This paper was accepted for publication on Wireless Networks in 2008.

DOI: 10.1007/s11276-008-0136-z hyperlink: <http://dx.doi.org/10.1007/s11276-008-0136-z>

Allocating data for broadcasting over wireless channels subject to transmission errors

Paolo Barsocchi * Alan A. Bertossi † M. Cristina Pinotti ‡ Francesco Potortì *

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 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 the clients. Two different channel error models are considered: the Bernoulli model and the simplified 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. Particular cases are detected where optimal solutions can be found in polynomial time. For general cases, simulations show that good sub-optimal solutions can be found on benchmarks whose item popularities follow Zipf distributions.

Keywords Wireless communication, Data broadcasting, Multiple channels, Flat scheduling, Average expected delay, Channel transmission errors, Bernoulli model, Gilbert-Elliot model, Heuristics.

1 Introduction

In wireless asymmetric communications, broadcasting is an efficient way of simultaneously disseminating data to a large number of clients [17]. 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 and may be influenced by transmission errors. Although data are usually encoded using *error correcting codes* (*ECC*) allowing some recoverable errors to be corrected by the client, there are several transmission errors which still cannot be corrected using ECC. Such *unrecoverable* errors affect the client expected delay, because the resulting corrupted item has 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]–[7],[9]–[11],[13, 15, 16, 19, 21, 22].

The database community usually partitions the data among the channels and then adopts a *flat* broadcast schedule on each channel [5, 15, 22], which consists in cyclically broadcasting in an arbitrary fixed order, that is once at a time in a round-robin fashion, the items assigned to the same channel [1]. 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 so-

*ISTI-CNR, via G. Moruzzi 1, 56124 Pisa, Italy, {Paolo.Barsocchi,potorti}@isti.cnr.it

†Dept. of Computer Science, University of Bologna, 40127 Bologna, Italy, bertossi@cs.unibo.it

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

lution that minimizes the average expected delay can be found in polynomial time in the case of *unit lengths* [22], that is when the transmission time is equal to one for all items, whereas the problem becomes NP-complete for non-unit lengths [10]. In this latter case, several heuristics have been developed in [4, 22], 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]. Relaxing the assumptions that each channel has to broadcast a flat schedule and that no item is present on two channels, the NP-completeness of the problem is still open.

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 mainly concentrates on finding broadcast schedules for a single channel to minimize the average expected delay [6, 10, 11, 19], since it usually assumes all items replicated over all channels. 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 near-optimal schedules [3]. In particular, the solution proposed by [19] adapts the SRR solution to the case of unrecoverable errors.

The present paper extends the data allocation problem first studied by the database community under the assumptions of multiple channels and flat data schedule per channel [4, 5, 22], to cope with the presence of erroneous transmissions, under the same assumptions of [19], namely unrecoverable errors. Two different error models will be considered to describe the behavior of wireless channels [20]. First, as in [19], the Bernoulli channel error is assumed, 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 so called simplified Gilbert-Elliot channel error model will be considered, which was not previously studied in [19]. 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 [18, 23]. As in [19], the erroneous transmissions are taken into account in the problem parameters and they are compensated by properly modifying the allocation of data items to channels.

Briefly, this paper is so organized. The rest of this section gives basic definitions and recalls the main dynamic programming algorithms for error-free channels. Sections 2 and 3 consider the Bernoulli and the Gilbert-Elliot channel error models, respectively, and illustrate how to adapt the recurrences in the previously recalled algorithms to cope with channel errors, for both items of unit and non-unit lengths. In particular, it is also shown how to find optimal solutions for some special cases. Experimental tests are reported at the end of both sections on benchmarks whose items probabilities are characterized by Zipf distributions, showing that good sub-optimal solutions are found. Finally, conclusions are offered in Section 4.

1.1 Background

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 *popularity* p_i and a *length* z_i, with $1 \leq i \leq N$. The popularity p_i represents the probability of item d_i to be requested by the clients. 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 correcting 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 unit, i.e., $z_i = 1$ for $1 \leq i \leq N$, the lengths are called *unit* lengths, otherwise they are said to be *non-unit* lengths.

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:

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

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, 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

Algorithm	Recurrence	Complexity	Lengths	Solution
DP	$sol_{1,n} = C_{1,n}$	$O(N^2K)$	unit	optimal
[22]	$sol_{k,n} = \min_{1\leq\ell\leq n-1} \{sol_{k-1,\ell} + C_{\ell+1,n}\}$		non-unit	sub-opt
Dichotomic	$sol_{1,n} = C_{1,n}$	$O(NK\log N)$	unit	optimal
[5]	$sol_{k,\lceil\frac{l+r}{2}\rceil} = \min_{B^l_{k-1}\leq\ell\leq B^r_{k-1}} \{sol_{k-1,\ell} + C_{\ell+1,\lceil\frac{l+r}{2}\rceil}\}$		non-unit	sub-opt
Dlinear	$sol_{k,n} = sol_{k-1,m} + C_{m+1,n}$	$O(N(K+\log N))$	unit	sub-opt
[4]	$m = \min_{B^{n-1}_k\leq\ell\leq n-1} \{\ell \ : sol_{k-1,\ell} + C_{\ell+1,n} < sol_{k-1,\ell+1} + C_{\ell+2,n}\}$		non-unit	sub-opt
Knapsack	$M_{i,j} = M_{i-1,j}$ if $j < z_i$	$O(NZ)$	unit	optimal
[5]	$M_{i,j} = \max\{M_{i-1,j},\ M_{i-1,j-z_i} + p_i\}$ if $j \geq z_i$	$K = 2$ only	non-unit	optimal

Table 1: *Main algorithms for the Data Allocation problem of N items on K error-free channels (Z is the sum of all item lengths and B^n_{k-1} is the $(k-1)$-th right border of the optimal solution with k channels and n items).*

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

$$\text{AED} = \sum_{k=1}^{K}\sum_{d_i\in G_k}\frac{Z_k}{2}p_i = \frac{1}{2}\sum_{k=1}^{K} Z_k P_k \qquad (2)$$

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

Thus, 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.

Almost all the algorithms proposed so far on error-free channels are based on dynamic programming and restrict the search for the solutions to *segmentations*, that is, partitions obtained by considering the items ordered by their indices, and by assigning items with consecutive indices to each channel. 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.

The recurrences for the four main dynamic programming algorithms for the data allocation problem, called *DP*, *Dichotomic*, *Dlinear*, and *Knapsack*, are reported in Table 1 along with their time complexity and solution optimality. All the algorithms, except the last one, work for K channels, assume that the items $d_1, d_2, \ldots, d_N$ are indexed by non-increasing $\frac{p_i}{z_i}$ ratios, that is $\frac{p_1}{z_1} \geq \frac{p_2}{z_2} \geq \cdots \geq \frac{p_N}{z_N}$, and denote with $C_{i,j}$ the cost of assigning consecutive items $d_i, \ldots, d_j$ to a single channel:

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

In contrast, Knapsack works for 2 channels only, assumes the items in an arbitrary order, and finally selects the entry $M_{N,\overline{\jmath}}$ which minimizes $\frac{1}{2}\left(\overline{\jmath}M_{N,\overline{\jmath}} + (Z-\overline{\jmath})(1-M_{N,\overline{\jmath}})\right)$.

2 Bernoulli error model

In this section, unrecoverable channel transmission errors modeled by a geometric distribution are taken into account. Each packet transmission over channel k has the same probability q_k to fail and $1-q_k$ to succeed, and each transmission error is independent from the others, with $1 \leq k \leq K$ and $0 \leq q_k \leq 1$. If a client receives a corrupted item d_i, it discards the item and then has to wait for a whole period Z_k, until the next transmission of d_i scheduled by the server.

2.1 Unit length items

Assume that the item lengths are unit, i.e., $z_i = 1$ for $1 \leq i \leq N$. If a client wants to receive item d_i and there are h bad transmissions of d_i followed by a good one, the client average delay for receiving item d_i is $\frac{N_k}{2} + hN_k$ time units with probability $q_k^h(1-q_k)$. Thus, the expected delay is:

$$t_i = \sum_{h=0}^{\infty}\left(\frac{N_k}{2} + hN_k\right)q_k^h(1-q_k) = \frac{N_k}{2}\frac{1+q_k}{1-q_k}$$

Using the property that $\sum_{i=1}^{n} a_i b_i$ is maximized when both sequences $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are sorted in the same order, one can prove by contradiction that there is an optimal solution where the items are sorted by non-increasing popularities.

Lemma 1. *Let G_h and G_j be two groups in an optimal solution. Let d_i and d_k be items with $d_i \in G_h$ and $d_k \in G_j$. If $N_h \frac{1+q_h}{1-q_h} < N_j \frac{1+q_j}{1-q_j}$, then $p_i \geq p_k$. Similarly, if $p_i > p_k$, then $N_h \frac{1+q_h}{1-q_h} \leq N_j \frac{1+q_j}{1-q_j}$.*

Proof. By contradiction, let $G_1, G_2, \ldots, G_K$ be an optimal solution for which there exist G_h and G_j such that $N_h \frac{1+q_h}{1-q_h} < N_j \frac{1+q_j}{1-q_j}$ and $p_i < p_k$. Consider now another solution obtained by exchanging d_i with d_k in the two groups G_h and G_j. The AED difference between the optimal solution and the other one is $\left(N_h \frac{1+q_h}{1-q_h} - N_j \frac{1+q_j}{1-q_j}\right)(p_i - p_k) > 0$ because $p_i - p_k < 0$ and $N_h \frac{1+q_h}{1-q_h} - N_j \frac{1+q_j}{1-q_j} < 0$. Hence, a better solution is achieved contradicting the optimality assumption. The last part of the lemma is proved similarly. □

Likewise, one can show that an optimal solution exists where the channels are indexed by non-decreasing channel error probabilities.

Lemma 2. *Let G_h and G_j be two groups in an optimal solution. If $N_h P_h > N_j P_j$, then $q_h \leq q_j$. Similarly, if $q_h < q_j$, then $N_h P_h \geq N_j P_j$.*

Unfortunately, one can easily realize that an optimal solution which is a segmentation and takes the channels by non-decreasing error probabilities does not always exist. However, in the special case where there are only two channels, an optimal solution can be found in $O(N \log N)$ time by a single scan of the data items, exploiting the following result.

Corollary 1. *Assume $K = 2$ and the items sorted by non-increasing popularities, and let (B_1) be an optimal segmentation. Then, $B_1 \leq (N - B_1) \frac{1+q_{max}}{1-q_{max}} \frac{1-q_{min}}{1+q_{min}}$, where q_{max} and q_{min} are the larger and the smaller error probabilities, respectively. Moreover, if $B_1 \geq \lceil \frac{N}{2} \rceil$ then the items $d_1, \ldots, d_{B_1}$ are assigned to the channel with error probability q_{min}.*

Proof. By contradiction, let $B_1 > (N - B_1) \frac{1+q_{max}}{1-q_{max}} \frac{1-q_{min}}{1+q_{min}}$. Then $N_1 \frac{1+q_{min}}{1-q_{min}} > N_2 \frac{1+q_{max}}{1-q_{max}}$. By Lemma 1, the item popularities are non-decreasing contradicting the assumption. To show the remaining property, observe that, since $B_1 \geq \lceil \frac{N}{2} \rceil$ and the items are sorted by non-increasing popularities, then $N_1 \geq N_2$, $P_1 \geq P_2$, and hence $N_1 P_1 \geq N_2 P_2$. By Lemma 2, the channels must be taken by increasing error probabilities. Therefore, the first group of items will be assigned to the channel with minimum error probability q_{min}. □

In the particular case that all the channels have the same probability to fail, that is, $q_1 = \cdots = q_K = q$, the problem can still be optimally solved in polynomial time by just using the same algorithm as where there are no errors. In fact the objective function is just the same up to the constant factor $\frac{1+q}{1-q}$.

Another particular case that can be optimally solved in polynomial time arises when all the channels, but one, have the same probability to fail, namely, $q_1 = \cdots = q_{K-1} = q$ and $q_K = q'$. Let $C_{i,j} = \frac{j-i+1}{2} \frac{1+q}{1-q} \sum_{h=i}^{j} p_h$ and $C'_{i,j} = \frac{j-i+1}{2} \frac{1+q'}{1-q'} \sum_{h=i}^{j} p_h$ be the cost of assigning consecutive items $d_i, \ldots, d_j$ to a channel with error probability q and q', respectively. Moreover, let $opt_{k,n}$ be the cost of an optimal segmentation for the first n items using k channels all having the same error probability q. Similarly, let $opt'_{k,n}$ be the cost of an optimal segmentation when one of the k channels has error probability q'. Clearly, $opt_{1,n} = C_{1,n}$ and $opt'_{1,n} = C'_{1,n}$. The optimal solution $opt'_{K,N}$ can be derived in $O(N^2 K)$ time applying the following recurrence, which exploits the fact that there is exactly one channel with different error probability q', with $1 < k \leq K$:

$$opt'_{k,n} = \min_{1 \leq \ell \leq n-1} \left\{ \min \left\{ \begin{array}{l} opt_{k-1,\ell} + C'_{\ell+1,n}, \\ opt'_{k-1,\ell} + C_{\ell+1,n} \end{array} \right\} \right\}$$

where, for $1 < k \leq K - 1$:

$$opt_{k,n} = \min_{1 \leq \ell \leq n-1} \{ opt_{k-1,\ell} + C_{\ell+1,n} \}$$

In the general case that the error probabilities of the K channels are not the same, both the Dichotomic and Dlinear algorithms can be used but with no guarantee that the so found solutions are optimal. Indeed, since it is not known which order of the channels will lead to the optimal solution, a reasonable greedy criterion can be that of assigning the most popular items to the most reliable channels, that is, indexing the channels so that

$q_1 \leq q_2 \leq \cdots \leq q_K$. Thus, letting the cost $C_{i,j;k}$ of assigning consecutive items $d_i, \ldots, d_j$ to channel k be $C_{i,j;k} = \frac{j-i+1}{2} \frac{1+q_k}{1-q_k} \sum_{h=i}^{j} p_h$, the recurrences of the Dichotomic and Dlinear algorithms shown in Table 1 can be applied by using the above $C_{i,j;k}$'s in place of the $C_{i,j}$'s defined in Equation 3. All the $C_{i,j;k}$'s can be calculated in $O(NK)$ time via proper prefix-sum computations, assuming that the items are already sorted, and thus the time complexities of the Dichotomic and Dlinear algorithms remain the same.

2.2 Non-unit length items

Consider now items with non-unit length 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 $(1 - q_k)^{z_i}$. Hence, the error probability for item d_i on channel k is $Q_{z_i,k} = 1 - (1 - q_k)^{z_i}$.

Since 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 $Q^h_{z_i,k}(1 - Q_{z_i,k})$, the expected delay becomes

$$t_i = \sum_{h=0}^{\infty} \left(\frac{Z_k}{2} + hZ_k \right) Q^h_{z_i,k}(1 - Q_{z_i,k}) = \frac{Z_k}{2} \frac{1 + Q_{z_i,k}}{1 - Q_{z_i,k}}$$

Recalling that the items are indexed by non-increasing $\frac{p_i}{z_i}$ ratios, the recurrences of Dichotomic and Dlinear algorithms can be used once the channels are indexed so that $q_1 \leq q_2 \leq \cdots \leq q_K$ and each $C_{i,j}$ is replaced by $C_{i,j;k} = \frac{1}{2} \left(\sum_{h=i}^{j} z_h \right) \left(\sum_{h=i}^{j} \frac{1+Q_{z_h,k}}{1-Q_{z_h,k}} p_h \right)$. All the $C_{i,j;k}$'s can be computed in $O(KH)$ time via prefix-sums, where $H = \min\{N \log z, z\}$ and $z = \max_{1 \leq h \leq N}\{z_h\}$. The time complexities of the Dichotomic and Dlinear algorithms become, respectively, $O(K(H + N \log N))$ and $O(KH + KN + N \log N)$. Note that in such a case optimality is not guaranteed since the problem is computationally intractable already for error-free channels.

However, when there are only two channels having the same error probability $q = q_1 = q_2$, an optimal solution can be found in $O(NZ)$ time applying the Knapsack algorithm simply replacing each popularity p_i with $p'_i = \frac{1+Q_{z_i}}{1-Q_{z_i}} p_i$, where $Q_{z_i} = 1 - (1 - q)^{z_i}$ (see [5] for the details of the algorithm).

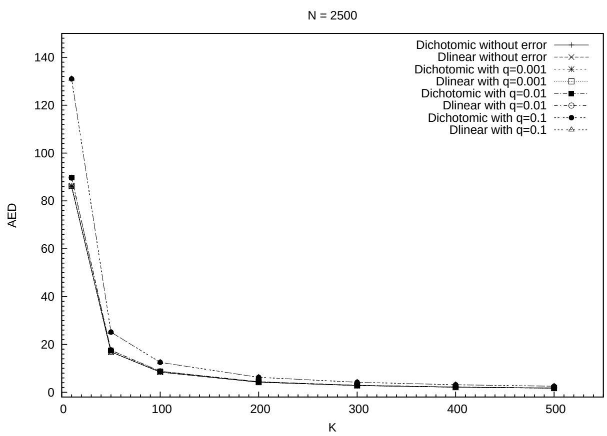

Figure 1: *Results for unit lengths when the channels are partitioned into three groups of the same size with error probability q, $2q$, and $3q$, respectively.*

2.3 Simulation experiments

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

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

Note that the item popularities are already sorted in non-increasing order. 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 a higher θ implies a higher skew, namely the difference among the p_i values becomes larger. In the experiments, θ is chosen to be 0.8, as suggested in [22], while N is set to 2500 and K varies in the range $10 \leq K \leq 500$. The channel error probabilities can assume the values 0.001, 0.01 and 0.1.

Figure 1 exhibits the AED obtained in the case that the data lengths are unit and the error probabilities are not identical for all channels. In particular, the channels are partitioned into three equally-sized groups with error probability q, $2q$, and $3q$, respectively. In other words, $q_1 = \cdots = q_{\lfloor \frac{K}{3} \rfloor} = q$,

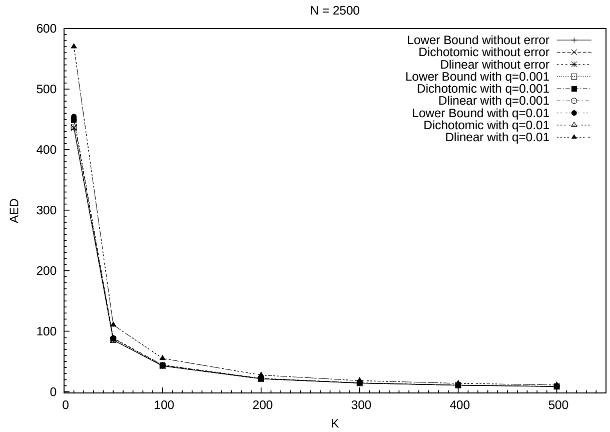

Figure 2: *Results for non-unit lengths when the channels are partitioned into three groups of the same size with error probability* q*,* $2q$*, and* $3q$*, respectively.*

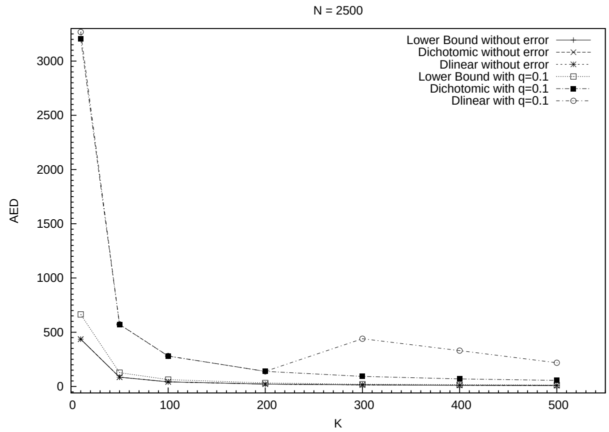

Figure 3: *Results for non-unit lengths when the channels are partitioned into three groups of the same size with error probability* 0.1*,* 0.2*, and* 0.3*, respectively.*

$q_{\lfloor \frac{K}{3} \rfloor + 1} = \cdots = q_{\lfloor \frac{2}{3}K \rfloor} = 2q$, and $q_{\lfloor \frac{2}{3}K \rfloor + 1} = \cdots = q_K = 3q$. One can observe that, when $q = 0.001$ and 0.01, the reported AEDs almost coincide with those where the channels are error-free. In other words, such small error probability values scarcely affect the average expected delay, which remains the optimal one found by the Dichotomic algorithm in the case of channels with no error. Whereas, the larger value $q = 0.1$ worsens the AED when the number K of channels is small with respect to the number N of items. Noting that all the channels have at least an error probability of $q = 0.1$, the AED in presence of errors must be at least $\frac{1+q}{1-q} = 1.22$ times the AED without errors. This is consistent with the AED reported in Figure 1, which is about 1.44 times the AED without errors, as computed by both the Dlinear and Dichotomic algorithms.

Consider now data items whose lengths are non-unit. In the experiments, the item lengths z_i are integers randomly generated according to a uniform distribution in the range $1 \leq z_i \leq 10$, for $1 \leq i \leq N$, as suggested in [19]. In addition, the reported results are averaged over 3 independent experiments. Moreover, since the data allocation problem is computationally intractable when data lengths are non-unit, lower bounds for a non-unit length instance are derived by transforming it into a unit length instance as follows. Each item d_i of popularity p_i and length z_i is decomposed into z_i items of popularity $\frac{p_i}{z_i}$ and length 1. Then, the AED obtained running the Dichotomic algorithm on the transformed instance gives a lower bound for the original non-unit instance.

Figures 2 and 3 plot the AEDs obtained for non-unit lengths and three equally-sized channel groups with error probability q, $2q$, and $3q$. When $q = 0.001$, the AEDs in Figure 2 almost coincide with those where the channels are error-free, as happened in the case of unit lengths. When $q = 0.01$, since the average data item length is 5 and the average channel error probability is 0.02, the AED of the original instance in the presence of error should be about $\frac{1+Q}{1-Q} = 1.22$ times the AED of the same original instance in the absence of error, where $Q = 1 - (1 - 0.02)^5 = 0.10$. In Figure 2, the largest ratios between the two above mentioned AEDs occur for small values of K, e.g., when $K = 10$ such a ratio is about $\frac{570}{440} = 1.29$. When $q = 0.1$, a similar reasoning leads to $Q = 1 - (1 - 0.2)^5 = 0.68$ and $\frac{1+Q}{1-Q} = 5.25$, while the largest ratio, for $K = 10$, is about $\frac{3200}{450} = 7.11$, as one can see in Figure 3. Moreover, one notes that the Dlinear algorithm has a bump for $K = 300$ because the selection of m in its recurrence (see Table 1) could be trapped in a bad local minimum.

3 Gilbert-Elliot error model

In this section, the channel error behavior is assumed to follow a simplified Gilbert-Elliot model, which is a two-state time-homogeneous discrete time Markov chain [20]. At each time instant, a channel can be in

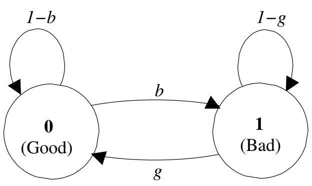

Figure 4: *The Gilbert-Elliot channel error model.*

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 4, 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, $Prob[X_{u+1} = 0 | X_u = 0] = 1-b$, $Prob[X_{u+1} = 0 | X_u = 1] = g$, $Prob[X_{u+1} = 1 | X_u = 1] = 1-g$, and $Prob[X_{u+1} = 1 | X_u = 0] = b$.

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. Recall that $r(\nu)$ is the probability of being in the same state after ν time units. 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 $\frac{1}{g}$, that are separated by gaps of consecutive zeros whose mean length is $\frac{1}{b}$.

3.1 Unit length items

Assume that the item lengths are unit, i.e., $z_i = 1$ for $1 \le i \le N$. If 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 $P_B(r(N_k))^{h-1}(1-r(N_k))$. Indeed, P_B is the probability of being in the bad state at the first transmission of d_i, $r(N_k)^{h-1}$ is the probability of remaining in the same state for the next $h-1$ transmissions, each at pairwise distance of N_k, and finally $1-r(N_k)$ is the probability of changing state at the h-th transmission. Thus, the expected delay is equal to

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

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

The following result, analogous to Lemma 1, shows that there is an optimal solution where the items are sorted by non-increasing popularities.

Lemma 3. *Let G_h and G_j be two groups in an optimal solution. Let d_i and d_k be items with $d_i \in G_h$ and $d_k \in G_j$. If $N_h\left(1+\frac{2P_B}{1-r(N_h)}\right) < N_j\left(1+\frac{2P_B}{1-r(N_j)}\right)$, then $p_i \ge p_k$. Similarly, if $p_i > p_k$, then $N_h\left(1+\frac{2P_B}{1-r(N_h)}\right) \le N_j\left(1+\frac{2P_B}{1-r(N_j)}\right)$.*

By the above lemma, there is an optimal solution which is a segmentation and can be found in $O(N^2K)$ time by the DP algorithm, setting

$$C_{i,j} = \frac{j-i+1}{2}\left(1+\frac{2P_B}{1-r(j-i+1)}\right)\sum_{h=i}^{j} p_h$$

In the general case where the steady state probabilities of being in the bad state are not identical for all channels, both the Dichotomic and Dlinear algorithms can still be applied to find sub-optimal solutions, after indexing the channels by non decreasing P_B's, namely $P_{B_1} \le \cdots \le P_{B_K}$, and replacing $C_{i,j}$ with

$$C_{i,j;k} = \frac{j-i+1}{2}\left(1+\frac{2P_{B_k}}{1-r_k(j-i+1)}\right)\sum_{h=i}^{j} p_h$$

where $r_k(\nu) = P_{B_k} + (1-P_{B_k})(1-b_k-g_k)^\nu$. As usual, all the $C_{i,j;k}$'s can be computed in $O(NK)$ time via prefix-sums.

In the special case where there are only two channels, an optimal solution can be efficiently found by exploiting the properties of the AED objective function. Indeed, the problem is to find a partition G_1 and G_2 such that

$\frac{1}{2}\left(N_1\left(1+\frac{2P_{B_1}}{1-r_1(N_1)}\right)P_1+N_2\left(1+\frac{2P_{B_2}}{1-r_2(N_2)}\right)P_2\right)$

is minimized. Since $P_2 = 1 - P_1$ and $N_2 = N - N_1$, when N_1 is fixed to a particular value, the AED is minimized by assigning to group G_1 the N_1 items with either the smallest or largest popularities, depending on whether

$N_1\left(2+\frac{2P_{B_1}}{1-r_1(N_1)}\right)+(N_1-N)\frac{2P_{B_2}}{1-r_2(N-N_1)}-N$

is positive or not, respectively. Such a property implies that there is an optimal solution which is a segmentation and which can be found in $O(N \log N)$ time by scanning all the possible values of N_1 once the items have been sorted by non-increasing popularities.

3.2 Non-unit length items

Let us now deal with items having non-unit 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 first transmission of item d_i heard by a client. Let $\hat{P}_B(s)$ denote the probability that in such a transmission the s-th packet is the first erroneous packet, where $1 \leq s \leq z_i$. Formally, $\hat{P}_B(1) = P_B$, and, for $2 \leq s \leq z_i$:

$$\hat{P}_B(s) = (1 - P_B)(1 - b)^{s-2}b$$

Consider now two consecutive transmissions of item d_i heard by a client, the first of which is 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, $\bar{P}_B(1, \sigma) = r(Z_k + 1 - \sigma)$ and, for $2 \leq s \leq z_i$:

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

Indeed, observe that $Z_k + 1 - \sigma$ is the distance between the erroneous packets in the two consecutive transmissions of d_i. Hence, when $s = 1$, the required probability coincides with $r(Z_k + 1 - \sigma)$. Otherwise, if $s > 1$, $1 - r(Z_k + 1 - \sigma)$ takes into account that receiving the first packet the system has changed state, $(1-b)^{s-2}$ aggregates the probability of receiving further $s - 2$ good packets, and b that of receiving the s-th corrupted packet.

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:

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

As before, $1 - r(Z_k + 1 - \sigma)$ considers that the system state changed between the σ-th packet of the first transmission and the first packet of the second transmission. Moreover, $(1 - b)^{z_i-1}$ gives the probability of remaining in the same good state for the next $z_i - 1$ packets.

Note that all the $\hat{P}_B(s)$ and $\bar{P}_B(s, \sigma)$'s can be computed in pseudo-polynomial time, that is in a time polynomial in the parameters Z and z, where $Z = \sum_{i=1}^{N} z_i$ and $z = \max_{1\leq i\leq N}\{z_i\}$.

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

$$\pi_0 = (1 - P_B)(1 - b)^{z_i-1}.$$

Instead, the client waits on the average for $\frac{Z_k}{2} + Z_k$ time units with probability

$$\pi_1 = \sum_{s_0=1}^{z_i} \hat{P}_B(s_0)\bar{P}_G(s_0)$$

in the case that the first transmission of d_i is erroneous and the second one is error-free. Indeed, $\hat{P}_B(s_0)\bar{P}_G(s_0)$ is the probability that the second transmission of d_i is good given that in the previous one the s_0-th packet was the first erroneous packet. Hence, π_1 is obtained varying s_0 among the z_i packets of d_i.

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

$$\pi_2 = \sum_{s_0=1}^{z_i}\left[\hat{P}_B(s_0)\sum_{s_1=1}^{z_i}\bar{P}_B(s_1, s_0)\bar{P}_G(s_1)\right].$$

When h bad transmissions of d_i are followed by a good one, the delay is $\frac{Z_k}{2} + hZ_k$ time units with probability

$$\pi_h = \sum_{s_0=1}^{z_i}\left[\hat{P}_B(s_0)\sum_{s_1=1}^{z_i}\left[\bar{P}_B(s_1, s_0)\sum_{s_2=1}^{z_i}\left[\bar{P}_B(s_2, s_1)\cdots\right.\right.\right.$$

$$\left.\left.\left.\cdots\sum_{s_{h-1}=1}^{z_i}\left[\bar{P}_B(s_{h-1}, s_{h-2})\bar{P}_G(s_{h-1})\right]\cdots\right]\right]\right]$$

The expected delay t_i is obtained summing up the above expressions over all h's:

$$t_i = \sum_{h=0}^{\infty} \left(\frac{Z_k}{2} + hZ_k \right) \pi_h$$

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 resulting series at the m-th term, for a given constant value m, that is considering $0 \leq h \leq m$. Indeed, experimental tests show that the series converges already for small values of m, as it will be checked in Subsection 3.3. Recalling that the items are indexed by non-increasing $\frac{p_i}{z_i}$ ratios, the recurrences of Dichotomic and Dlinear algorithms can be applied once each $C_{i,j}$ is computed as $\sum_{h=i}^{j} t_h^m p_h$. The time for computing $C_{i,j}$ is derived as follows. Assuming a proper prefix-sum has been done as a preprocessing, $Z_k = \sum_{h=i}^{j} z_h$ can be retrieved in $O(1)$ time, while the computation of t_h^m requires $O(z_h^m)$ time. Therefore, in the worst case, the computation of $C_{i,j}$ takes $O(Nz^m)$ time, and that of all the $C_{i,j}$'s costs $O(N^3 z^m)$ time, which is pseudo-polynomial. Hence, the time for computing the $\hat{P}_B(s)$'s, $\bar{P}_B(s,\sigma)$'s, and $C_{i,j}$'s leads to a pseudo-polynomial time complexity for both the Dichotomic and Dlinear algorithms.

As in the unit length case, if the steady state probabilities of being in the bad state are not identical for all channels, then Dichotomic and Dlinear can be run after the channels are indexed so that $P_{B_1} \leq \cdots \leq P_{B_K}$ and each $C_{i,j}$ is replaced with $C_{i,j;k} = \sum_{h=i}^{j} t_h^m(k) p_h$, where $t_h^m(k)$ is computed as t_h^m by substituting P_{B_k} for P_B. Clearly, the computation of all the $C_{i,j;k}$'s takes $O(N^3 K z^m)$ time.

3.3 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 unit length, the item popularities follow a Zipf distribution with $\theta = 0.8$, while $N = 2500$ and $10 \leq K \leq 500$. 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 5, for

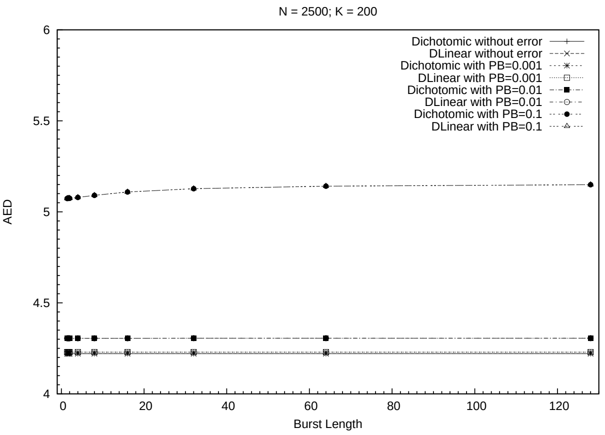

Figure 5: *The AED behavior versus the mean error burst length.*

$1 < \frac{1}{g} \leq 130$. 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.

Figure 6 exhibits the AED obtained in the case where the data lengths are unit and the steady state probabilities are not identical for all channels. As in the Bernoulli error model, the channels are indexed in such a way that $P_{B_1} = \cdots = P_{B_{\lfloor \frac{K}{3} \rfloor}} = P_B$, $P_{B_{\lfloor \frac{K}{3} \rfloor+1}} = \cdots = P_{B_{\lfloor \frac{2}{3}K \rfloor}} = 2P_B$, and $P_{B_{\lfloor \frac{2}{3}K \rfloor+1}} = \cdots = P_{B_K} = 3P_B$. One can observe that, when $P_B = 0.001$ and 0.01, the reported AEDs almost coincide with those where the channels are error-free, whereas the AED worsens when $P_B = 0.1$. Noting that in this latter case the steady state probability is 0.2 on the average, and thus $1 + \frac{2P_B}{1-r(N_k)} \simeq 1.40$, one expects that the AED in the presence of errors should be about 40% larger than that in the absence of errors. This is confirmed by the results reported in Figure 6, where the experimental AED is about 44% larger than in the error-free case.

Consider now data items whose lengths are non-unit. Since the algorithms take pseudo-polynomial time, a restricted set of experiments is performed. 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

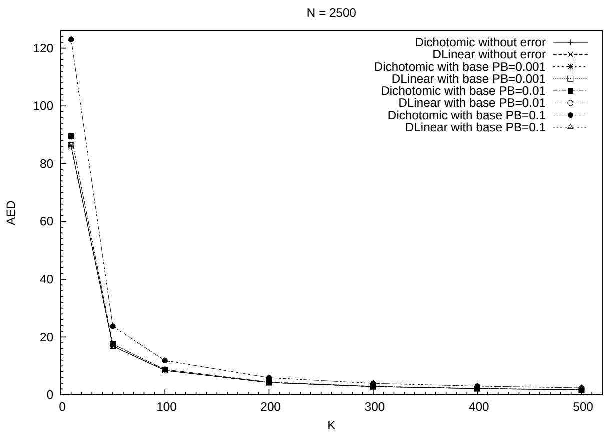

Figure 6: *Results for unit lengths when the channels are partitioned into three groups of the same size with steady state probability* P_B*,* $2P_B$*, and* $3P_B$*, respectively.*

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

(a)

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

(b)

Table 2: *Values of* t_i^m *when: (a)* $z_i = 10$*,* $Z_k = 50$*,* $\frac{1}{g} = 10$*, and* $P_B = 0.01$*; and (b)* $z_i = 5$*,* $Z_k = 50$*,* $\frac{1}{g} = 10$*, and* $P_B = 0.16$*.*

in the range $1 \leq z_i \leq 10$, for $1 \leq i \leq N$. All the K channels have the same steady state probability P_B, which assumes the values 0.001, 0.01, and 0.1. The reported results are averaged over 3 independent experiments. The expected delay of item d_i is evaluated by computing t_i^5, that is truncating at the fifth term the series giving t_i. Indeed, as shown in Table 2 for $z_i = 10$, $Z_k = 50$, $\frac{1}{g} = 10$, and $P_B = 0.01$ and for $z_i = 5$, $Z_k = 50$, $\frac{1}{g} = 10$, and $P_B = 0.1$, 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-unit, lower bounds for non-unit length instances are derived by transforming them into unit length instances, as explained in Subsection 2.3. Moreover, since the steady state probability P_B is the same for all channels, the AEDs giving the lower bounds are obtained by running the DP algorithm as explained in Subsection 3.1.

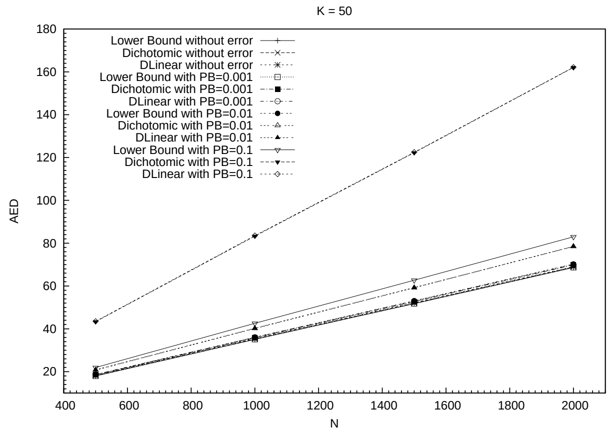

Figure 7: *Results for non-unit lengths when all the channels have the same steady state probability* P_B*, which assumes the values* 0.001*,* 0.01*, and* 0.1*.*

Figure 7 shows the experimental results for non-unit lengths, where P_B assumes the values 0.001, 0.01 and 0.1. In this figure, lower bounds are shown for both error-free and error-prone channels. One notes that, for every value of P_B, the behavior 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 than the lower bound in the 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 (e.g. see [12]).

4 Conclusions

This paper studied the problem of allocating N data to K channels, assuming 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 behavior of two dynamic programming algorithms previously presented for error-free channels has been experimentally tested, modelling the

Channel error model	Channel error probabilities	Unit lengths		Non-unit lengths	
		$K = 2$	$K > 2$	$K = 2$	$K > 2$
Bernoulli	$q_1 = \cdots = q_K$	$O(N \log N)$	$O(NK \log N)$	$O(NZ)$	Strong NP-hard
	$q_1 < \cdots < q_K$	$O(N \log N)$	open	NP-hard	Strong NP-hard
Gilbert-Elliot	$P_{B_1} = \cdots = P_{B_K}$	$O(N \log N)$	$O(N^2K)$	NP-hard	Strong NP-hard
	$P_{B_1} < \cdots < P_{B_K}$	$O(N \log N)$	open	NP-hard	Strong NP-hard

Table 3: *Complexity results for finding optimal solutions of the Data Allocation problem of N items on K faulty channels (Z is the sum of all item lengths).*

channel error by means of the Bernoulli model as well as the Gilbert-Elliot one. Simulations showed that such algorithms provide good sub-optimal solutions when tested on benchmarks whose item popularities follow Zipf distributions. In particular, for small channel error probabilities, the average expected delay is almost the same as the optimal one found in the case of channels without errors. However, some subcases have been detected where an optimal solution can be found in polynomial or pseudo-polynomial time. All the complexity results, proved in the present paper, are summarized in Table 3 (however, experiments showed that near optimal solutions are found by the algorithms even when the problem complexity is open or intractable).

As a matter of further research, it might be interesting to investigate other metrics on the schedules in presence of errors, like the edit distance between the schedule computed and the error-free schedule, or more generally to study the stability of the solutions over a range of error values. Another model to explore would let the client retrieve the packets of the items that do not have errors and let other packets be retrieved in the subsequent transmissions. Finally, an interesting open question is that of determining whether a closed formula for computing the item expected delays exists or not when the lengths are non-unit and the Gilbert-Elliot model is adopted.

Acknowledgement

This work has been supported by ISTI-CNR under the BREW research grant. The C++ code used in the simulations was written by G. Spagnardi.

References

[1] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik. Broadcast disks: data management for asymmetric communication environments. In *Proc. SIGMOD*, pp. 199–210, 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*, 14(2):219–231, 2008.

[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)*, pp. 11–20, 1998.

[7] A.A. Bertossi, M.C. Pinotti, and R. Rizzi. Scheduling data broadcasts on wireless channels: Exact solutions and heuristics. Chapter 73 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*, pp. 126–134, 1999.

[9] T. Imielinski, S. Viswanathan, and B.R. Badrinath. Energy efficient indexing on air. In *Proc. SIGMOD*, pp. 25–36, 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)*, pp. 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)*, pp. 659–666, 2000.

[12] P. Koutsakis, Scheduling and call admission control for burst-error wireless channels. In *Proc. of the 10th IEEE Symposium on Computers and Communications (ISCC)*, pp. 767–772, 2005.

[13] 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)*, pp. 293–302, 2000.

[14] S. Martello and P. Toth. *Knapsack Problems*. Wiley, Chichester, 1990.

[15] 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.

[16] 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)*, pp. 167–176, 2000.

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

[18] W. Turin. *Performance Analysis of Digital Transmission Systems*. Computer Science Press, New York, 1990.

[19] 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)*, pp. 90–99, 1997.

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

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

[22] 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.

[23] 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.