

Performance of UMTS Code Sharing Algorithms in the Presence of Mixed Web, Email and FTP Traffic

Doru Calin, Santosh P. Abraham, Mooi Choo Chuah

Abstract— The paper presents a performance study of two algorithms for dynamic allocation of the Dedicated Channels (DCH) in UMTS over the air interface, namely Least Recently Used (LRU) algorithm and an Adaptive algorithm. The algorithms are designed to efficiently share the dedicated channels among users whose traffic patterns are characterized by bursty packet transfers followed by long inactivity periods. Specifically, Web browsing, FTP and Email applications were considered in order to evaluate the performance of the above mentioned resource allocation algorithms in the context of bursty traffic with relaxed delay constraints and in the presence of delays introduced by the backhaul network.

Index Terms—UMTS, Inactivity timer, dynamic DCH allocation, traffic burstiness, Web, Email, FTP

I. INTRODUCTION

UMTS systems were originally designed to achieve a peak data rate of 2 Mbps within 5 MHz of spectrum [1, 2]. This makes possible to offer a variety of new data services over wireless communication channels. Data can be delivered either by using dedicated channels (DCH) that operate at higher data rates or by implementing the concept of dedicated shared channels (DSCH), which allow multiple users to be scheduled over the same communication channel by employing a time division mechanism to control the access to the radio resources. Furthermore, the HSDPA (High Speed Downlink Packet Access) evolution of UMTS will allow theoretical peak data rates to be pushed to 10 Mbps by making the system more robust to errors caused by the radio channel thanks to some advanced features such as Adaptive Modulation and Coding (AMC), Node-B based scheduling, fast physical layer hybrid ARQ (H-ARQ) and shorter Transmission Time Interval (TTI) (2 ms) [3].

The paper analyzes the potential of a UMTS system to carry

data services over dedicated channels. In this context, packet data service users must be either in the Cell_DCH or Cell_FACH state to transmit data [4]. In the Cell_DCH state, the user is provided with a dedicated channel (DCH) with a fixed data rate. In the Cell_FACH state the user shares the Forward Access Channel (FACH) with other users. The data rate available to a user in the Cell_DCH is much higher than that available in the Cell_FACH state. However, due to limitations in the number of available orthogonal codes and transmit power, the number of users that can be kept in the Cell_DCH state is limited. This number depends on the data rate per user. For example, if all users require 384Kb/s then only 7 codes are available per sector, and thus a maximum of 7 users can be kept simultaneously in the Cell_DCH state.

Applications such as Email, File Transfer Protocol (FTP) and Web browsing can be carried out over packet data services in UMTS. These applications have what is usually called relaxed delay requirements. The relaxed delay requirement is due to the non-real time nature of the content and the fact that most users will tolerate a few seconds of delay for the download of a Web page or Email. Hence a few hundreds of ms of delay to obtaining a DCH may be tolerable. In addition, data streams produced by these applications are not continuous. They consist of bursts of data (of varying sizes) followed by relatively large idle times that are usually of the order of seconds, which calls for intelligent allocation of network resources to maximize the system efficiency.

These two properties i.e., relaxed delay requirements and a non-contiguous data stream, can and should be exploited to share the limited pool of DCHs among a large number of competing users. Since users activity is in idle mode for several seconds for the data applications considered by the paper, they may be switched out of the DCH reservation mode during their idle periods. This switching would not affect the quality perceived by the user, as long as the user is granted access (with acceptable delay) to a DCH when its next burst of data arrives. The released DCH could subsequently be used to serve other active users. This method of dynamic allocation of DCHs on a need basis would allow many more users of data services to be admitted into the UMTS network while

D. Calin is with the Networking Technologies and Performance Analysis Department of Bell Labs, Lucent Technologies, Holmdel, NJ 07733 USA (corresponding author: e-mail: calin@bell-labs.com).

S. P. Abraham is now with Qualcomm Inc., San Diego, CA, USA.

M. C. Chuah is now with the Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA.

guaranteeing reasonable quality of service to all of them.

This work introduces and evaluates two algorithms for the dynamic allocations of DCHs among the data users of a UMTS network. The study has been conducted to account for delays on the backhaul segment of the network. As such, a random delay jitter was modeled to affect packet transmissions over the backhaul. A method for tuning the parameters of an adaptive algorithm to increase its robustness to the backhaul delay is presented. Section II describes the algorithms used by this study and Section III explains the network model that has been used. Relevant performance metrics such as page channel occupancy, throughput and power usage were collected and reported in section IV. Section V is devoted to some concluding remarks.

II. DYNAMIC ALLOCATION ALGORITHMS

The two algorithms studied in this work are named:

1. Least Recently Used (LRU) algorithm
2. Adaptive algorithm

A. Least Recently Used (LRU) Algorithm

The LRU algorithm uses an inactivity timer for each user. However, there is no preset threshold for switching out of the Cell_DCH state a user. Instead, a user is allowed to remain in the Cell_DCH state in the absence of uplink or downlink transmissions, as long as there are no other users requiring a DCH. When another user requires DCH bandwidth, the inactivity timers of all users in Cell_DCH are scanned. The user that has the highest inactivity timer value is switched to the Cell_FACH state. Then the freed DCH is allocated to the user that requires it.

B. Adaptive Algorithm

To describe the adaptive algorithm we require the following notations:

n : The index of the algorithm updates

u_n : Number of uplink RLC PDU (Radio Link Control Packet Data Unit) arrivals between the n^{th} update and $n+1^{th}$ update

d_n : Number of downlink RLC PDU arrivals between the n^{th} update and $n+1^{th}$ update

q_n : Queue length of the RLC buffer (in terms of RLC PDUs) at the n^{th} update epoch

At update epoch $n+1^{th}$ do:

1. $x_{n+1} = u_n + d_n + q_n$
2. Compute the Estimated Bandwidth Requirement, denoted by y_n , as follows:

$$y_{n+1} = a y_n + (1-a)x_n + I : a < 1$$

3. **if** ($y_n + I < \alpha$ && user_state = Cell_DCH)

Switch user to FACH state

else if ($y_n + I > \beta$ && user_state = Cell_FACH && DCH available)

Switch user to DCH state.

Note that $dn+qn$ is a measure of downlink data rate required in the n th interval and un is a measure of uplink rate. The parameter x_n is an approximation of the sum of required uplink and downlink data rates. If a user is in Cell_FACH and has a high value x_n then a DCH may be needed. In order to avoid over-reacting to sudden changes in x_n a “low pass” filtered version of x_n i.e., y_n is used in making the switching decision. Criterion to be used for selecting the parameters a , α , β are discussed later.

Neither the RLC buffer level nor the downlink arrival rate taken alone are sufficient for the efficient operation of the algorithm. Note that the buffer level remains low as long as user arrival rate is sufficiently lower than the transmission rate. Thus a user could be switched from Cell_DCH to Cell_FACH during its page download phase if switching decisions were made purely based on queue lengths. Using only downlink arrival rate can be problematic due to TCP. Since TCP has a slow start process, the arrival rate at the beginning of a download can be small. Hence using measurements of arrival rate only may lead to large delays in detecting the beginning of a page download, thus adversely affecting user perceived page download time.

Note that the metric chosen for measurements combines queue length and data rate, thus requiring a single filtering operation and a single set of thresholds. This choice makes the implementation simple and cost attractive.

III. SIMULATION SCENARIOS

The network simulated is shown in the Figure 1. The path from the server to the RNC is through a network of random delay. To introduce variability in this delay we assume that it is randomly distributed. Results presented in this paper were generated assuming an exponential distribution for the backhaul delay.

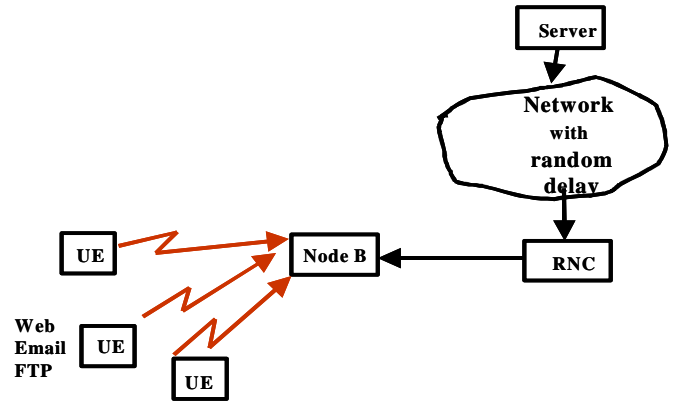


Fig. 1: Schematic of the simulated network

The traffic through the network consists of sessions brought by each user. Users arrive according to a Poisson process. Each user brings a session of one of the following types:

HTTP web browsing, Email or FTP. The parameters used for the data model are presented in Table 1.

TABLE I: CHARACTERISTICS OF THE TRAFFIC MODELS

Traffic Type	Characteristics	Fraction
HTTP Web browsing	Each session consists of an exponentially distributed number of pages with mean 23. The think time between pages is exponentially distributed with mean 30. The mean page size is 60KB and consists of a set of objects whose size distribution is taken from an empirical distribution [[5]].	39%
Email	The download and upload message size is exponentially distributed with mean 40 KB. The "idle period" consisting of writing, reading time is modeled as an exponential with mean 80s.	48%
FTP	Consists of a single file transfer whose size is hyper exponentially distributed	13%

Once all the traffic for a user session has been transmitted, the user departs. The user arrival rate is 0.01 users/sec. The permitted user data rates over the air link are 384 Kbps, 128 Kbps, 64 Kbps.

A. Description of a user Web Browsing Session

A user initiates the download of a page by sending a request. The response to this request is generally an HTML or Java script file that contains requests for several objects that make up the page (see Figure 2). For example, consider browsing through "cnn.com". The first file downloaded consists of text, some HTML code and some Java script. The execution of the HTML code and Java scrip initiates the download of other objects such as pictures, flash media files or advertisements.

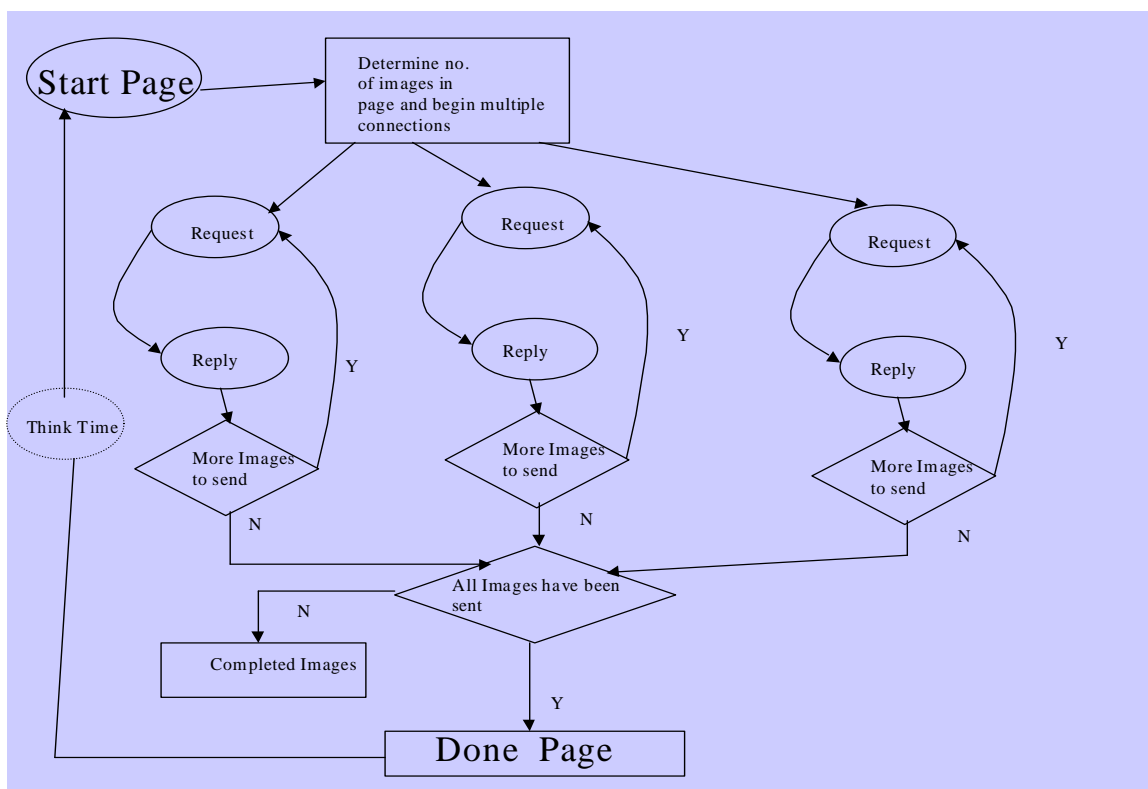


Fig. 2: Flow chart of events in a Web browsing session

Once the objects of a page are completely received, the user moves into the think time where no traffic is generated. The download of the next page begins after the think time elapses. In our study we assume that user cycles through this download think time sequence a random but finite number of times.

A partial plot of the empirical distribution of object sizes is shown in Figure 3.

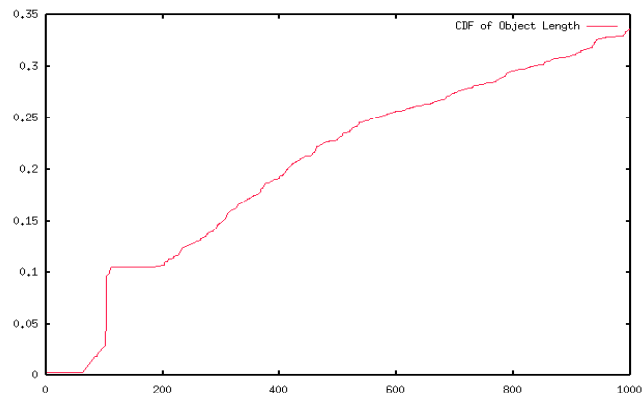


Fig. 3: Distribution of HTTP object sizes

A. Power Allocation for Users in DCH State

The power allocated to a user is updated at regular intervals. The required power allocation is based on an empirical joint distribution of I_{or} (power received by a user from its own sector) and I_{oc} (power received by the user from other sectors) that was obtained from field tests. The empirical distribution is given in Figure 4 [5]. Using this joint distribution we obtain the unconditional distribution of I_{or} and the conditional distribution of I_{oc} given I_{or} . When a new user arrives, a value of I_{or} is chosen for the user (based on the unconditional distribution of I_{or}). This value of I_{or} remains unchanged for the user throughout the simulation. At the power update interval, a value of I_{oc} is chosen using the conditional distribution of I_{oc} given I_{or} .

The power allocation is based on the ratio (I_{or}/I_{oc}) and the data rates of the user. If a user cannot be allocated its maximum data rate due to power limitations, an attempt to allocate a lower bandwidth is made, i.e., if a user cannot be allocated 384Kb/s, then one attempts to allocate power for 128Kb/s and so on. If the available power cannot satisfy the requirements of the user's permitted data rates, the user rate is dropped to zero. One of the parameters reported in the simulation study is the mean downlink power usage across all users. In calculating the downlink power usage for the user, we assume that a user consumes its full power allocation during TTIs where it sends data and 10% of allocated power in TTIs where no data is sent.

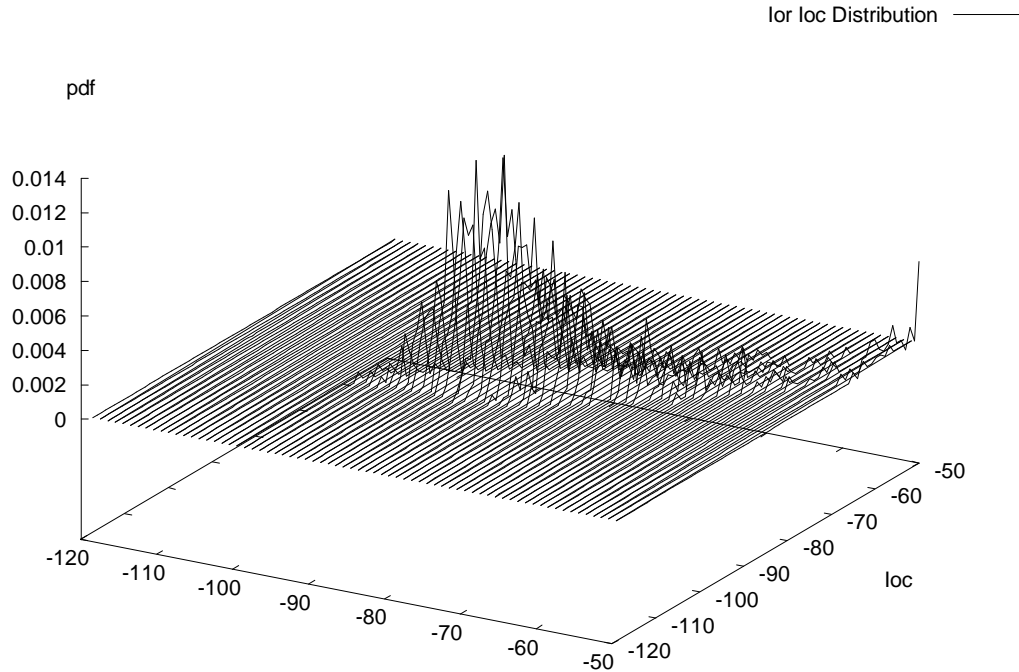


Fig. 4: Joint I_{or}, I_{oc} distribution

A. LRU timer Implementation

For the LRU algorithm, decisions for switching users are made periodically at intervals of 200ms. The power requirements for the DCHs are also updated at these intervals.

In order to determine which users are eligible for switching from Cell_FACH to Cell_DCH state in LRU

algorithm, the following criterion is used. Let define n , u_n , d_n , q_n , x_{n+1} as in the section II.E. x_{n+1} is the measurement used for determining user activity.

If $x_{n+1} > 10$ for a user in Cell_FACH, then the user is eligible for changing state to Cell_DCH. For the LRU algorithm, the user whose inactivity timer is the largest is released if there are users in Cell_FACH waiting for a DCH. In order to prevent excessive switching in the LRU due to short intervals of empty buffer, one requires that the

user's buffer should stay empty for at least 100ms when it hits the empty state. In addition, one also requires that a user remains for at least one second (1s) in a state before being eligible to switch to another state. The time required for switching between states is 0.5s.

B. Adaptive Algorithm

For the adaptive algorithm, a switching decision is made every 200ms. One also requires that a user must remain in a given state for at least 1s before being eligible to switch to another state. The most critical issue in the implementation of the adaptive algorithm is the choice of the parameters a , α , β . These parameters directly impact the time taken to detect the beginning of a new page and the end of the page download. Clearly, the value of a , α , β depends on the measurement update interval. We do not provide guidelines for the choice of the measurement update interval since it depends on the available processing power. We present a technique for determining a , α , β given the update interval.

Our goals in setting the parameters are as follows:

- i) Switch from FACH to DCH quickly when the download of a new page starts.
- ii) Prevent transitions due to the jitter on the packet interarrival times.
- iii) Trigger transitions from DCH to FACH when the utilization is low.

Note that the adaptive algorithm with an update period of T operates like an RC filter with "time constant" ($T/\ln a$). In order to satisfy condition ii) we use the following heuristic rule. We obtain a suitable delay threshold, D , such that

$Prob(RTT > D) < 0.05$, where RTT stands for network round trip time (that includes the backhaul delay).

We then choose a such that $T / |\ln a| = T \lceil D / T \rceil$

In order to satisfy condition i) let us assume that we desire to detect a new page arrival within n updates. First, we choose a suitable packet size k such that most packets from page downloads are greater than k . Assuming the value of the update parameter y is 0 before a page begins, we choose β such that y_n becomes larger than β if the buffer is of size k and there are no arrivals or departures for n consecutive updates.

To satisfy condition iii) we choose α such that a user is switched from DCH to FACH when the update y falls to a value below the number of RLC PDUs that can be transmitted within an update interval with 1/10th of the FACH bandwidth, (f), i.e:

$$\frac{y_n * 320}{t} \leq \frac{f}{10}$$

$$\alpha = \frac{ft}{3200}$$

In our simulation studies we considered three different update intervals, 5 TTIs, 10 TTIs, and 15 TTIs (one TTI lasts 20ms). The FACH data rate was 32Kb/s. We have considered two values for the average of the random network delay on the backhaul. In the first case the delay is exponentially distributed with a mean of 100ms and in the second case, the mean is 200ms. The parameters used for these cases are given in Table 2 and Table 3, respectively. The assumptions used in the selection of the parameters are: page arrival is signaled by the arrival of an object of size at least 10 PDUs. The desired time to detect the arrival of a page is 500ms or 600ms.

TABLE 2: ADAPTIVE ALGORITHM PARAMETER CHOICE FOR 100 MS NETWORK DELAY

Update Interval (T)	$T / \ln a $	D	a	β	α
100 ms	500 ms	500 ms	0.8187	6.3	1
200 ms	600 ms	500 ms	0.7165	6.3	2
300 ms	600 ms	500 ms	0.6065	6.3	3

TABLE 3: ADAPTIVE ALGORITHM PARAMETER CHOICE FOR 200 MS NETWORK DELAY

Update Interval (T)	$T / \ln a $	D	a	β	α
100 ms	1000 ms	1000 ms	0.905	4	1
200 ms	1000 ms	1000 ms	0.8187	4.51	2
300 ms	1200 ms	1000 ms	0.7788	4	3

IV. SIMULATION RESULTS

In Table 4 and Table 5 we present the results with a mean network delay of 100ms and 200ms, respectively. The throughput of all the three applications considered by this study, Web, FTP and Email decreases at higher backhaul delays. The backhaul delay increase also increases in general the number of switches between Cell_DCH and Cell_FACH state.

The adaptive algorithm shows robustness with respect to the variations on the backhaul network delay. The throughput performance for the adaptive algorithm is also close to that observed with the LRU algorithm. The number of switches is around 2 per page, which is the desired amount for the non-LRU algorithms. In terms of required power, the LRU consumes more than the adaptive algorithm: in the 100 ms backhaul delay case the UE consumed power is as much as 2 dBm higher than for the adaptive algorithm. This difference in the required dedicated power per user is reduced to less than 1 dBm for the 200 m backhaul delay case.

TABLE 4: SIMULATION RESULTS WITH BACKHAUL DELAY OF MEAN 100 MS

Parameter	LRU	5 TTI	10 TTI	15 TTI
Occupancy of DCH	4.63%	4.27%	5.16%	4.18%
Mean total power allocated per UE [dBm]	-20.91	-22.89	-22.16	-22.94
HTTP number of switches per page	0.48	1.81	1.70	1.70
Email number of switches per page	0.86	3.24	1.68	1.59
HTTP Mean page size (KB)	61.72	57.54	57.42	60.84
Email Mean page size (KB)	41.37	41.04	40.92	41.22
HTTP Mean user perceived bandwidth (Kbps)	25.84	24.51	24.42	24.40
Email Mean User perceived bandwidth (Kbps)	65	55.90	61.87	64.62
FTP Mean user perceived bandwidth (Kbps)	124.11	122.16	124.11	122.16

TABLE 4: SIMULATION RESULTS WITH BACKHAUL DELAY OF MEAN 100 MS

Parameter	LRU	5 TTI	10 TTI	15 TTI
Occupancy of DCH	4.01%	4.13%	5.04%	4.06%
Mean total power allocated per UE [dBm]	-21.10	-22.19	-21.61	-21.81
HTTP number of switches per page	0.7	1.96	2.17	1.9
Email number of switches per page	1.03	1.95	1.67	1.66
HTTP Mean page size (KB)	60.2	57.37	61.17	62.58
Email Mean page size (KB)	41.29	40.52	40.99	41.12
HTTP Mean user perceived bandwidth (Kbps)	15.54	14.55	14.57	15.13
Email Mean User perceived bandwidth (Kbps)	35.45	33.87	33.58	33.57
FTP Mean user perceived bandwidth (Kbps)	66.62	67.55	63.07	66.54

V. CONCLUDING REMARKS

The paper focused on the performance of dynamic DCH allocation algorithms in UMTS networks in the presence of backhaul network delay. The algorithms have been studied for the case where users generate a mixed of three different applications: Web-browsing sessions, Email and FTP. Two algorithms, namely LRU (Least Recently Used) and Adaptive, were studied. We developed a method for setting the parameters for the Adaptive algorithm using the page object size statistics, update interval and desired delay.

Dynamic DCH allocation algorithms operate by

switching a user out of the DCH state when extended periods of inactivity are detected. In the case of the adaptive algorithm, a moving average metric based on the packet arrival process and queue length is used to determine user inactivity. When the metric drops below a certain threshold, the user is switched out of the DCH state into the Cell_FACH state.

When these algorithms are used on DCH's carrying bursty traffic such as Web browsing traffic, Email and FTP, it is desirable that a DCH remains allocated to a user during the download period and be released only during an extended page viewing period. However, in the case where there is delay and delay jitter, it is possible for the dynamic DCH allocation algorithms to switch a user out of the DCH state during a page download due to a high interarrival time between packets of the page caused by network delay jitter. Such scenario is especially possible when the parameters of the algorithms are not chosen properly. This switching process places additional computational burden and increases the user perceived latency of the packets. In addition, the user's TCP throughput suffers especially in the case where a downloaded page consists of a number of small objects.

The power consumption of the LRU algorithm was found to be the highest among all the approaches. The adaptive algorithms perform well if the filter coefficients and thresholds are chosen properly. However, due to the sensitivity of these parameters to traffic statistics, the choice of a set of parameters that would work for all traffic types may not be easy to determine.

REFERENCES

- [1] H. Holma and A. Toskala, "WCDMA for UMTS", Radio Access for Third Generation Mobile Communications, 2nd Edition, John Wiley & Sons, 2002.
- [2] "Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS", ETSI/SMG2 TR.101.112, UMTS30.03 version 3.2.0, Apr. 1998.
- [3] 3GPP Technical Report 25.848, Physical layer aspects of UTRA High Speed Downlink Packet Access, version 4.0.0, March 2001.
- [4] 3GPP Technical Report. 25.331, "Radio Resource Control (RRC) Protocol Specification", Release 6, v6.5.0 (2005-03).
- [5] B. Mah, "An Empirical Model of HTTP Network Traffic", IEEE Infocom 97