Performance of Code Allocation Algorithms on UMTS Uplink with Mixed Voice/Data Traffic

Balakrishna J. Prabhu, A. Chockalingam
Department of ECE
Indian Institute of Science, Bangalore 560012
e-mail: achockal@.ece.iisc.ernet.in

D. Jayaram
Communications and Embedded Systems Group
L & T Infotech Ltd., Bangalore 560001
e-mail: Jayaram.D@lntinfotech.com

ABSTRACT

In this paper, we are concerned with the performance evaluation of code allocation algorithms in Universal Mobile Telecommunication System (UMTS) networks. UMTS networks will offer multiple services (voice, data, video, etc.) with different quality-of-service (QoS) requirements to mobile users. In this paper, we evaluate the performance of different code (rate) allocation algorithms on the UMTS uplink in a mixed voice/data traffic scenario. Two different code allocation algorithms are considered: one based on the overall buffer occupancy at the user terminal, and the other based on dividing the available codes equally among the requesting users. Further, for a data-only system, we evaluate the performance of two algorithms for rate and power allocation based on received signal-to-interference ratio (SIR) at the base station.

I. INTRODUCTION

UMTS networks will offer multiple services (voice, data, video, etc.) with different quality-of-service (QoS) requirements to mobile users [1]-[3]. UMTS networks support high speed radio access (up to 2 Mbps) based on Wideband Code Division Multiple Access (WCDMA). The users, during call establishment, request the network the desired QoS for the connection in terms of data rate, delay, priority, reliability, etc. The network allocates the available resources (transmission rate and transmit power) to different users based on certain resource allocation policy. Rate allocation to each user can be varied either by allocating multiple spreading codes with constant spreading factor or by varying the spreading factor. In WCDMA, the physical channels follow a layered structure of *radio frames* and *time slots*. Each radio frame is of 10ms duration and consists of 15 slots. The maximum number of bits sent per radio frame depends on the rate allocated to the user. The network can dynamically vary the rate allocation to different users on a frame-by-frame basis.

The radio interface protocol layers which are involved in the radio resource allocation are shown in Fig. 1 [4]. The radio resource control (RRC) in Layer 3 (L3) is responsible for the signaling and control information exchange between the user and network to effect allocation and deallocation of radio resources [5]. The data transport services offered by the physical layer (Layer 1 – L1) [6] is achieved through the use of *transport channels* via the media access control (MAC) [7] sublayer in Layer 2 (L2). Each user terminal may simultaneously have multiple transport channels multiplexed on to one or more physical channels [8]. For example, a 64 Kbps packet switched (PS) data transport channel and a 12.2 Kbps circuit switched (CS) voice transport channel can be multiplexed on to a 128 Kbps physical channel. Transport channels are defined by how and with what features (e.g., transport block size, transmission time interval, etc., which are defined in the following paragraph) data is transferred over the air interface.

This work was supported in part by L & T Infotech Ltd., Bangalore 560001 under project No. SID/PC14032.

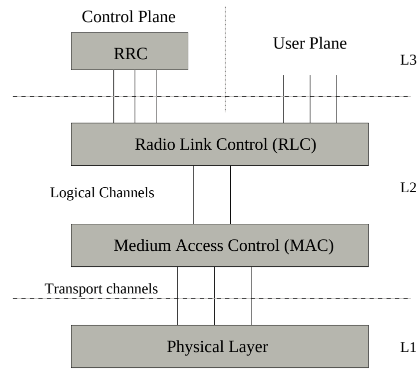

Fig. 1. Radio interface protocol architecture

Refer Fig. 2 which illustrates the following definitions. The basic unit of data exchange between L1 and MAC for L1 processing is called a *transport block* (TB). The transport block size is the number of bits in a TB. A set of TBs exchanged between L1 and MAC during the same frame using the same transport channel is called a *transport block set* (TBS). The TBS size is the number of bits in the TBS. The TB size and the TBS size are chosen by the MAC at the user terminal. The periodicity at which a TBS is transferred by L1 on to the air interface is called *transmission time interval* (TTI). The allowed values of TTI are 10, 20 40, and 80 ms. The MAC delivers one TBS to L1 every TTI. The format (TB size, TBS size, type/rate of coding, size of CRC) offered by L1 to MAC (and vice versa) for the delivery of TBS during a TTI on a given transport channel is called *transport format* (TF). The transport format determines the transport channel bit rate before L1 processing. For example, let TB size = 336 bits (320 bits payload + 16 bits header), TBS size = 2 TBs per TTI, and TTI = 10 ms. Then the transport channel bit rate (with header) is given by 336*2/10 = 67.2 Kbps. The user bit rate (without header) is given by 320*2/10 = 64 Kbps. Thus, it can be seen that variable bit rate transmission on a single transport channel can be achieved by changing (from one TTI to the other) either the TBS size only, or both the TB size and the TBS size. A set of transport formats associated with a transport channel is called a *transport format set* (TFS). A combination of TFs on different transport channels in a given TTI is called *transport format combination* (TFC). A set of transport format combinations allowed by the network is called the *transport format combination set* (TFCS). The network informs the TFCS to the user terminal to be used on the uplink transmission by the user terminal. The MAC at the user terminal then chooses between the different TFCs specified in the TFCS. The user terminal MAC can (based on certain criteria) choose different TFCs for different TTIs.

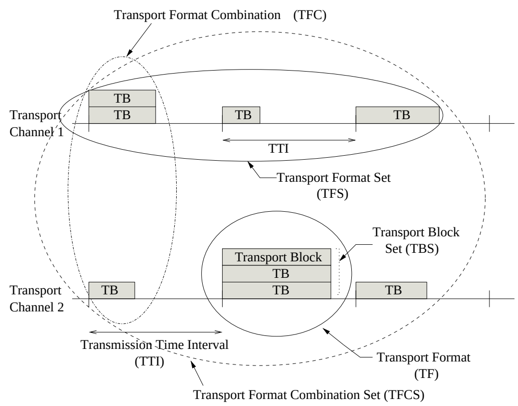

Fig. 2. Transport channel definitions

In this paper, we are interested in evaluating the performance of different code (rate) allocation algorithms on the UMTS uplink in a mixed voice/data traffic scenario. Two different code allocation algorithms are considered: one based on the overall buffer occupancy at the user terminal, and the other based on dividing the available codes equally among the requesting users. In the buffer occupancy based algorithm, the rate allocation is done is such a way that the user with a larger buffer occupancy will be assigned a higher rate. We also evaluate the performance of SIR based rate and power allocation algorithms in a data-only system.

II. SYSTEM MODEL

We consider a mixed voice/data system in which there are N mobile user terminals. Each user can generate a CS voice call and/or a PS data call. That is, at any given time, a terminal can have a) neither a voice nor a data call, b) a voice call and no data call, c) a data call and no voice call, and d) simultaneously both a voice as well as a data call. Voice call arrivals are assumed to follow a Poisson distribution, i.e., the voice call inter-arrival time is exponentially distributed with mean $T_i^{(v)}$. The voice call holding time is also assumed to be exponential with mean call holding time $T_h^{(v)}$. Since speech consists of alternating active and silence periods, we model the voice source as a ON/OFF source where the ON and OFF periods are assumed to be exponentially distributed with means T_{on} and T_{off}, respectively. During a PS data call, packets arrive in bursts where the burst inter-arrival time is assumed to be exponentially distributed with mean $T_i^{(d)}$, and the number of transport blocks (TB) per burst is assumed to be geometrically distributed with mean M_d.

A. Code Allocation Algorithms

The network allocates channel rates to users by assigning different number of spreading codes with different spreading factors. The spreading factor can be in the range 4 to 256 on the UMTS uplink. In other words, with a 3.84 Mcps chip rate, the minimum channel rate can be $3.84 \times 10^6/256 = 15$ Kbps. This channel rate will typically be larger than the user information rate because of overhead bits due to coding, CRC, etc.

In our model, we consider that the network allocates channel rates to users in integer multiples of R_{min}, where R_{min} is the minimum channel rate that can be assigned to a user. Here, we take $R_{min} = 60$ Kbps which corresponds to a spreading factor of 64. This 60 Kbps channel rate can carry a 12.2 Kbps CS voice call or 16 Kbps PS data traffic, both with their associated overhead bits due to coding, CRC, etc [9]. The values of the available channel rates then are 60, 120, 180, and 240 Kbps. We assume that N_c spreading codes each corresponding to rate R_{min} are available with the network for allocation. The network can change the allocation of spreading codes to different users every TTI. We assume that the network gives priority to voice calls while allocating spreading codes, i.e., available codes are first allotted to voice calls and the remaining codes are allotted to data calls. The algorithms presented in the following are for assigning codes for the data calls. A code allotted to a voice call is held for the entire call duration. In user terminals with simultaneous voice and data calls, silence periods in voice can be used to send data packets. For terminals with only data calls, codes can be allocated/deallocated every TTI.

A.1 Algorithm I

In this algorithm, the network allots codes to users based on the data buffer occupancy at the individual user terminal. The network allocates codes in such a way that the user with a larger buffer occupancy will be assigned a higher rate.

Let N be the number of users in the system. Let N_c be the total number of codes of rate corresponding to R_{min} available in the system. Let M, $1 \leq M \leq N_c$, be the maximum number of codes of rate corresponding to R_{min} that can be allotted to a given user. It is assumed that the network has the knowledge of buffer occupancy in all the participating user terminals. The network first assigns codes to newly arrived voice calls. The network then arranges the users in descending order of data buffer occupancy. Let N_i be the number of transport blocks in the buffer of the i^{th} user in the ordered list. Let V_i be the number of codes allotted to the i^{th} user in this list. Note that V_i can be either 1 or 0 depending on whether the i^{th} user has a voice call or not. The number of the remaining codes available for allocation to data calls (including data calls in user terminals with simultaneous voice and data), P_0, is then given by

$$P_0 = N_c - \sum_{i=1}^{N} V_i. \quad (1)$$

The algorithm performs code allocation to the largest buffer occupancy user first, the second largest buffer occupancy user next, and so on. This process is continued until the least buffer occupancy user gets the code allocation or the available codes are exhausted, whichever happens first. Accordingly, the number of codes assigned to the i^{th} user in the list, C_i, is obtained as

$$C_i = min\Big(N_i, P_{i-1}, M - V_i\Big), \quad i = 1, 2, ..., N, \tag{2}$$

where $P_i, i = 1, 2, ..., N$, is the number of codes remaining after C_i codes have been allotted to the i^{th} user in the list. With the definition of P_0 in Eqn. (1) as the initial condition, the P_is are recursively obtained using the relation

$$P_i = P_{i-1} - C_i. \tag{3}$$

A.2 Algorithm II

In this algorithm, the network allots codes to different users based on dividing the available codes equally among the requesting users. Like in Algorithm I, codes are allocated to voice calls first. Following the same notation used in Algorithm I, the assignment of the remaining codes to the data calls is done as follows.

Let N_a be the number of users with non-zero buffer occupancy. If $N_a \geq P_0$, then P_0 out of N_a users are randomly chosen and one code is allotted to each of them. If $N_a < P_0$, then the N_a users are arranged in ascending order of their buffer occupancies, and the number of codes allotted to the i^{th} user, C_i, is obtained as

$$C_i = min\left(N_i, \left\lfloor \frac{R_{i-1}}{N_a - (i-1)} \right\rfloor, M - V_i\right), \quad i = 1, 2, ..., N_a. \tag{4}$$

III. PERFORMANCE EVALUATION

The transport channel parameters (TB size, TTI, TFs) for a voice call are taken to be as shown in Table I [9]. The TTI is 20 ms and 244 bits per TTI are generated by the vocoder, resulting in an information rate of 12.2 Kbps. These 244 bits are divided into three subflows each having different number of bits per TTI. The allowed transport formats are TF0$_v$ and TF1$_v$ as shown in Table I. Note that the TB set size is either 0 or 1 TBs. We consider that the user terminal MAC chooses the transport format combination (TFC) of the subflows to be (TF0$_v$, TF0$_v$, TF0$_v$) during silence periods and (TF1$_v$, TF1$_v$, TF1$_v$) during active periods of voice.

	Subfbw #1	Subfbw #2	Subfbw #3
TB size, bits	81	103	60
TTI, ms	20	20	20
TF0$_v$	0×81	0×103	0×60
TF1$_v$	1×81	1×103	1×60

TABLE I

TRANSPORT CHANNEL PARAMETERS FOR 12.2 KBPS SPEECH

The transport channel parameter for a data call are taken to be as shown in Table II [9]. The TTI is 20 ms and the transport block (TB) size is 336 bits (which includes 320 bits of data and 16 bits of header). The TB set size can be 0, 1, 2, 3, or 4 corresponding to transport formats TF0$_d$, TF1$_d$, TF2$_d$, TF3$_d$, TF4$_d$, respectively. Accordingly, for TB set size of 4×336, the maximum information rate possible for the data call is $4 \times 320/0.02 = 64$ Kbps.

Table III shows the allowed transport format combination set (TFCS) and the number of codes required for each of the TFC

TB size, bits	336
TTI, ms	20
TF0$_d$	0×336
TF1$_d$	1×336
TF2$_d$	2×336
TF3$_d$	3×336
TF4$_d$	4×336

TABLE II

TRANSPORT CHANNEL PARAMETERS FOR 64 KBPS INTERACTIVE/BACKGROUND DATA TRAFFIC

for the mixed voice/data traffic scenario model that we have considered. It is seen that voice call is allotted one code and data call can be allotted up to four codes. It is noted that the combination TFC10 allows more bits to be sent per code allotted through the use of increased puncturing (puncturing limit = 0.76) [9].

	Voice Trch[2]	Data Trch	No. of R_{min} codes required
TFC1	TF0$_v$, TF0$_v$, TF0$_v$	TF0$_d$	0
TFC2	TF0$_v$, TF0$_v$, TF0$_v$	TF1$_d$	1
TFC3	TF0$_v$, TF0$_v$, TF0$_v$	TF2$_d$	2
TFC4	TF0$_v$, TF0$_v$, TF0$_v$	TF3$_d$	3
TFC5	TF0$_v$, TF0$_v$, TF0$_v$	TF4$_d$	4
TFC6	TF1$_v$, TF1$_v$, TF1$_v$	TF0$_d$	1
TFC7	TF1$_v$, TF1$_v$, TF1$_v$	TF1$_d$	2
TFC8	TF1$_v$, TF1$_v$, TF1$_v$	TF2$_d$	3
TFC9	TF1$_v$, TF1$_v$, TF1$_v$	TF3$_d$	4
TFC10	TF1$_v$, TF1$_v$, TF1$_v$	TF4$_d$	4

TABLE III

TRANSPORT FORMAT COMBINATION SET (TFCS) FOR MIXED VOICE/DATA TRAFFIC

N	16
M	4
$T_h^{(v)}$	100 secs
T_{ON}	1 sec
T_{OFF}	1.35 secs
M_d	10

TABLE IV

SIMULATION PARAMETERS

We evaluated the performance of voice and data calls through simulations. The voice call performance evaluated is the voice call blocking probability. Since voice calls are given priority, the voice call blocking probability follows the Engset formula given by $P(N, N_c, \rho) = \frac{\binom{N-1}{N_c} \cdot \rho^{N_c}}{\sum_{i=0}^{N_c} \binom{N-1}{N_c} \cdot \rho^i}$, where $\rho = \frac{T_h^{(v)}}{T_i^{(v)}}$. For data calls, the mean data burst transfer delay performance is evaluated. The values used for various parameters in the simulations are shown in Table IV.

Figure 3 shows the mean data burst delay performance as a function of data burst arrival rate for the two code allocation algorithms described before. The curves are parameterized by the mean voice call inter-arrival times, T_i^v. When the voice call arrival rate is very low (e.g., $T_i^{(v)} = 1500$ secs), both the algorithms result in good delay performance for data. For example, for data burst arrival rate less than 3.5 bursts/sec/node, both the algorithms result in a mean delay of less than 50 frames (i.e., 500 ms) duration. Also, the buffer occupancy based Algorithm I performs marginally better than the equal share based Algorithm II. This is because by allocating more codes to a higher buffer

[2] This includes all the three subfbws.

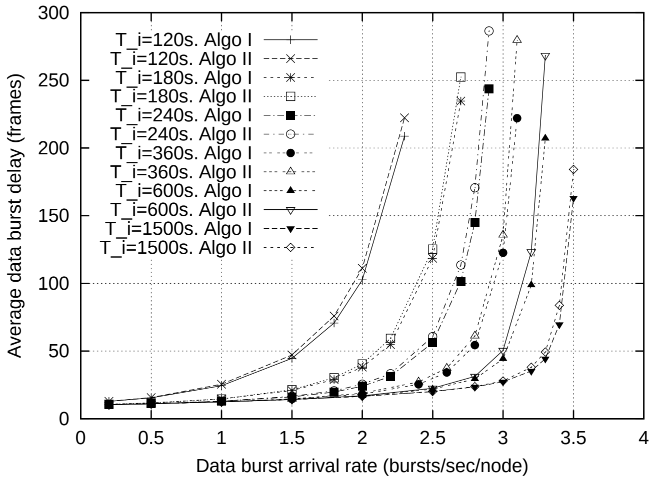

Fig. 3. Average data burst delay vs. data burst arrival rate. $N = 16$. $N_c = 12$.

occupancy user the chances of picking the combination TFC10 gets increased compared to the equal share allotment Algorithm II, which results in more number of bits sent per code allotted. This effect will be more as the burst arrival rate gets increased. Hence, the mean delay performance improvement of Algorithm I over Algorithm II is more visible at higher arrival rates. Also, as the voice call arrival rate increases (e.g., $T_i^{(v)} = 120$ secs), the delay performance of data bursts degrades as the voice calls get priority in code allocation.

IV. SIR BASED ALGORITHMS

In the previous section, the code allocation algorithms did not consider the received SIR at base station. Also, power allocation to the user terminals was not considered while allocating the transmission rate. In this section, we evaluate the performance of two rate and power allocation algorithms based on received SIR at the base station for a data-only system. In the first algorithm, the user terminal is allocated a transmission rate as well as transmit power by the base station. In the second algorithm, the base station allocates only the transmission rate and the user terminal transmits at the maximum allowable transmit power, P_{max}. In both the algorithms allocations are made in such a way that a certain BER performance criterion is met at the base station for each user. The received SIR at the base station for mobile i is given by

$$SIR_i = \frac{P_i G_i}{\sum_{j=0, j \neq i}^{N_d} P_j G_j + \eta}, \tag{5}$$

where P_i is the transmit power of mobile i, G_i is the channel gain for mobile i, and η is the thermal noise power. The channel gain, G_i, is given by

$$G_i = d_i^{-\eta} \cdot 10^{-\zeta_i/10}, \tag{6}$$

where d_i is the distance of mobile i from the base station, η is the path loss exponent, and ζ_i is the normally distributed shadow fading random variable for mobile i. The received E_b/N_o for mobile i is then given by,

$$\left(\frac{E_b}{N_o}\right)_i = SIR_i \frac{W}{R_i}, \tag{7}$$

where R_i is the rate of the code assigned to user i. In order to meet a given BER, the received E_b/N_o should be greater than a given threshold value, Ω_{th}.

A. SIR based algorithm with power control

The user terminals with backlogged data packets are ordered according to their received SIR at the base station. The code allocation is then done in descending order of the received SIR, i.e., the codes are allotted to the user with the highest SIR first, and so on. The SIR estimates of user terminals which transmitted packets in the previous frame can be measured on the traffic channel and made available at the base station. The SIR estimates of the user terminals with new message arrivals can be measured from the resource request transmission made by them on the random access channel. The resource request transmission for new message arrivals is done using the maximum permissible transmit power, P_{max}.

Once the rate allocation is done, the base station calculates the minimum overall transmit power and the individual transmit power for each mobile in order to meet the given E_b/N_o threshold value, Ω_{th}. The objective of the base station is to find the power allocations to each mobile so that

$$min \sum_{i \in C} P_i \tag{8}$$

subject to

$$P_i \geq \Omega_{th} \cdot \frac{R_i}{W} \cdot \frac{1}{G_i} \cdot \Big(\sum_{j \in C, j \neq i} P_j G_j + \eta \Big), \forall i \in C \tag{9}$$

$$P_i \leq P_{max}. \tag{10}$$

where C is the set of user terminals which have been allotted rates in this iteration. In order to minimize $\sum P_i$, we use the equality constraint in Eqn. (9) to get

$$P_i = \Omega_{th} \cdot \frac{R_i}{W} \cdot \frac{1}{G_i} \cdot \Big(\sum_{j \in C, j \neq i} P_j G_j + \eta \Big), \forall i \in C \tag{11}$$

In matrix notation,

$$\mathbf{GP}' = \eta \mathbf{1}, \tag{12}$$

where **G** is given by

$$\begin{bmatrix} G_{i_0} \frac{W}{R_{i_0}} \frac{1}{\Omega_{th}} & -G_{i_1} & \cdots & -G_{i_l} \\ -G_{i_0} & G_{i_1} \frac{W}{R_{i_1}} \frac{1}{\Omega_{th}} & \cdots & -G_{i_l} \\ \cdots & \cdots & \cdots & \cdots \\ -G_{i_0} & -G_{i_1} & \cdots & G_{i_l} \frac{W}{R_{i_l}} \frac{1}{\Omega_{th}} \end{bmatrix},$$

and **P** = $[\ P_{i_0} \quad P_{i_1} \quad \cdots \quad P_{i_l}\ ]$. l is the number of user terminals which have been a code in this iteration.

If the power allocation to any user terminal exceeds the maximum permissible transmit power level, P_{max}, its rate of transmission is reduced and the power vector is again calculated. This process is continued till a feasible power vector is found.

B. SIR based algorithm without power control

In this case, the user terminals are allowed to transmit with the maximum power level, P_{max}, and no power optimization is done at the base station. The requesting user terminals are arranged in descending order of received SIR. Maximum possible rate is allotted to the user terminal with the best received SIR. Rate is then assigned to the user terminal with the next best SIR, and so on till either the codes are exhausted or all the user terminals have been assigned codes. The base station then calculates the received E_b/N_o for each user terminal using this rate allocation and transmit power P_{max}. If the received E_b/N_o for any user terminal is less than Ω_{th}, the rate allotted to that user terminal is reduced till the received E_b/N_o exceeds Ω_{th}.

Cell radius	1 km
σ^2	4 dB
P_{max}	20 dBm
N_c	12
N_d	16
Ω_{th}	9 dB

TABLE V

SIMULATION PARAMETERS FOR SIR BASED ALGORITHMS

We evaluate the performance of the above SIR based rate and power algorithms through simulations. The transport format set available to the mobile is given in Table II. The shadow fading experienced by each node is assumed to be Gaussian with mean 0 and variance σ^2 and is correlated with correlation of 0.8. The simulation parameters are given in Table V. The path loss model is taken to be

$$loss(dB) = 128.1 + 37.6 \cdot log_{10}(R), \tag{13}$$

where R is the distance of the user terminal from the base station in km [10].

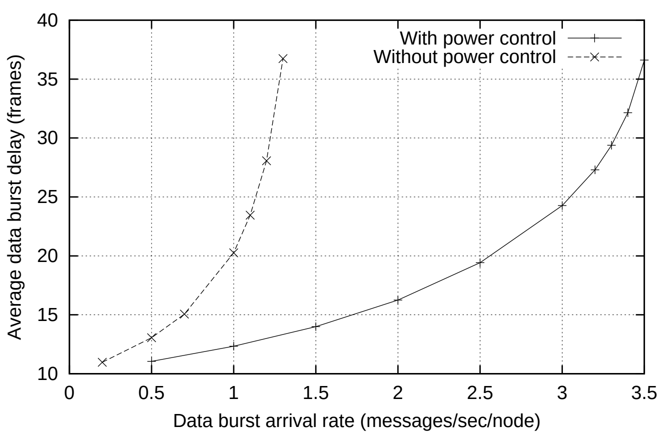

Fig. 4. Average data burst delay vs. data burst arrival rate. $N = 16$. $N_c = 12$.

C. Results and Discussion

Fig. 4 shows the average data burst delay versus the data burst arrival rate for the SIR based algorithms. It is observed that the SIR based algorithm with power control performs better in terms of average data burst delay for a given arrival rate. The interference from the transmitting nodes is lesser when power control is used than when there is no power control. This allows more nodes to transmit at the same time resulting in better resource (code) utilization and hence lower delays. Fig. 5 shows the average data burst delay for individual nodes as a function of their distance from the base station for a particular simulation scenario. It is observed that the algorithm with power control ensures that nodes which are far away see the same delay as the nodes which are closer to the base station. In the no power control case, all the nodes transmit at P_{max} resulting in a high interference from the nearby nodes. This causes the far off nodes to wait till the end of the transmission from closer nodes before being able to send their data bursts. Hence, the far off nodes experience a higher delay than the nearby nodes.

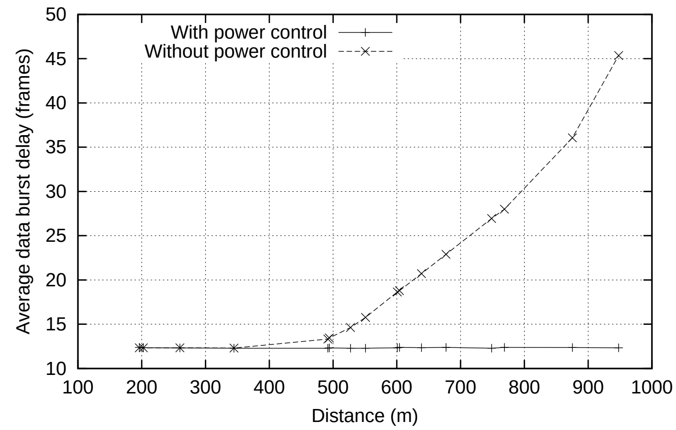

Fig. 5. Average data burst delay vs. distance. $N = 16$. $N_c = 12$. Data burst arrival rate = 1 message/sec/node.

V. CONCLUSION

In this paper, we evaluated the performance on two code allocation algorithms for mixed voice/data traffic on the UMTS uplink. Two different code allocation algorithms were considered: one based on buffer occupancy at the user terminal, and the other based on dividing the available codes equally among the requesting users. We showed that the allocation based on buffer occupancy performs marginally better than the equal share algorithm. For a data-only system, we evaluated the performance of SIR based algorithms with and without power control. It was shown that the system with power control performs better than the system without power control.

REFERENCES

[1] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for 3rd Generation Mobile Communications*, John Wiley & Sons, 2000.

[2] H. Kaaranen *et al*, *UMTS Networks, Architecture, Mobility and Services*, John Wiley & Sons, Chichester, 2002.

[3] J. Laiho, A. Wacker, and T. Novasad, *Radio Network Planning and Optimisation for UMTS*, John Wiley & Sons, Chichester, 2002.

[4] J. P. Castro, *The UMTS Network and Radio Access Technology, Air Interface Techniques for Future Mobile Systems*, John Wiley & Sons, Chichester, 2001.

[5] 3GPP TS 25.331, “Radio Resource Control (RRC) Protocol Specification”, version 4.3.0, Release 4, 2001.

[6] 3GPP TS 25.201, “Physical Layer - General Description”, version 4.1.0, Release 4, 2001.

[7] 3GPP TS 25.321, “Medium Access Control (MAC) Protocol Specification”, version 4.3.0, Release 4, 2001.

[8] 3GPP TS 25.211, “Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD)”, version 4.3.0, Release 4, 2001.

[9] 3GPP TS 34.108, “Common Test Environments for User Equipment (UE) Conformance Testing”, version 4.1.0, Release 1999.

[10] 3GPP TR 101.112 (UMTS 30.03): “Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS”, version 3.2.0, Release 1998.