the transmission times. In the following, we define $t_{n+1} = \infty$.

**Iterative Algorithm for Off-line Optimization**

$i = 1$;
`while` $i \leq n$ `do begin`
  $\tau_i = y_0$;
  $k = 1$;
  `while` $\tau_i + \ldots + \tau_{i+k-1} > t_{i+k} - t_i$ `begin`
    `for` $j = 0$ `to` $k$
      $\tau_{i+j} = y_{k-j}$;
    $k = k + 1$;
  `end`
  $i = i + k$;
`end`

**Theorem 1** *The iterative algorithm above produces the optimal solution $^*$ to the problem in (1)-(2).*

PROOF: The theorem can be proved using the Kuhn-Tucker conditions. We skip the proof due to space constraints.

## IV. Conclusions and Extensions

In this work, we have formulated an optimization problem which characterizes the optimal trade-off between energy efficiency, measured in minimum energy per transmitted bit, and average delay, measured in average delay per bit. We showed that the optimization is convex, and produced an iterative off-line algorithm which solves the problem exactly when packet arrival times and packet lengths are known at time 0. The structure of our solution contrasts sharply with that of the solution given in [1, 2] for the deadline problem.

The off-line iterative algorithm naturally extends to situations where the different packets contain different numbers of bits $b_i$. It also extends to the multi-access (uplink) and broadcast (downlink) communication scenarios, under the additional assumption that, at any given time, only one user is granted channel access (time-sharing). In this case, different users may have different energy functions. Thus, we replace $e(\tau_i)$ in (1)-(2) by $e_{j_i}(\tau_i)$, where $e_{j_i}(\tau_i)$ is the energy function of the user sending (or receiving) packet $i$. As long as the energy function of each user satisfies assumptions 1 and 2, the optimization remains convex, and the iterative algorithm can be modified to give the optimal solution. Finally the off-line iterative algorithm suggests an on-line version where the arrival times (and/or the packet lengths) are not known at time 0. Our current work aims at assessing the performance of this on-line algorithm.

## References


[1] B. Prabhakar, E. Uysal-Biyikoglu, and A. El Gamal, "Energy efficient transmission over a wireless link via lazy packet scheduling," in *Proc. Infocom 2001*, (Anchorage, Alaska), 2001.

[2] A. El Gamal , C. Nair, B. Prabakhar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proc. Infocom 2002*, (New York City), 2002.

[3] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.