A New Quality of Service Measure for Cellular Radio Systems

Tak-Shing Peter Yum, *Senior Member, IEEE,* and Kwan Lawrence Yeung, *Member, IEEE*

***Abstract*—A new "quality of service" (QOS) measure called rms excessive blocking, denoted as $EB(2)$, is proposed as a composite measure of the blocking rate and the cell-to-cell variation of blocking rates. $EB(2)$ belongs to a class of measures $EB(p), p \geq 1$, which is unbiased, makes reference to the QOS requirement, is one-sided, and takes on the form of a penalty function. The effectiveness of this rms excessive blocking measure is compared to the conventional mean plus a standard deviation type of measure and evaluated on two efficient dynamic channel-assignment strategies, namely, borrowing with directional channel-locking strategy and compact-pattern (CP)-based dynamic channel assignment (DCA).**

***Index Terms*—Dynamic channel assignment, quality of service, service deviation.**

I. Introduction

THE TERM "quality of service" (QOS) in the communications context refers to certain characteristics of network services as observed by transport users. These characteristics describe aspects of services attributable to the network provider. The future B-ISDN will require a unified support of a broad spectrum of QOS demands by the diverse traffic types. The QOS design, therefore, has attracted much attention [1]–[5]. Research in this area is mainly concerned with how to define a set of QOS parameters and monitor and control the communication system under consideration to guarantee the QOS.

Like conventional telephone networks, the generally accepted QOS parameter in cellular radio systems is the call-blocking probability. Owing to the need of switching to a new frequency channel when a mobile unit crosses a cell boundary, additional blocking due to this handoff operation will be introduced. The probability that a handoff call will be dropped is known as the dropped-call probability. The cell-to-cell variation of blocking probabilities (also known as the service deviation) is another important attribute of QOS. It has not received much attention in the literature. Service deviation is caused by the nonuniform traffic distribution among the cells and can usually be reduced by a nonuniform assignment of nominal channels that matches the nonuniform traffic distribution. Further reduction can be obtained by the use of dynamic channel allocation.

Many channel-assignment strategies were proposed in the literature [6]–[13]. These strategies aim at maximizing the overall traffic-carrying capacity of a system, while keeping the overall average blocking probability of the system at a prescribed level. The resulting system, however, may have some cells failing to satisfy the QOS requirement, while others have unnecessarily low blocking probabilities. A well-designed cellular mobile system should have an acceptable overall average blocking probability as well as a very small service deviation among cells.

In this paper, we start by stating the requirements for a satisfactory QOS measure. Then, we define a class of measures $EB(p), p \geq 1$. Next, we specialize our investigation on $EB(2)$, which can be interpreted as the "rms excessive blocking" and test its effectiveness by using it to compare the relative merits of two well-performing dynamic channel-assignment strategies.

II. QOS Measures

The overall average blocking probability is the most intuitive QOS measure for cellular mobile radio systems. This measure, however, cannot reflect the cell-to-cell variation of blocking probabilities. To design a measure that can reflect this variation, let us start from the requirements for a good measure of QOS. Let the requirement on the blocking probability B of the customers be $B \leq B^*$. (B^* is usually chosen as 0.01 for telephone networks and 0.02 for cellular networks.) We stipulate that a good QOS measure must be unbiased, make reference to B^*, and be one-sided. In addition, we let the measure take on the form of a penalty function, i.e., the smaller the better (the overall average blocking probability takes on such form).

1) Unbiased: What we mean by an unbiased measure is that the measure should reflect the QOS as seen by a randomly picked customer in the system. As different cells have different traffic rates, the cells with higher traffic rates should therefore be weighted heavier in the computation of service deviation.
2) Make reference to B^*: The QOS requirement is that the blocking probability of the customers be no larger than B^*. Therefore, an acceptable QOS measure must make reference to B^*.
3) One-sided: Blocking probabilities of individual cells often deviate from B^*. The penalty should be only on those that deviate above B^*. Note that those that deviate above B^* are the excess blockings and should be minimized, whereas those that deviate below B^* should not enter into the penalty function.

Let N be the total number of cells in the system, and let λ_i be the call-arrival rate and B_i the blocking probability at cell

Manuscript received December 15, 1995; revised April 25, 1996.

T-S. P. Yum is with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: yum@ie.cuhk.hk).

K. L. Yeung is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: eekyeung@cityu.edu.hk).

Publisher Item Identifier S 0018-9545(97)04623-9.

i. Let $\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_N$. The overall average blocking probability of the system $\bar{B}$ is defined as

$$\bar{B} = \Lambda^{-1} \sum_{i=1}^{N} B_i \lambda_i. \tag{1}$$

Twenty years ago, Anderson [7] defined a measure of service deviation A as

$$A = \sqrt{\frac{\sum_{i=1}^{N}(B_i - \bar{B})^2}{N-1}}. \tag{2}$$

This is a measure of the cell-to-cell variation of blocking probability, but this measure is easily seen to violate the requirements stated above. In addition, it has to be used together with $\bar{B}$. Note that this measure differs from the conventional standard deviation measure in that $\bar{B}$ is a weighted average of B_i's.

We now define a spectrum of QOS measures $EB(p)$, $p \geq 1$ and call them an excessive blocking of order p. They make reference to B^*, are one-sided, and weighted by the traffic rates. Let I be the set of cells with blocking probabilities exceeding B^*. Then, $EB(p)$ is defined as

$$EB(p) = \left[\sum_{i \in I} (B_i - B^*)^p \frac{\lambda_i}{\Lambda} \right]^{1/p}, \qquad p = 1, 2, \ldots. \tag{3}$$

Each $EB(p)$ is a composite measure that takes both the overall average blocking probability and service deviation among cells into consideration. Each is equal to zero when all cells satisfy the QOS requirement as only the B_i's that are larger than B^* contribute to the penalty function. In particular, $EB(1)$ is easily identified as the overall average excessive blocking or mean excessive blocking. It penalizes the mean deviation from B^*. $EB(2)$ is called the rms excessive blocking. It has the property of penalizing larger deviations of B_i from B^* as compared to $EB(1)$. As we increase p, we are penalizing more and more on such deviations. In the limit $p \to \infty$, only the maximum deviation term is penalized. We believe the exact choice of p would depend on applications and can be decided by the system operators. On the other hand, the degree of customer dissatisfaction, in general, grows more than linearly with the increase of blocking probability. We choose $p = 2$ as a typical measure with a well-defined intuitive meaning of being the rms excessive blocking and focus our example and case studies on the study of $EB(2)$.

We have stipulated earlier that a good QOS measure should satisfy three criteria: be unbiased, make reference to B^*, and be one-sided. Thus, comparing $EB(p)$ and the typical conventional (mean and Anderson) measure, we can see their differences in the following three aspects.

1) Consider two cells in a network, where one has call-arrival rate 2λ and the other λ. Let the blocking probabilities of both cells be the same [say, with the use of nonuniform allocation of nominal channels or dynamic channel assignment (DCA)]. Both cells therefore give the same contribution to the computation of Anderson's measure, but it is clear that the cell with arrival rate 2λ should contribute twice as much in the averaging process when compared to the cell with arrival rate λ. Having the unbiased property, $EB(p)$ does exactly that.
2) It is generally difficult to compare the relative performance of two systems without specifying the performance requirement. With the use of $EB(p)$, different relative performance is obtained with different choices of B^* (the example in the next section will elaborate on this). The (mean and Anderson) measure cannot reflect the change of performance relative to the change of performance objective.
3) The distribution of the blocking probabilities in a cellular system is more often asymmetrically distributed around the blocking requirement B^*. Under such conditions, the Anderson's measure cannot provide an accurate picture of the service deviation. By requiring the deviation penalty be one-sided, $EB(p)$ captures only those that violate the blocking requirement.

III. An Example

In this Section, we use an example to illustrate the properties of $EB(2)$ and compare it to a typical conventional measure S, where $S = \bar{B} + 2A$. Consider a 30-cell cellular system. Assume the fixed channel assignment (FCA) is used and ten channels are assigned to each cell. Let the call-holding time be exponentially distributed with a mean of 3 min. Let

$$\begin{aligned} T_1 = [&20, 40, 60, 80, 140, 10, 20, 40, 80, 5,\\ &80, 20, 140, 20, 10, 20, 10, 60, 40, 80,\\ &120, 40, 20, 80, 100, 60, 40, 20, 60, 40]\\ T_2 = [&110, 110, 130, 110, 110, 110, 110, 110, 110, 130,\\ &110, 110, 110, 130, 110, 110, 110, 130, 110, 130,\\ &130, 110, 110, 130, 110, 110, 130, 110, 110, 130] \end{aligned}$$

be two traffic-rate distributions in the 30 cells (in calls/h).

Using the Erlang-B formula, the blocking probability at each cell can be found. With that, $\bar{B}_1$ and $\bar{B}_2$, the overall average blocking probabilities of the system under traffic distributions T_1 and T_2, are found to be 0.020 and 0.040, respectively, from (1). From (2), we have $A_1 = 0.024$ and $A_2 = 0.014$. Then, $S_1 = \bar{B}_1 + 2A_1 \approx 0.068$ and $S_2 = \bar{B}_2 + 2A_2 \approx 0.068$. Since $S_1 = S_2$, one can conclude that both distributions give comparable performance. This, however, is not true as revealed by $EB(2)$.

Fig. 1 shows $EB(2)$ versus two traffic distributions with the same S. We can see that for $B^* < 0.016$, traffic distribution T_1 gives a smaller rms excessive blocking than T_2. For $B^* > 0.016$, the reverse is true. Distribution T_1 reaches $EB(2) = 0$ at $B^* = 0.08$, while for distribution T_2, it is at $B^* = 0.06$. This example shows that $EB(2)$ has the property of discouraging overengineering. This is because the performance of the "overengineered cells" cannot be used to compensate the performance of the "underengineered cells" with the use of $EB(2)$.

IV. Case Studies

The increasing demand for mobile telephone services calls for a new generation of technology to meet future demand. To cope with this challenge, many new modulation and multiple-access techniques have been developed, but for a given spectrum and specific technology used, the traffic-

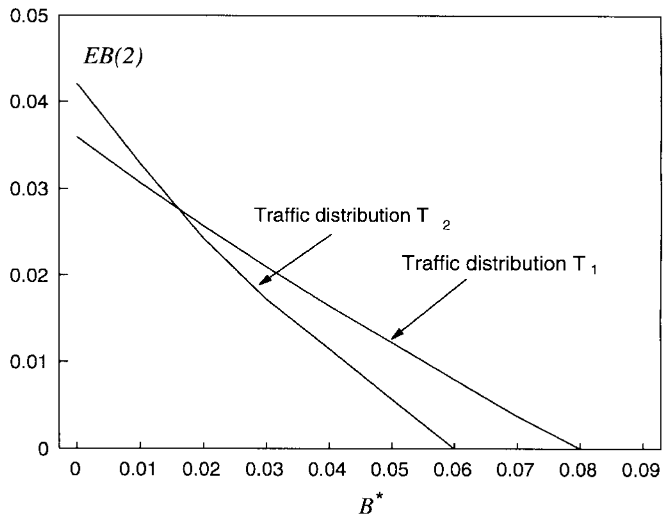

Fig. 1. $EB(2)$ for two traffic distributions with the same S.

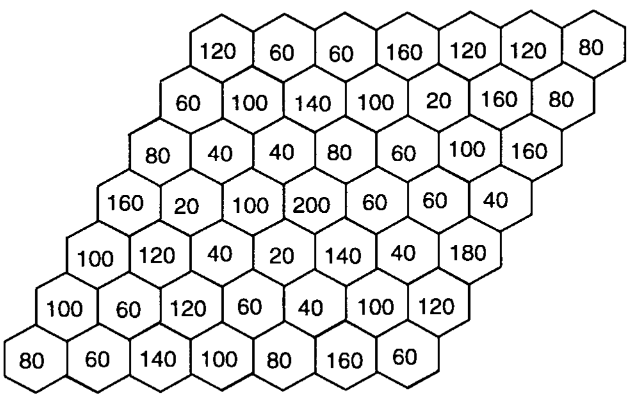

Fig. 2. Cellular system with nonuniform traffic distribution.

carrying capacity of a cellular system depends on how the frequency channels are managed. Conventionally, FCA is used, where each cell is assigned with a fixed set of nominal channels. If a new call finds that no free nominal channel is available, the call is blocked. In contrast to FCA, there is no definite relationship between the cells of the system and the channels that are used in them in DCA [14]–[16]. Any cell can use any channel as long as the interference constraints are satisfied. Channels are assigned for use in cells only for the duration of the call. After the call is over, the channels are returned to a central pool.

Simulation studies [13], [17] of some DCA algorithms showed that they can all improve network performance at low to normal traffic loads, but the heavy load network performance is no better than that of the FCA. This undesirable behavior occurs because cells that are assigned with the same channels are, on the average, spaced apart larger than the minimum cochannel reuse distance. As a result, under heavy load conditions, the throughput of DCA's becomes lower than that of FCA. To improve this situation, channel reassignment can be used to pack the cochannel cells. Channel reassignment means switching calls in progress to other channels wherever possible to reduce the distance between cochannel cells. In [14], an idealized DCA called "maximum packing" (MP) was proposed. With the use of MP, a new call will be blocked only if there is no possible reallocation of channels to allow the call to be carried. The MP strategy requires system-wide information, and the complexity of searching all possible reallocations renders this strategy impractical for implementation.

In this Section, we evaluate the effectiveness of $EB(2)$ on two very efficient, but different DCA strategies. They are: 1) borrowing with directional channel locking (BDCL) [8], [18] strategy and 2) compact-pattern (CP)-based DCA [10] strategy.

BDCL is a channel-borrowing-based DCA strategy [11]. In channel-borrowing-based strategies, channels are allocated to each cell on a nominal basis. When a call request arrives and finds all nominal channels busy, a channel is borrowed from a neighboring cell, provided that the borrowing will not violate the cochannel-interference constraints. By incorporating channel reassignment (or intracell handoff), the channel-borrowing-based DCA strategies can give a superior performance over the FCA, even at overload conditions by keeping the channel-reuse distance between cochannel cells at a minimum. BDCL is shown to give the lowest blocking probabilities among such strategies. Recently, a new DCA based on Hopfield neural networks was proposed [13]. From the numerical examples in [13], we again can see that BDCL consistently outperforms the other DCA's under the nonuniform traffic distributions. To further enhance the performance of the BDCL strategy, nonuniform nominal channel allocations [9] can be used for allocating nominal channels to cells for systems with nonuniform traffic distributions. In this study, the hybrid allocation with composite backtracking [9] is chosen.

The CP-based DCA strategy [10] takes a totally different approach to dynamic channel allocation. There is no definite relationship between channels and cells. Thus, no nominal channel is defined. All channels are kept in a central pool and are assigned to a cell on a call-by-call basis. Channels in specific CP's are used whenever possible. A CP of a cellular network is defined as the channel-allocation pattern with the minimum average distance between cochannel cells. CP-based DCA consists of two phases: channel allocation and channel packing. Channel allocation is used to assign an optimal idle channel to a new call. Channel packing is responsible for the restoration of the CP's and is performed only when a CP channel is released.

The QOS of the above two DCA strategies, together with the fixed channel-assignment (FCA) strategy, are studied on a 49-cell hypothetical cellular system with a seven-cell channel-reuse pattern (as shown in Fig. 2). Let there be a total of 70 channels in the system. Let the arrival of calls follow a Poisson process and the call-holding time be exponentially distributed with a mean of 3 min. Let $B^* = 0.02$.

First, we consider a uniform traffic distribution with a base load of 100 calls/h/cell. Each cell has ten nominal channels. Figs. 3 and 4 show the overall average blocking probability $\bar{B}$ and A versus the percentage increase of the base traffic load. We can see from Fig. 4 that the CP-based strategy has the highest value of A, and FCA has $A = 0$. The BDCL strategy has slightly larger $\bar{B}$, but a lower A value when compared to the CP-based strategy. These two figures give no indication as

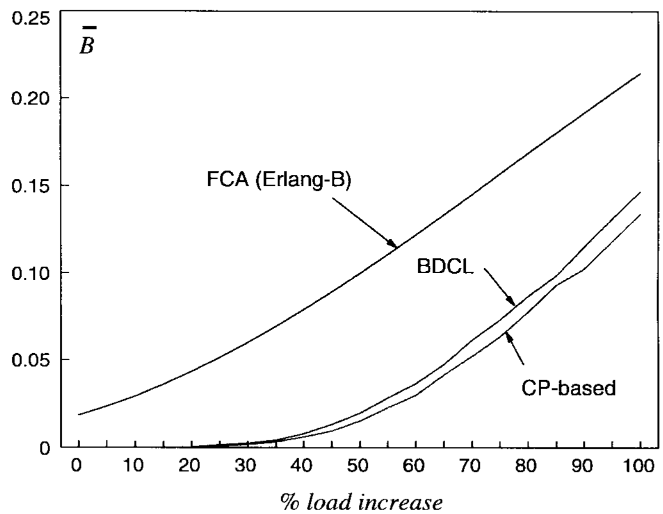

Fig. 3. Blocking performance under uniform traffic distribution.

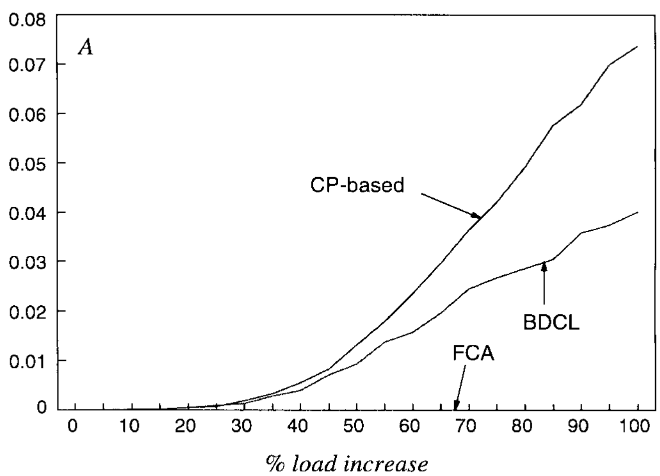

Fig. 4. Service deviations A under uniform traffic distribution.

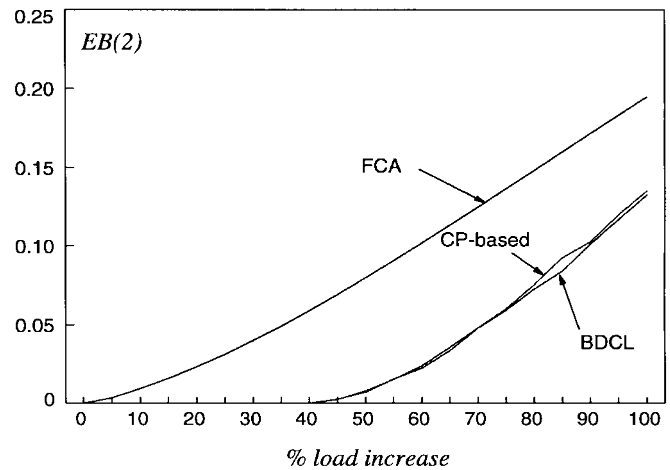

Fig. 5. $EB(2)$ under uniform traffic distribution.

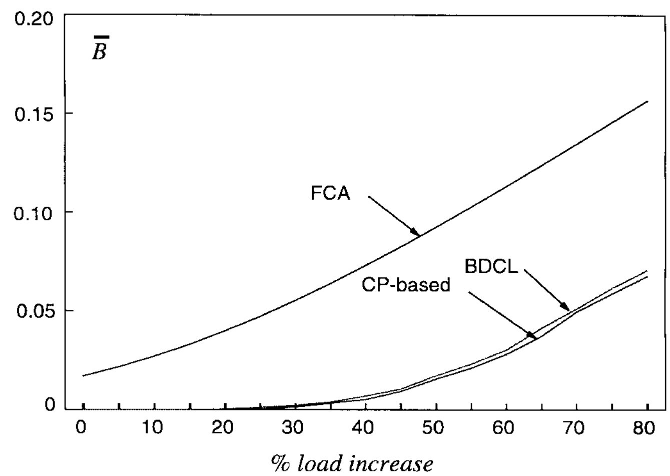

Fig. 6. Blocking performance under nonuniform traffic distribution.

to which strategy is better. Let us now take a look at $EB(2)$ (shown in Fig. 5). It shows that CP-based DCA and BDCL give similar QOS, while that of FCA is significantly inferior. It further shows that no cell will violate the QOS requirement at 40% overload for the two DCA's, while for FCA, any traffic that is higher than the base load causes some cells to have blocking probability larger than B^*. At $EB(2) = 0.01$, a 53% overload can be tolerated.

Next, let us consider a nonuniform traffic distribution with the base traffic load in calls per h shown inside each cell in Fig. 2. Figs. 6 and 7 show $\bar{B}$ and $EB(2)$ versus the percentage increase of traffic load. Note that the hybrid allocation with composite backtracking [9] has been used for allocating nominal channels to cells. As expected, FCA gives the highest $\bar{B}$ as well as $EB(2)$ value. It is interesting to note the resemblance of the curves in these two figures. It is also interesting to see that the two very different approaches to DCA give virtually identical QOS performance in terms of $EB(2)$. At $EB(2) = 0.01$, 45% overload can be tolerated. From (3)

$$EB(\infty) = \begin{cases} 0, & \text{if } B_i < B^* \text{ for all } i \\ B_{\max} - B^*, & \text{otherwise} \end{cases}$$

where $B_{\max} = \max\{B_i\}$. Fig. 8 shows that the CP-based strategy gives a higher $EB(\infty)$ than that of BDCL. This is because in BDCL strategy, each cell owns a set of nominal channels, and these channels are always returned after each lending. This limits the channel-starving situations. As a result, the blocking probabilities of the cells have a narrower spread. CP-based strategy, however, does not have nominal channels assigned to individual cells. So, the probability distribution has a wider spread, causing $EB(\infty)$ to be larger.

V. Conclusion

A new QOS measure called rms excessive blocking, denoted as $EB(2)$, was proposed as a composite measure of the blocking rate and the cell-to-cell variation of blocking rates. $EB(2)$ belongs to a class of measures $EB(p)$, $p \geq 1$, which is unbiased, makes reference to the QOS requirement, is one-sided, and takes on the form of a penalty function. The effectiveness of this rms excessive blocking measure was compared to the conventional mean plus the standard-deviation-type of measure and was evaluated on two dynamic channel-assignment strategies: BDCL strategy and CP-based DCA.

The rms excessive blocking measure can be exploited for improving the performance of the DCA strategies. For

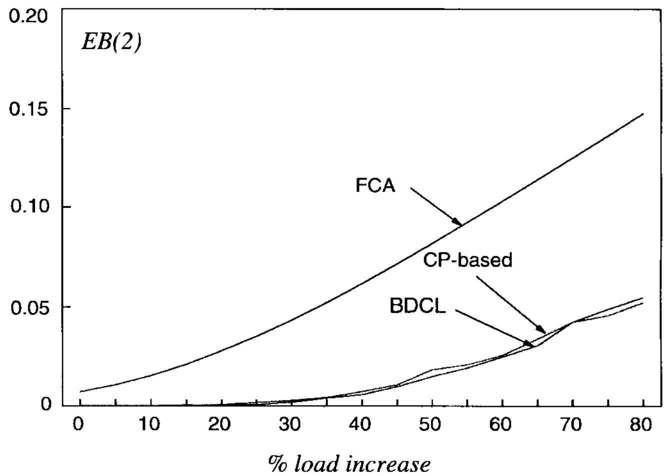

Fig. 7. $EB(2)$ under nonuniform traffic distribution.

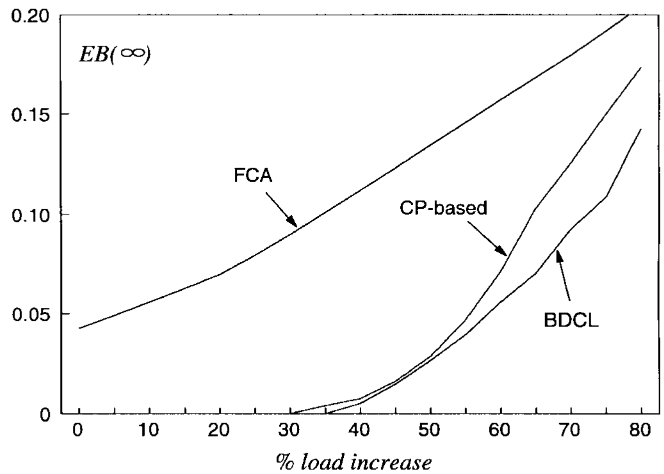

Fig. 8. $EB(\infty)$ under nonuniform traffic distribution.

example, one simple method for improving the BDCL strategy is to limit the number of channels that a hotspot cell can lend to its neighboring cells. To determine the optimal limit as well as the extent the performance can be improved needs more detailed studies. Future work in this area should take factors such as handoff failure probability, channel quality, etc. into account.

ACKNOWLEDGMENT

The authors would like to thank Dr. R. Yeung, Dr. H. S. Wang, and the reviewers of this paper for their many helpful suggestions.

REFERENCES

[1] "Terms and Definitions Related to Quality of Service and Network Performance Including Dependability," ITU-T Recommendations E.800, Aug. 1994.
[2] A. Campbell, G. Coulson, F. Garcia, D. Hutchison, and H. Leopold, "Integrated quality of service for multimedia communications," in *Proc. IEEE Infocom'93*, pp. 732–739.
[3] J. Kurose, "Open issues and challenges in providing quality of service guarantees in high-speed networks," *Comput. Commun. Rev.*, vol. 23, no. 1, 1993.
[4] L. Georgiadis, R. Guerin, V. Peris, and K. N. Sivarajan, "Efficient network QOS provisioning based on per node traffic shaping," in *Proc. IEEE INFOCOM'96*, pp. 102–110.
[5] J. Jung and D. Seret, "Translation of QOS parameters into ATM performance parameters in B-ISDN," in *Proc. IEEE Infocom'93*, pp. 748–755.
[6] V. H. MacDonald, "Advanced mobile phone service: The cellular concept," *Bell Syst. Tech. J.*, vol. 58, pp. 15–41, Jan. 1979.
[7] L. G. Anderson, "A simulation study of some dynamic channel assignment algorithms in a high capacity mobile telecommunications system," *IEEE Trans. Veh. Technol.*, vol. VT-22, pp. 210–217, Nov. 1973.
[8] M. Zhang and T. S. Yum, "Comparisons of channel assignment strategies in cellular mobile telephone systems," *IEEE Trans. Veh. Technol.*, vol. 38, pp. 211–215, Nov. 1989.
[9] K. L. Yeung and T. S. Yum, "The optimization of nominal channel allocation in cellular mobile systems," in *Proc. IEEE ICC*, Geneva, Switzerland, May 1993, pp. 915–919.
[10] ——, "Compact pattern based dynamic channel assignment for cellular mobile systems," *IEEE Trans. Veh. Technol.*, vol. VT-33, pp. 892–896, Nov. 1994.
[11] S. M. Elnoubi, R. Singh, and S. C. Gupta, "A new frequency channel assignment in high capacity mobile communication systems," *IEEE Trans. Veh. Technol.*, vol. VT-31, pp. 125–131, 1982.
[12] E. D. Re, R. Fantacel, and G. Giambene, "Handover and dynamic channel allocation techniques in mobile cellular networks," *IEEE Trans. Veh. Technol.*, vol. 44, no. 2, pp. 229–237, 1995.
[13] E. D. Re, R. Fantacci, and L. Ronga, "A dynamic channel allocation technique based on Hopfield neural networks," *IEEE Trans. Veh. Technol.*, vol. 45, no. 1, pp. 26–32, 1996.
[14] D. E. Everitt and N. W. Macfadyen, "Analysis of multicellular mobile radiotelephone systems with loss," *Brit. Telecom Technol. J.*, vol. 1, no. 2, pp. 37–45, 1983.
[15] K. N. Sivarajan, R. J. McEliece, and J. W. Ketchum, "Dynamic channel assignment in cellular radio," in *IEEE Veh. Technol. Conf. VTC'90*, pp. 631–637.
[16] D. Everitt and D. Manfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1172–1180, 1989.
[17] D. C. Cox and D. O. Reudink, "Increasing channel occupancy in large-scale mobile radio systems: Dynamic channel reassignment," *IEEE Trans. Veh. Technol.*, vol. VT-22, pp. 218–222, Nov. 1973.
[18] K. L. Yeung and T. S. Yum, "Cell group decoupling analysis of a channel borrowing based dynamic channel assignment strategy in linear radio systems," *IEEE Trans. Commun.*, vol. 43, pp. 1289–1292, Apr. 1995.

Tak-Shing Peter Yum (S'76–A'78–SM'86) received the B.Sc., M.Sc., M.Ph., and Ph.D. degrees from Columbia University, New York, in 1974, 1975, 1977, and 1978, respectively.

He worked at Bell Telephone Laboratories for two and a half years and taught at National Chiao Tung University, Taiwan, for two years before joining the Chinese University of Hong Kong, Shatin, Hong Kong, where he has been since 1982. He has published original research on packet-switched networks with contributions in routing algorithms, buffer management, deadlock-detection algorithms, message-resequencing analysis, and multiaccess protocols. In recent years, he branched out to work on the design and analysis of cellular, lightwave, and video-distribution networks. He believes that the next challenge is designing an intelligent network that can accommodate the needs of individual customers.

Kwan Lawrence Yeung (S'93–M'95) received the B.Eng. and Ph.D. degrees in information engineering from the Chinese University of Hong Kong, Shatin, Hong Kong, in 1992 and 1995, respectively.

In May 1995, he joined the City University of Hong Kong as an Assistant Professor, where he is currently with the Department of Electronic Engineering. During the summer of 1993, he was with AT&T Bell Laboratories, Holmdel, NJ. His research interests include personal and mobile communication systems, high-speed networking, broadband packet switches, and lightwave networks.