# TECHNICAL RESEARCH REPORT

Push-Based Information Delivery in Two Stage Satellite Systems-Terrestrial Wireless

*by Ozgur Ercetin, Leandros Tassiulas*

**CSHCN T.R. 2000-2**
**(ISR T.R. 2000-4)**


*The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.*

**Web site http://www.isr.umd.edu/CSHCN/**

# Push-based Information Delivery in Two Stage Satellite-Terrestrial Wireless Systems

**Özgür Erçetin and Leandros Tassiulas**
Institute for Systems Research, and Electrical Engineering Department
University of Maryland, College Park, MD 20742.
Tel: (301) 405-6620, Fax: (301) 314-9920
*ercetin@isr.umd.edu, leandros@isr.umd.edu*

May 16, 1999

**ABSTRACT**

*Satellite broadcast data delivery has inherent advantages in providing global access to information to everyone. However, users of the satellite communications need expensive and cumbersome equipment to receive and transmit satellite signals. Furthermore, as the amount of information being broadcast increases, average user latency increases as well. In many situations, users in a locality may have similar interests and hence they can be better served by a local broadcast schedule. A two stage satellite-terrestrial wireless broadcast system can provide more efficient service. In such a system, main server broadcasts information via satellite to the geographically distributed local ground stations. Every station has limited buffer capacity to store the items broadcast by the satellite. According to their cache content, and the interests of their users, local stations deliver the information to their users via terrestrial wireless channel. We develop novel methods for the joint cache management and scheduling problem encountered in these systems. Our results demonstrate that two stage systems can provide more efficient data delivery compared to the single stage systems.*

# 1 Introduction

Broadcast data delivery is an effective way to disseminate information to massive user population. The two foremost advantages of broadcast data delivery are scalability to large user populations, and availability to geographically distributed mobile users. Applications that employ broadcast data delivery include Hughes' DirecPC broadcast data delivery system [13], PointCast 's webcasting [12], Marimba's Castanet [14], and Intel's Intercast[15]. The types of information that is broadcast include news and weather information, traffic information, schedule information in airports, and train stations, stock quotes, and so on.

Under the broadcasting approach as depicted in Figure 1, a server continuously, and repetitively broadcasts data to a user community without any feedback about the user's needs due to the limited uplink communication capability from the user to the server. The data is delivered in terms of *items*, where an item is a fixed sized information unit. The items are broadcast in a particular order that is called a *schedule*. Time is slotted, and the length of each slot is the transmission time of the item on the broadcast channel. When a user needs a certain item, it monitors the broadcast channel until the desired item is detected, and captures it for use. There is some latency for the fulfillment of the needs of a user, since the server can only broadcast a single item per broadcast channel at any time instant. This latency depends on the broadcast schedule, as well as the user access pattern.

The recent attention to the broadcast data delivery has been motivated by the inherent communication asymmetry observed in the data communication networks. The communication asymmetry may arise either as physical asymmetry, or as application asymmetry. The physical asymmetry is observed, when the bandwidth or power requirements for the uplink and the downlink channels are not the same. The application asymmetry is said to be present, when the flow of the information is much higher in one direction (e.g. from the

server to the clients).

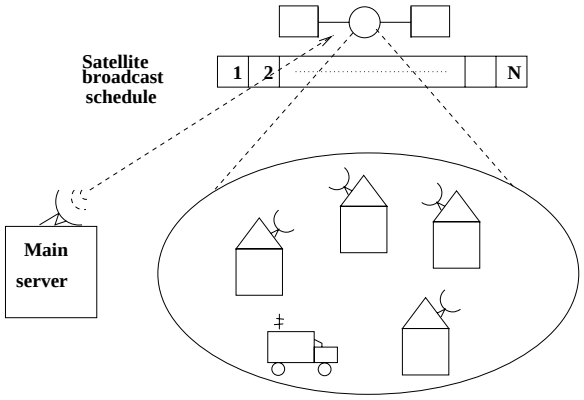


Figure 1: *A Single-stage Broadcast Data Delivery System in a Wireless Communication Environment*

One of the main aspects of the proposed National/Global Information Infrastructure is to provide global access to information to everyone. Satellite communications have inherent advantages for this purpose, especially for providing services in the remote areas where the communications infrastructure is inadequate. However, due to the system implications, users of the satellite communications systems need expensive and cumbersome equipment to receive and transmit satellite signals. In order for such a system to reach the common people, the user equipment must be affordable and portable.

At the same time, for the information to be useful, it should be delivered in a reasonable amount of time. One disadvantage of satellite broadcast data delivery, is as the amount of information being broadcast increases the average user latency increases as well. In many situations, the users in a locality may have specific interests, that differ from the overall user interests. In such cases, a local broadcast schedule, which matches the users interests will result in better overall average user latency.

In this study, we propose a two- stage broadcast data delivery system, which addresses both of these issues, and provides efficient information delivery to the users with smaller and

cheaper equipment.

In our two-stage broadcast data delivery system, the main server broadcasts information to the local ground stations via the satellite. The local ground stations act as intermediate stages, and transfer this information to the users in their coverage area via the terrestrial wireless channel. In many situations, the users in a locality may have similar interests. The satellite schedule, however; is designed according to the interests of the whole user population. The overall user interests may be quite different, compared to the interests of the individual groups. Thus, a user group can be better served, if the local ground station, which is servicing that user group, can create a schedule that is more suited for those users.

Broadcast data delivery has been investigated extensively. The focus of the work in this area can be mainly divided into two categories: Broadcast scheduling, and cache management. Broadcast schedule design appeared to be a complex problem, and many heuristics have been developed focusing on different aspects of the system. For example [2], [4], [7] developed heuristics that minimizes the average latency, while [8] investigated more energy efficient scheduling schemes.

The cache management received much attention recently, not only from the satellite communications community, but also from the Internet community. If a user has local storage, it can retrieve items from the broadcast and store them in the memory for future requests. By selectively prefetching items from the broadcast, the user is effectively able to minimize the mismatch between its access needs and the server's broadcast schedule and the average latency of the users is reduced. [9], [10], [11], has investigated this problem.

In our work, we investigate the design of the satellite schedule, the cache management at the ground stations, and the design of the terrestrial schedule, that arise in joint fashion in the context of our problem. Previous studies have treated these problems separately. In this paper, we introduce joint design of the cache management and the scheduling policies.

The paper is organized as follows: Section 2 describes the system model. In section 3, we discuss the joint cache management and broadcast scheduling techniques for the local stations. Specifically, we introduce an intuitive sub-optimal solution to this complex problem, and discuss the performance of this solution over the numerical examples. In section 5, we derive a lower bound for the mean response time. In section 6, we look into the satellite schedule design problem in two stage systems, and develop intuitive satellite schedule design schemes. We evaluate the resulting schemes with the numerical examples.

# 2 System Model

The following figure 2 depicts the two-stage satellite-terrestrial system. The main server is transmitting information items via satellite to geographically distributed local ground stations with transmission rate of $R_s$ items per unit time. Each ground station is receiving the information items broadcast by the satellite and has a local storage capacity of $C$ items. Each ground station retransmits the items stored in the local cache to a local user community with transmission rate $R_g$ items per unit time, where $R_s = r \cdot R_g$, and $r$ is a positive integer. This is a reasonable assumption, since typically, the capacity of the satellite broadcast channel is higher than the terrestrial broadcast channel.

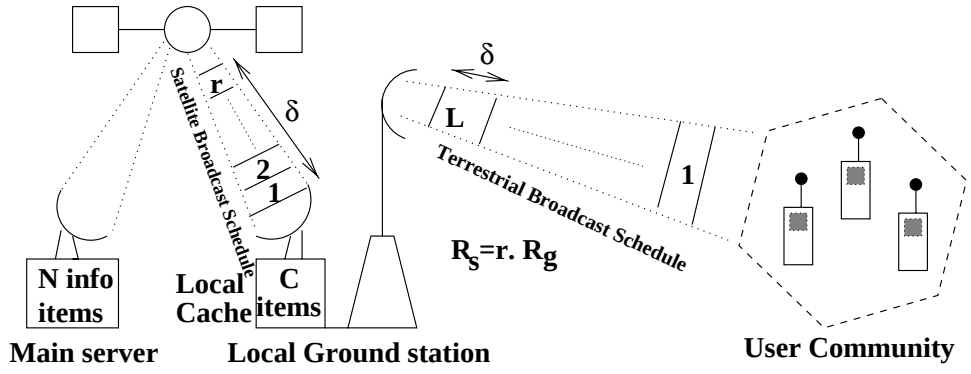


Figure 2: *Multistage Information Distribution*

Time is slotted, and the length of each satellite *(terrestrial)* slot is determined by the transmission time of the items on the satellite *(terrestrial)* broadcast channel. If the terrestrial slot has a length $\delta$, then the satellite slot length is $\delta/r$, and during a terrestrial time slot, $r$ items are broadcast by the satellite. Also notice that, the beginning of the terrestrial slot $n$ coincides with the beginning of the satellite slot $(n-1)\cdot r+1$.

Denote the satellite broadcast schedule by $\{s(n)\}_{n=0}^{\infty}$. Let $B(n)$ denote the set of items residing in the cache of an arbitrary local station at satellite slot $n$. Let $C$ be the size of the cache. The cache content is updated potentially at every satellite slot $n$, and

$$B(n+1) \in \{B' | B' \subset \{s(n)\} \cup B(n), |B'| = C\} \tag{1}$$

At every terrestrial slot $n$, an item that resides in the cache at that time, that is in $B((n-1)\cdot r+1)$ is broadcast on the terrestrial channel.

The following figure 3 summarizes the functionality of the local ground stations' cache management policy and both the main server's and the local servers' broadcast scheduling policy.

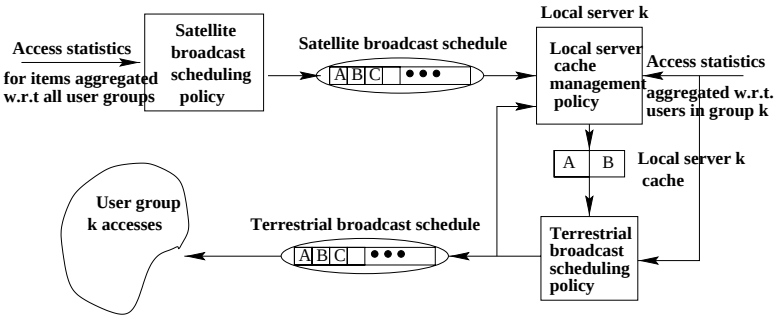


Figure 3: *Functionality of satellite scheduling, cache management and terrestrial scheduling in two-stage satellite-terrestrial broadcast system.*

The problem investigated in this paper is, the minimization of the overall mean response time by intelligent satellite and terrestrial broadcast schedule and local server cache management schemes. The terrestrial schedule should try to match the user access characteristics

as much as possible, given the satellite schedule, and the available items saved in the cache. The cache management policy should save the items broadcast by the satellite, that are going to be broadcast by the terrestrial schedule. The satellite schedule should try to deliver the items in the order that the cache management and terrestrial scheduling techniques can minimize the local average user latency.

# 3 Joint Cache Management and Terrestrial Broadcast Scheduling

In this section, we focus on the local server's cache management and terrestrial scheduling. For this reason, the satellite schedule is assumed to be known by the local ground stations in advance. The design of the satellite schedule is treated in Section 6.

## 3.1 A Motivating Example for Joint Cache Management and Scheduling

Even if we use naive cache management and terrestrial scheduling techniques, the resulting performance improvement with the two-stage systems is significant. Let's illustrate this with the following example. Assume that there are 4 groups of users that are interested in only 4 items. Each item is requested with the same probability. Neither of the items requested by the users of different groups are the same. Therefore, the main server broadcasts 16 items to the local ground stations of these 4 groups. Assume that, the satellite schedule is uniform, that is, these 16 items are broadcast sequentially, and periodically. Let each local ground station be able to store 2 items at any given time instant, and the satellite channel be twice as fast as the terrestrial channel. By an intuitive but simple cache management procedure, which stores any new item of interest received from the satellite schedule, by disposing the item that is the oldest in the cache, the cache content evolves as depicted in figure 4. By

equally simple terrestrial scheduling policy, which chooses an item from the cache that is broadcast least recently, the terrestrial schedule evolves as depicted in figure 4. It is easy to see that the resulting terrestrial schedule has a mean response time of 4.25 satellite slots. If we used single-stage broadcast scheduling policies, a single schedule would have to service all four user groups and the best schedule would then be the uniform satellite schedule that is depicted in figure 4. The mean response time of this schedule is 8 satellite slots. The improvement in the average latency by the multi-stage system is almost 50%.

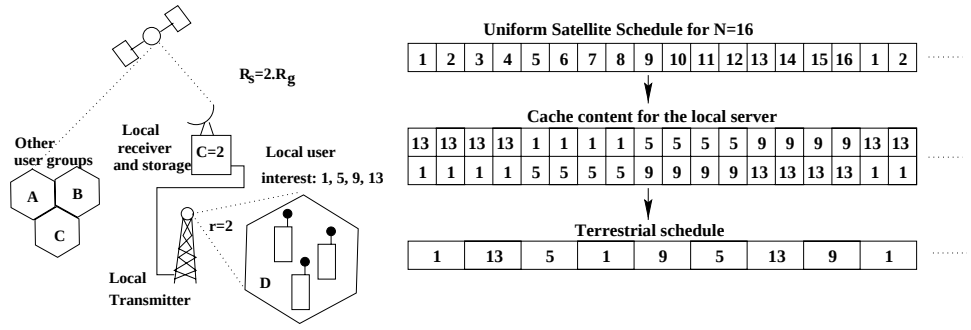


Figure 4: *The cache management, and terrestrial schedule at one of the four ground stations. The users of this ground station are equally interested in items 1, 5, 9, 13. The terrestrial channel is half as fast as the satellite channel.*

## 3.2 Cache Management at Local Stations

If we assume that all information items that are requested by the users of a locality are available at the local ground station, then the problem reduces to the determination of a terrestrial schedule which matches the user access characteristics as much as possible. This problem is studied extensively in the past in the context of single-stage satellite broadcast systems.

In any broadcast scheduling system, there are two quantities related to each item $i$, that affect the scheduling decision at each slot $n$. The elapsed time $w_i(n)$ since the last

transmission of item $i$ and the rate $\lambda_i$ of request generation for item $i$. The likelihood of item $i$ being transmitted at $n$ increases with $\lambda_i$ and $w_i(n)$. In [2], Su and Tassiulas considered a set of policies for a single-stage system, where the broadcast scheduling is determined based on priority indices of the items. The index of item $i$ is the product $\lambda_i^{\gamma} w_i(n)$, where $\gamma$ is an exponent that reflects relative importance of $\lambda_i$ versus $w_i(n)$. It is shown that $\gamma = 0.5$ gives the best performance among these priority index policies. $\gamma = 0.5$ refers to a special case, where the index is $\sqrt{\lambda_i w_i^2(n)}$. Note that $\frac{1}{2}\lambda_i w_i^2(n)$ is the *aggregate expected delay* experienced by item $i$ requests since the last transmission of item $i$ before slot $n$. Su et al., showed that this, so called *Mean Aggregate Delay (MAD)*, algorithm gives performances close to the lower bound.

The single stage and the proposed two-stage systems differ in the availability of items for scheduling at a given slot. In two-stage system, only the items that are saved in the cache, are available for scheduling. Intuitively, we want the terrestrial schedule in the two stage system to be same as the one determined in the single-stage system. For this reason, we use MAD scheduling policy as the terrestrial scheduling policy and develop a good cache management policy to achieve this goal.

Therefore, with this approach, our objective is to determine a cache management policy that feeds the scheduler at every slot with the items, that are going to be broadcast in the following slots according to the MAD policy. We now describe two cache management algorithms that comply to this intuition.

The cache management algorithms, that we propose, determine the mean aggregate delay of each item at every slot, similar to the MAD algorithm. The algorithms determine a cost that is going to be induced to the system, if an item that is in the cache or being broadcast by the satellite schedule at the current slot, is not fetched. This cost, depends mainly on the fact that the MAD scheduler will be unable to broadcast an item, that is not fetched, at

least until that item is rebroadcast at the satellite schedule. Thus, by not fetching an item, we may be delaying the broadcast of that item. The objective of the cache management algorithms, is to minimize this delay.

For the design of the cache management algorithms, we need to define a few new parameters. Let $l_i(n)$ denote the number of satellite slots from the end of satellite slot $n$, until the end of the next satellite slot after $n$ at which page $i$ is broadcast on the satellite channel. Also, let $w_i(n)$ denote the elapsed number of satellite slots since the last terrestrial transmission of item $i$ before the current satellite slot $n$. These parameters are illustrated in Figure 5. $l_i(\cdot)$, and $w_i(\cdot)$ can be interpreted as the expected latency and the waiting time of item $i$ at the satellite and terrestrial schedules, respectively.

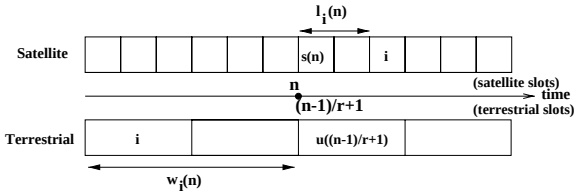


Figure 5: *Illustration of $l_i(.)$, and $w_i(.)$, for both satellite and terrestrial broadcasts of item $i$. In this example $r = 3$.*

Assume that the user access characteristics can be represented by the user item request probability distribution, $\{\beta_i^k\}_{i=1}^N$, where $N$ is the total number of items broadcast by the satellite. That is, at any time instant a user that is serviced by the $k$ th local station would like to access an item $i$ with probability $\beta_i^k$. If the user population of each group $k = 1, \ldots, M$, is large enough, then we may assume that the request generation process of each group is stationary with constant rate $\lambda^k$ requests per slot. Hence, requests for item $i$ from each group $k$ are generated according to a stationary process with rate $\lambda_i^k = \lambda^k \beta_i^k$. We will drop the superscript $k$ in the definition of the user access rate, when we describe the results corresponding to a single arbitrary local user group.

Due to the lack of uplink request channel, the only information about the user accesses, that the local stations can determine, is the expected number of users waiting for service at a given slot. The expected user backlog for item $i$ observed by an arbitrary local station at satellite slot $n$, is defined as $\overline{X_i}(n) = \lambda_i w_i(n)$. The evolution of the expected user backlog for item $i$ at a local station is depicted in Figure 6. The aggregate average delay of item $i$ at slot $k$ is defined as $\frac{1}{2}\lambda_i(w_i(k))^2$, and is the lined area between $t_0$ and $k$ in Figure 6.

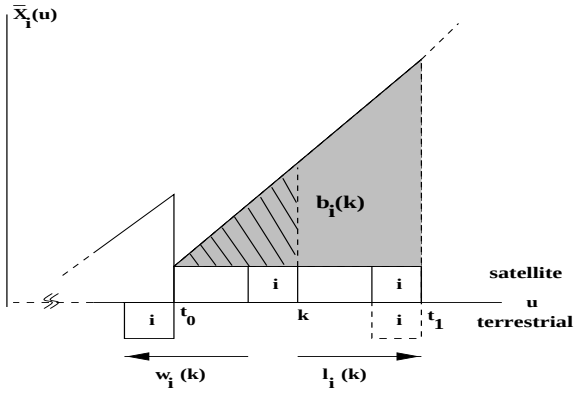


Figure 6: *The evolution of the expected user backlog of item $i$ observed by a local station. The upper side of the u-axis refers to the satellite transmissions, while the lower side refers to the terrestrial transmissions.*

At terrestrial slot $n$, $u(n)$ is broadcast on the terrestrial channel, that is

$$u(n) = \arg \max_{i \in B((n-1)\cdot r+1)} \lambda_i^{0.5} w_i((n-1)\cdot r+1) \tag{2}$$

It is also assumed that, the transmission of the scheduled item $u(n)$ takes place prior to the cache management at every terrestrial slot.

### 3.2.1 Total Waiting Time (TWT) Cache Management Algorithm

Assume that, at satellite slot $k$, item $i$ is broadcast by the satellite. The local station has two options; to fetch item $i$ and replace an item in the cache, or to keep the cache content

unchanged. If item $i$ is not fetched at $k$, the earliest time it can be scheduled at the terrestrial schedule is at satellite slot $k + l_i(k)$, i.e. when item $i$ is retransmitted by the satellite. At that time, the expected item $i$ backlog in terms of satellite slots will be given by,

$$b_i(k) = \overline{X_i}(k + l_i(k)) = \frac{1}{2}\lambda_i[w_i(k) + l_i(k)]^2 \tag{3}$$

Observe that $b_i(k)$ refers to the shaded area in Figure 6.

Since, we would like to minimize the mean aggregate delay, which depends on the average backlog evolutions of the items, we do not want to delay the broadcast of an item at the terrestrial schedule for too long. Thus, TWT algorithm fetches an item, if that item will have a very high total waiting time at its next satellite retransmission.

The TWT algorithm can be stated more concisely as follows.

> **TWT Algorithm:** *At every satellite slot $n$, calculate the total minimum average delay $b_i(n)$ for every item in the cache, i.e. $i \in B(n)$, and the item currently transmitted by the satellite, i.e. $s(n)$. If $b_{s(n)}(n) > \min_{j \in B(n)} b_j(n)$, then replace $\arg\min_{j \in B(n)} b_j(n)$ with $s(n)$, otherwise the memory state remains unchanged.*

This algorithm has low time complexity. However, it does not consider the future desired terrestrial schedule, so it does not allow the scheduler to match this desire as much as possible. In fact, there may be a situation, where an item $i$ may have a large total minimum expected delay, $b_i(k)$, at the current slot $k$, but it is not optimal to have a terrestrial transmission until a short time before its next satellite retransmission. In such a situation, discarding item $i$ at slot $k$, and keeping another item with sooner optimal terrestrial transmission may be a better solution. Our second cache management algorithm relies on this intuition.

### 3.2.2 Estimated Terrestrial Transmissions (ETT) Cache Management Algorithm

The improvement of this algorithm over TWT is, it estimates the future evolution of the terrestrial schedule, and according to this estimate, it determines the cache content that leads to the lowest possible mean aggregate delay. According to the estimated terrestrial schedule, the ETT algorithm determines at satellite slot $k$, whether an item $i \in B(k)$, is broadcast before its next satellite transmission or not. If it is not broadcast, then there is no need to keep item $i$ in the cache. If it is broadcast, the ETT algorithm determines the reward of keeping item $i$ in the cache at $k$. Let $t_1$ and $t_2$ correspond to the satellite slot, the estimated terrestrial transmission of item $i$, and the satellite slot, the next satellite transmission of item $i$ after $k$, takes place respectively (see Figure 7). From the previous arguments, if item $i$ is not stored at $k$ the earliest time it can be scheduled at the terrestrial schedule, is at $t_2$. Since item $i$ is not supposed to be broadcast until $t_1$ anyway (and we have determined $t_1$ by estimating the future terrestrial schedule), the reward of keeping item $i$ in the cache is only the additional waiting time from $t_1$ to $t_2$, that is eliminated by the broadcast of item $i$ at $t_1$. This reward is illustrated as $d_i(k)$ in Figure 7. Therefore, ETT algorithm stores an item $i$ at slot $k$, if the reward for keeping item $i$, $d_i(k)$, is sufficiently large.

The ETT algorithm addresses the weakness of the TWT algorithm, by estimating the future evolution of the terrestrial schedule. This is not a difficult task, since it is known that a version of the MAD policy is used as the scheduler. One way to create a simple estimate of the terrestrial schedule, is to use MAD as the scheduler and assume that all items are available for scheduling. However, such an assumption is not valid for our system, and will not provide a good estimate of the terrestrial schedule. In this algorithm we use a better estimator, which may schedule only the items in the cache at the current slot or the items

broadcast by the satellite between the current slot and the scheduling time. Let's illustrate this procedure with an example. Assume that $r = 1$, and let $k$ in Figure 7 denote the current slot we are in. We would like to estimate the terrestrial schedule from this slot onwards. The items that $MAD$ can schedule at $k$ are the items that are in the cache at that slot, i.e. $O(0) = B(k)$. At the same slot, we receive another item $s(k)$ from the satellite, which may or may not be fetched. Since we are not sure about the caching decision, at the next slot we assume that MAD chooses $u(k+1)$ from a larger set that is, $O(1) = B(k) \cup \{s(k)\}$. At $k+1$, we may also fetch $s(k+1)$, then MAD should choose the item to broadcast at slot $k+2$ from the set $O(2) = B(k) \cup \{s(k), s(k+1)\}$, and so on.

Since, we are trying to decide whether to fetch item $s(k)$, or not, the only part of the terrestrial schedule, we need to estimate is the part, where we have at least one terrestrial broadcast for all the items in $O(1)$. Additionally, for the purpose of the algorithm, we are only interested in the first terrestrial broadcast of the items. Thus, abovementioned estimation process ends, when for every item in $O(1)$, the first terrestrial transmission of that item, after $k$, has been determined.

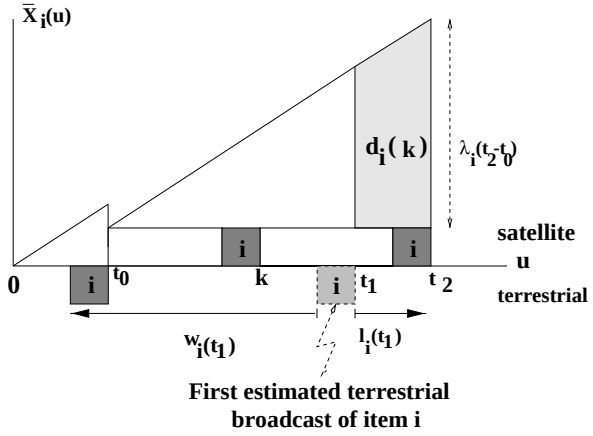


Figure 7: *Illustration for ETT algorithm. The upper side of the u-axis refers to the satellite transmissions, while the lower side refers to the terrestrial transmissions.* $t_1$ *denotes the first estimated terrestrial transmission slot of item* $i$ *after* $k$.

The algorithm in detail (for the general case of different satellite and terrestrial channel speeds) is given as follows:

**ETT Algorithm:**

- *do for all items, $i$, that are in the cache or in transmission at satellite schedule at the current satellite slot $k$, i.e. $i \in B(k) \cup \{s(k)\}$.*

  1. $\alpha = 0$, $next \neq i$.
  2. $O(0) = B(k)$.
  3. *do for all items broadcast, i.e.* $j = 1, \ldots, N$, $a_j(k) = w_j(k)$.
  4. *If* $(k + \alpha - 1) mod\ r = 0$
     - *Calculate* $c_j(k + \alpha) = \lambda_j^{0.5} a_j(k + \alpha)$, *for* $j \in O(\alpha)$.
     - $next = \arg\max_{j \in O(\alpha)} c_j(k + \alpha)$
     - $a_{next}(k + \alpha) = 0$
  5. $\alpha + +$
  6. $a_j(k + \alpha) = a_j(k + \alpha - 1) + +$, *for* $j = 1, \ldots, N$.
  7. *Determine the set of the observed items,* $O(\alpha) = B(k) \cup \{s(k), \ldots, s(k + \alpha - 1)\}$, *i.e. the items that are in the cache at $k$ and the items broadcast at the satellite schedule between $k$ and $k + \alpha - 1$.*
  8. *if* $next \neq i$ *repeat from 4.*
  9. $t_1[i] = k + \alpha - 1$.

- *Calculate the total additional delay $d_i(k)$ for each page from*

$$d_i(k) = \frac{\lambda_i}{2} l_i(t_1[i]) \{2w_i(t_1[i]) + l_i(t_1[i])\} \tag{4}$$

*If* $k + l_i(k) < t_1[i]$, *then* $d_i(k) = 0$.

- *If* $d_{s(k)}(k) > \min_{j \in B(k)} d_j(k)$, *then replace* $\arg\min_{j \in B(k)} d_j(k)$ *by* $s(k)$, *otherwise, cache content remains unchanged.*

# 4 Performance Comparisons and the Impact of the Base Station Buffer

In this section, we present the performance results of the proposed algorithms. The satellite-terrestrial broadcast system that we have considered, involves many important parameters. In our numerical studies we tried to address many of those parameters. Specifically we investigated the performances of the algorithms, with respect to different satellite schedules, user access probabilities, satellite-terrestrial broadcast speeds, and more importantly cache sizes.

In our simulations we modeled the user access probability distribution by Zipf 1 distribution [5]. In order to provide some generality, we modify Zipf 1 distribution according to two parameters: $c$, the center and $l$, the length of the distribution. The center $c$, denotes the item, for which the users have the most interest, i.e. the starting point of the Zipf 1 distribution. The length $l$, denotes the total number of items that are accessed by the users. For the sake of briefness, let $z1(c, l, N)$ denote the Zipf 1 distribution centered at item $c$ with a length $l$. That is,

$$z1(c,l,N)[i] = \begin{cases} \frac{1/(N-c+i)}{\sum_{j=1}^{l} 1/j} & , \; i \leq \max\{0, c+l-N\} \\ 0 & , \; \max\{0, c+l-N\} < i < c \\ \frac{1/(i-c+1)}{\sum_{j=1}^{l} 1/j} & , \; \min\{c+l, N\} \geq i \geq c \end{cases} \tag{5}$$

As before, $r$ denotes the ratio of the satellite and terrestrial broadcast channel speeds. In this study, we evaluate the performance of TWT and ETT with respect to the differ-

ence between the mean response times *(MRT)* of the terrestrial schedules created by these algorithms, and the lower bound. The resulting mean response times of the systems with different terrestrial broadcast channel speeds, are measured in terms of the satellite slots. Recall that, MAD policy is used as terrestrial scheduling policy. If the cache size is as large as the number of items requested by the users, then the MAD schedule will be created on the terrestrial channel. Thus, the performance of TWT and ETT approaches to the MRT of the MAD schedule as the cache size increases.

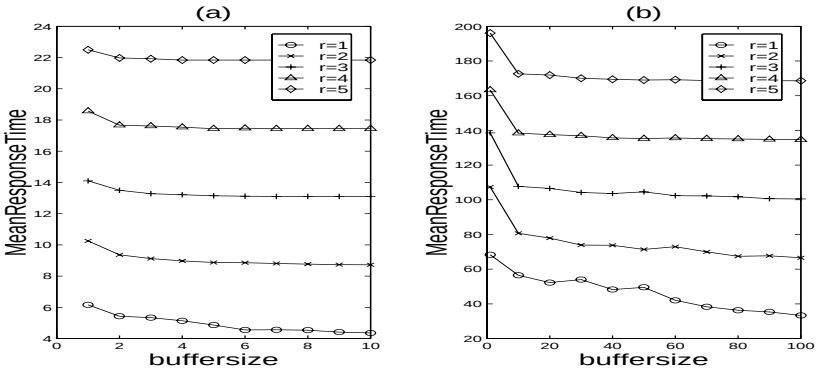


Figure 8: *The performance characteristics of TWT. In both parts the satellite schedule is created by the MAD algorithm over (a)* $z1(5, 10, 10)$*, and (b)* $z1(50, 100, 100)$*. The user access probability distributions are also Zipf 1, (a)* $z1(1, 10, 10)$*, and (b)* $z1(1, 100, 100)$*.*

Figure 8 depicts the performance of TWT with respect to different terrestrial broadcast channel speeds and local server cache sizes. In this example, we observed the case where the satellite schedule and the user accesses do not match. There is always the possibility that the satellite schedule is created matching the access characteristics of a group of users, and thus the MRT will be minimized without any need to any local server. However, while minimizing the MRT of this lucky user group, we may be increasing the MRT of the other user groups listening to the same satellite channel. Therefore, we need to find a satellite schedule that is more or less suitable for all user groups, at least in the sense that overall

MRT is lowered. Such a satellite schedule is explored in section 6.

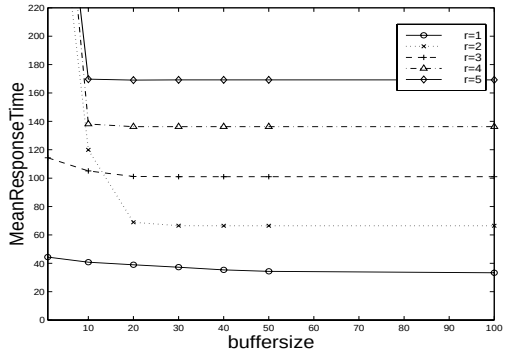


Figure 9: *The performance characteristics of the* ETT *algorithm. The satellite schedule is uniform schedule. The total number of pages broadcast by the satellite is 100. The user access probability distribution is* $z1(1, 100, 100)$.

Figure 9 displays the characteristics of the ETT algorithm for different satellite and terrestrial channel speeds, and for different cache sizes. In this example, the satellite schedule is an uniform schedule, that is, each item is broadcast exactly once in a period. The user access probability distribution is $z1(1, 100, 100)$.

It is also interesting to note that, the system with a slower terrestrial channel, approaches the near-optimal terrestrial schedule faster than a system with a faster terrestrial channel. Notice that, as the terrestrial channel gets slower, the length of the each terrestrial slot (in terms of the satellite slots) gets larger. Thus, at every terrestrial slot, the number of items broadcast by the satellite is higher. This situation helps our cache management algorithms, to better design the cache content according to the desires of the users, and to feed the scheduler with the optimal item to broadcast at every terrestrial slot. Hence, even with small caches, the proposed algorithms, can create terrestrial schedules, which are near-optimal for that terrestrial channel speed.

In Figure 10, we compared the performances of the $TWT$ and the $ETT$ algorithms for the

worst case, i. e. for the case where the satellite and the terrestrial broadcast channel speeds are the same. This case is considered to be the worst, since cache management algorithms should feed the scheduler with correct items to broadcast at every satellite slot, rather than every $r$, $r > 1$ satellite slots, which is the case with a slower terrestrial channel. In Figure 10, the satellite schedule is created by the MAD algorithm for the distribution z1(50,100,100). The user access probability distribution for this case is uniform distribution.

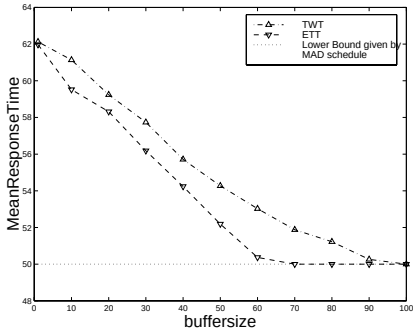


Figure 10: *The comparison of performance of TWT and ETT. The satellite schedule is determined by the MAD algorithm for the distribution z1(50,100,100), while the user access probability distribution is uniform.*

From our numerical studies, we have also noticed that both TWT and ETT can create optimal terrestrial schedule for a given user access probability distribution, regardless of what the satellite schedule, and the initial cache content are, provided that the local station's cache size is $\min\{\frac{T_0}{r}, N\}$, where $T_0$ is the period of the satellite schedule. We believe this is correct, for any satellite schedule, and the proof of this fact is among our future research goals.

# 5 A Lower Bound for Mean Response Time

## 5.1 Background and Assumptions for the Lower Bound

A broadcast schedule can be completely and uniquely defined by the inter-transmission gaps of the items that are broadcast. Let $T_p^i$ denote the inter-transmission gap between $p-1$ th and $p$ th item $i$ transmissions. Assume that initially, $t = 0$, item backlogs are empty. (This assumption becomes irrelevant as we observe the long term behavior of the system.)

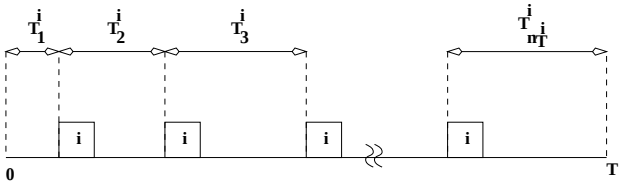


Figure 11: *Illustration of $T_p^i$ on a broadcast schedule of length $T$.*

Let $n_T^i$, $s_T^i$ be the number of times item $i$ is scheduled at the terrestrial and the satellite channels respectively, during an interval of length $T$. We relax the requirement of slotted system behavior, by assuming that the transmissions do not need to occur at integer multiples of time. Obviously, the optimal mean response time of a continuous time system is less than the optimal mean response time of its slotted time counterpart.

The mean response time of item $i$ in a schedule of length T, is then given by [4]:

$$\frac{1}{2T}\sum_{p=1}^{n_T^i}(T_p^i)^2 \tag{6}$$

The optimization problem that we are looking into, is the minimization of the aggregate of the individual item mean response times as stated above, subject to some conditions on the cache management, and the satellite and the terrestrial broadcast schedules.

For the single-stage systems a lower bound for the mean response time has been derived

by Ammar [4]. This lower bound is given by

$$S \geq \frac{1}{2}\left(\sum_{i=1}^{N}\sqrt{\lambda_i}\right)^2 \tag{7}$$

where, $N$ refers to the total number of items broadcast at the broadcast channel.

This lower bound is also applicable to the two-stage systems. However, in most cases this lower bound will not be a tight one, since in the derivation of above equation, it is assumed that any item may be scheduled at any instant. In two-stage systems, only the items that are stored in the cache of the local station at a particular instant can be scheduled at that instant.

For this reason, in the following we derive a non-trivial lower bound which also considers the cache management at the local stations.

## 5.2 A Condition on the Cache Size

The minimum of the mean response time of an item as given in Eq. (6) is $\frac{T}{2(n_T^i+1)}$, and is achieved by a terrestrial schedule with equal inter-transmission gaps (for proof see [4]). In two-stage broadcast systems we can store an item in the cache for only a limited amount of time, so we may not be able to create this best schedule. Our aim in the design of two-stage broadcast system, is to create a terrestrial schedule which has inter-transmission gaps as uniform as possible for every user group (Figure 12).

Let $h_T^i$ be the maximum total duration of time, item $i$ can be stored in the cache of the local station during an interval of $T$. Following lemma, discusses the minimum of $h_T^i$ that is required to create the abovementioned best terrestrial schedule.

**Lemma 1** *$h_T^i = \frac{n_T^i - s_T^i}{n_T^i+1}T$ is the sufficient duration of time item $i$ should be kept in the cache, in order to create a terrestrial schedule with $n_T^i$ equidistant item $i$ transmissions, from a*

*satellite schedule with $s_T^i$ ($s_T^i \leq n_T^i$) item $i$ transmissions.*

**Proof**

The sufficient caching capacity that is needed to create the desired terrestrial schedule depends not only on $n_T^i$, but also on $s_T^i$. Item $i$ can be fetched at the local station, only when it is broadcast at the satellite schedule. Since there are $s_T^i$ satellite item $i$ transmissions, the total caching duration $h_T^i$ can be divided into as many as $s_T^i$ different caching intervals. The terrestrial item $i$ transmissions may take place only when item $i$ is stored in the cache, i.e. during the caching intervals.

From this understanding and the continuous time approximation of the system, one can easily see that it is best to divide the total caching duration into $s_T^i$ caching intervals, and to begin and end a caching interval with a terrestrial transmission (For example see Figure 12).

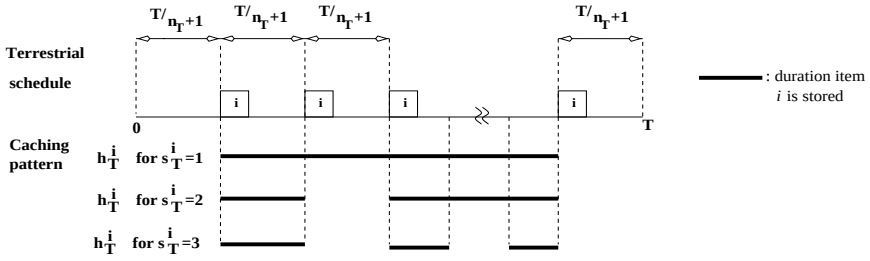


Figure 12: *Illustration of a terrestrial schedule with $n_T^i$ equidistant terrestrial transmissions. Also note an example of caching and satellite scheduling patterns for creating the desired terrestrial schedule.*

Consider the following greedy algorithm, which tries to minimize $h_T^i$ by arranging the satellite schedule and the caching intervals:

- *If $s_T^i > n_T^i$, then place $n_T^i$ satellite transmission at the same instant as $n_T^i$ terrestrial transmissions of the desired schedule. No caching is needed. $h_T^i = 0$*

- *Otherwise, place $s_T^i$ satellite transmissions at the same instant as $s_T^i$ arbitrarily chosen terrestrial transmissions in the desired terrestrial schedule. Each caching interval should cover time duration between a satellite transmission, and the last terrestrial transmission before the next satellite transmission.*

  - *Special case: After the $s_T^i$*th *satellite transmission, if there are still terrestrial transmissions left, then caching interval should cover all of those as well.*

It is easy to see that when we use above algorithm, it is sufficient to cache item $i$ for a period of $h_T^i = \frac{n_T^i - s_T^i}{n_T^i + 1} T$ if, $s_T^i \leq n_T^i$.

Thus, we found a satellite schedule, and cache management technique, that creates the desired terrestrial schedule, with a cache of size $h_T^i = \frac{n_T^i - s_T^i}{n_T^i + 1} T$, and so the suffiency condition is satisfied. □

**Lemma 2** *$h_T^i = \frac{n_T^i - s_T^i}{n_T^i + 1} T$ is also the necessary duration of time item $i$ should be kept in the cache, in order to create a terrestrial schedule with $n_T^i$ equidistant item $i$ transmissions, from a satellite schedule with $s_T^i$ ($s_T^i \leq n_T^i$) item $i$ transmissions.*

**Proof**

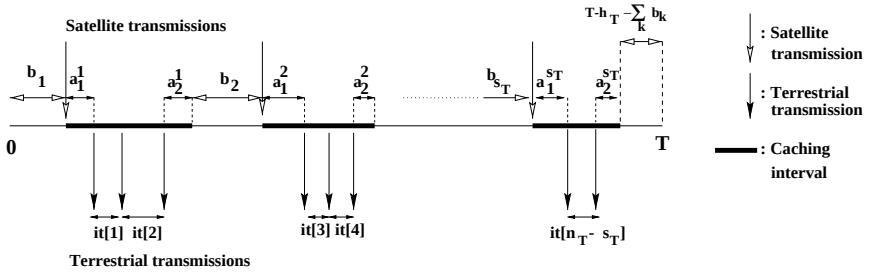


Figure 13: *Illustration of the parameters used in the proof.*

It is clear that, for a terrestrial schedule with equal inter-transmission gaps, item $i$ will have a mean response time of $\frac{1}{2T} \cdot \frac{T^2}{n_T^i + 1}$. Assume that the total caching duration is divided

into $s_T^i$ caching intervals, with the possibility that a caching interval may have a length of zero. We need a few new parameters to describe the necessity condition. Let $b_k$ be the duration between the end of the $k-1$ th caching interval and the beginning of the next caching interval. Let $a_1^k$ be the duration between $k$ th satellite transmission and the first terrestrial transmission in the $k$ th caching interval. Let $a_2^k$ be the duration between the last terrestrial transmission in the $k$ th caching interval, and the ending time of the $k$ th caching interval. Let $it[j]$ be the inter-transmission gap between two consecutive terrestrial transmissions that are both inside the same caching interval. Finally, let $B_k$ be the set of terrestrial transmissions occurring in the $k$ th caching interval. These parameters are illustrated in Figure 13.

Thus,

$$T_p^i = \begin{cases} b_1 + a_1^1 & p = 1 \\ it[l] & p \in B_k,\, p \neq \min\{j | j \in B_k\},\, l = f(p) \\ a_2^{k-1} + b_k + a_1^k & p \in B_k,\, p = \min\{j | j \in B_k\} \\ a_2^{s_T^i} + T - h_T^i - \sum_{k=1}^{s_T^i} b_k & p = n_T^i + 1, \text{ i.e. between } p = n_T^i \text{ and } T \end{cases} \tag{8}$$

where $f(\cdot)$ is a function that maps '$p$' into the correct index for $it[\cdot]$.

The best *possible* terrestrial and satellite schedule for an arbitrary $h_T^i$ can be determined by the following optimization problem $OP$.

$$OP: \min_{b_k, a_1^k, a_2^k, it[\cdot]} (b_1 + a_1^1)^2 + \sum_{k=2}^{s_T^i} (a_2^{k-1} + b_k + a_1^k)^2 + \sum_{l=1}^{n_T^i - s_T^i} it^2[l] + \left( T - h_T^i - \sum_{k=1}^{s_T^i} b_k + a_2^{s_T^i} \right)^2 \tag{9}$$

$$\text{s.t.} \qquad \sum_{k=1}^{s_T^i} a_1^k + a_2^k + \sum_{l=1}^{n_T^i - s_T^i} it[l] \leq h_T^i$$

The solution of $OP$ for the case $a_1^k \leq 0$ and $a_2^k \leq 0$ for $k = 1, 2, \ldots, s_T^i$, results in

$$it[l] = \frac{h_T^i}{n_T^i - s_T^i} \text{ , for } l = 1, 2, \ldots, n_T^i - s_T^i \tag{10}$$

$$b_k = \frac{T - h_T^i}{s_T^i + 1} \text{ , for } k = 1, 2, \ldots, s_T^i \tag{11}$$

$$a_1^k = 0 \text{ , } a_2^k = 0 \text{ , for } k = 1, 2, \ldots, s_T^i \tag{12}$$

However, the condition $a_1^k \leq 0$ and $a_2^k \leq 0$ for $k = 1, 2, \ldots, s_T^i$, is true if $h_T^i \leq \frac{n_T^i - s_T^i}{n_T^i + 1} T$ (See appendix for details). Thus, we can create the terrestrial schedule with equidistant transmissions if and only if $h_T^i \geq \frac{n_T^i - s_T^i}{n_T^i + 1} T$. □

If the caching condition given in the previous lemmas is not satisfied, the resulting best terrestrial schedule is going to be similar to the one given in Figure 14, and will have a mean response time given by the following equation (13). Notice that, this mean response time is higher than the mean response time of the terrestrial schedule with equidistant transmissions:

$$\frac{1}{2T}(s_T^i + 1)\frac{(T - h_T^i)^2}{(s_T^i + 1)^2} + \frac{1}{2T}(n_T^i - s_T^i)\frac{(h_T^i)^2}{(n_T^i - s_T^i)^2} \tag{13}$$

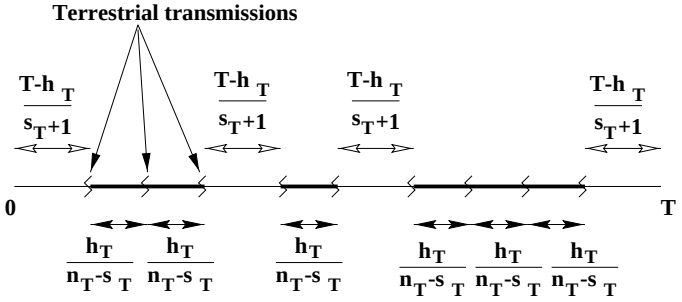


Figure 14: *Illustration of the optimal continuous time terrestrial schedule for* $h_T^i \leq \frac{n_T^i - s_T^i}{n_T^i + 1} T$.

## 5.3 Long-term Average Aggregate Response Time

Let $l_i = \lim_{T\to\infty} \frac{n_T^i}{T}$, $m_i = \lim_{T\to\infty} \frac{s_T^i}{T}$, and $h_i = \lim_{T\to\infty} \frac{h_T^i}{T}$. Assume that $l_i$, $m_i$, and $h_i$ exists. The capacity of the satellite channel is typically higher than the terrestrial channel. Let $r$ be a positive integer that corresponds to the ratio of the satellite channel rate and the terrestrial channel rate. Then, the mean response time of item $i$ in a particular system where the terrestrial transmission rate is $l_i$, the satellite transmission rate is $m_i$ and the portion of time item $i$ is cached in the memory is $h_i$, can be lower bounded by:

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{p=1}^{n_T^i} (T_p^i)^2 \geq \begin{cases} \frac{(1-h_i)^2}{2m_i} + \frac{h_i^2}{2(l_i/r - m_i)} & \text{, for } l_i/r > m_i, \text{ and } h_i < 1 - \frac{m_i}{l_i/r} \\ \frac{1}{2l_i/r} & \text{, otherwise} \end{cases} \tag{14}$$

Assume that the satellite schedule item transmission rates are $\{m_i\}_{i=1}^N$ and the users of a particular local station are interested in these items with probabilities $\{\lambda_i\}_{i=1}^N$. Also assume that the local station can store at most $C$ items at any particular instant. Then, observe the following optimization problem:

$$LB : \min_{\{h_i\}_i, \{l_i\}_i} \left\{ \sum_{i:h_i \geq 1 - \frac{m_i}{l_i/r}} \frac{\lambda_i}{2l_i} + \sum_{i:h_i < 1 - \frac{m_i}{l_i/r}, l_i/r > m_i} \frac{\lambda_i}{2} \left[ \frac{(1-h_i)^2}{m_i} + \frac{h_i^2}{l_i/r - m_i} \right] \right\} \tag{15}$$

$$\text{s.t.} \qquad (1) \quad \sum_{i=1}^N l_i \leq 1$$

$$(2) \quad \sum_{i=1}^N h_i \leq C$$

In the above optimization problem the cost function is the lower bound for the mean response time. The first constraint suggests that no two item is scheduled for transmission at the same instant in the terrestrial schedule, and the second constraint suggests that at any time instant there are no more than C items in the cache of the local server.

**Theorem 1** *The long term average aggregate delay of all the requests experienced by a user under any terrestrial scheduling and cache management policy for a given satellite schedule, is lower bounded by the solution of the above optimization problem* LB.

**Proof**

The minimum of the lower bound over all possible values of these parameters is a lower bound of any terrestrial broadcast scheduling and cache management policy working under the given satellite schedule. □

### 5.4 Demonstration of Lower Bound

In Figure 15, the satellite schedule is uniform, while the user access probability distribution is z1(1,100,100). In this figure, we compared the performances of the proposed algorithms, with the lower bound that we have derived in the section 5. The nonlinear optimization problem *LB* that is given by equation 15, is solved by using the Feasible Sequential Quadratic Programming (FSQP) software developed in part by Tits, Lawrence, and Zhou [6]. The cost function in our optimization problem, is a quasi-convex function. For this reason, the minima determined by FSQP is also a global minima for the corresponding cache size.

## 6 Satellite Schedule Design

In general, the users in different local stations have different interests, i. e. different access probability distributions. Therefore, a satellite schedule that is suitable for an user group may result in the worst possible performance for another user group. The joint design of the satellite schedule, the cache management strategies, and the terrestrial schedules, may provide a better overall performance. Until now, we have discussed the joint design of the cache management and the terrestrial scheduling techniques. In this section, we look into the satellite schedule design.

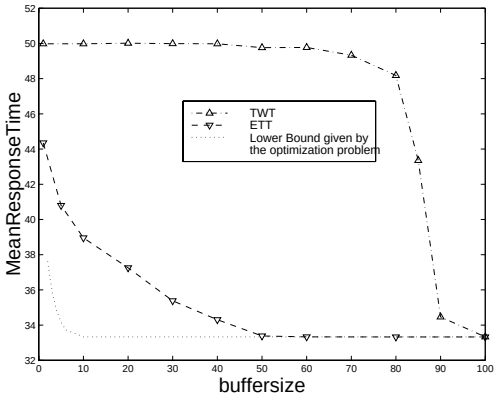


Figure 15: *The comparison of performance of TWT and ETT. The satellite schedule is uniform with 100 items to broadcast, and the user access probability distribution is* $z1(1, 100, 100)$.

We assume that, the proposed cache management algorithms, and the MAD scheduling policy are used at the local stations. In particular, we assume that the ETT algorithm is used as the cache management algorithm. Our cache management algorithms require that the satellite schedule is determined apriori, and is known by the local stations in advance. For this reason, we only consider the off-line design of the satellite schedule. As a heuristic, we consider the case, where the satellite schedule is determined by the MAD policy. With these assumptions, the only question that remains to be answered is, which access probability distribution should be used, to determine the satellite schedule that minimizes the overall average response time observed by all of the users. The system that we consider consists of multiple user groups where each user group has an arbitrary user request rate functions, and is serviced by a local station with an arbitrary cache size.

The intuitive answer to above question is to use an aggregate of the user distributions to determine the satellite schedule. That is,

$$p_s(i) = \frac{1}{\sum_{k=1}^{M} \lambda^k} \sum_{k=1}^{M} \lambda_i^k \ , \forall i = 1 \ldots N \tag{16}$$

where $p_s$ denotes the probability mass function used for the determination of the satellite schedule. Let $M$ be the total number of user groups in the satellite-terrestrial system.

It is clear that, the satellite probability distribution depends on the individual item access rates of the user groups. However, the user groups, that are serviced by a local station with a large cache, may observe a low mean response time, even with a satellite schedule, which does not match the user accesses. At the same time, a user group which is serviced by a local station with a small or no cache, require that the satellite schedule matches the user accesses as much as possible. Therefore, for these groups, the importance of their access statistics, should not be the same in the determination of the satellite probability distribution. Thus, one might define, in general, the probability distribution that can be used to determine the satellite probability distribution as,

$$p_s(i) = \frac{1}{\sum_{k=1}^{M} \lambda^k} \sum_{k=1}^{M} w_k \lambda_i^k \ , \forall i = 1 \ldots N \tag{17}$$

where $w_k$ denotes the weight of the user distribution $k$ in the overall satellite probability distribution. Now, the question is, how we shall determine these weights.

In order to have a feeling of what is important in the determination of the weights of individual user access probability distributions, we performed a simulation with a system with two user groups. In this simulation the first user group requests items according to Zipf 1 distribution which is centered at item 1. The second user group requests items according to Zipf 1 distribution which is centered at item 50. Both user groups are interested in all 100 items that are broadcast by the satellite. The local station, which is servicing the first user group, has a cache of size 30. The cache of the local station, which is servicing the second user group, is the variable of our simulation. The satellite schedule is determined by the MAD policy with respect to the distribution

$$p_s(i) = \kappa z1(1, 100, 100) + (1 - \kappa) z1(50, 100, 100)$$

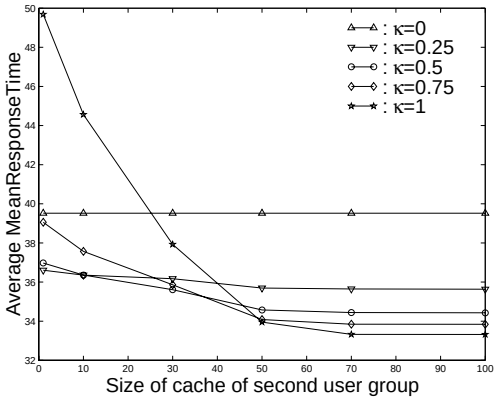


Figure 16: *Illustration of the overall average mean response time of two group system with changing weights and cache sizes.*

where $\kappa = \{0, 0.25, 0.5, 0.75, 1\}$. The results of this simulation is depicted in Figure 16.

From Figure 16, we see that the intuitive solution to use an aggregate of the user access probability distributions ($\kappa = 0.5$) provides a very reasonable and acceptable average performance. However, it is clear that, as the difference between the sizes of the caches gets larger, the intuitive solution is not the best one. From this simulation, one may notice that, the weight of the probability distribution of an user group is inversely proportional to the size of the cache of its local station. Thus, we come up with the *WSAT* (Weighted Satellite Schedule) algorithm, that is stated as follows.

**WSAT Algorithm:** *The WSAT satellite schedule is created by the MAD policy over the distribution,*

$$p_{WSAT}(i) = \frac{1}{\sum_{k=1}^{M} \lambda^k} \sum_{k=1}^{M} w_k \lambda_i^k \ , \ \forall i = 1 \ldots N \tag{18}$$

*where $w_k$ is defined as,*

$$w_k = \frac{\frac{N_k}{C_k}}{\sum\limits_{l=1}^{M} \frac{N_l}{C_l}} \tag{19}$$

*$N_k$ and $C_k$, are the total number of items, the users in group $k$ are interested in, and the size of the cache of the local station servicing user group $k$, respectively.*

We have performed numerical analysis comparing the average response time of the system with WSAT, and the system with the satellite schedule determined by the aggregate probability distribution. We provide three examples for these simulations. In each example, we observed a system with five user groups, with different Zipf 1 distributions and local station cache sizes, $C_k$ 's. It is assumed that, $\lambda^k$'s and $r$'s are equal.

| | | | MRT | |
|---|---|---|---|---|
| Group | Distribution | Cache size | WSAT | Aggregate |
| 1 | z1(1,20,100) | $C_1 = 5$ | 24.48 | 28.31 |
| 2 | z1(20,20,100) | $C_2 = 5$ | 24.83 | 28.30 |
| 3 | z1(40,20,100) | $C_3 = 5$ | 25.64 | 30.20 |
| 4 | z1(60,20,100) | $C_4 = 10$ | 24.65 | 21.72 |
| 5 | z1(80,20,100) | $C_5 = 20$ | 8.07 | 8.07 |
| | | Overall | 21.534 | 23.32 |

Table 1: *Example I*

| | | | MRT | |
|---|---|---|---|---|
| Group | Distribution | Cache size | WSAT | Aggregate |
| 1 | z1(1,100,100) | $C_1 = 5$ | 36.45 | 42.08 |
| 2 | z1(20,100,100) | $C_2 = 10$ | 39.03 | 40.72 |
| 3 | z1(40,100,100) | $C_3 = 20$ | 41.96 | 38.83 |
| 4 | z1(60,100,100) | $C_4 = 30$ | 40.54 | 36.90 |
| 5 | z1(80,100,100) | $C_5 = 40$ | 36.37 | 35.07 |
| | | Overall | 38.87 | 38.72 |

Table 2: *Example II*

As seen from these three examples, the intuitive solution to use the aggregate of the user access probability distribution as the satellite probability distribution performs quite well. There are also some cases, where the aggregate solution outperforms the WSAT algorithm. However, as the skewness of the user access probability distributions and the difference

| | | | MRT | |
|---|---|---|---|---|
| Group | Distribution | Cache size | WSAT | Aggregate |
| 1 | z1(1,100,100) | $C_1 = 5$ | 35.21 | 39.45 |
| 2 | z1(10,100,100) | $C_2 = 10$ | 36.84 | 38.75 |
| 3 | z1(20,100,100) | $C_3 = 20$ | 38.39 | 37.75 |
| 4 | z1(30,100,100) | $C_4 = 30$ | 39.08 | 36.53 |
| 5 | z1(40,100,100) | $C_5 = 40$ | 37.86 | 36.50 |
| | | Overall | 37.476 | 37.796 |

Table 3: *Example III*

between the cache sizes of the local stations increase, WSAT performs much better than the aggregate solution. More examples can be found in [1].

# 7 Conclusions and Future Work

In this paper, we have looked into the efficient delivery of information in two-stage broadcast systems. Multi-stage systems such as the one investigated in this paper, can provide both cost- and time- efficient access to the information. Such systems can also provide hierarchical information distribution, in the ad-hoc wireless military battlefield networks.

In order to provide low-latency service to the users, we have assumed that local servers are equipped with caches, and developed cache management and scheduling techniques to improve the observed latency. In our numerical work, we have shown that, if an individual user group's interests are quite different than the overall user interests, then our proposed solution, can substantially improve the observed latency of that user group.

In this study, we have assumed that there is no uplink between the users and the local servers or the local servers and the main server. The results of this paper should be extended to more general hybrid push- and pull- based systems. This problem is also an interesting problem that may gather further attention.

In this paper, we have focused on the delivery of information in the minimal average

time. There is also an equally important problem of delivery of time-sensitive information with minimum average number of deadline misses. This problem, although similar to ours, requires special attention, especially for the real time applications.

# Appendix: Solution of OP

The Lagrangian function for the optimization problem OP is,

$$\begin{aligned} L(\mathbf{b}, \mathbf{a_1}, \mathbf{a_2}, \mathbf{it}, \gamma) &= (b_1 + a_1^1)^2 + \sum_{k=2}^{s_T^i} (a_2^{k-1} + b_k + a_1^k)^2 + \sum_{l=1}^{n_T^i - s_T^i} it^2[l] \\ &+ \left( T - h_T^i - \sum_{k=1}^{s_T^i} b_k + a_2^{s_T^i} \right)^2 + \gamma \left( \sum_{k=1}^{s_T^i} a_1^k + a_2^k + \sum_{l=1}^{n_T^i - s_T^i} it[l] - h_T^i \right) \end{aligned} \quad (20)$$

The solution of OP is by Lagrange optimization. Taking the derivative of the Lagrangian function with respect to $b_1$, $b_l$ where $l \neq 1$, $a_1^1$, $a_1^l$ where $l \neq 1$, $a_2^{s_T^i}$, $a_2^l$ where $l \neq s_T^i$, and $it[l]$ we obtain the following seven equations. $\gamma$ is the Lagrangian constant.

$$2b_1 = T - h_T^i - \sum_{k=2}^{s_T^i} b_k + a_2^{s_T^i} - a_1^1 \quad (21)$$

$$2b_l = T - h_T^i - \sum_{k=1, k\neq l}^{s_T^i} b_k + a_2^{l-1} - a_1^l \quad (22)$$

$$a_1^1 = -\gamma/2 - b_1 \quad (23)$$

$$a_1^l = -\gamma/2 - b_l - a_2^{l-1} \quad (24)$$

$$a_2^{s_T^i} = -\gamma/2 - \left( T - h_T^i - \sum_{k=1}^{s_T^i} b_k \right) \quad (25)$$

$$a_2^l = -\gamma/2 - b_{l+1} - a_1^{l+1} \quad (26)$$

$$it[l] = -\gamma/2 \quad (27)$$

If $a_1^k \leq 0$ and $a_2^k \leq 0$ for $k = 1, 2, \ldots, s_T^i$, then $a_1^k = 0$, $a_2^k = 0$, for $k = 1, 2, \ldots, s_T^i$, and

$$it[l] = \frac{h_T^i}{n_T^i - s_T^i} \text{ , for } l = 1, 2, \ldots, n_T^i - s_T^i \tag{28}$$

$$b_k = \frac{T - h_T^i}{s_T^i + 1} \text{ , for } k = 1, 2, \ldots, s_T^i \tag{29}$$

$a_1^l \leq 0$, if $-\gamma/2 \leq b_l$, and $a_2^l \leq 0$, if $-\gamma/2 \leq b_{l+1}$. That is, if $\frac{h_T^i}{n_T^i - s_T^i} \leq \frac{T - h_T^i}{s_T^i + 1}$, then both $a_1^l$, and $a_2^l$ are non-positive. This condition can easily be rephrased as $h_T^i \leq \frac{n_T^i - s_T^i}{n_T^i + 1} T$.

# References


[1] Ö. Erçetin. "Information Delivery in Two-stage Satellite-Terrestrial Wireless Systems", *Master's Thesis*, University of Maryland at College Park, May 1998.

[2] C.J. Su, L. Tassiulas and V. Tsotras. "Broadcast Scheduling for Information Distribution," *ACM Journal on Wireless Networks*, vol5, no2, pp 137-147, 1999.

[3] Ö. Erçetin, L. Tassiulas. "Information Delivery in Two stage Satellite-Terrestrial Systems", *Proceedings of CISS'98*, Princeton, NJ.

[4] M. H. Ammar, J.W. Wong."The design of Teletext Broadcast Cycles." *Performance Evaluation*, 5(4):235-242, Dec.85.

[5] G. K. Zipf *Human Behaviour and the Principle of Least Effort*, Addison-Wesley Press, Cambridge MA, 1949.

[6] C. Lawrence, J.L. Zhou, A. L. Tits, CFSQP version 2.5, http://www.isr.umd.edu/Labs/CACSE/FSQP/fsqp.html

[7] N.H. Vaidya, S. Hameed. "Data Broadcast in Asymmetric Wireless Environments." *Proc. of WOSBIS'96*, Rye, NY, 1996.

[8] S. Acharya, M. Franklin, S. Zdonik. "Dissemination-based Data Delivery using Broadcast Disks." *IEEE Personal Communications*, 2(6):50-60, Dec 1995.

[9] S. Acharya, M. Franklin, S. Zdonik. "Prefetching from a Broadcast Disk." *Proc. 12th Int'l Conf. Data Eng.*, New Orleans, LA, Feb 1996.

[10] M.H. Ammar "Response Time in a Teletext System: An Individual User's perspective." *IEEE Trans. on Communications*, COM-35(11):1159-1170, Nov 1987.

[11] L. Tassiulas, C.J. Su. "Optimal Memory Management Strategies for a Mobile User in a Broadcast Data Delivery System." *IEEE JSAC Special Issue on Networking and Perf. Issues of Personal Mobile Comm.*, 15(7):1226-1238, Sep 1997.

[12] http://www.pointcast.com

[13] http://www.direcpc.com

[14] http://www.marimba.com

[15] http://www.intercast.com