Analysis of Energy-Conserving Access Protocols for Wireless Identification Networks

Imrich Chlamtac[a], Chiara Petrioli[b], Jason Redi[c]
[a] The University of Texas at Dallas, Richardson, TX, U.S.A., {*chlamtac,radha*}*@utdallas.edu*
[b] Università di Roma “La Sapienza”, Italy, *petrioli@dsi.uniroma1.it*
[c] Boston University, Boston, MA, U.S.A., *redi@bu.edu*

Abstract

A myriad of applications such as radio frequency identification (RFID) and smart card networks are emerging in which nodes are designed for extremely low-cost, large scale applications such that the replacement of batteries is not feasible. Energy conservation therefore becomes a major constraint. Classical access protocols are either not energy conserving or lead to unacceptable delays. In a previous paper we presented three classes of energy-conserving protocols. In this paper we describe analytic models to describe and evaluate their performance.

1 Introduction

Radio Frequency Identification Devices (RFIDs), such as warehouse identification tags and intelligent ID cards, referred to here as IDNETs, are examples of a new world of applications which use small, inexpensive devices for which battery conservation is a critical system parameter.

A typical IDNET is composed of a number of interconnected base stations communicating over a shared wireless channel to a large number of small, low-cost, wireless nodes or “tags”. These tags usually contain some sort of microprocessor, power source in the form of a battery, capacitor or solar cell, a radio frequency receiver, possibly a transmitter, and some support logic.

The range of potential uses for RFID tags is extremely large. By the year 2000, the RFID market is expected to expand 20 fold to over $5 billion [3]. Some examples of current uses for RFID tags are: tags attached to the ears or worn around the necks of livestock for use in location tracking, smart tags used in warehouses to automatically and quickly track inventory, and the numerous tag companies targeting the retail market for electronic shelf labels and automatic purchasing.

There are four fundamental characteristics of RFID wireless nodes which make RFID networks distinct from other wireless systems:

1. *Scale:* There may be a very large number, perhaps thousands, of wireless nodes per base station.
2. *Cost:* Due to the large number of extremely inexpensive items to “tag”, nodes must be extremely inexpensive, often on the order of only a few dollars.
3. *Size:* Nodes must be very small. The size of a pack of cards will be the maximum size for many applications.
4. *Traffic:* Communication is typically based on short, simple messages. Transmission speeds are usually low, on the order of tens of kilobits per second in order to minimize cost and power of transmission.

From these characteristics follow a number of important observations which further define the unique constraints of an RFID network. Most importantly, as in all mobile computing applications, the battery’s energy is a limited and scarce resource, which is not expected to increase in potential more than 30% in the near future [5]. Especially demanding of energy is any uplink transmission as it can typically use twice as much energy as reception [4]. It is possible to use spread spectrum modulation such that the base station indirectly provides the energy for limited uplink communication, but this requires that any uplink traffic be base station initiated. Furthermore, limited unlicensed bandwidth and simplicity of the tag means that all tags must share the same broadcast band.

For all of the above considerations, these types of systems require new access protocols which are designed around these unique constraints and provide a combination of two important factors: low delay and low energy requirements. The allowable delay is an application dependent constraint. For example, tracking the movement of tags across the cells within a system requires updates to be performed within a short, bounded, amount of time. The system is already constrained by the speed of the shared channel and has to manage a potentially large number of tags. Therefore, it is extremely important for access protocols to not add significantly to the transmission delay.

Low energy consumption is the second requirement that the access protocol must satisfy. The large number of nodes makes it economically impossible to replace or recharge the batteries. Therefore tags must be designed so as to require a minimum of energy for operation. In order to conserve battery life, the tag can enter a ”sleep” state where the CPU is in a low power mode and radio reception is disabled. In contrast with this, the ”awake” state, in which the CPU operates at full energy and the radio frequency circuitry is active, can typically use 100 times as much energy.

Among existing protocols, classical random access protocols are not energy conserving. While deterministic protocols lead to unacceptable delays.

In [2], we addressed the problem of designing *communications protocols which operate under an energy constraint*, in which the fraction of timeslots in which tags need to be in the active (awake) state is minimized and the access delay meets the applications constraints. We introduced three classes of protocols: grouped-tag TDMA, directory and pseudo-random. In this paper we present analytic models which use infinite Markov chains and the theory of M/D/1 queues with vacations for further evaluating the performance of these protocols. All three classes of protocols represent a dramatic improvement over classical approaches.

In the remainder of the paper we first describe the tag network model, followed by a description of the protocols. In the ensuing analysis of energy consumption and access delay, we derive the system behavior for both uniformly and non-uniformly distributed traffic destinations. We follow with a detailed evaluation of the protocols and conclude with a summary.

2 Network Model and Protocols Description

We consider a single cell system where a base station communicates with N tags through a radio channel of bandwidth B. The communication is packet-oriented. We assume the time to be slotted and the base station's transmissions to be synchronized to the beginnings of slots. The packet length c is constant, and exactly one packet can be transmitted during one slot. In this analytic model, we do not explicitly treat transmission errors.

We define an *access protocol* as consisting of two components: a *transmission scheduling strategy* at the base station which in each slot selects a packet for transmission from the arrival queue, and a *wake-up schedule* at each tag which determines the slots in which the tag is awake. In general, the transmission scheduling strategy can take into account different parameters: the number of packets in the queue, the packets' ages, as well as the wake-up schedules of their destinations. In the protocols discussed here the "oldest packet" criterion is generally adopted to help meet the application delay requirements. We next present and compare three classes of protocols for constructing efficient wake-up schedules: grouped-tag TDMA protocols, directory protocols, and pseudo-random protocols.

2.1 Grouped-Tag TDMA Protocols

Classical TDMA can be adapted for use in energy-conserving environments in the following way: we divide tags into m disjoint groups, with the cardinality of each group differing by at most one tag, and assign (reserve) each slot of the TDMA cycle to a unique group. This increases the average energy consumption per slot by the cardinality of the groups, but decreases the average delay since there is a greater probability that a tag will be awake soon after a packet for it has arrived at the base station. The optimal selection of this group size will ensure the best energy and delay performance of this class of protocols.

2.2 Directory Protocols

In our directory protocol, the base station always waits for a group of k packets in the queue to accumulate. The base station then transmits a list or directory of the k packet destinations before transmitting the packets in the subsequent slots. The tags are all awake during the transmission of the directory, and can therefore schedule their wake-up slots to coincide with the broadcast of their packets. When there is no group being currently transmitted, the tags wake up periodically every v slots in order to give the base station an opportunity to start the transmission of a new group.

The choice of the parameter k depends on the load and must take into account the trade-off between the increase in the delay due to a larger k and the energy savings from more infrequent broadcasting of directories. Additionally, the parameter v should depend upon k and the load. A tag system with small value of v and a low load will have the tags waking up frequently until enough packets have accumulated at the base station, while a large value of v will incur an increase in delay before the start of group's transmission.

2.3 Pseudo-Random Protocols

The pseudo-random protocols are a class of protocols based on deterministic (pseudo-random) schedules which preserves the power of randomization for fairness, while providing the advantages of determinism, i.e., the base station's ability to predict tags' state in each slot. In this class of protocols all tags run the same pseudo-random number generator and determine their state (awake or asleep) at each slot based on a probability p and the stored state of the random number generator. In order to avoid a complete overlap of the wake-up schedules, the pseudo-random generator of each tag is initialized using a unique seed, which is known at the base station. Therefore, by using the same pseudo-random number generator it is possible for the base station to determine the schedules of the tags it wants to transmit to. The base station can initiate changes in the value of p as a function of the load, the number of tags, etc.

3 Analysis

To compare the performance of the various protocols proposed, we consider the behavior of a single cell tag system. Since the time needed to successfully receive a packet, given that the tag is awake, is exactly one slot, then the slot duration is $\sigma = b + \tau$, with b denoting the packet transmission time ($\frac{c}{B}$) and τ the propagation delay. The presented evaluation utilizes the following definitions:

$T-$ the average waiting time in slots experienced by a packet in the system from arrival at the base station to successful reception at the tag.

$E-$ the average percentage of slots in which a tag is awake.

$L-$ the average number of packets in the system.

The energy measure proposed does not take into account the contribution during slots where a tag is asleep. The energy used while in this state is consumed over the lifetime of the tag, whether the tag is being used or not. Therefore only the percentage of awake slots is necessary for comparing different energy-conserving access protocols.

We assume that the packets arrive at the base station according to a Poisson process with interarrival rate λ'. Each packet is addressed to a single destination selected from either a uniform or a Gaussian distribution. The latter is introduced in order to model the heterogeneous nature of the traffic which is common to many tag applications. We use the following notations:

$\lambda-$ the mean arrival rate expressed in packets per slot ($\lambda = \lambda'(b+\tau)$).

$\mu-$ the mean service rate expressed in packets per slot.

$\rho-$ the utilization factor $\frac{\lambda}{\mu}$.

3.1 Analysis of the Grouped-Tag TDMA

The access delay for packets whose destinations are in different groups is independent. Therefore, for each group i, we can associate a different queuing system whose Poisson traffic stream has rate λ_i. Every time the queue is empty, the server goes "on vacation" for one TDMA cycle (m slots). Otherwise the service time of a packet is constant and equals 1. Thus, the average waiting time in the queue for a packet destined for group i is equivalent to that of an $M/D/1$ queuing system with vacations [1]:

$$\begin{aligned} W_i &= \frac{\rho_i}{2\mu_i(1-\rho_i)} + \frac{m}{2} \\ &= \frac{m}{2(1-\lambda_i m)} \end{aligned}$$

Let T_x denote the average number of slots, as a function of the parameter x, during which a packet is waiting in the system. Then:

$$\begin{aligned} T_x &= \sum_{i=1}^{m} \frac{\lambda_i}{\lambda}(W_i+1) \\ &= 1 + \sum_{i=1}^{m} \frac{\lambda_i m}{2\lambda(1-\lambda_i m)} \end{aligned}$$

It remains to calculate the λ_i. These quantities depend on the destination distribution utilized. Under a uniform destination distribution the percentage of the global traffic dedicated to each group is proportional to the number of tags in that group. Thus,

$$\lambda_i = \begin{cases} \frac{\lambda\lceil \frac{N}{m} \rceil}{N} & i \le N mod m \\ \frac{\lambda\lfloor \frac{N}{m} \rfloor}{N} & i > N mod m \end{cases}$$

We next consider the case of when the packets destinations are chosen according to a Gaussian distribution with a density function f_x, average μ and variance σ^2. $\lambda_i = \lambda p_i$, where p_i denotes the probability of choosing a packet in the i-th group. p_i is the area of the region bounded above by f_x and the interval associated to the i-th group, normalized by the area of the *usable* part of the Gaussian (i.e., the one corresponding to the interval $[1...N]$). So

$$p_i = \begin{cases} \frac{F_x(i\lceil \frac{N}{m} \rceil) - F_x((i-1)\lceil \frac{N}{m} \rceil)}{F_x(N)-F_x(0)} & i \le N mod m \\ \frac{F_x(N mod m\lceil \frac{N}{m} \rceil + (i - N mod m)\lfloor \frac{N}{m} \rfloor)}{F_x(N)-F_x(0)} - & i > N mod m \\ \frac{F_x(N mod m\lceil \frac{N}{m} \rceil + (i-1-N mod m)\lfloor \frac{N}{m} \rfloor)}{F_x(N)-F_x(0)} & \end{cases}$$

where $F_x(y)$ is the cumulative distribution function of a Gaussian distribution.

3.2 Analysis of the Directory Protocol

Let $a = \left\lfloor \frac{c}{\lceil log N \rceil} \right\rfloor$ be the number of different records which fit into a slot. Then, the service time of a group is $k + k'$ slots, where k' indicates the time needed for transmitting the directory and is given by $\lceil \frac{k}{a} \rceil$. When a cycle ends and no completed groups are in the queue, the tags go to sleep for an idle interval of v slots. Thus, a server can start processing a new group if and only if there is at least one completed group in the queue and either an idle or a service cycle ended in the previous slot. We compute the average waiting time T of a packet in the system as:

$$\begin{aligned} T &= W_p + T_g - W_s \\ &\cong \frac{k-1}{2\lambda} + T_g - (\frac{k-1}{2}) \end{aligned}$$

$-W_p$ is the average waiting time of a packet in the queue before its group is completed and is $\cong \frac{k-1}{2\lambda}$.

$-T_g$ is the average waiting time of a group in the system.

$-W_s$ is the average time between the successful reception of a packet and the completion of its group's transmission. $W_s = \frac{k-1}{2}$.

To compute T_g we construct the following analytic model. We observe the system at the regeneration points embedded at the beginning of each slot and model the system as a discrete time, infinite Markov chain $M =< S_{<i,j,h>}, P_{<i,j,h>,<i',j',h'>} >$. Each state $S_{<i,j,h>}$ denotes the number i of completed groups in the system, the number j of non-grouped packets in the queue, and the index h of the current slot in either the idle ($0 \le h \le v-1$) or the service ($v \le h \le v+k+k'-1$) cycle. Note that within either a service or idle cycle we have only to increase the position of the slot and keep track of the arrivals, while during the last slot of a cycle we also have to decide the nature (busy/idle) of the next interval and possibly indicate the exit of a group from the system. We obtain the following transition probabilities:

Case $h \in [0...v-1) \bigcup [v...v+k+k'-1)$

$$P_{<i,j,h>,<i',j'h'>} = \begin{cases} 0 & (h' \neq h+1) \vee (i' < i) \\ & \vee((i' = i) \wedge (j' < j)) \\ A_{j'-j+(i'-i)k} & otherwise \end{cases}$$

Case $h = v-1$

$$P_{<i,j,v-1>,<i',j',h'>} = \begin{cases} 0 & ((h' \neq 0) \wedge (i' = 0)) \\ & \vee((h' \neq v) \wedge (i' \geq 1)) \\ & \vee(i' < i) \vee ((i' = i) \wedge (j' < j)) \\ A_{j'-j+(i'-i)k} & otherwise \end{cases}$$

Case $h = v+k+k'-1$

$$P_{<i,j,v+k+k'-1>,<i',j',h'>} = \begin{cases} 0 & ((h' \neq 0) \wedge (i' = 0)) \\ & \vee((h' \neq v) \wedge (i' \geq 1)) \\ & \vee(i' < i-1) \vee ((i' = i-1) \\ & \wedge(j' < j)) \\ A_{j'-j+(i'-i)k+k} & otherwise \end{cases}$$

$-A_r$ is the probability of r messages originated by a Poisson process during a slot and is given by:

$$A_r = e^{-\lambda} \frac{(\lambda)^r}{r!}$$

The steady state probability $\pi_{<i,j,h>}$ of being in the $< i, j, h >$ state is obtained by solving the following system of equations:

$$\begin{cases} \Pi = P^T \Pi \\ \sum_i \sum_j \sum_h \Pi = 1 \end{cases}$$

Let us denote $\pi_g(i)$ the probability of having i completed groups in the system:

$$\pi_g(i) = \sum_j \sum_h \pi_{<i,j,h>}$$

The average number of completed groups in the system L_g can then be given by:

$$L_g = \sum_i i \pi_g(i)$$

Finally, applying Little's theorem we can compute the average waiting time of a group in the system:

$$T_g = \frac{L_g k}{\lambda}$$

The energy consumption E can be computed as follows:

$$E = Pr_i E_i + Pr_b E_b$$

$-Pr_i$ is the percentage of slots belonging to idle cycles and is given by:

$$Pr_i = \sum_i \sum_{j=0}^{k-1} \sum_{h=0}^{v-1} \pi_{<i,j,h>}.$$

$-Pr_b$ is the percentage of busy slots, $Pr_b = 1 - Pr_i$.

$-E_i$ is the average energy consumption per tag per slot during an idle period and is equal to $\frac{1}{v}$.

$-E_b$ is the average energy consumption per tag per slot during a service cycle.

$$E_b = \frac{k' + \frac{k}{N}}{k + k'}.$$

3.3 Approximate Analysis of the Pseudo-Random Protocol

We analyze the protocol under the assumption that all tags use the same awake probability parameter p. To obtain an exact description of the system behavior through a Markov chain, the states should include the number of packets addressed to each tag, since the probability of successfully transmitting depends on the number of unique packet destinations in the queue. It is therefore impossible to analyze systems of a realistic size. We solve this problem by making use of Stern's independence assumption [6], stating that at the beginning of a time slot each message draws a new destination from a given uniform distribution. The Markov chain describing the system is then represented by the total number of packets at the beginning of a slot. The transition probability matrix P is given, similarly to the random access case, by

$$P_{0,j} = A_j$$

$$P_{i,j} = \begin{cases} 0 & j < i-1 \\ \sum_{k=1}^{min(i,N)} R_{i,k}(1-(1-p)^k)A_0 & j = i-1 \\ \sum_{k=1}^{min(i,N)} R_{i,k}(1-(1-p)^k)A_{j-i+1} + & \\ \sum_{k=1}^{min(i,N)} R_{i,k}(1-p)^k A_{j-i} & otherwise \end{cases}$$

$-R_{i,k}$ is the probability that i packets are addressed to k different destinations and is given by

$$R_{i,k} = \frac{\binom{N}{k} \sum_{j=0}^{k-1} (-1)^j \binom{k}{j} (k-j)^i}{N^i}.$$

The steady-state probabilities π_i are the solutions of the following set of equations

$$\begin{cases} \Pi = P^T \Pi \\ \sum_i \Pi = 1 \end{cases}$$

We can now compute the average number L of packets in the system

$$L = \sum_i i \pi_i$$

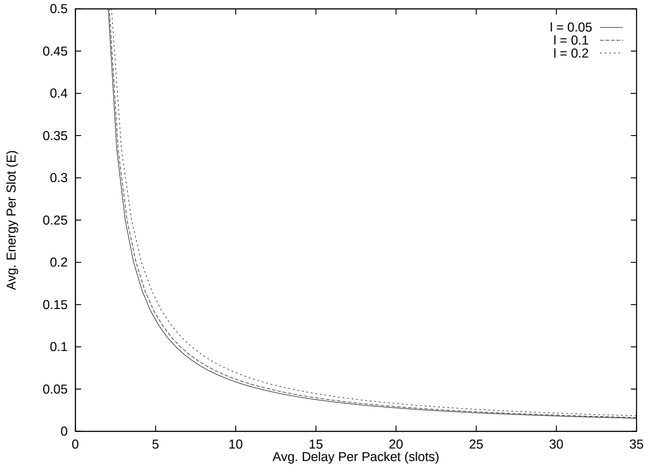

Figure 1: Grouped TDMA Protocol, Uniform Destination Distribution, Optimal group size

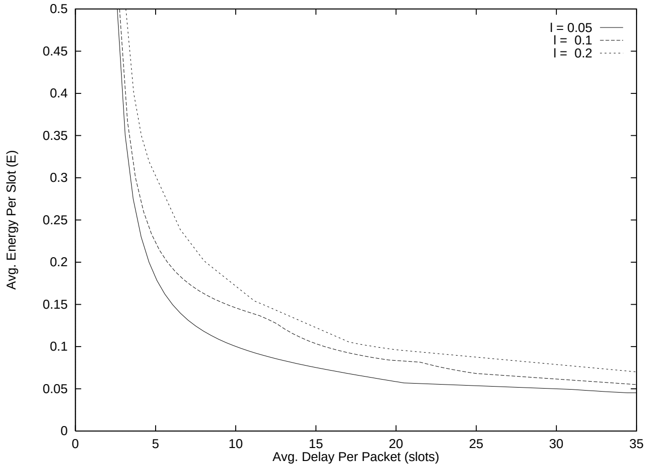

Figure 2: Directory Protocol, Uniform Destination Distribution

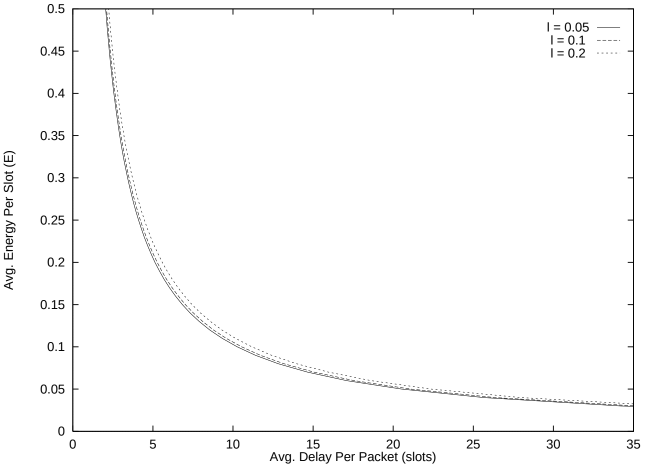

Figure 3: Pseudo-Random Protocol, Uniform Destination Distribution

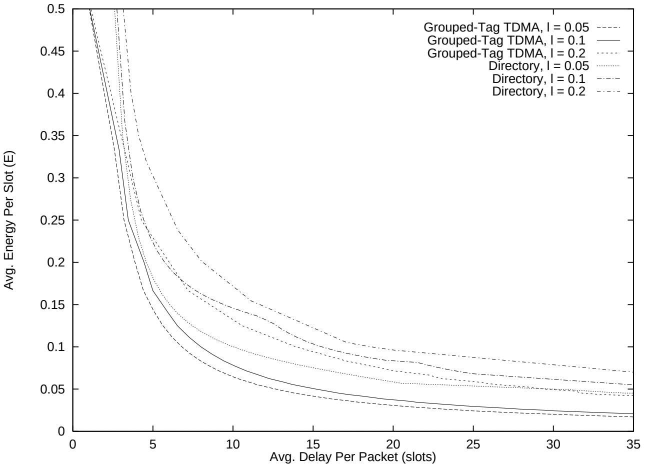

Figure 4: Energy conserving protocols, Gaussian destination distribution ($\mu = \frac{N}{2}, \sigma^2 = 10N$)

and, using the Little's result:

$$T = \frac{L}{\lambda}$$

Finally, as in the previous section, the average energy consumption is obviously $E = p$.

4 Performance

Figures 1 through 5 plot quantitative results of the equations derived in previous sections. These plots were all generated by assuming 1000 tags with consecutive sequential ids are assigned to a single base station. All of the plots include sets of data with interarrival rate parameters equal to 0.05, 0.1 and 0.2 arrivals per slot. To show the effect of heterogeneous destination distributions we use two Gaussians, with mean $\frac{N}{2}$ and variance $10N$ and N respectively. The infinite Markov chain of the pseudo-random protocol was approximated by truncating the transition probability matrix after the first 100 states while up to 2500 states were considered for the directory protocol with the limitation $k \in \{1, \ldots, 10\}$ and $v \in \{1, \ldots, 25\}$.

Fig.1,2, 3 compare the performance of the three protocols proposed using a uniform destination distribution. Fig.1 shows the trade-off between energy and delay for optimum values of group size using the grouped-tag TDMA protocol. We note that the performance is better for low loads because there are fewer packets available for transmission for the same slot in a cycle. Since tags' schedules are cyclic and independent of packets in the base station's queue (a tag does not know when the base station may have a packet destined to it) some unnecessary energy can be used due to tags continuing to wake up cyclically.

The performance of the directory protocol (Fig.2) depends on the choice of the parameters k and v for each particular arrival rate.

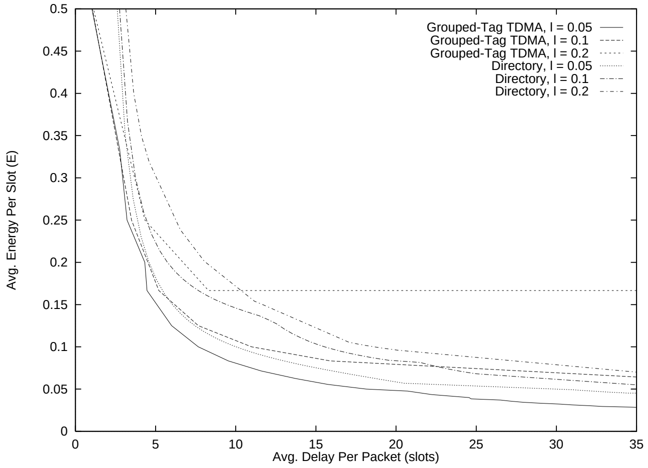

Figure 5: Energy conserving protocols, Gaussian destination distribution ($\mu = \frac{N}{2}, \sigma^2 = N$)

If we assume an optimum value for v, an increase in the arrival rate λ will reduce the amount of time it takes for a complete group of K packets to arrive, reducing the average delay for the packets in that group. However, this requires the tags to be awake more often in order to receive this group, thereby increasing the energy. When N and group size k are large, the size of the directory becomes prohibitively large, and therefore the amount of energy used by the tags just to read the directory can become a major factor.

Fig.3 shows the energy vs. delay graph for the pseudo-random protocol given a uniform destination distribution. The performance achieved is better than that of the directory protocol but slightly worse than that of the grouped-tag TDMA although this difference decreases as λ increases.

Fig.4, and 5 compare protocol performance for the case of heterogeneous traffic. The directory protocol is not affected by the destination distribution. Instead, we show how the performance of the grouped-tag TDMA degrades rapidly under heterogeneous traffic. Packets belonging to a group with a high probability of traffic are severely delayed due to the high volume of localized destinations. The reorganization of the groups, even if possible, can be extremely time-consuming. Since the number of packets in the queue addressed to the same group increases as either the load increases or the Gaussian distribution becomes narrower, the performance of the protocol decreases in both these cases.

For neither of these Gaussians is the performance of the pseudo-random protocol discussed. Intuitively, we do not expect the performance to change dramatically from the case of a uniform destination distribution. Since the tag's schedules are the same, the energy will not change, and the probability of at least one packet in the queue with an awake destination should not significantly change. This rational is backed by the simulation results in [2] which shows only a slight increase in delay for even the tighter Gaussian distribution.

5 Conclusions

In this paper, we addressed the analysis of three types of wireless access protocols which include an energy constraint: grouped-tag TDMA, directory and pseudo-random. Careful selection of protocol parameters addresses the goal of minimizing the energy required for reception of packets while meeting the application delay constraints. A detailed analysis and plots of quantitative data from this showed a variety of performances under varying loads. In particular, the grouped-tag TDMA protocol achieves the best performance for either a low traffic load or a uniform destination distribution. When the destination distribution becomes more realistically clustered, other protocols such as the directory or pseudo-random protocols need to be adopted.

The pseudo-random protocol consistently out-performs the directory protocol in both energy and delay for all loads and most destination distributions. However, for cases where the destination distribution becomes extremely clustered, it is apparent that the directory protocol is the best solution, unless dynamic handling of the protocol parameters is introduced.

References

[1] Bertsekas, D. and Gallager, R. *Data Networks*, Prentice Hall, 1992.

[2] Chlamtac, I. and Petrioli, C. and Redi, J. Energy-Conserving Access Protocols for IDNETs Technical Report 96-006, Boston University.

[3] Dunn, D. RFID Engine Chugs Onto Scene. *Electronic Buyer's News*, February 5, 1996, pp 14.

[4] Harris, E.P., Depp, S.W., Pence,E., Kirkpatrick, S., Sri-Jayantha, M., and Troutman, R.R. Technology directions for portable computers. *Proceedings of the IEEE*, 83(4):636–58, April 1995.

[5] Powers, R.A. Batteries for low power electronics. *Proceedings of the IEEE*, 83(4):687–93, April 1995.

[6] Stern, T. E. Packet Scheduling protocols in multi-beam communication satellites. *Proceedings of the International Symposium on Information Theory*, Italy, June 1979, pp. 130–131.