Energy Efficiency of Media Access Protocols for Mobile Data Networks

A. Chockalingam, *Senior Member, IEEE*, and Michele Zorzi, *Senior Member, IEEE*

***Abstract*—As mobile terminals are powered by a finite battery source, energy constraints play a major role in the design of wireless communications systems. In this letter, based on a general analytical framework, we study the energy efficiency of a class of multiple access schemes, using the average number of correctly transmitted packets for a given amount of allocated energy as an appropriate metric. We show that a good choice of the protocol rules can significantly improve the energy efficiency.**

***Index Terms*—Energy efficiency, media access.**

I. Introduction

PORTABLE user terminals for mobile communications must rely on limited battery energy for their operation. The design of protocols for such applications must consider judicious use of the available energy resources, and should exploit the characteristics of the wireless environment toward improved efficiency. In particular, error correlations (naturally present in wireless channels) introduce memory which can, in principle, be exploited for a variety of purposes, including energy conservation.

Recent research results which take a broad view of energy management are presented in [1]–[5]. It has been recognized that energy conservation is a task which should be performed at all levels of the protocol stack (and not only limited to the search for better batteries or lower-power circuits), so that it should be an objective in the design of a communications system as a whole. In [1], Bambos and Rulnick study the optimization of power control strategies to maximize the battery life under QoS constraints. Energy performance of error control schemes is studied by Zorzi and Rao [2], and by Lettieri *et al.* [3].

In this letter, we focus on media access protocols for wireless data networks. The issue of energy consumption of media access protocols has been addressed in [4], and a protocol designed based on energy conservation principles has been proposed in [5]. Unlike those papers, where a detailed analysis of the energy consumption performance of some protocols is carried out in an ad hoc fashion, here we consider a general framework for such a study, based on Markov analysis and the theory of renewal reward processes, which can be applied to a very broad class of protocols. Also, energy efficiency (as defined in [2]) is used here as a more appropriate metric than just battery life. As a simple example of application, we compare various versions of a specific media access protocol and are able to assess the relative advantages of each version depending on the parameters that characterize the wireless channel.

Paper approved by A. Goldsmith, the Editor for Wireless Communication of the IEEE Communications Society. Manuscript received January 21, 1998; revised May 10, 1998. This work was supported in part by the Center for Wireless Communications, University of California at La Jolla, San Diego, CA. This paper was presented in part at the IEEE 48th Vehicular Technology Conference (VTC'98), Ottawa, Canada, May 1998.

A. Chockalingam is with Qualcomm, Inc., San Diego, CA 92121-2779 USA (e-mail: chockal@qualcomm.com).

M. Zorzi is with the Center for Wireless Communications and the Department of Electrical and Computer Engineering, University of California at La Jolla, San Diego, CA 92093-0407 USA (e-mail: zorzi@ece.ucsd.edu).

Publisher Item Identifier S 0090-6778(98)08129-X.

II. Energy Efficiency Analysis

In an energy constrained environment, when the channel is bad over a long period of time, continued transmission of data packets in a message or repeated transmission of data packets until success may lead to wasted energy due to many unsuccessful data packet transmissions. On the other hand, in the presence of a rapidly fading channel, error recovery through retransmission may prove to be beneficial. In order to compare the energy performance of the various protocols under different fading scenarios using a unified metric, we study the *energy efficiency* of a protocol, which was introduced in [2] as

$$\frac{\text{total amount of data delivered}}{\text{total energy consumed}}. \quad (1)$$

We assume here that the protocol evolution can be tracked by means of a Markov chain with finite state space Ω. For example, this is the case for a protocol with finite-state machine in the presence of Markovian errors, a situation which is a good approximation of reality in a number of situations [6], [7]. By appropriately defining metrics on the transitions of this chain, renewal reward analysis allows to compute throughput and energy performance [2], [8].

Let P_{ij} be the transition probability from state i to state j, and let π_i be the steady-state probability of the chain being in state $i \in \Omega$. It is possible to define various semi-Markov processes in which this Markov chain is embedded [8]. In general, consider two reward functions, $R^{(1)}$ and $R^{(2)}$, where $R_{ij}^{(1)}, R_{ij}^{(2)}$ are quantities associated with transition ij, and let $R^{(1)}(\tau), R^{(2)}(\tau)$ be the cumulative values of those functions, i.e., the total reward earned through the system evolution in the time interval $[0, \tau]$. From renewal theory [9], we have the following fundamental result:

$$\lim_{\tau\to\infty}\frac{R^{(1)}(\tau)}{R^{(2)}(\tau)} = \frac{\sum_{i\in\Omega}\pi_i\sum_{j\in\Omega}P_{ij}R_{ij}^{(1)}}{\sum_{i\in\Omega}\pi_i\sum_{j\in\Omega}P_{ij}R_{ij}^{(2)}} \quad (2)$$

which can be easily computed for a number of cases of interest. For example, let S_{ij}, C_{ij}, and D_{ij} be the average number of successfully received packets, amount of consumed energy, and time delay associated with transition ij. Then, if $R^{(2)} = D$, evaluation of (2) for $R^{(1)} = S$ and $R^{(1)} = C$ gives the average throughput and energy consumption, respectively.[1] On the other hand, the choice $R^{(1)} = S$ and $R^{(2)} = C$ yields the energy efficiency of the protocol. Therefore, once the Markov chain for the protocol evolution has been found, all the relevant performance metrics can be easily computed from the above.

III. Example of Application

The access protocol considered in this letter can be viewed as a hybrid protocol employing the slotted ALOHA and reservation concepts (a detailed description of the protocol can be found in [10]). Before transmission of a message, a header packet is sent on a contention basis. The base station notifies mobiles about the current channel reservation status by broadcasting a busy/idle flag, which indicates that the channel is reserved if busy and that the channel is available for contention if idle. A more articulated feedback information may be needed for proper protocol operation, especially if the capture effect is exploited (see [10] for details).

In the *basic* protocol, following a header packet success, all data packets in the message are continuously sent regardless of whether the data packets are received correctly or not. The recovery of such data packet errors is left to the higher layer protocols. The *error-detect* (*ED*) protocol reacts to data packet errors by aborting the ongoing message transmission (thereby releasing the channel to other users), and yields better throughput when the channel fading is highly correlated. The *retransmission* protocol, on the other hand, tries to recover erroneous data packets by adopting a "persist-until-success" retransmission strategy, and provides better throughput compared to *basic* and *ED* protocols under fast fading conditions. Detailed throughput and delay performance analyzes of these protocols can be found in [10], [11]. The focus of this contribution is rather on the energy efficiency performance which, even though mostly overlooked in the literature, is crucial in a mobile radio environment.

In order to adequately track the protocol evolution over a channel characterized by Markovian packet error process [6], [7] and instantaneous and error-free feedback, the following five states are sufficient: 1) idle; 2) header packet success; 3) header packet failure; 4) data packet success; and 5) data packet failure.

Let N be the number of users and let λ be the message generation probability of each user in each slot. Also, let p and q be the conditional probabilities of a packet success in slot $t+1$ given a success in slot t, and of a packet failure in slot $t+1$ given a failure in slot t, respectively, so that $P_E = (1-p)/(2-p-q)$ is the average packet error rate [10]. Consider the *basic* protocol. If slot t is idle or contains a header failure, an idle flag will be transmitted at the beginning of slot $t+1$ and no data packet transmission can occur in slot $t+1$. Assuming no capture in header packet transmission, the system state in slot $t+1$ will be *idle* with probability $X_0 = (1-\lambda)^N$ (no arrivals), *header packet success* with probability $X_1 = N\lambda(1-\lambda)^{N-1}(1-P_E)$ (one arrival and packet success), or *header packet failure* with probability $X_2 = 1 - X_0 - X_1$ otherwise. If a header success occurred in slot t, then slot $t+1$ will be reserved for data transmission, which will be a success or a failure with probability p and $1-p$, respectively. Finally, if a data packet transmission occurred in slot t, then the message will end or continue with probability g_d and $\overline{g}_d = 1 - g_d$, respectively (i.e., geometic distribution with mean $1/g_d$ is assumed for the message length). In the former case, transitions will occur as from the *idle* state. In the latter case, the next packet transmission will be successful or unsuccessful according to the channel transition probabilities, p and q. The transition probability matrix for the *basic* protocol can then be written as

$$\boldsymbol{P} = \begin{bmatrix} X_0 & X_1 & X_2 & 0 & 0 \\ 0 & 0 & 0 & p & 1-p \\ X_0 & X_1 & X_2 & 0 & 0 \\ g_d X_0 & g_d X_1 & g_d X_2 & \overline{g}_d p & \overline{g}_d (1-p) \\ g_d X_0 & g_d X_1 & g_d X_2 & \overline{g}_d (1-q) & \overline{g}_d q \end{bmatrix} \tag{3}$$

The transition matrices of the *ED* and the *retransmission* protocols are identical to (3), except for the last row. In the *ED* protocol, the slot after a data packet failure is made available for contention, and therefore the fifth row of the matrix is the same as the first row. In the *retransmission* protocol, a data packet failure is always followed by its retransmission so that the last row of the matrix is to be changed to $(0, 0, 0, 1-q, q)$.

In all cases, only visits to state 4 correspond to successful transmissions, so that $S_4 = 1$ and $S_1 = S_2 = S_3 = S_5 = 0$.[2] A busy flag corresponds to a single packet transmission (the channel is reserved), whereas an idle flag corresponds to a contention slot with an average number of transmissions equal to λN. Therefore, since a busy flag will result in the system being in state 4 or 5 and an idle flag will result in the system being in state 1, 2, or 3, we can assign the energy consumption metrics: $C_4 = C_5 = 1$ and $C_1 = C_2 = C_3 = \lambda N$. Finally, since the model tracks the slot-by-slot evolution of the protocol, we have for the delay metrics $D_i = 1$ for all i. The energy efficiency of different protocols is the calculated by using the appropriate transition matrix $\boldsymbol{P}$ in (2).

Figs. 1 and 2 show the energy efficiency as a function of the achievable throughput for the different versions of the access protocol. The parameters considered in the plots are: number of mobiles $N = 10$, fading margin $F = 5$ dB, no capture, average message length 10 packets. Each curve is generated by varying the arrival rate, λ. Fig. 1 gives the performance when the fading process is *slow*, i.e., the normalized Doppler frequency $f_D T = 0.02$. Fig. 2 gives the results when the fading is *fast*, i.e., $f_D T = 0.64$. The relationship between the physical channel parameters F and $f_D T$, and the Markov parameters of the packet error process is detailed in [6]. Note that at a carrier frequency of 900 MHz and a packet duration of 10 ms, the $f_D T$ values of 0.02 and 0.64 correspond to mobile user speeds of about 2.5 km/h (e.g., pedestrian user) and 80 km/h (e.g., vehicular user), respectively. The figures show that

[1] Ergodicity of all processes involved will be assumed throughout.

[2] Note that here we can use a simplified approach where states instead of transitions are labeled, i.e., the transition metrics only depend on either the origin or the destination. Also, header packet successes are not counted as useful throughput.

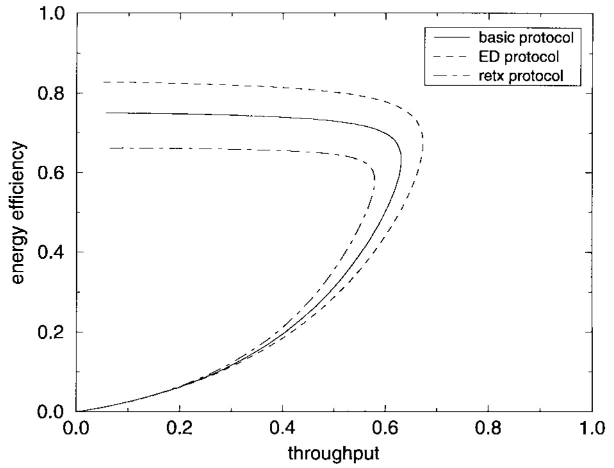

Fig. 1. Energy efficiency versus throughput for various protocols in *slow fading*. Normalized Doppler frequency $f_D T = 0.02$, number of users $N = 10$, average message length ten packets, no capture, fading margin $F = 5$ dB.

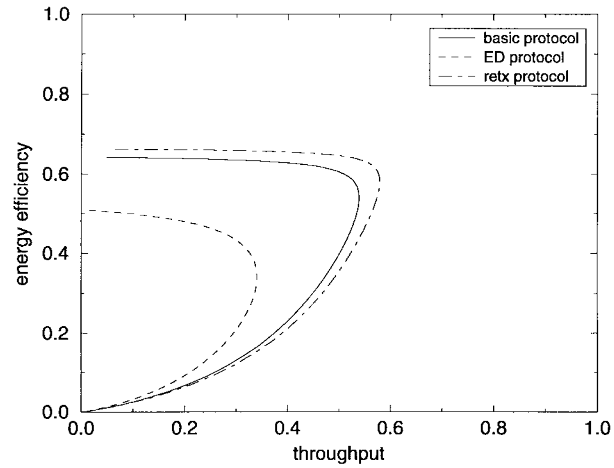

Fig. 2. Energy efficiency versus throughput for various protocols in *fast fading*. Normalized Doppler frequency $f_D T = 0.64$, number of users $N = 10$, average message length ten packets, no capture, fading margin $F = 5$ dB.

for very small arrival rate $(\lambda \to 0)$ the throughput goes to zero whereas the energy efficiency is maximum (no collisions). As λ is increased, corresponding to traveling clockwise along the curves, the energy efficiency is degraded, whereas throughput increases up to some optimal value of λ (equal to $1/N$, as discussed in [10]), after which it decreases due to too many collisions. The knee of the curves is the desired operating point for the system, and it is seen from Fig. 1 that for slow fading the *ED* protocol has the best throughput and energy efficiency, whereas the worst performance is achieved by the *retransmission* protocol. This was to be expected, since, in the presence of significant correlation between successive errors, aborting the transmission may be the best thing to do. For example, if the average length of an error burst is comparable to the average message length, completing the message transmission after an error may lead to unsuccessful transmissions with consequent waste of bandwidth and energy. Insisting on retransmission is the worst thing to do in this case. On the other hand, when the packet errors are almost independent, a single retransmission may lead to successful message completion, whereas the *ED* strategy may unnecessarily abort messages. In this case, as illustrated in Fig. 2, the *retransmission* protocol shows the best performance, and the *ED* protocol the worst. Therefore, we may conclude that the *ED* protocol is energy efficient for pedestrian user speeds, whereas the *retransmission* protocol is more efficient for vehicular user speeds.

As a refinement to the *ED* and the *retransmission* protocols, instead of terminating the data transmission at the first instance of a data packet failure (as in the *ED* protocol), or repeatedly sending a data packet until success (as in the *retransmission* protocol), the base station could allow the mobile to resend a lost data packet only up to a certain number of times (defined as a parameter, n_r), after which the mobile is asked to abort the data transmission. This is a generalized form of the protocol, which includes *ED* $(n_r = 0)$ and *retransmission* $(n_r \to \infty)$ as special cases.

In Fig. 3, the energy efficiency curves, for the maximum throughput performance point corresponding to $\lambda = 1/N$,

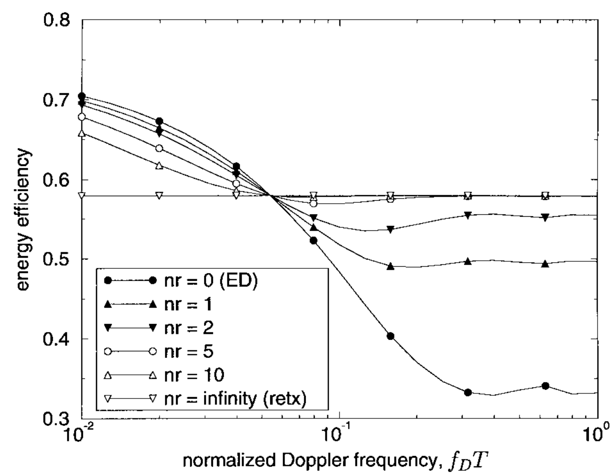

Fig. 3. Energy efficiency versus normalized Doppler frequency, $f_D T$. Parameterized retransmission with $n_r = 0, 1, 2, 5, 10, \infty$, number of users $N = 10$, average message length 10 packets, no capture, fading margin $F = 5$ dB, arrival rate $\lambda = 1/N$.

are plotted for the parameterized retransmission strategy as a function of $f_D T$. Note that this figure is the same as given in [10] for the throughput performance, since in this case where $\lambda = 1/N$ we have $C_i = 1$ for all i, and throughput and energy efficiency coincide. The energy efficiency performance of the *ED* protocol is best when the fading is slow, and worst when fading is fast. Under fast fading conditions, the energy efficiency improves significantly compared to the *ED* protocol even if only few retransmission attempts are allowed. Even in slow fading, the parameterized retransmission strategy performs well, close to the *ED* protocol's performance which is best in this case. In summary, from an energy consumption point of view, a protocol which allows a limited number of retransmission attempts results in good energy efficiency performance over a range of normalized Doppler frequencies (or equivalently, mobile user speeds) of interest. For example, the curve for $n_r = 2$ is very close to the envelope of all curves in Fig. 3, and can be seen as a reasonable compromise.

REFERENCES

[1] N. Bambos and J. M. Rulnick, "Mobile power management for wireless communication networks," *Wireless Networks*, vol. 3, pp. 3–14, 1997.

[2] M. Zorzi and R. R. Rao, "Energy constrained error control for wireless channels," *IEEE Personal Commun. Mag.*, vol. 4, pp. 27–33, Dec. 1997.

[3] P. L. Lettieri, C. Fragouli, and M. B. Srivastava, "Low power error control for wireless links," in *Proc. Mobicom'97*, Budapest, Hungary, Sept. 1997, pp. 139–150.

[4] J.-C. Chen, K. M. Sivalingam, P. Agrawal, and S. Kishore, "A comparison of MAC protocols for wireless local networks based on battery power consumption," in *Proc. IEEE INFOCOM'98*, San Francisco, CA, Apr. 1998, pp. 150–157.

[5] K. M. Sivalingam, M. B. Srivastava, P. Agrawal, and J.-C. Chen, "Low-power access protocols based on scheduling for wireless and mobile ATM networks," in *Proc. IEEE ICUPC'97*, San Diego, CA, Oct. 1997, pp. 429–433.

[6] M. Zorzi, R. R. Rao, and L. B. Milstein, "On the accuracy of a first-order Markov model for data block transmission on fading channels," in *Proc. IEEE ICUPC'95*, Nov. 1995, pp. 211–215.

[7] ——, "Error statistics in data transmission over fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1468–1477, this issue.

[8] R. A. Howard, *Dynamic Probabilistic Systems*. New York: Wiley, 1971.

[9] S. H. Ross, *Stochastic Processes* New York: Wiley, 1983.

[10] A. Chockalingam, M. Zorzi, L. B. Milstein, and P. Venkataram, "Performance of a Wireless Access Protocol on a Correlated Rayleigh fading channel with capture," *IEEE Trans. Commun.*, vol. 46, pp. 644–655, May 1998.

[11] W. Xu, A. Chockalingam, and L. B. Milstein, "Performance analysis of a multichannel wireless access protocol in the presence of bursty packet losses," in *Proc. IEEE MILCOM'98*, Bedford, MA, Oct. 1998.