Data broadcasting over error-prone wireless channels *

P. Barsocchi[1], A.A. Bertossi[2], M.C. Pinotti[3], F. Potortí[1]

[1] ISTI
National Research Council
56124 Pisa, Italy
{Paolo.Barsocchi,potorti}@isti.cnr.it

[2] Dept. of Computer Science
University of Bologna
40127 Bologna, Italy
bertossi@cs.unibo.it

[3] Dept. of Comp. Science and Math.
University of Perugia
06123 Perugia, Italy
pinotti@unipg.it

Abstract

Broadcasting is an efficient and scalable way of transmitting data over wireless channels to an unlimited number of clients. In this paper the problem of allocating data to multiple channels is studied, assuming skewed allocation of most popular data items to less loaded channels, flat data scheduling per channel, and the presence of unrecoverable channel transmission errors. The objective is that of minimizing the average expected delay experienced by clients. Two different channel error models are considered: the geometric model and the Gilbert-Elliot one. In the former model, each packet transmission has the same probability to fail and each transmission error is independent from the others. In the latter one, bursts of erroneous or error-free packet transmissions due to wireless fading channels are modeled. For the geometric channel error model and uniform data item lengths, an optimal solution can be found in polynomial time when all the channels have the same probability to fail. Heuristic algorithms are exhibited for the geometric model and non-uniform data item lengths as well as for the Gilbert-Elliot error model and both uniform and non-uniform data lengths. Extensive simulations show that such heuristics provide good sub-optimal solutions when tested on benchmarks whose item popularities follow Zipf distributions.

*This work has been supported by ISTI-CNR under the BREW research grant.

1 Introduction

In wireless asymmetric communication, broadcasting is an efficient way of simultaneously disseminating data to a large number of clients [15]. Consider data services on cellular networks, such as stock quotes, weather infos, traffic news, where data are continuously broadcast to clients that may desire them at any instant of time. In this scenario, a server at the base-station repeatedly transmits data items from a given set over wireless channels, while clients passively listen to the shared channels waiting for their desired item. The server has to pursue a data allocation strategy for assigning items to channels and a broadcast schedule for deciding which item has to be transmitted on each channel at any time instant. Efficient data allocation and broadcast scheduling have to minimize the client expected delay, that is, the average amount of time spent by a client before receiving the item he needs. The client expected delay increases with the size of the set of the data items to be transmitted by the server. Indeed, the client has to wait for many unwanted data before receiving his own data. Moreover, the client expected delay may be influenced by transmission errors because items are not always received correctly by the client. Athough data are usually encoded using *error control codes* (*ECC*) allowing some recoverable errors to be corrected by the client without affecting the average expected delay, there are several transmission errors which still cannot be corrected using ECC. Such *unrecoverable* errors heavily affect the client expected delay. Indeed, the resulting corrupted items have to be discarded and the client must wait until the same item is broadcast again by the server.

Several variants for the problem of data allocation and broadcast scheduling have been proposed in the literature [1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 16, 18, 19]. The database community usually partitions the data among the channels and then adopts a *flat* broadcast schedule on each channel [5, 13, 19]. In such a way, the allocation of data to channels becomes critical for reducing the average expected delay, while the flat schedule on each channel merely consists in cyclically broadcasting the items assigned to the same channel in an arbitrary fixed order, that is, once at a time in a round-robin fashion [1]. In order to reduce the average expected delay, *skewed* data allocations are used where items are partitioned according to their popularities so that the most requested items appear in a channel with shorter period. Assuming that each item transmitted by the server is always received correctly by the client, a solution that minimizes the average expected delay can be found in polynomial time in the case of *uniform lengths* [19], that is when all the items have the same transmission time, whereas the problem becomes computationally intractable for non-uniform lengths [5]. In this latter case, several heuristics have been developed in [4, 19], which have been tested on some benchmarks where item popularities follow Zipf distributions. Such distributions are used to characterize the popularity of one item among a set of similar data, like a web page in a web site [8]. Thus far, the data allocation problem has not been investigated by the database community when the wireless channels are subject to transmission errors.

In contrast, a wireless environment subject to errors has been considered by the networking community which, however, concentrates only on finding broadcast scheduling for a single channel to minimize the average expected delay [6, 10, 11, 16]. Indeed, the networking community assumes all items replicated over all channels, and therefore no data allocation to the channels is needed. Although it is still unknown whether a broadcast schedule on a single channel with minimum average delay can be found in polynomial time or not, almost all the proposed solutions follow the *square root rule (SRR)*, a heuristic which in practice finds almost optimal schedules [3]. The aim of SRR is to produce a broadcast schedule where each data item appears with equally spaced replicas, whose frequency is proportional to the square root of its popularity and inversely proportional to the square root of its length. In particular, the solution proposed by [16] adapts the SRR solution to the case of unrecoverable errors. In such a case, since corrupted items must be discarded worsening the average expected delay, the spacing among replicas has to be properly recomputed.

The present paper extends the data allocation problem first studied by the database community under the assumptions of skewed data allocation to channels and flat data schedule per channel [4, 5, 19], to cope with the presence of erroneous transmissions, under the same assumptions of [16], namely unrecoverable errors. Two different error models will be considered to describe the behavior of wireless channels [17]. First, as in [16], the channel error is modeled by a geometric distribution, where each packet transmission has the same probability q to fail and $1 - q$ to succeed, and each transmission error is independent from the others. Then, the sligthly more sophisticated Gilbert-Elliot channel error model will be considered, which was not previously studied in [16]. Such a model is able to capture burstiness, that is sequences of erroneous or error-free packet transmissions, and well approximates the error characteristics of certain wireless fading channels [20]. Specifically, in the case of geometric channel error model, it will be shown that an optimum solution, namely one minimizing the average expected delay, can be found in polynomial time for the data allocation problem when the data items have uniform lengths and all the channels have the same failure probability. Instead, heuristic algorithms will be exhibited which find sub-optimal solutions for the geometric error model and non-uniform data lengths as well as for the Gilbert-Elliot error model and both uniform and non-uniform data lengths. Extensive simulations will show that such heuristics find good solutions when tested on benchmarks whose item probabilities are characterized by Zipf distributions.

The rest of this paper is so organized. Section 2 first gives notations, definitions as well as the problem statement, and then recalls the basic dynamic programming algorithms known so far in the case of error-free channel transmissions. Sections 3 and 4 consider the geometric and the Gilbert-Elliot channel error model, respectively, and illustrate heuristics for items of uniform and non-uniform lengths. Such heuristics are derived by properly redefining the recurrences in the dynamic programming algorithms previously presented for error-free channels. Experimental tests are reported at the end of both Sections 3 and 4. Finally, conclusions are offered in Section 5.

2 Background on error-free channels

Consider a set of K identical error-free channels, and a set $D = \{d_1, d_2, \ldots, d_N\}$ of N data items. Each item d_i is characterized by a *probability* p_i and a *length* z_i, with $1 \leq i \leq N$. The probability p_i represents the popularity of item d_i, namely its probability to be requested by the clients, and it does not vary along the time. Clearly, $\sum_{i=1}^{N} p_i = 1$. The length z_i is an integer number, counting how many packets are required to transmit item d_i on any channel and it includes the encoding of the item with an error control code. For the sake of simplicity, it is assumed that a packet transmission requires one time unit. Each d_i is assumed to be non preemptive, that is, its transmission cannot be interrupted. When all data lengths are the same, i.e., $z_i = z$ for $1 \leq i \leq N$, the lengths are called *uniform* and are assumed w.l.o.g. to be unit, i.e. $z = 1$. When the data lengths are not the same, the lengths are said *non-uniform*.

The *expected delay* t_i is the expected number of packets a client must wait for receiving item d_i. The *average expected delay* (AED) is the number of packets a client must wait on the average for receiving any item, and is computed as the sum over all items of their expected delay multiplied by their probability, that is

$$\mathrm{AED} = \sum_{i=1}^{N} t_i p_i \tag{1}$$

When the items are partitioned into K groups $G_1, \ldots, G_K$, where group G_k collects the data items assigned to channel k, and a flat schedule is adopted for each channel, that is, the items in G_k are cyclically broadcast in an arbitrary fixed order, Equation 1 can be simplified. Indeed, if item d_i is assigned to channel k, and assuming that clients can start to listen at any instant of time with the same probability, then t_i becomes $\frac{Z_k}{2}$, where Z_k is the schedule *period* on channel k, i.e., $Z_k = \sum_{d_i \in G_k} z_i$. Then, Equation 1 can be rewritten as

$$\mathrm{AED} = \sum_{k=1}^{K} \sum_{d_i \in G_k} \frac{Z_k}{2} p_i = \frac{1}{2} \sum_{k=1}^{K} Z_k P_k \tag{2}$$

where P_k denotes the sum of the probabilities of the items assigned to channel k, i.e., $P_k = \sum_{d_i \in G_k} p_i$. Note that, in the uniform case, the period Z_k coincides with the cardinality of G_k, which will be denoted by N_k.

Summarizing, given K error-free channels, a set D of N items, where each data item d_i comes along with its probability p_i and its integer length z_i, the *Data Allocation Problem* consists in partitioning D into K groups $G_1, \ldots, G_K$, so as to minimize the AED objective function given in Equation 2. Note that, in the special case of uniform lengths, the corresponding objective function is derived replacing Z_k with N_k in Equation 2.

Almost all the algorithms proposed so far for the data allocation problem on error-free channels are based on dynamic programming. Such algorithms restrict the search for the solutions to the so called *segmentations*, that is, partitions obtained by considering the items ordered by their indices, and by assigning items with consecutive indices to each channel. Formally, a segmentation is a partition of the ordered sequence $d_1, \ldots, d_N$ into K adjacent segments $G_1, \ldots, G_K$, each of consecutive items, as follows:

$$\underbrace{d_1, \ldots, d_{B_1}}_{G_1}, \underbrace{d_{B_1+1}, \ldots, d_{B_2}}_{G_2}, \ldots, \underbrace{d_{B_{K-1}+1}, \ldots, d_N}_{G_K}$$

A segmentation can be compactly denoted by the $(K-1)$-tuple

$$(B_1, B_2, \ldots, B_{K-1})$$

of its *right borders*, where border B_k is the index of the last item that belongs to group G_k. Notice that it is not necessary to specify B_K, the index of the last item of the last group, because its value will be N for any solution.

Three main dynamic programming algorithms for the data allocation problem are now briefly surveyed, called *DP*, *Dichotomic*, and *Dlinear*. All the algorithms assume that the items $d_1, d_2, \ldots, d_N$ are indexed by non increasing $\frac{p_i}{z_i}$ ratios, that is $\frac{p_1}{z_1} \geq \frac{p_2}{z_2} \geq \cdots \geq \frac{p_N}{z_N}$. Observe that in the uniform case this means that the items are sorted by non increasing probabilities. Let $SOL_{k,n}$ denote a segmentation for grouping items $d_1, \ldots d_n$ into k groups and let $sol_{k,n}$ be its corresponding cost, for any $k \leq K$ and $n \leq N$. Moreover, let $C_{i,j}$ denote the cost of assigning to a single channel the consecutive items $d_i, \ldots, d_j$:

$$C_{i,j} = \sum_{h=i}^{j} t_h p_h = \frac{1}{2} \left(\sum_{h=i}^{j} z_h \right) \left(\sum_{h=i}^{j} p_h \right) \qquad (3)$$

Note that, for uniform lengths, the above formula simplifies as $C_{i,j} = \frac{1}{2}(j-i+1) \sum_{h=i}^{j} p_h$.

The DP algorithm is a dynamic programming implementation of the following recurrence, where k varies from 1 to K and, for each fixed k, n varies from 1 to N. When $k = 1$, $sol_{1,n} = C_{1,n}$, whereas, when $k > 1$, one has:

$$sol_{k,n} = \min_{1 \leq \ell \leq n-1} \{ sol_{k-1,\ell} + C_{\ell+1,n} \} \qquad (4)$$

For any value of k and n, the DP algorithm selects the best solution obtained by considering the $n-1$ solutions already computed for the first $k-1$ channels and for the first ℓ items, and by combining each of them with the cost of assigning the last $n-\ell$ items to the single k-th channel. The DP algorithm requires $O(N^2K)$ time. It finds an optimal solution in the case of uniform lengths and a sub-optimal one in the case of non-uniform lengths [19].

To improve on the time complexity of the DP algorithm, the Dichotomic algorithm has been devised. Let B_h^n denote the h-th border of $SOL_{k,n}$, with $k > h \geq 1$. Assume that $SOL_{k-1,n}$ has been found for every $1 \leq n \leq N$. If $SOL_{k,l}$ and $SOL_{k,r}$ have been found for some $1 \leq l \leq r \leq N$, then one knows that B_{k-1}^c is between B_{k-1}^l and B_{k-1}^r, for any $l \leq c \leq r$. Thus, choosing $c = \lceil \frac{l+r}{2} \rceil$ as the middle point between l and r, Recurrence 4 can be rewritten as:

$$sol_{k,c} = \min_{B_{k-1}^l \leq \ell \leq B_{k-1}^r} \{ sol_{k-1,\ell} + C_{\ell+1,c} \} \qquad (5)$$

where B_{k-1}^l and B_{k-1}^r are, respectively, the final borders of $SOL_{k,l}$ and $SOL_{k,r}$. The Dichotomic algorithm reduces the time complexity of the DP algorithm to $O(NK \log N)$. As for the DP algorithm, the Dichotomic algorithm also finds optimal and sub-optimal solutions for uniform and non-uniform lengths, respectively [5].

Finally, fixed k and n, the Dlinear algorithm selects the feasible solutions that satisfy the following Recurrence:

$$sol_{k,n} = sol_{k-1,m} + C_{m+1,n} \qquad (6)$$

where m is the minimum ℓ in the range $B_k^{n-1} \leq \ell \leq n-1$ such that $sol_{k-1,\ell} + C_{\ell+1,n} < sol_{k-1,\ell+1} + C_{\ell+2,n}$, and $sol_{1,n} = C_{1,n}$.

In practice, Dlinear adapts Recurrence 4 by exploiting the property that, if $SOL_{k,n-1}$ is known, then one knows that B_k^n is no smaller than B_k^{n-1}, and by stopping the trials as soon as the cost $sol_{k-1,\ell} + C_{\ell+1,n}$ of the solution starts to increase. The overall time complexity of the Dlinear algorithm is $O(N(K + \log N))$. Thus the Dlinear algorithm is even faster than the Dichotomic one, but the solutions it provides are always sub-optimal, both in the uniform and non-uniform case [4].

3 Geometric channel error model

In this section, unrecoverable channel transmission errors modeled by a geometric distribution are taken into account. Under such an error model, each packet transmission over every channel has the same probability q to fail and $1-q$ to succeed, and each transmission error is independent from the others, with $0 \leq q \leq 1$. Since the environment is asymmetric, a client cannot ask the server to immediately retransmit an item d_i which has been received on channel k with an unrecoverable error. Indeed, the client has to discard the item and then has to wait for a whole period Z_k, until the next transmission of d_i scheduled by the server. Even the next item transmission could be corrupted, and in such a case an additional delay of Z_k has to be waited. Therefore, the expected delay t_i has to take into account the extra waiting time due to a possible sequence of independent unrecoverable errors.

Assume that the item lengths are uniform, i.e., $z_i = 1$, for $1 \leq i \leq N$. Recall that in such a case the period of channel k is N_k. If a client wants to receive item d_i, which is trasmitted on channel k, and the first transmission he can hear of d_i is error-free, then the client waits on the average $\frac{N_k}{2}$ time units with probability $1-q$. Instead, if the first transmission of d_i is erroneous, but the second one is error-free, then the client experiences an average delay of $\frac{N_k}{2} + N_k$ time units with probability $q(1-q)$. Generalizing, if there are h bad transmissions of d_i followed by a good one, the client average delay for receiving item d_i becomes $\frac{N_k}{2} + hN_k$ time units with probability $q^h(1-q)$. Thus, summing up over all h, the expected delay t_i is bounded by

$$\sum_{h=0}^{\infty}(\frac{N_k}{2} + hN_k)q^h(1-q) = \frac{N_k}{2} + N_k\frac{q}{1-q}$$

because $\sum_{h=0}^{\infty} q^h \leq \frac{1}{1-q}$ and $\sum_{h=0}^{\infty} hq^h \leq \frac{q}{(1-q)^2}$. Therefore, one can set the expected delay as

$$t_i = \frac{N_k}{2}\frac{1+q}{1-q} \tag{7}$$

By the above setting, the objective function to be minimized becomes

$$\text{AED} = \sum_{i=1}^{N} t_i p_i = \frac{1}{2}\frac{1+q}{1-q}\sum_{k=1}^{K} N_k P_k \tag{8}$$

Therefore, for items with uniform lengths, the data allocation problem can be optimally solved in polynomial time. This derives from Lemmas 1 and 2 of [5] which prove optimality in the particular case of error-free channels, that is, when $q = 0$. Indeed, when $q > 0$, similar proofs hold once the cost $C_{i,j}$ of assigning consecutive items $d_i, \ldots, d_j$ to the same channel is defined as $C_{i,j} = \frac{j-i+1}{2}\frac{1+q}{1-q}\sum_{h=i}^{j} p_h$. In words, Lemmas 1 and 2 of [5] show that, whenever the items $d_1, d_2, \ldots, d_N$ are sorted by non-increasing probabilities, there always exists an optimal solution which is a segmentation and which can be found by the Dichotomic algorithm.

Consider now items with non-uniform lengths and recall that Z_k is the period of channel k. In order to receive an item d_i of length z_i over channel k, a client has to listen for z_i consecutive error-free packet transmissions, which happens with probability $(1-q)^{z_i}$. Hence, the failure probability Q_{z_i} for item d_i on channel k is $1-(1-q)^{z_i}$.

In the case that the first transmission of d_i heard by the client is error-free, the client has to wait on the average $\frac{Z_k}{2}$ time units with probability $1-Q_{z_i}$. Instead, the client waits on the average for $\frac{Z_k}{2} + Z_k$ time units with probability $Q_{z_i}(1-Q_{z_i})$ in the case that the first transmission of d_i is erroneous and the second one is error-free. In general, h bad transmissions of d_i followed by a good one lead to a delay of $\frac{Z_k}{2} + hZ_k$ time units with probability $Q_{z_i}^h(1-Q_{z_i})$. Therefore, summing up over all h as seen in the uniform case, the expected delay becomes

$$t_i = \frac{Z_k}{2}\frac{1+Q_{z_i}}{1-Q_{z_i}} \tag{9}$$

Thus, the average expected delay to be minimized is

$$\text{AED} = \frac{1}{2}\sum_{k=1}^{K} Z_k \sum_{d_i \in G_k} \frac{1+Q_{z_i}}{1-Q_{z_i}} p_i \tag{10}$$

Recalling that the items are indexed by non-increasing $\frac{p_i}{z_i}$ ratios, the new recurrences for the Dichotomic and Dlinear algorithms are derived from Recurrences 5 and 6, respectively, once each $C_{i,j}$ is defined as $C_{i,j} = \frac{1}{2}\left(\sum_{h=i}^{j} z_h\right)\left(\sum_{h=i}^{j} \frac{1+Q_{z_h}}{1-Q_{z_h}} p_h\right)$. Note that in such a case optimality is not guaranteed since the problem is computationally intractable already for error-free channels.

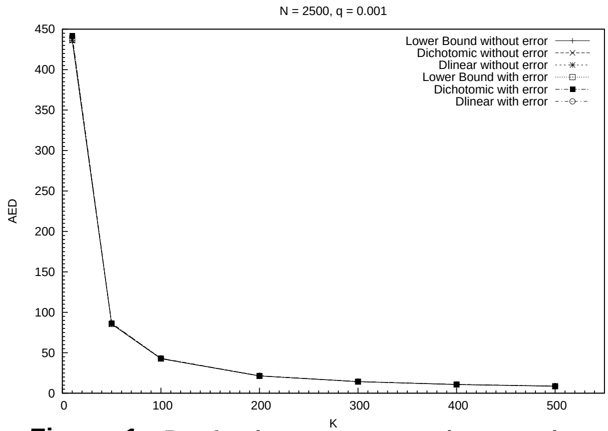

Figure 1. *Results for* 2500 *items of non-uniform lengths when the* K *channels have the same failure probability* $q = 0.001$.

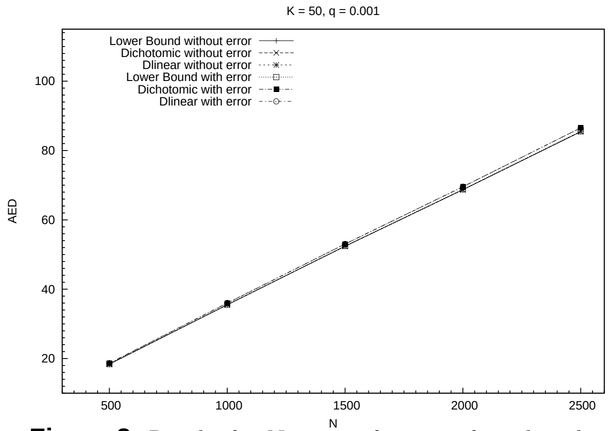

Figure 2. *Results for* N *items of non-uniform lengths when the* 50 *channels have the same failure probability* $q = 0.001$.

3.1 Simulation experiments

In this subsection, the behaviour of the Dichotomic and Dlinear heuristics is tested in the case of geometric channel error model. The heuristics were written in $C++$ and the experiments were run on an AMD Athlon X2 4800+ with 2 GB RAM. The above algorithms have been experimentally tested on benchmarks where the item probabilities follow a Zipf distribution. Specifically, given the number N of items and a real number $0 \leq \theta \leq 1$, the item probabilities are defined as

$$p_i = \frac{(1/i)^\theta}{\sum_{h=1}^{N}(1/h)^\theta} \qquad 1 \leq i \leq N$$

In the above formula, θ is the *skew* parameter. In particular, $\theta = 0$ stands for a uniform distribution with $p_i = \frac{1}{N}$, while $\theta = 1$ implies a high skew, namely the difference among the p_i values becomes larger. In the experiments, θ is chosen to be 0.8, as suggested in [19], while either N is set to 2500 and K varies in the range $10 \leq K \leq 500$, or K is set to 50 and N varies in the range $500 \leq N \leq 2500$. The channel failure probabilities can assume the values 0.001 and 0.01. The experiments are conducted only for the non-uniform case because in the case of data items with identical lengths the Dichotomic algorithm finds the optimal solution. The item lengths z_i are integers randomly generated according to a uniform distribution in the range $1 \leq z_i \leq 10$, for $1 \leq i \leq N$.

Moreover, since the data allocation problem is computationally intractable when data lengths are non-uniform, lower bounds for a non-uniform instance are derived by transforming it into a uniform instance as follows. Each item d_i of probability p_i and length z_i is decomposed in z_i items of probability $\frac{p_i}{z_i}$ and length 1. Since more freedom has been introduced, it is clear that the optimal AED for the so transformed problem is a lower bound on the AED of the original problem. Since the transformed problem has uniform lengths, when all the channels are either error-free or have the same failure probability, the optimal AED can be obtained by running the polynomial time Dichotomic algorithm.

Figures 1-4 show the experimental results for non-uniform lengths in the case that the failure probability q is 0.001 and 0.01. One can note that the two above mentioned lower bounds almost coincide. Referring to Figures 3 and 4, the AED of the transformed uniform length instance in the presence of errors is $\frac{1+q}{1-q} = 1.02$ times the AED of the same transformed instance without errors. One can also note that, since the average data item length is 5, the AED of the original instance in the presence of errors should be about $\frac{1+Q}{1-Q} = 1.10$ times the AED of the same original instance in the absence of errors, where $Q = 1-(1-0.01)^5 = 0.05$. This can be easily checked in Figure 3, e.g., for

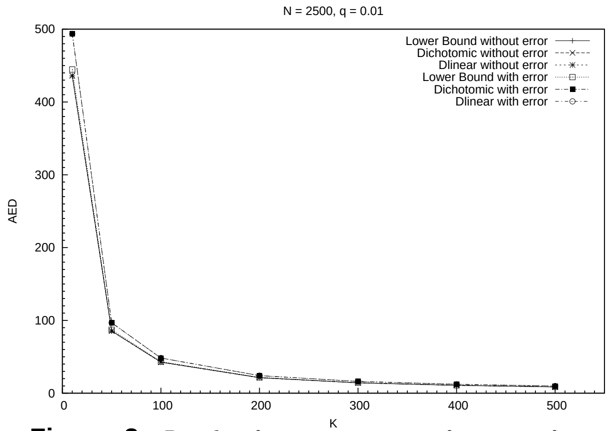

Figure 3. *Results for* 2500 *items of non-uniform lengths when the* K *channels have the same failure probability* $q = 0.01$.

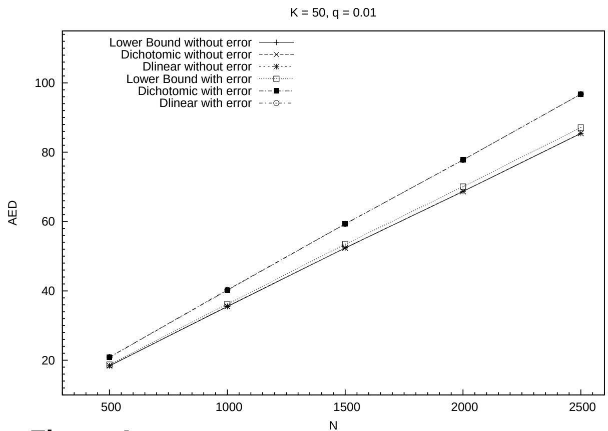

Figure 4. *Results for* N *items of non-uniform lengths when the* 50 *channels have the same failure probability* $q = 0.01$.

$K = 10$, where the ratio between the two AEDs is about $\frac{500}{450} = 1.11$.

4 Gilbert-Elliot channel error model

In this section, the channel error behavior is assumed to follow a Gilbert-Elliot model, which is a two-state time-homogeneous discrete time Markov chain [17]. At each time instant, a channel can be in one of two states. The state 0 denotes the *good* state, where the channel works properly and thus a packet is received with no errors. Instead, the state 1 denotes the *bad* state, where the channel is subject to failure and hence a packet is received with an unrecoverable error. Let $X_0, X_1, X_2, \ldots$ be the states of the channel at times $0, 1, 2, \ldots$. The time between X_u and X_{u+1} corresponds to the length of one packet. The initial state X_0 is selected randomly. As depicted in Figure 5, the probability of transition from the good state to the bad one is denoted by b, while that from the bad state to the good one is g. Hence, $1-b$ and $1-g$ are the probabilities of remaining in the same state, namely, in the good and bad state, respectively. Formally, $Prob[X_{u+1} = 0 | X_u = 0] = 1 - b$, $Prob[X_{u+1} = 0 | X_u = 1] = b$, $Prob[X_{u+1} = 1 | X_u = 1] = 1 - g$, and $Prob[X_{u+1} = 1 | X_u = 0] = g$.

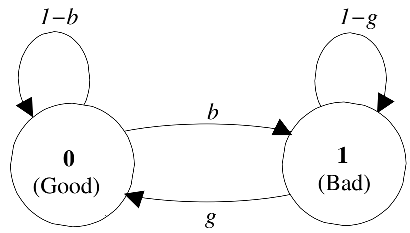

Figure 5. *The Gilbert-Elliot channel error model.*

It is well known that the *steady-state* probability of being in the good state is $P_G = \frac{g}{b+g}$, while that of being in the bad state is $P_B = \frac{b}{b+g}$. This markovian process has mean $\mu = P_B$, variance $\sigma^2 = \mu(1-\mu) = \frac{bg}{(b+g)^2}$, and autocorrelation function $r(\nu) = P_B + (1 - P_B)(1 - b - g)^{\nu}$, where $b + g < 1$ is assumed. Since the system is memoryless, the state holding times are geometrically distributed. The mean state holding times for the good state and the bad state are, respectively, $\frac{1}{b}$ and $\frac{1}{g}$. This means that the channel exhibits error bursts of consecutive ones whose mean length is $\frac{1}{g}$, separated by gaps of consecutive zeros whose mean length is $\frac{1}{b}$.

4.1 Uniform data item lengths

Assume that the item lengths are uniform, i.e., $z_i = 1$, for $1 \leq i \leq N$. Recall that in such a case the period of channel k is N_k. If a client waits for item d_i on channel k, and no error occurs in the first transmission of d_i, then the client waits on the average $\frac{N_k}{2}$ time units with probability $P_G = 1 - P_B$. Instead, if an error occurs during the first transmission of d_i and there is no error in the second trasmission, then the average delay experienced by the client is $\frac{N_k}{2} + N_k$ time units with probability $P_B(1-r(N_k))$. In general, when there are h erroneous transmissions of d_i followed by an error-free one, the client average delay is $\frac{N_k}{2} + hN_k$ time units with probability $P_B(r(N_k))^{h-1}(1-r(N_k))$. Thus, the expected delay is equal to

$$\frac{N_k}{2}P_G + P_B(1-r(N_k))\sum_{h=1}^{\infty}(\frac{N_k}{2} + hN_k)(r(N_k))^{h-1}$$

$$= \frac{N_k}{2}P_G + P_B\frac{N_k}{2} + P_B\frac{N_k}{1-r(N_k)}$$

because $\sum_{h=1}^{\infty}(r(N_k))^{h-1} = \frac{1}{1-r(N_k)}$ and $\sum_{h=1}^{\infty} h(r(N_k))^{h-1} = \frac{1}{(1-r(N_k))^2}$. Hence, the expected delay t_i and the objective function AED become, respectively:

$$t_i = \frac{N_k}{2}\left(1 + \frac{2P_B}{1-r(N_k)}\right) \quad (11)$$

$$\text{AED} = \frac{1}{2}\sum_{k=1}^{K}\left(N_k\left(1 + \frac{2P_B}{1-r(N_k)}\right)\sum_{d_i \in G_k} p_i\right) \quad (12)$$

The new recurrences for the Dichotomic and Dlinear algorithms are derived from Recurrences 5 and 6, respectively, by setting $C_{i,j} = \frac{j-i+1}{2}\left(1 + \frac{2P_B}{1-r(j-i+1)}\right)\sum_{h=i}^{j} p_h$.

4.2 Non-uniform data item lengths

Let now deal with non-uniform item lengths. Recall that Z_k is the period of channel k and that a client has to listen for z_i consecutive error-free packet transmissions in order to receive the item d_i over channel k.

Consider the case that the first transmission of d_i heard by a client be erroneous. Let $\hat{P}_B(s)$ denote the probability that in such a transmission the s-th packet is the first erroneous packet, where $1 \le s \le z_i$. Formally,

$$\hat{P}_B(s) = \begin{cases} P_B & \text{if } s = 1 \\ (1-P_B)(1-b)^{s-2}b & \text{if } 2 \le s \le z_i \end{cases}$$

Consider now the case that two consecutive transmissions of d_i heard by a client are erroneous. Let $\bar{P}_B(s,\sigma)$ denote the probability that, in the second transmission, the first erroneous packet is the s-th one given that in the previous transmission the first erroneous packet was the σ-th one. Thus, when $s = 1$, $\bar{P}_B(1,\sigma) = r(Z_k + 1 - \sigma)$, whereas when $2 \le s \le z_i$:

$$\bar{P}_B(s,\sigma) = (1 - r(Z_k + 1 - \sigma))(1-b)^{s-2}b$$

Finally, let $\bar{P}_G(\sigma)$ denote the probability that a whole transmission of d_i is error-free given that in the previous transmission of d_i the first erroneous packet was the σ-th one:

$$\bar{P}_G(\sigma) = (1 - r(Z_k + 1 - \sigma))(1-b)^{z_i-1}$$

To evaluate the expected delay t_i, observe that if the first transmission of d_i heard by the client is error-free, the client has to wait on the average $\frac{Z_k}{2}$ time units with probability $(1-P_B)(1-b)^{z_i-1}$. Instead, the client waits on the average for $\frac{Z_k}{2} + Z_k$ time units with probability $\sum_{s_0=1}^{z_i}\hat{P}_B(s_0)\bar{P}_G(s_0)$ in the case that the first transmission of d_i is erroneous and the second one is error-free. Moreover, two bad transmissions of d_i followed by a good one lead to a delay of $\frac{Z_k}{2} + 2Z_k$ time units with probability $\sum_{s_0=1}^{z_i}\left[\hat{P}_B(s_0)\sum_{s_1=1}^{z_i}\bar{P}_B(s_1,s_0)\bar{P}_G(s_1)\right]$. Thus, in general, the expected delay t_i is

$$\frac{Z_k}{2}(1-P_B)(1-b)^{z_i-1} + \sum_{h=1}^{\infty} \left[\left(\frac{Z_k}{2} + hZ_k\right)\sum_{s_0=1}^{z_i}\left[\hat{P}_B(s_0)\sum_{s_1=1}^{z_i}\left[\bar{P}_B(s_1,s_0)\sum_{s_2=1}^{z_i}\left[\bar{P}_B(s_2,s_1)\cdots\sum_{s_{h-1}=1}^{z_i}\left[\bar{P}_B(s_{h-1},s_{h-2})\bar{P}_G(s_{h-1})\right]\cdots\right]\right]\right]\right]$$

Since finding a closed formula for t_i seems to be difficult, an approximation t_i^m of the expected delay can be computed by truncating the above series at the m-th term, for a given constant value m. Indeed, experimental tests show that the series converges already for small values of m, as it will be checked in Subsection 4.3. Thus, the average expected delay becomes AED $= \sum_{i=1}^{N} t_i^m p_i$. Recalling that the items are indexed by non-increasing $\frac{p_i}{z_i}$ ratios, the Dichotomic and Dlinear algorithms can be applied once each $C_{i,j}$ is computed as $\sum_{h=i}^{j} t_h^m p_h$.

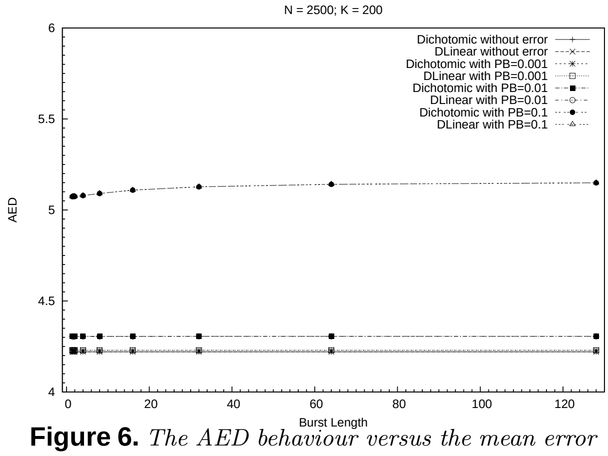

Figure 6. *The AED behaviour versus the mean error burst length, when $N = 2500$ and $K = 200$.*

4.3 Simulation experiments

This subsection presents the experimental tests for the Dichotomic and Dlinear heuristics in the case of the Gilbert-Elliot channel error model. In the experiments for items of uniform length, the item popularities follow a Zipf distribution with $\theta = 0.8$, as in Subsection 3.1, while either $N = 2500$ and $10 \leq K \leq 500$, or $K = 50$ and $500 \leq N \leq 2500$. Moreover, the steady-state probability P_B of being in the bad state can assume the values 0.001, 0.01 and 0.1, while the mean error burst length $\frac{1}{g}$ is fixed to 10. Note that b is derived as $g\frac{P_B}{1-P_B}$ once P_B and $\frac{1}{g}$ are fixed. However, the choice of $\frac{1}{g}$ is not critical because the sensitivity of the AED to $\frac{1}{g}$ is low, as depicted in Figure 6 for $N = 2500$, $K = 200$, and $1 < \frac{1}{g} \leq 130$. Note that the choice of such an upper bound on $\frac{1}{g}$ is not restrictive because the probability of having a burst with length n is $g(1-g)^{n-1}$, which is negligible as n grows.

Figures 7 and 8 report the experimental results in the case of uniform lengths and all channels with the same steady-state probability P_B. The graphics show that the impact of P_B is irrelevant when $P_B = 0.001$ and 0.01 because the AED values are almost the same as for error free channels, which in their turn are optimal. When $P_B = 0.1$, the AED value may increase up to the 20%, with respect to the error-free case.

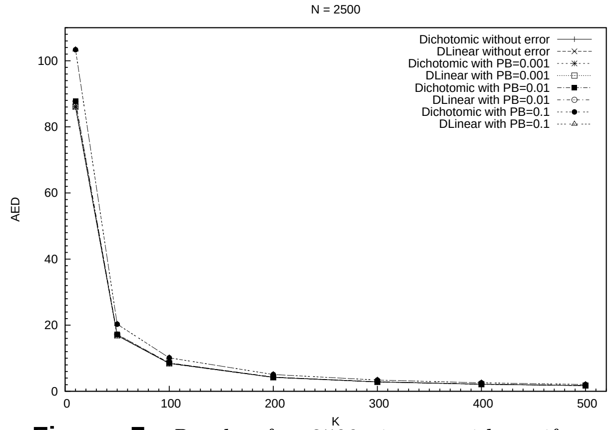

Figure 7. *Results for 2500 items with uniform lengths when the K channels have the same steady-state probability P_B, which assumes the values 0.001, 0.01, and 0.1.*

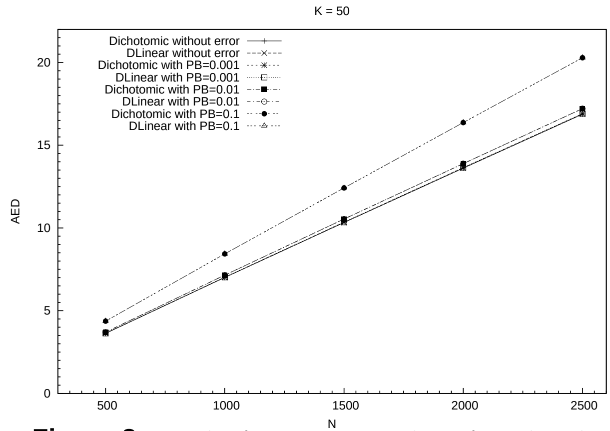

Figure 8. *Results for N items with uniform lengths when the 50 channels have the same steady-state probability P_B, which assumes the values 0.001, 0.01, and 0.1.*

Consider now data items whose lengths are non-uniform. In the experiments, the number K of channels is set to 50, the number N of items varies between 500 and 2000, the item popularities follow a Zipf distribution with $\theta = 0.8$, and the item lengths z_i are integers randomly generated according to a uniform distribution in the range $1 \leq z_i \leq 10$, for $1 \leq i \leq N$. The expected delay of item d_i is evaluated by computing t_i^5, that

m	t_i^m
1	25.9150699
2	25.9382262
3	25.9388013
4	25.9388156
5	25.9388160
6	25.9388167

Table 1. *Values of* t_i^m *when* $z_i = 10$*,* $Z_k = 50$*,* $\frac{1}{g} = 10$*, and* $P_B = 0.01$*.*

m	t_i^m
1	25.1989377
2	25.2537833
3	25.2689036
4	25.2730723
5	25.2745215
6	25.2745384

Table 2. *Values of* t_i^m *when* $z_i = 5$*,* $Z_k = 50$*,* $\frac{1}{g} = 10$*, and* $P_B = 0.16$*.*

is truncating at the fifth term the series giving t_i. Indeed, as shown in Table 1 for $z_i = 10$, $Z_k = 50$, $\frac{1}{g} = 10$, and $P_B = 0.01$ and in Table 2 for $z_i = 5$, $Z_k = 50$, $\frac{1}{g} = 10$, and $P_B = 0.16$, at the fifth term the series giving t_i is already stabilized up to the fourth decimal digit.

Since the data allocation problem is computationally intractable when data lengths are non-uniform, lower bounds for non-uniform instances are derived by transforming them into uniform instances, as explained in Subsection 3.1, and by running the Dichotomic algorithm. In particular, since the steady-state probability P_B is the same for all channels, the AEDs giving the lower bounds are obtained from Equation 12.

Figure 9 shows the experimental results for non-uniform lengths when P_B is identical for all channels and assumes the values 0.001, 0.01 and 0.1. In the figure, lower bounds are shown for both error-free and error-prone channels. One notes that, for every value of P_B, the behaviour of both the Dichotomic and Dlinear algorithms is identical. When $P_B = 0.001$, both algorithms provide optimal solutions because their AEDs almost coincide with the lower bound for channels without errors. When $P_B = 0.01$, the AEDs of both the Dichotomic and Dlinear algorithms are 12% larger

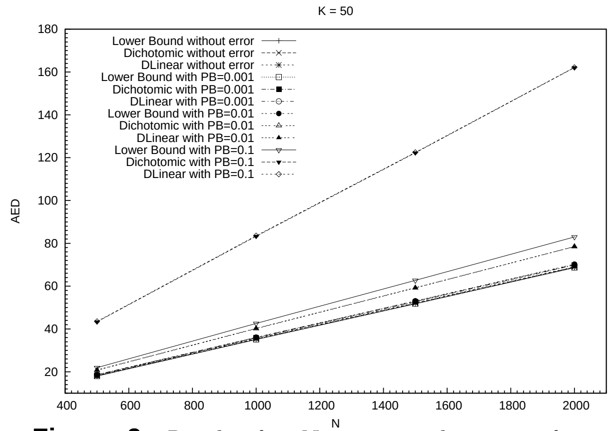

Figure 9. *Results for* N *items with non-uniform lengths when the* 50 *channels have the same steady-state probability* P_B*, which assumes the values* 0.001*,* 0.01*, and* 0.1*.*

than the lower bound in presence of errors. In the last case, namely $P_B = 0.1$, the AEDs found by the algorithms are as large as twice those of the lower bound in presence of errors. However, such a value of P_B represents an extremal case which should not arise in practice.

5 Conclusions

This paper studied the problem of allocating data to multiple channels, assuming skewed allocation of most popular data items to less loaded channels, flat data scheduling per channel, and the presence of unrecoverable channel transmission errors. The objective was that of minimizing the average expected delay experienced by clients. The behaviour of two polynomial time heuristics has been experimentally tested modelling the channel error by means of the geometric model as well as the Gilbert-Elliot one. Extensive simulations showed that such heuristics provide good sub-optimal solutions when tested on benchmarks whose item popularities follow Zipf distributions. Since the problem is computationally intractable (that is, NP-hard) for non-uniform data lengths and error-free channels, the computational complexity of the problem in the presence of errors remains an open issue only for uniform data lengths and the Gilbert-Elliot model. As regard to the non-uniform case, an interesting open question is that of determining whether a closed formula for computing the item expected delays exists or not when the Gilbert-Elliot model is adopted.

References

[1] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik. Broadcast disks: data management for asymmetric communication environments. In *Proc. SIGMOD*, May 1995.

[2] M.H. Ammar and J.W. Wong. The design of teletext broadcast cycles. *Performance Evaluation*, 5(4):235–242, 1985.

[3] M.H. Ammar and J.W. Wong. On the optimality of cyclic transmission in teletext systems. *IEEE Transactions on Communications*, 35(11):1159–1170, 1987.

[4] S. Anticaglia, F. Barsi, A.A. Bertossi, L. Iamele, and M.C. Pinotti. Efficient heuristics for data broadcasting on multiple channels. *Wireless Networks*, 2007, to appear.

[5] E. Ardizzoni, A.A. Bertossi, M.C. Pinotti, S. Ramaprasad, R. Rizzi, and M.V.S. Shashanka. Optimal skewed data allocation on multiple channels with flat broadcast per channel. *IEEE Transactions on Computers*, 54(5):558–572, 2005.

[6] A. Bar-Noy, R. Bhatia, J.S. Naor, and B. Schieber. Minimizing service and operation costs of periodic scheduling. In *Proc. Ninth ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 11–20, 1998.

[7] A.A. Bertossi, M.C. Pinotti, and R. Rizzi. Scheduling data broadcasts on wireless channels: Exact solutions and heuristics. In T. Gonzalez (Editor). *Handbook of Approximation Algorithms and Metaheuristics.* Taylor & Francis Books (CRC Press), Boca Raton, 2007.

[8] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and Zipf-like distributions: evidence and implications. In *Proc. IEEE INFOCOM*, 1999.

[9] T. Imielinski, S. Viswanathan, and B.R. Badrinath. Energy efficient indexing on air. In *Proc. SIGMOD*, May 1994.

[10] C. Kenyon and N. Schabanel. The data broadcast problem with non-uniform transmission time. In *Proc. Tenth ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 547–556, 1999.

[11] C. Kenyon, N. Schabanel, and N. Young. Polynomial time approximation scheme for data broadcast. In *Proc. ACM Symp. on Theory of Computing (STOC)*, pages 659–666, 2000.

[12] S.-C. Lo and A.L.P. Chen Optimal index and data allocation in multiple broadcast channels. In *Proc. Sixteenth IEEE Int'l Conf. on Data Engineering (ICDE)*, February 2000.

[13] W.C. Peng and M.S. Chen. Efficient channel allocation tree generation for data broadcasting in a mobile computing environment. *Wireless Networks*, 9(2):117–129, 2003.

[14] K.A. Prabhakara, K.A. Hua, and J. Oh. Multi-level multi-channel air cache designs for broadcasting in a mobile environment. In *Proc. Sixteenth IEEE Int'l Conf. on Data Engineering (ICDE)*, February 2000.

[15] I. Stojmenovic (Editor). *Handbook of Wireless Networks and Mobile Computing.* Wiley, Chichester, 2002.

[16] N. Vaidya and S. Hameed. Log time algorithms for scheduling single and multiple channel data broadcast. In *Proc. Third ACM-IEEE Conf. on Mobile Computing and Networking (MOBICOM)*, September 1997.

[17] A. Willig. Redundancy concepts to increase transmission reliability in wireless industrial LANs. *IEEE Transactions on Industrial Informatics*, 1(3): 173-182, 2005.

[18] W.G. Yee, Efficient data allocation for broadcast disk arrays. *Technical Report*, GIT-CC-02-20, Georgia Institute of Technology, 2001.

[19] W.G. Yee, S. Navathe, E. Omiecinski, and C. Jermaine. Efficient data allocation over multiple channels at broadcast servers. *IEEE Transactions on Computers*, 51(10):1231–1236, 2002.

[20] M. Zorzi, R. Rao, and L.B. Milstein. Error statistics in data transmission over fading

channels. *IEEE Transactions on Communications*, 46(11):1468–1477, 1998.