Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China, April 16-18, 2006 (pp745-749)

Analysis of Delay Distribution of Data call in CDMA Systems Supporting Voice and Delay-tolerant Data Calls

Insoo Koo
University of Ulsan
Republic of Korea

Jeongrok Yang
KIPO
Republic of Korea

Kiseon Kim
GIST
Republic of Korea

Abstract: In this paper, we present a procedure to analyze the delay distribution of data traffic in CDMA systems supporting voice and delay-tolerant data services with a finite buffer. The queueing method using a buffer for a delay-tolerant traffic can be used to improve the system utilization or the availability of system resources. Under the first-come and first-serve (FCFS) service discipline, we present a numerical procedure for the calculation of delay distribution that is defined as the probability that a new data call get a service within the maximum tolerable delay requirement, based on a two-dimensional Markov model. It is shown that the buffer size is one of the important parameters to appropriately balance the availability of system resources for each traffic.

Key–Words: CDMA, Delay distribution, Voice/data

1 Introduction

Next generation mobile communications systems are primarily designed to provide users with multimedia services such as voice, interactive data, file transfer, internet access and image, in the affordable way as today's wired communication systems do. Multimedia traffic have different and multiple quality of service (QoS) requirements, which results in diverse amounts of required bandwidth according to traffic classes [1, 2, 3]. Large bandwidth traffic classes are generally more restricted than small bandwidth traffic classes in the availability of system resources. Therefore, the system capacity is mainly determined by large bandwidth traffics [3]. In order to improve the resource availability of large bandwidth traffics, resource management schemes such as queueing and channel reservation have been considered [3, 4]. As performance measure of queueing and channel reservation methods, the blocking probability and the average delay performance have been mainly considered. Specially, the queueing method to exploit the delay-tolerant characteristic of traffic can be used to improve the system utilization or the availability of system resources for the traffic [3, 5]. More meaningful measurement for delay traffic however is the delay distribution rather than the average delay performance where the delay distribution is defined as the probability that a new data call get a service within the maximum tolerable delay requirement. To provide the flexibility in the consideration of delay requirement, it is necessary to analyze the 'delay distribution' of data call. Subsequently, in the paper we present a numerical procedure for the calculation of the delay distribution of data call in a CDMA system supporting voice and data-tolerant data services with a finite buffer. We also investigate the effect of offered load of each traffic type and the buffer size on the system performance through a numerical example.

The paper is organized as follows. Following the introduction section, the system model is explained in Section 2. In Section 3, the numerical procedure to analyze the delay distribution is presented, based on the Markov model. In Section 4, a numerical example is considered. Finally, conclusion and remarks are drawn.

2 System Model

In CDMA systems, although there is no hard limit on the number of concurrent users, there is a practical limit on the number of supportable concurrent users in order to control the interference among users having the same pilot signal; otherwise the system can fall into the outage state where QoS requirements of users cannot be guaranteed. In order to satisfy the QoS requirements for all concurrent users, the capacity of CDMA system supporting voice and data services in the reverse link should be limited with following equation [6]

$$\gamma_v i + \gamma_d j \leq 1 \,, \qquad i \; and \; j \; \geq 0 \qquad (1)$$

where

$$\gamma_v = \left(\frac{W}{R_v q_v} + 1\right)^{-1} \text{ and } \gamma_d = \left(\frac{W}{R_d q_d} + 1\right)^{-1} \quad (2)$$

γ_v and γ_d are the amount of system resources that are used by one voice and one data user, respectively. i and j denote the number of supportable users in the voice and data service groups, respectively. W is the allocated frequency bandwidth. q_v and q_d are the bit energy to interference power spectral density ratio for voice and data calls, respectively, which is required to achieve the target bit error rate at the base station. R_v and R_d are the required information data rates of the voice and data calls, respectively. Each user is classified by QoS requirements such as the required information data rate and the required bit energy to interference spectral density ratio, and all users in same service group have same QoS requirements. Eqn. (1) indicates that the calls of different services take different amount of system resources according to their QoS requirements.

Further, we consider the queue with the finite length of K for delay-tolerant data traffic, to exploit its delay-tolerant characteristic, and assume that the service discipline is First-Come-First-Serve (FCFS). Based on these assumptions and system model, the call admission control (CAC), for the case $\gamma_d > \gamma_v$ can be summarized as follows.

- If $\gamma_v i + \gamma_d j \le 1 - \gamma_d$, then both new voice and new data calls are accepted.
- If $1 - \gamma_d < \gamma_v i + \gamma_d j \le 1 - \gamma_v$, then new voice calls are accepted, and new data calls are queued.
- If $1 - \gamma_v < \gamma_v i + \gamma_d j \le 1 + (K-1)\gamma_d$, then new voice calls are blocked, and new data calls are queued.
- If $\gamma_v i + \gamma_d j > 1 + (K-1)\gamma_d$, then both new voice and new data calls are blocked.

The arrivals of voice and data calls are assumed to be distributed according to independent Poisson processes with average arrival rate λ_v and λ_d, respectively. The service times of voice and data calls are assumed to be exponentially distributed with average service time $1/\mu_v$ and $1/\mu_d$, respectively. Then, the offered traffic loads of voice and data calls are expressed as $\rho_v = \lambda_v/\mu_v$ and $\rho_d = \lambda_d/\mu_d$, respectively. The call-level state diagram is given in Figure 1 for the possible region of states, based on the call admission rule. The possible region of states can be expressed

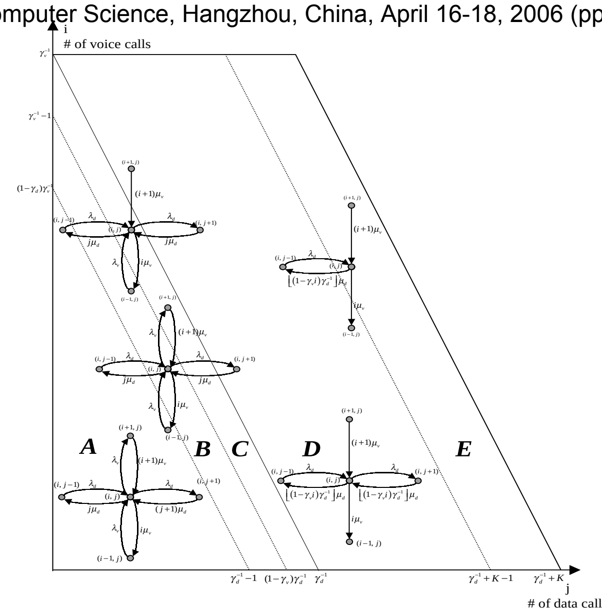

Figure 1: State transition diagram for the case of $\gamma_d > \gamma_v$.

as

$$\Omega_S = \left\{(i,j) | 0 \le i \le \gamma_v^{-1}\ j \ge 0,\ \gamma_v i + \gamma_d j \le 1 + \gamma_d K\right\} \quad (3)$$

Noting that total rate of flowing into a state (i,j) is equal to that of flowing out, we can get the steady-state balance equation for each state. Figure 2 summarizes the steady-state balance equations for the state transition diagram according to the region that the current state belongs to. If the total number of all possible states is n_s, the balance equations become $(n_s - 1)$ linearly independent equations. With these $(n_s - 1)$ equations and the normalized equation, $\sum_{(i,j)\in\Omega_S} P_{i,j} = 1$, a set of n_s linearly independent equations for the state diagram can be formed as

$$\mathbf{Q}\pi = \mathbf{P} \quad (4)$$

where $\mathbf{Q}$ is the coefficient matrix of the n_s linear equations, π is the vector of state probabilities, and $\mathbf{P} = [0, \cdots, 0, 1]^T$. The dimensions of $\mathbf{Q}$, π and $\mathbf{P}$ are $n_s \times n_s$, $n_s \times 1$, and $n_s \times 1$, respectively. By solving $\pi = \mathbf{Q}^{-1}\mathbf{P}$, we can obtain the steady-state probabilities of all states [3].

Based on the CAC rule, a new voice call will be blocked if the channel resources are not enough to accept the call, and the corresponding blocking probability for voice calls is given by

$$P_{b_v} = \sum_{(i,j)\in\Omega_{(nv,blo)}} P_{i,j} \quad (5)$$

A :

$$(\lambda_v + \lambda_d + i\mu_v + j\mu_d)P_{i,j} = \lambda_v P_{i-1,j} + \lambda_d P_{i,j-1} + (i+1)\mu_v P_{i+1,j} + (j+1)\mu_d P_{i,j+1}$$
$$\text{for } (i,j) \in \left\{ (i,j) \middle| 0 \le \begin{pmatrix} \gamma_v \\ \gamma_d \end{pmatrix} (i \;\; j) \le 1 - \max\{\gamma_v, \gamma_d\} \right\}$$

⇨ Both new voice and new data calls are accepted.

B : If $\gamma_v < \gamma_d$,

$$(\lambda_v + \lambda_d + i\mu_v + j\mu_d)P_{i,j} = \lambda_v P_{i-1,j} + \lambda_d P_{i,j-1} + (i+1)\mu_v P_{i+1,j} + j\mu_d P_{i,j+1}$$
$$\text{for } (i,j) \in \left\{ (i,j) \middle| 1 - \gamma_d < \begin{pmatrix} \gamma_v \\ \gamma_d \end{pmatrix} (i \;\; j) \le 1 - \gamma_v \right\}$$

⇨ New voice calls are accepted & new data calls are queued.

C :

$$(\lambda_d + i\mu_v + j\mu_d)P_{i,j} = \lambda_v P_{i-1,j} + \lambda_d P_{i,j-1} + (i+1)\mu_v P_{i+1,j} + j\mu_d P_{i,j+1}$$
$$\text{for } (i,j) \in \left\{ (i,j) \middle| 1 - \min\{\gamma_v, \gamma_d\} < \begin{pmatrix} \gamma_v \\ \gamma_d \end{pmatrix} (i \;\; j) \le 1 \right\}$$

⇨ New voice calls are blocked & new data calls are queued.

D :

$$\left(\lambda_d + i\mu_v + \lfloor (1-\gamma_v i)\gamma_d^{-1} \rfloor \mu_d\right)P_{i,j} = \lambda_d P_{i,j-1} + (i+1)\mu_v P_{i+1,j} + \lfloor (1-\gamma_v i)\gamma_d^{-1} \rfloor \mu_d P_{i,j+1}$$
$$\text{for } (i,j) \in \left\{ (i,j) \middle| 1 < \begin{pmatrix} \gamma_v \\ \gamma_d \end{pmatrix} (i \;\; j) \le 1 + \gamma_d (K-1) \right\}$$

⇨ New voice calls are blocked & new data calls are queued.

E :

$$\left(i\mu_v + \lfloor (1-\gamma_v i)\gamma_d^{-1} \rfloor \mu_d\right)P_{i,j} = \lambda_d P_{i,j-1} + (i+1)\mu_v P_{i+1,j}$$
$$\text{for } (i,j) \in \left\{ (i,j) \middle| 1 + \gamma_d (K-1) < \begin{pmatrix} \gamma_v \\ \gamma_d \end{pmatrix} (i \;\; j) \le 1 + \gamma_d K \right\}$$

⇨ Both new voice and new data calls are blocked.

Figure 2: Steady-state balance equations corresponding to the Voice/data CDMA system.

where

$$\Omega_{(nv,blo)} = \{(i,j)|\gamma_v i + \gamma_d j > 1 - \gamma_v\} \quad (6)$$

$\Omega_{(nv,blo)}$ is composed of the regions C, D and E in Figure 1. Similarly, a new data call will be blocked if the queue is full, and the blocking probability for data calls is given by

$$P_{b_d} = \sum_{(i,j)\in\Omega_{(nd,blo)}} P_{i,j} \quad (7)$$

where

$$\Omega_{(nd,blo)} = \{(i,j)\,|\gamma_v i + \gamma_d j > 1 + \gamma_d(K-1)\} \quad (8)$$

$\Omega_{(nd,blo)}$ corresponds to the region E in Figure 1.

3 Delay Distribution

In this section, we analyze the delay distribution of data calls based on the Markov chain model where the delay is defined as the time that a data call waits in a queue until being accepted in the system. Let us denote τ as the delay, and we separate the CDF of τ into two parts which correspond to discrete and continuous parts of the random variable τ respectively such that

$$F_d(t) = \Pr\{\tau \le t\} = F_d(0) + G(t) \quad (9)$$

where $F_d(0) = \Pr\{\tau \le 0\}$, and $G(t)$ represents the continuous part of the delay. Firstly, the discrete part $F_d(0)$ represents the case when the delay is zero, and it can be calculated as follows:

$$\begin{aligned} F_d(0) &= \Pr\{\tau \le 0\} = \Pr\{\tau = 0\} \\ &= \sum_{(i,j)\in\Omega_{(nd,acc)}} P'_{i,j} \end{aligned} \quad (10)$$

where $\Omega_{(nd,acc)}$ is the acceptance region of new data calls, which is given as

$$\Omega_{(nd,acc)} = \{(i,j)\,|\gamma_v i + \gamma_d j \le 1 - \gamma_d\} \quad (11)$$

and $P'_{i,j}$ represents the probability that there are i voice and j data calls in the system just before a new data call is admitted, and is given as

$$P'_{i,j} = \frac{P_{i,j}}{1 - P_{b_d}} \quad (12)$$

To investigate the continuous part of delay $G(t)$, let (i', j') denote the number of calls excluding the number of service-completed calls within time τ from (i,j). Consider the case that (i,j) belongs to the queueing region of new data calls just before a new data call is admitted where the queueing region of new data calls is given as

$$\Omega_{(nd,que)} = \{(i,j)\,|1 - \gamma_d < \gamma_v i + \gamma_d j \le 1 + (K-1)\gamma_d\} \quad (13)$$

In order for a new data call to be accepted within the time t according to the FCFS service discipline, (i', j') should fall into the acceptance region of new data calls within the time t. $G(t)$ is the sum of the probabilities of all cases that a state (i,j) in $\Omega_{(nd,que)}$ changes into (i', j') in $\Omega_{(nd,acc)}$ within the time, t, which can be expressed as

$$\begin{aligned} G(t) =& \\ &\textstyle\sum_{(i,j)\in\Omega_{(nd,que)}} \Pr\{(i', j') \in \Omega_{(nd,acc)} \text{within time } t | \\ &\qquad \text{the system state is } (i,j)\} \cdot P'_{i,j} \\ &= \textstyle\sum_{(i,j)\in\Omega_{(nd,que)}} \int_0^t w_{(i,j)}(\tau)d\tau \cdot P'_{i,j} \end{aligned} \quad (14)$$

where $w_{(i,j)}(\tau)$ is the delay distribution for the state (i,j), and it represents the probability of a new data call being accepted within time τ, given that the system state is (i,j) just before the call is admitted. Let us denote k as the number of service-completed voice calls during the change from (i,j) in $\Omega_{nd,que}$ to (i', j') in $\Omega_{nd,acc}$. Then, the delay distribution for the state (i,j) can be expressed as

$$w_{(i,j)}(\tau) = \sum_{k=0}^{I} w_{(i,j)_k}(\tau) \quad (15)$$

Table 1: System parameters for the numerical example; a CDMA system supporting voice and delay-tolerable data services.

Item	Value
transmission bandwidth (W)	1.25 MHz
the required information data rate for voice call (R_v)	9.6 kbps
the required information data rate for data call (R_d)	19.2 kbps
the required bit-energy to interference-spectral-density ratio for voice call (q_v)	7 dB
the required bit-energy to interference-spectral-density ratio for data call (q_d)	7 dB
average arrival rate for voice calls (λ_v)	variable
average arrival rate for data calls (λ_d)	variable
average service time for voice calls ($1/\mu_v$)	200 sec
average service time for data calls ($1/\mu_d$)	20 sec

where

$$I = \min\left(i, i - \left\lfloor \frac{1-\gamma_d(1+j)}{\gamma_v} \right\rfloor\right) \qquad (16)$$

$w_{(i,j)_k}(\tau)$ represents the delay distribution multiplied by the probability that k voice calls get service-completed, given that the system state is (i,j) just before a new data call is admitted. I is the maximum number of service-completed voice calls during the change, which happens when only voice calls are service-completed. Since the service time distribution is memoryless and the delay distribution is independent of the current arrival, $w_{(i,j)_k}(\tau)$ is the convolution of k independent, exponential random variables where k corresponds to the number of service-completed voice calls [7].

Substituting $w_{(i,j)_k}(\tau)$ into $w_{(i,j)}(\tau)$, and then successively substituting $w_{(i,j)}(\tau)$ into $G(t)$, the CDF of delay can be calculated as

$$F_d(t) = \sum_{(i,j)\in\Omega_{(nd,acc)}} P'_{i,j} + \sum_{(i,j)\in\Omega_{(nd,que)}} \int_0^t \sum_{k=0}^{I} \mathcal{L}^{-1}\left\{W_{(i,j)_k}(s)\right\} \cdot P'_{i,j}\ d\tau \qquad (17)$$

where $\mathcal{L}^{-1}$ denotes the inverse Laplace transform.

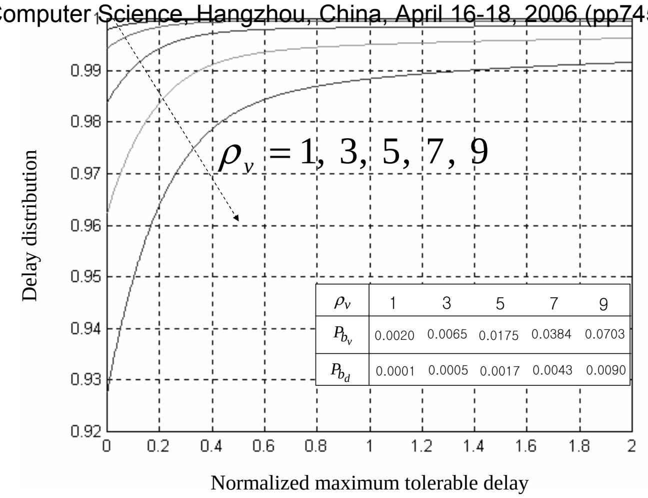

Figure 3: Delay distribution according to the voice traffic load when data traffic load is given as 5.

4 Numerical Example

For a numerical example, we consider a CDMA system supporting voice and delay-tolerant data services. Table 1 shows the considered system parameters.

Figure 3 shows the delay distribution as a function of the maximum tolerable normalized delay, $\tau_{n_{max}}$ for different offered traffic loads of voice call when the offered traffic load of data is given as 5 and the queue size is 3 where normalized delay τ_n means that the delay τ is normalized by average service time, $1/\mu_d$. The discontinuity at $\tau_{n_{max}} = 0$ comes from the fact that the probability that new data calls can be accepted without being blocked is non-zero. The delay distribution decreases for a fixed value of $\tau_{n_{max}}$ as the offered traffic load of voice increases, which means that the probability of a new data call to be accepted within a certain delay decreases. The voice and data blocking probabilities increase as the offered load of voice traffic increases. The delay distribution increases and gradually approaches 1 as the normalized maximum tolerable delay increases, which indicates that delay requirement of all data calls except the blocked calls is satisfied if the maximum tolerable delay of data calls is unlimited.

Figure 4 shows the delay distribution for different offered traffic loads of data when the offered traffic load of voice is given as 5 and the queue size is 3. We observe that the delay distribution decreases as the offered traffic load of data increases, for a fixed value of $\tau_{n_{max}}$. It is noteworthy that the probability that a new data call is accepted within the maximum tolerable delay without being blocked decreases, as the offered traffic load of voice or data increases. Comparing Figure 4 with Figure 3, we can observe that the variation

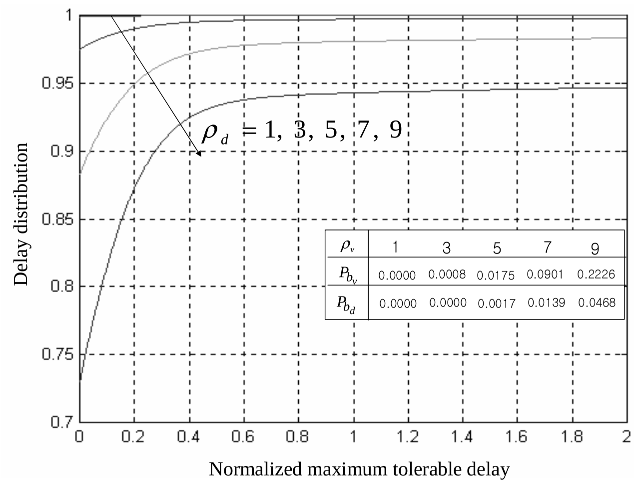

Figure 4: Delay distribution according to the data traffic load when the voice traffic load is given as 5.

of delay distribution for the offered traffic load of data is greater than that for the offered traffic load of voice. It comes from the fact that one data call requires more system resources than one voice call in the numerical example, that is $\gamma_d > \gamma_v$.

5 Conclusions

In this paper, we presented a procedure to analyze the delay distribution in a CDMA system supporting voice and delay-tolerant data services. Under the FCFS service discipline we presented the numerical procedure for the formulation of delay distribution based on a two-dimensional Markov model. The effect of offered traffic load and the buffer size on the system performance was investigated through a numerical example. It was shown that the blocking probability of each call and the delay distribution of data traffic are degraded as the offered traffic load of voice and data call increases, and the data call requiring more resource than voice call has more effect on the performance. Through the delay distribution according to the buffer size, it was shown that the improvement of data blocking probability comes from the aggravation of voice blocking probability and the degradation of delay distribution.

Acknowledgements:
This work was supported by grant No. B1220-0501-0045 from the University fundamental Research Program of the Ministry of Information & Communication in Republic of Korea

References:

[1] W. Andrews, K. Kumaran, K. Ramanan, A. Stolyar and P. Whiting, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, 39(2), Feb. 2001, pp.150-154, 2001.

[2] A. W. Berger and W. Whitt, "Extending the effective bandwidth concept to networks with priority classes," *IEEE Communications Magazine*, 36(8), Aug. 1998, pp. 78-83, 1998.

[3] Koo, I., E. Kim and K. Kim, "Erlang Capacity of Voice/Data DS-CDMA Systems with Prioritized Services," *IEICE Transactions on Communications*, vol. E84-B, no. 4, pp. 716-726, Apr. 2001.

[4] Cruz-Perez, F. A., D. Lara-Rodriguez, M. Lara. "Performance Analysis of the Fractional Channel Reservation in TDMA and CDMA Integrated Services Networks," *Proc. IEEE Vehicular Technology Conference - Spring*, pp. 1007-1011, 2001.

[5] J. Wu and J. Lin, "Performance analysis of voice/data integrated CDMA system with QoS constraints," *IEICE Transactions on Communications*, E79-B(3), March 1996, pp. 384-391, 1996.

[6] Yang, J. R., Y. Y. Choi, J. H. Ahn and K. Kim, "Capacity Plane of CDMA System for Multimedia Traffic," *Electronics Letters*, vol. 33, no. 17, pp. 1432-1433, Aug. 1997.

[7] Gross, D. and C. M. Harris, *Fundamentals of Queueing Theory*, John Wiley & Sons, Inc., 1998.