ON HANDOFF PERFORMANCE FOR AN INTEGRATED VOICE/DATA CELLULAR SYSTEM

Hailiang Wu [1], Lizhong Li [1], Bin Li[1], Li Yin [2], Imrich Chlamtac [3], Bo Li [1]

[1]Hong Kong University of Science and Technology, Clear water Bay, Kowloon, Hong Kong, China
[2]University of California at Berkeley, USA
[3]University of Texas at Dallas, USA

Abstract - One of the key performance in wireless cellular system is the handoff dropping probability. This paper studies the handoff performance for an integrated voice/data wireless cellular system. Specifically, we propose a new handoff control policy, called *Dual Trunk Reservation with Queuing (DTR-Q)*. It uses *two thresholds*, one to reserve bandwidth for voice handoff, and the other for managing the data traffic. By taking advantage of the adaptiveness of data traffic, it uses a single queue to further buffer the data request in case the requested bandwidth is not available. The unique features of the proposed *DTR-Q* scheme are: 1) it can provide the necessary service guarantee and service differentiation for voice and data traffic; 2) It adopts a *Complete Sharing* approach that can maximize the channel utilization. We propose an analytical model to calculate the key performance measures, and thoroughly investigate system performances under a variety of system parameters.

Keywords – bandwidth allocation, guarded channel policy, handoff, mobile cellular networks

1. INTRODUCTION

We have recently witnessed phenomenal growth in the development and deployment of wireless services, evident from the proliferation of cellular data services and the emerging wireless multimedia applications. This opens up a new research avenue and calls for the re-examination of some of the fundamental issues in wireless cellular networks, in particular various issues such as QoS guarantee have to be carefully examined. One of the key elements in providing QoS guarantee is an effective *bandwidth allocation policy*, which not only has to ensure that the system guarantee potentially different QoS requirements and provide the necessary service differentiation from diverse applications, but at the same time has to fully utilize the scarce wireless bandwidth available in the wireless cellular networks. Moving from a voice-centric macro-cell wireless network to a multi-services micro-cell or/and pico-cell wireless network brings several new challenges: 1) the characteristics of data and multimedia applications are typically different from that of voice, e.g., data traffic usually consumes more bandwidths that of voice, but can be adaptive, thus has less stringent service requirement; 2) The limited wireless resources have to be utilized effectively, and fairly allocated to different types of traffic with potentially different service requirements and service differentiation; 3) smaller cell potentially results in more handoffs, making it more difficult to provide the necessary QoS guarantee.

Bandwidth allocation has been extensively studied in single-service wireless cellular networks. The *Guarded Channel (GC)* schemes have been shown to be effective for providing the necessary QoS guarantee in terms of both call termination and call blocking probabilities [1-3]. One of the challenges in moving to a multi-service system is that the limited bandwidth has to be shared among multiple traffics. In [4] a *Complete Sharing (CS)* and *Complete Partition (CP)* schemes were in investigated for two types of traffic, namely *narrow-band* and *wideband.* It assumed that wideband traffic does not have handoff. Huang *et al* proposed a *movable boundary* bandwidth allocation scheme for voice and data traffic [5], in which bandwidth is divided into two sub-pools by two thresholds that can be dynamically adjusted. This facilitates the bandwidth provisioning for different QoS requirement and is adaptive to the changing traffic. The limitation is that there is no service differentiation between voice new calls and handoff. On the other hand, queuing the new or handoff calls can further reduce the handoff dropping probability as indicated in [1, 6].

In this paper, we propose a new handoff control scheme called *Dual Trunk Reservation with Queuing (DTR-Q)* scheme for a voice/data integrated cellular system. It uses *two thresholds*, one to reserve bandwidth for handoff voice calls, and the other for managing data traffics. By taking advantage of the adaptiveness of the data traffic, a single queue is used to further buffer the data requests in case the requested bandwidth is not available. The unique features of the proposed *DTR-Q* scheme are: 1) it can provide the necessary service guarantees and service differentiation for voice and data traffic; 2) It adopts the scheme of *Complete Sharing,* which can maximize the channel utilization. We propose an analytical model to calculate the key performance measures, and thoroughly investigate system performances under a variety of system parameters.

The rest of the paper is organized as follow. We describe the DTR-Q schemes in Section 2. We present the analytical models in Section 3, following by the numerical studies in Section 4. We conclude the paper in Section 5 with discussions on possible avenue for further study.

2. THE DTR-Q CONTROL POLICY

A good bandwidth allocation scheme should provide different *QoS* guarantees for different traffic types, while at the same time has to fully utilize the scarce wireless bandwidth available in the wireless cellular networks. A variety of bandwidth allocation schemes have been proposed to support multiple traffics, which can be classified as *Complete Partition (CP), Complete Sharing (CS)* or hybrid schemes, depending on how the bandwidth are allocating among diverse traffic. Compared with *CP* scheme, *CS* scheme can usually achieve much higher channel utilization [4]. This is because in *CS* scheme, users from different traffic types are allowed to share all the available channels statistically. Thus in this paper, we adopt *CS* control scheme in our *Dual Trunk Reservation with Queuing (DTR-Q)* scheme to maximize system efficiency.

In the DTR-Q scheme, total C channels in each cell are divided into *three* regions by two thresholds K_1 and K_2 (see Fig.1). When the number of channels occupied is less than the threshold K_2, then both data and voice traffic can be admitted into the system; when the number of channels occupied is over the threshold K_2, data calls will be placed into the queue if it is not full (assume the size of the finite queue is $K_{\max}$), otherwise it will be dropped; when the number of channels occupied is more than the threshold K_1, then only voice handoff calls can be allowed. The voice handoff call will be dropped only if there is no channel available. Under this basic control model, the voice handoff gets highest priority, while data receives lowest service. The reason is that the data traffic can tolerate certain degree of delay/delay-jitter, while voice cannot. On the other hand, buffering data requests can lead to a relative lower data blocking probability as shown in [6].

3. THE ANALYTICAL MODEL

We consider a homogeneous wireless network where all cells have the same number of channels and experience the same new and handoff call arrival rates. In each cell, the arrivals of new voice calls, new data calls, handoff voice calls and handoff data calls are Poisson distributed with arrival rate λ_{vn}, λ_{dn}, λ_{vh} and λ_{dh} respectively. Thus the total voice call arrival rate and data call arrival rate are $\lambda_v = \lambda_{vn} + \lambda_{vh}$ and $\lambda_d = \lambda_{dn} + \lambda_{dh}$ respectively. Since data can usually tolerate some degree of service degradation, new data calls and handoff data calls are not distinguished. Call duration times or call holding times of voice and data are exponentially distributed with the average call duration time $1/\mu_{vr}$ and $1/\mu_{dr}$. In addition, the cell residence time for voice and data calls is exponentially distributed with mean $1/\mu_{vh}$ and $1/\mu_{dh}$ respectively. Thus, the channel occupancy times for voice and data calls are exponentially distributed with mean $1/\mu_v = 1/(\mu_{vr} + \mu_{vh})$ and $1/\mu_d = 1/(\mu_{dr} + \mu_{dh})$ respectively. This set of assumptions have been found to be reasonable as long as the number of mobiles in a cell is much greater than the number of channels and have been widely used in literature [7, 8, 2, 3]. The exponential channel holding time assumption is shown to be valid for a wide range of system under certain conditions [9], which allows us to conduct the mean value analysis [10]. We further assume that both voice and data occupy one unit of bandwidth.

Scheme DTR-Q can be modelled as a three-dimensional Markov chain. Let $P_{i,j,k}$ be the steady state probability that there are *i* voice calls, *j* data calls in the system, and *k* data calls in the data buffer. The steady-state balance-equations of DTR-Q scheme are shown as below. Due to the length limitation, detail descriptions of the analytical models are given in [11].

Case 1: If $i + j = 0$, then

$$(\lambda_v + \lambda_d)P_{0,0,0} = \mu_v P_{1,0,0} + \mu_d P_{0,1,0} \tag{1}$$

Case 2: *If* $0 < i + j < K_2$, then

$$\begin{aligned}(\lambda_v + \lambda_d + i\mu_v + j\mu_d)P_{i,j,k} &= (i+1)\mu_v P_{i+1,j,0} \\ &+ (j+1)\mu_d P_{i,j+1,0} + \lambda_v P_{i-1,j,0} + \lambda_d P_{i,j-1,0}\end{aligned} \tag{2}$$

Case 3: *If* $i + j = K_2$ *and* $k = 0$, then

$$\begin{aligned}(\lambda_v + \lambda_d + i\mu_v + j\mu_d)P_{i,j,0} &= (i+1)\mu_v P_{i+1,j-1,1} \\ &+ (j+1)\mu_d P_{i,j+1,0} + (i+1)\mu_v P_{i+1,j,0} + j\mu_d P_{i,j,1} \\ &+ \lambda_v P_{i-1,j,0} + \lambda_d P_{i,j-1,0}\end{aligned} \tag{3}$$

Case 4: *If* $i + j = K_2$ *and* $k > 0$, then

$$\begin{aligned}(\lambda_v + \lambda_d + i\mu_v + j\mu_d)P_{i,j,k} &= (i+1)\mu_v P_{i+1,j-1,k+1} \\ &+ (j+1)\mu_d P_{i,j+1,k} + (i+1)\mu_v P_{i+1,j,k} + j\mu_d P_{i,j,k+1} \\ &+ \lambda_d P_{i,j-1,k}\end{aligned} \tag{4}$$

Case 5: *If* $K_2 < i + j < K_1$, then

$$\begin{aligned}(\lambda_v + \lambda_d + i\mu_v + j\mu_d)P_{i,j,k} &= (i+1)\mu_v P_{i+1,j1,k} \\ &+ (j+1)\mu_d P_{i,j+1,k} + \lambda_v P_{i-1,j,k} + \lambda_d P_{i,j,k-1}\end{aligned} \tag{5}$$

Case 6: *if* $i + j = K_1$, then

$$\begin{aligned}(\lambda_v + \lambda_d + i\mu_v + j\mu_d)P_{i,j,k} &= (i+1)\mu_v P_{i+1,j,k} \\ &+ (j+1)\mu_d P_{i,j+1,k} + \lambda_v P_{i-1,j,k} + \lambda_d P_{i,j,k-1}\end{aligned} \tag{6}$$

Case 7: *if* $K_1 < i + j < C$, then

$$\begin{aligned}(\lambda_{vh} + \lambda_d + i\mu_v + j\mu_d)P_{i,j,k} &= (i+1)\mu_v P_{i+1,j,k} \\ &+ (j+1)\mu_d P_{i,j+1,k} + \lambda_{vh} P_{i-1,j,k} + \lambda_d P_{i,j,k-1}\end{aligned} \tag{7}$$

Case 8: *if* $i + j = C$, then

$$(\lambda_d + i\mu_v + j\mu_d)P_{i,j,k} = \lambda_{vh}P_{i-1,j,k} + \lambda_d P_{i,j,k-1} \quad (8)$$

After obtaining the steady-state transition equations for each state in the Markov chain, we can solve the linear equation together with the normalizing condition (see Eq. (9)) by using LU factorization.

$$\sum_{i+j\geq 0}^{C}\sum_{k\geq 0}^{K_{\max}} P_{i,j,k} = 1 \quad (9)$$

After obtaining all the steady state probabilities $P_{i,j,k}$, the voice call blocking probability P_{vb}, the handoff voice call dropping probability P_{vd}, the data call blocking probability P_{db} and the average queue length $\overline{Q}$ can be derived as below:

$$P_{vb} = \sum_{i+j\geq K_1}^{C}\sum_{k\geq 0}^{K_{\max}} P_{i,j,k} \quad (10)$$

$$P_{vd} = \sum_{i+j=C}\sum_{k\geq 0}^{K_{\max}} P_{i,j,k} \quad (11)$$

$$P_{db} = \sum_{i+j\geq K_2}^{C}\sum_{k\geq 0}^{K_{\max}} P_{i,j,k} \quad (12)$$

$$\overline{Q} = \sum_{i+j=K_2}^{C}\sum_{k=K_{\max}} kP_{i,j,k} \quad (13)$$

4. NUMERICAL RESULTS

In this section, we present numerical results in three cases: 1) performance under different voice call arrival rate λ_v; 2) performance under different data call arrival rate λ_d; 3) performance under different K_2. In order to validate the accuracy of the analysis, we also develop an event-driven simulation. We consider the following system configuration. The total channel number C of each cell is set to be 30, the voice call arrival rate λ_v can range from 0.04 to 0.29, the data call arrival rate λ_d can rage from 0.03 to 0.11. While other system parameters are fixed as: $\mu_{vr} = 0.005$, $\mu_{vh} = 0.01$, $\mu_{dr} = 0.001$, $\mu_{dh} = 0.0001$.

4.1 Performance under different λ_v

Figs. 2 to 4 are the results under $K_1 = 25$, $K_2 = 25$ and $\lambda_d = 0.03$. Figs. 2 and 3 show that voice call dropping and blocking probability increase with the increase of data buffer size. This is because that under the buffer case, we place the data calls that cannot be serviced into the data buffer instead of simply blocking them. The data calls queuing in the data buffer will enter the system for service immediately if there is a channel becoming available. This means potentially more channels are occupied by the data calls. As the buffer size increases, more data calls can be queued in the data buffer, and competes for channels with the voice call, thus the voice handoff dropping and blocking probability will be increased.

Observing from Fig. 4, it is clear that the data call blocking probability decreases with the increase of data buffer size. Under no buffer case, the data call will be blocked if there is no free channel available for data traffic, but in buffer case, such kinds of data calls can be buffered to wait for service, this result in the decrease of the data blocking probability. As the buffer size increases, more data calls can be queued in the buffer and less data calls will be blocked.

4.2 Performance under different λ_d

Figs. 5 and 6 presents the results obtained under $\lambda_{vn} = \lambda_{vh} = 0.09$ and $\mu_d = 0.006$. Fig. 6 shows that the voice dropping probabilities may increase with the increase of λ_d and data buffer size. The increase of data buffer size means more data calls can be buffered, and obtain service if there are channels available. The results obtained also reveal that the voice call dropping will not be increase after the buffer size increases to a certain value, such as 10 in this case. This is caused by the traffic control policy itself, where only when the channels occupied are less than K_2, data call can be accepted. Therefore not every data call queued in the buffer can obtain service. This is crucial for maintaining the performance of handoff dropping probabilities. Similar results can be obtained for voice blocking probability under these conditions.

As expected, data blocking probability will decrease with the increasing of buffer size and λ_d, this is shown in Fig. 6.

4.3 Performance under different K_2

Figs. 7 and 8 indicate the system performance under different K_2. In the *DTR-Q* scheme, K_2 is used to control the data traffic flow into the network. Fig. 7 shows that the voice dropping probabilities are increased with the increasing of K_2 and data buffer size. In case that K_2 is small when comparing with the channel capacity, such as $K_2 = 8$, the increment of voice dropping probability is not obvious with data buffer increase. This is because that more data calls will be blocked from entering the system if we set K_2 to be small. But in case that K_2 is large, the voice dropping probability will be increased significantly with the increasing data buffer size. This is caused by the fact that larger K_2 means more data traffic can be accepted by the network system, the chance for data calls buffered in the data buffer to obtain service increase. Similar results can be obtained for voice blocking probability under these conditions.

Fig. 8 shows that the data call blocking probability is decreased with the increasing of data buffer size, the increase of data buffer size apparently benefits the data call and results in the decrease of data call blocking probability.

The increase of K_2 causes more channels to be occupied by data and as a result, the data blocking probability is decreased.

5. CONCLUSION

In this paper, we propose a new handoff control policy for voice/data integrated wireless cellular networks, namely *Dual Trunk Reservation with Queuing (*DTR-Q*)*. The performances of DTR-Q are analyzed with three-dimensional Markov process, and an event-driven simulation is developed to validate the accuracy of our analyzing. From the results obtained from both analysis and simulation, we can conclude that: 1) DTR-Q scheme can guarantee different *QoS* requirements from different traffic types; 2) buffering data call requests may benefit data users while impact negatively the performance of voice users.

There are a number of issues that will be addressed in our future research: 1) the bandwidth allocation schemes discussed in this paper are static. However as shown in [3, 8], dynamic schemes can always achieve better performance; 2) the handoff rates for both voice and data following convention have been assumed given, as there lacks adequate model that can derive handoff rates based on other parameters such as new call arrival rates and mobility, similar to those in [7]. Finally, we will be considering more realistic data arrival using TCP traffic, and exploring the self-similar behavior of the data arrival observed in [12].

REFERENCES

[1] R. Guerin, "Queuing-blocking systems with two arrival streams and guarded channels," *IEEE Transactions on Communications*, 36(2), February 1988.

[2] R. Ramjee, R. Nagarajan and D. Towsley, "On optimal call admission control in cellular networks," *Proceedings of the IEEE Infocom*, March 1996, pp.43-50.

[3] S. Wu, K. Y. M. Wong and B. Li, "A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks," IEEE/ACM Transactions on Networking, 10(2), April 2002.

[4] B. Epstein and M. Schwartz, "Reservation strategies for multimedia traffic in a wireless environment," *Proceedings of the IEEE VTC,* July 1995, pp.165-169.

[5] Y.-R. Haung, Y.-B. Lin and J. M. Ho, "Performance analysis for voice/data integration on a finite mobile systems," *IEEE Transactions on Vehicular Technology*, 49(2), February 2000.

[6] C. -J. Chang, T.-T. Su and Y.-Y. Chiang, "Analysis of a cutoff priority cellular radio system with finite queuing and reneging/dropping," *IEEE/ACM Transactions on Networking,* 2(2), pp. 166-175, April 1994.

[7] Y.-B. Lin, Performance modeling for mobile telephone networks, IEEE Networks, Nov/Dec 1997.

[8] X. Luo, B. Li, I. Thng, Y.-B. Lin and I. Chlamtac, "A measurement-based pre-assignment scheme with connection-level QoS support for multiservice mobile networks," *IEEE Transactions on Wireless Communications*, 1(3), July 2002.

[9] Y. Fang, Y.-B. Lin and I. Chlamtac, "Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks," *IEEE Transactions on Computer*, 47(6), June 1998.

[10] E. Lazowska et al, Quantitative system performance: computer system analysis using queuing network models, Prentice-hall, 1984

[11] H.-L. Wu, L.-Z. Li, B. Li and B. Li, "Performance analysis of bandwidth allocations for multi-service wireless cellular networks: data buffer case," *Technical Report*, CS Dept. HKUST, March 2002.

[12] M. Cheng and L.-F. Chang, "Wireless dynamic channel assignment performance under packet data traffic," *IEEE Journal on Selected Areas in Communications*, 17(7), July 1999.

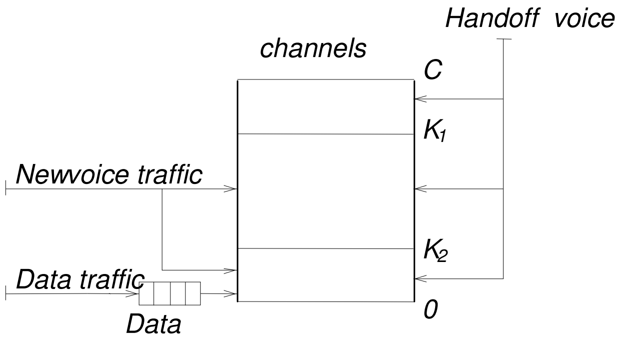

Fig. 1. The *DTR-Q* control scheme

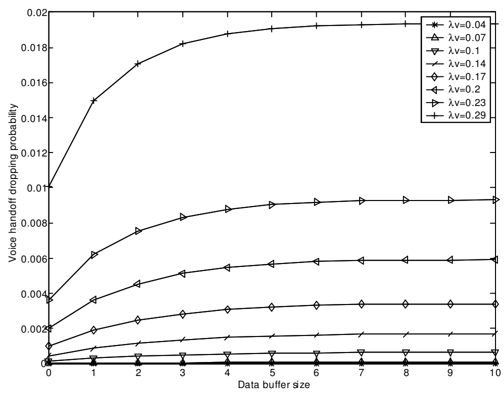

Fig. 2. The voice dropping probability vs. λ_v

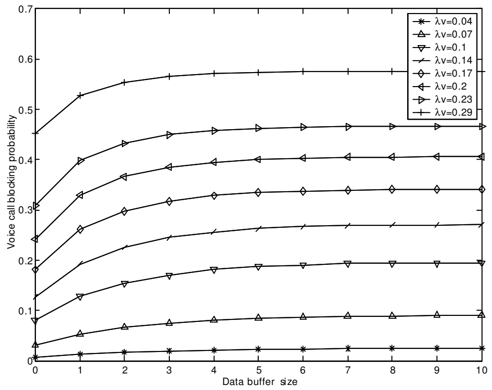

Fig. 3. The voice blocking probability vs. λ_v

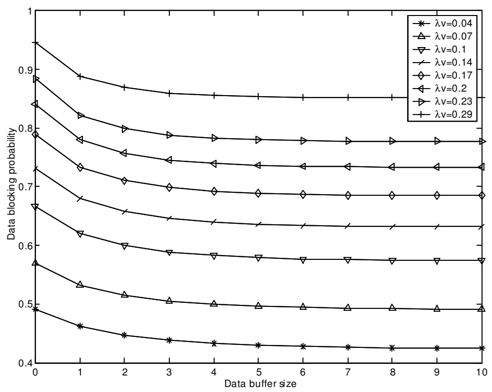

Fig. 4. The data blocking probability vs. λ_v

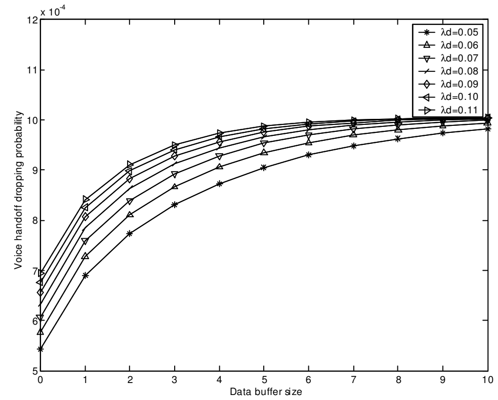

Fig. 5. The voice dropping probability vs. λ_d

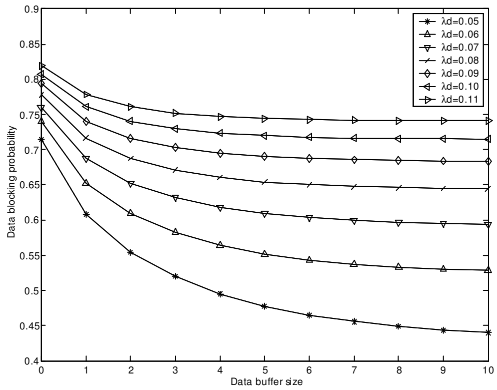

Fig. 6. The data blocking probability vs. λ_d

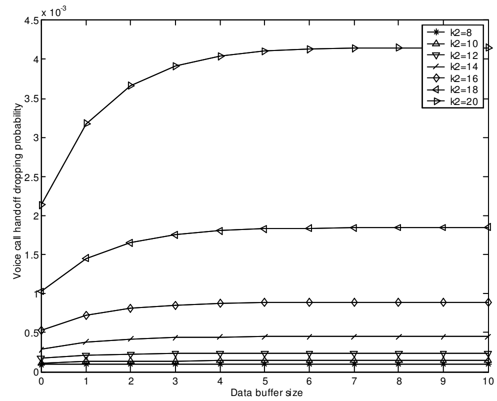

Fig. 7. The voice dropping probability vs. K_2

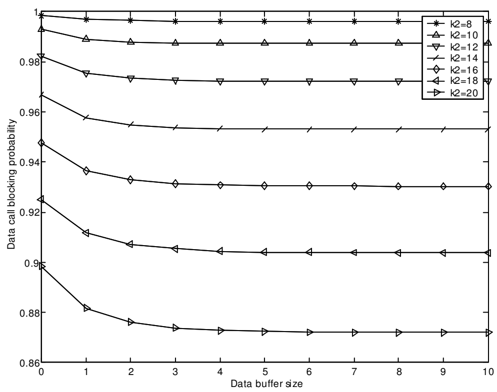

Fig. 8. The data blocking probability vs. K_2