Location Dependent Dynamic Load Balancing

Evşen Yanmaz and Ozan K. Tonguz

Department of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213-3890, USA

Abstract—The performance of most dynamic load balancing schemes, proposed to solve the *hot spot* problem in cellular systems, depends on the user distribution within a cell. In this paper, we first study the impact of user distribution on the call blocking probability performance that can be achieved by two key dynamic load balancing schemes (i.e., channel borrowing without locking (CBWL) and integrated cellular and ad hoc relay (iCAR) system [1]-[2]). We show that the achievable call blocking probability performance is location-dependent, and iCAR and CBWL schemes might not be able to achieve 2% call blocking probability for all user distributions, unless major modifications are made to CBWL or iCAR to overcome *local congestion* problems. Therefore, in this paper, we propose to use a joint dynamic load balancing scheme that combines the features of CBWL and iCAR schemes. Our objective is to guarantee a low call blocking probability in congested cells irrespective of user location.

I. INTRODUCTION

The substantial increase in demand for bandwidth in cellular networks necessitates the efficient allocation of resources (channels) to each cell due to limited frequency spectrum available for wireless applications. If a cell or group of cells has to serve high traffic for some time intervals, e.g., downtown areas during rush hour, the system should be able to cope with the service requests. To achieve this, the excess traffic of the *hot spots* can be transferred to other cells with available channels by dynamic load balancing strategies.

Several researchers have previously proposed interesting dynamic load balancing and channel assignment schemes to overcome the *congestion* problem [1]-[9]. One can classify the dynamic load balancing schemes in the literature into two groups: *i)* Strategies based on channel borrowing from cooler cells such as simple borrowing, channel borrowing without locking (CBWL) [1],[4], etc., and *ii)* Strategies based on traffic (call) transfer to cooler cells such as directed retry, integrated cellular and ad hoc relay (iCAR) system [2],[6], etc. The call blocking probability performance of all of these schemes have been studied in detail both analytically and via simulations, and it has been shown that dynamic load balancing can, by and large, alleviate the *congestion* problem in cellular networks.

Most of the analysis reported for dynamic load balancing schemes assumes *uniformly distributed traffic* within the cellular area. We show that the performance of dynamic load balancing schemes is location-dependent, and the predictions reported in the literature for uniform user distribution might not be realistic for other user distributions, and low call blocking probability guarantees cannot be generalized for all

possible distribution scenarios. In this paper, we study the call blocking probability performance of CBWL and iCAR schemes for user locations generated according to several probability density functions. We compute the probability that a user can be involved in load-balancing and use this probability to obtain the achievable call blocking probability by these schemes for different user distributions. Based on the analysis conducted, we observe that major modifications need to be made to the individual schemes to achieve acceptable levels of call blocking probability. Therefore, in this paper, we propose to use a joint dynamic load balancing scheme that combines the features of CBWL and iCAR schemes. Our objective is to guarantee a low call blocking probability in *hot cells* irrespective of user location. We aim to achieve this by increasing the likelihood that a mobile user can participate in load-balancing. While in this paper we concentrate on a given set of user distributions, note that if system parameters are carefully chosen, with the proposed scheme low call blocking probabilities can be achieved for several other user distributions, as well.

II. OVERVIEW OF CBWL AND ICAR SCHEMES

A. Channel Borrowing without Locking (CBWL)

Most channel borrowing and channel assignment schemes require channel locking, since a channel can be borrowed by a cell only when it is idle in all of the cells within the required channel reuse distance of the borrowing-cell, which limits the number of channels that are available for lending to a cell and reduces the capacity in the co-channel cells. An approach that does not require channel locking is proposed in [1]. With CBWL, a channel is borrowed from an adjacent cell if a new call arrives and all channels of the base station (BS) are occupied. The borrowed channel cannot be used by the original lending BS, but can still be used by nearby co-channel BSs. To prevent the increase of co-channel interference, borrowed channels are used with reduced transmitted power. Therefore, they can be accessed only in part of the borrowing cell. Fig. 1 shows the transferable traffic region (TTR) for CBWL. For CBWL, the normalized TTR size (p) with respect to cellular area were calculated in [1] for uniform distribution and are shown in Table I. Note that the value of p depends on the frequency reuse factor (N) employed in the cellular infrastructure.

There are two types of CBWL: CBWL without channel rearrangement (CBWL/NR) and CBWL with channel rear-

Fig. 1. Two cell system illustrating the TTR's for CBWL and iCAR schemes.

TABLE I

p VALUES FOR DIFFERENT FREQUENCY REUSE FACTORS FOR CBWL [1]

N	p	N	p
3	0.09	12	0.39
7	0.19	19	0.48
9	0.22	25	0.49

rangement (CBWL/CR). In CBWL/NR only the new call requests within TTR_c (such as user T) are granted service using a borrowed channel. On the other hand, in CBWL/CR, even if the new call request is not in TTR_c (such as user Y), it can still get service if there is at least one active user within the channel-borrowing area. In this case, the active user within TTR_c (such as user W) will borrow a channel from a neighboring cell and start transmission at a reduced power and the channel released by that user will be assigned to the new user requesting service.

B. Integrated Cellular and Ad hoc Relaying System (iCAR)

Dynamic load balancing in a *hot* cell can be achieved either by *borrowing channels* from the *cooler* cells in the system as with CBWL, or by *forwarding the excess traffic* of the *hot* cell to the *cooler* cells [2], [6]-[7]. iCAR systems employ the second approach. Basically in an iCAR system, a number of ad hoc relay stations (ARS's) is placed throughout the geographical coverage area, so that the signals between the mobile hosts (MH's) and base stations (BS's) can be relayed, and a MH requesting service in a *hot* cell can communicate with the BS's of the *cooler* neighbors of the *hot* cell [2]. An ARS is assumed to be a wireless communication device, which may have limited mobility under control of a mobile switching center (MSC), and it can communicate directly with an MH, a BS, or another ARS through air interfaces. Each ARS is assumed to have two air interfaces, one for communicating with the BS's (cellular interface) and the other for communicating with the MH's and other ARS's (relay interface). Similarly, each MH is assumed to have two air interfaces. It is also assumed that the ARS's use 2.4 GHz unlicensed Industrial, Scientific, and Medical (ISM) band channels, while relaying

the signal of the MH's and other ARS's.

The basic relaying mechanisms through ARS's can be described as follows:

Primary Relaying: If an MH cannot be assigned a voice (or data) channel in a congested cell, it can be *directly* relayed to a neighboring cell via ARS's if the MH is within the ARS coverage area [2]. For example, assuming that cell 2 in Fig. 1 is congested, a new user X will not be able to find a cellular-band channel in cell 2. However, since it is in the coverage area of an ARS, which is located at the cell border, it can use a channel of cell 1 via *primary relaying* through this ARS.

Secondary Relaying: If the MH requesting service is outside the ARS coverage area of the congested cell, an ongoing call within the ARS coverage area can be relayed to a neighboring cell via ARS's freeing up a channel in the congested cell to serve the new call [2]. Note that, when secondary relaying is employed, it is implied that primary relaying is also employed. For example, in Fig. 1, a new user Y can not be assigned a channel in cell 2 and it is not covered by an ARS. In this case, checking if there are any ongoing calls within the ARS coverage area, one realizes that user Z is in the coverage area of an ARS. User Z is then relayed to cell 1, and user Y can then use the channel released by user Z in cell 2.

Note from Fig. 1 that for iCAR systems TTR_i corresponds to the ARS coverage area within a cell.

III. IMPACT OF USER DISTRIBUTION ON DYNAMIC LOAD BALANCING PERFORMANCE

From Fig. 1 one can observe that the performance of both CBWL and iCAR schemes will depend on the size of the TTR and the distribution of the users within a cell (see Fig. 2 for different user distributions). Clearly, if most of the users are located around the BS, CBWL will perform effectively, whereas iCAR will not be able to achieve load balancing since there are very few users within ARS coverage area. Similarly, if most users are located closer to the border of the cell, CBWL will not help significantly. Or, in the extreme case where most users are located within an annulus neither scheme might be able to reduce the call blocking probability of a *hot* cell below 2%. Note that the three user distributions, in addition to the uniform distribution, shown in Fig. 2 are clearly not the only possible user distributions. These distributions are merely shown to illustrate the cases where iCAR or CBWL might not be sufficient to reduce the call blocking probabilities to acceptable levels within the *hot* cells.

A. Effective coverage for CBWL and iCAR schemes

In [1] and [2], it was shown that CBWL/CR and iCAR with secondary relaying perform significantly better than CBWL/NR and iCAR with primary relaying, respectively. Therefore, we focus our attention on CBWL/CR and iCAR with secondary relaying in this paper. The closed-form call blocking probability expression for these schemes were reported in [10], and it was shown that the call blocking probability depends on the number of available fixed and load balancing channels, the traffic intensities in the cells, the

Fig. 2. Example user distributions within a cell.

call blocking probabilities in the neighboring cells, and the probability that a user can be involved in load-balancing.

In this section, we explore this final parameter, which can be restated as the probability that a user is located within TTR, denoted by p . Clearly, for a uniform user distribution, p would be the ratio of the area of TTR within a cell to the area of a cell. For general distributions, the effective p for a given TTR is given by:

$$p = \int \int_{TTR} f_{XY}(x, y) dx dy \quad (1)$$

where $f_{XY}(x, y)$ is the joint probability density function (pdf) of the user locations, and is defined for the (x, y) coordinate pairs within the cellular region.

In the following, we show the effective p value for CBWL and iCAR systems for several user distributions (some of which are shown in Fig. 2). We assume hexagonal-shaped cells, where the range of the BS's is 2 km. We assume that the frequency reuse factor is 7, which corresponds to a p value of 0.19 (from Table I) and a TTR radius of 793 m for uniform distribution for CBWL. We assume that there are a total of six ARS's located at the cell borders. For a fair comparison, we assume that the range of each ARS is 458 m so that the p value for uniform distribution is 0.19 for iCAR system as well.

Fig. 3 shows the effective p value for uniform and normal (around BS and ARS's, and within annulus) distributions versus the standard deviation of the (x, y) coordinates. As expected, although for uniform distribution p is 0.19, depending on the user distribution the effective p value ranges between 0 and 1. This implies that the call blocking probability performance of each scheme will vary with the user distribution.

Using the expressions derived in [10], we can compute the call blocking probability that can be achieved by each dynamic

Fig. 3. Effective p for iCAR and CBWL for different user distributions.

load balancing scheme for the effective p values obtained above. Fig. 4 shows the call blocking probability of iCAR and CBWL for several user distributions. Unless otherwise noted, we use the following system parameter values for generating the results. We assume that the number of fixed and load-balancing channels per cell is 50 and 12, respectively. The traffic intensity in the *hot* cell is assumed to be 50 Erlangs, which corresponds to a 10% call blocking probability if no dynamic load balancing scheme is employed. As can be seen from the figure, while for uniform distribution the call blocking probability can be reduced to 2% with both iCAR and CBWL, for other distributions 2% blocking probability might not be guaranteed.

Fig. 4. Call blocking probability for iCAR and CBWL for different user distributions.

B. Methods to increase effective coverage for CBWL and iCAR schemes

As demonstrated in the previous section, the achievable call blocking probability performance is location-dependent, and iCAR and CBWL schemes might not be sufficient to achieve 2% call blocking probability for all user distributions. In this section, we illustrate the necessary modifications that need to be made to each scheme to achieve a certain effective coverage for given user distributions. Since Fig. 4 shows that call blocking probability can be reduced to 2% if $p = 0.19$

for the given system parameters and uniform distribution, we try to find the necessary modifications to achieve an effective p value of 0.19 for other user distributions.

First, we will examine CBWL scheme. Recall that the p value for CBWL can be changed by modifying the frequency reuse factor N . Fig. 5 shows the required frequency reuse value to achieve a minimum effective p value of 0.19 for several user distributions. Note that while $N = 7$ is sufficient for uniform and normal (within annulus) distributions, the value needs to be increased to 27 if users are mostly located near the borders, whereas a smaller N can be used if users are mostly located around the BS's.

Fig. 5. Necessary frequency reuse factor for CBWL to achieve an effective p of 0.19 for different user distributions.

Next, we examine the modifications necessary for iCAR system. The p value for iCAR is determined by the range and number of the ARS's. To increase the effective p value, we can either place more ARS's within the cellular area, or we can increase the range of the existing ARS's located at the borders.

Fig. 6 shows the necessary number of ARS's to achieve an effective p value of 0.19 for the system parameters given in the previous subsection, when the range of each ARS is fixed to 458 m. Observe that for an effective p value of 0.19 for uniform distribution, having 6 ARS's located at the borders is sufficient. However, if the users are mostly located closer to the BS, the number of necessary ARS's can increase to 19.

Alternatively, one can fix the number of ARS's to 6, and increase the range (i.e., transmit power) of the individual ARS's. Fig. 7 shows the necessary range for ARS's to achieve an effective p value of 0.19. Observe that, the necessary range might go up to 1.6 km if users are densely located at certain locations within the cellular area.

Although, *theoretically*, a given effective p value can be obtained for both CBWL and iCAR schemes by doing some modifications, *practically* these modifications might not be feasible. For CBWL, the frequency reuse needs to be increased significantly, which means introducing more BS's and changing the frequency assignments. This is a very costly and even unnecessary approach, since the user distributions within each cell is different from one another. For iCAR system, the number of ARS's or the range of individual ARS's

Fig. 6. Necessary number of ARS's for iCAR to achieve an effective p of 0.19 for different user distributions.

Fig. 7. Necessary range for ARS's for iCAR to achieve an effective p of 0.19 for different user distributions, if only 6 ARS's are located at the borders.

needs to be changed. Introducing more ARS's will increase the deployment cost. Moreover, a routing scheme needs to be employed in the resulting multi-hop overlay network, which will increase the signaling overhead. Channel assignment to each ARS will also be more complex. On the other hand, if we increase the range of ARS's (or have multiple power levels), the ARS's will need to be more intelligent, which will increase the cost. Higher power will reduce the lifetime of the network if the power is limited. Also, the power cannot be increased indefinitely due to interference limitations in the ISM-band.

IV. LOCATION DEPENDENT DYNAMIC LOAD BALANCING

Due to the practical limitations illustrated in the previous section, instead of making major modifications to CBWL or iCAR to overcome *local* congestion problems, we propose to use a location dependent dynamic load balancing scheme. The objective is to increase the effective p to guarantee a low call blocking probability wherever the user is located.

Since CBWL performs better if users are located closer to the cell centers, whereas iCAR performs better if the users are located closer to the cell borders, we propose to use a joint scheme that combines the features of CBWL and iCAR schemes. Clearly, fewer the number of ARS's, the less

expensive iCAR will be. Also, since voice communication has very stringent delay requirements, it is desirable to have as few hops between the MH's and the BS's as possible. Therefore, we assume that 6 ARS's are located at the cell borders. We also assume that the frequency reuse factor is 7 (which is widely used in practical cellular systems).

The location dependent dynamic load balancing algorithm works as follows. The MH, who could not find an available channel in a cell, will be relayed to an adjacent cell if it is covered by an ARS, or it will use a borrowed channel if it is within TTR_c . If the MH is not within TTR_i or TTR_c , an active MH within the joint TTR will be switched to a relay or borrowed channel, and the released channel will be assigned to the new MH. If there are no active MH's within the joint TTR, the MH will be blocked.

In the following, we compute the effective p value of the joint dynamic load balancing scheme. We assume that the range of the ARS's is 458 m. Fig. 8 shows that the effective p stays above 0.3 for all distributions under study.

Fig. 8. Effective p for the location dependent dynamic load balancing scheme for different user distributions.

The call blocking probability that can be achieved by the joint dynamic load balancing scheme is shown in Fig. 9. Results show that the call blocking probability for all distributions is around 7×10^{-3} . Recall from Fig. 4 that the individual schemes could not reduce the call blocking probability below 2% for some of these distributions.

Although the proposed scheme will require more control signaling, the service provider can use a small number of ARS's with a reasonable range, and does not need to change the original frequency planning to achieve acceptable call blocking probability levels. Also, fewer number of relay and borrowed channels than the individual schemes is required for the joint scheme. Clearly, the best call blocking probability performance can be achieved when p is close to 1 (i.e., when the location dependence is removed), though an effective p value of 1 might not be necessary. If there is prior knowledge about the user location pattern and hence the necessary effective p value, this information can be used to decide the range and location of the relay stations.

Fig. 9. Call blocking probability for the location dependent dynamic load balancing scheme for different user distributions.

V. CONCLUSION

In this paper, we have proposed to use a joint dynamic load balancing scheme that combines the attractive features of two key dynamic load balancing schemes: CBWL and iCAR. We have studied the impact of user distribution on the call blocking probability that can be achieved by these schemes. It is shown that the achievable call blocking probability performance is location-dependent, and iCAR and CBWL schemes might not be able to achieve 2% call blocking probability for all user distributions, unless major modifications are made to CBWL or iCAR to overcome *local* congestion problems. We have shown that with the joint scheme proposed in this paper acceptable levels of call blocking probability can be achieved in *hot* cells irrespective of the user location.

REFERENCES

- [1] H. Jiang and S.S. Rappaport, "CBWL: A New Channel Assignment and Sharing Method for Cellular Communication Systems," *IEEE Trans. Vehic. Tech.*, vol. 43, no.2, pp. 313-322, May 1994.
- [2] H. Wu, C. Qiao, S. De, and O. K. Tonguz, "Integrated Cellular and Ad-Hoc Relay Systems: iCAR," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2105-2115, Oct. 2001.
- [3] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey," *IEEE Personal Communications*, vol. 3, no. 3, pp. 10-31, Jun. 1996.
- [4] S. M. Elnoubi, R. Singh, and S. C. Gupta, "A New Frequency Channel Assignment Algorithm in High Capacity Mobile Communication Systems," *IEEE Trans. Vehic. Tech.*, vol. VT-31, no. 3, pp. 125-131, Aug. 1982.
- [5] T.-S. P. Yum and W.-S. Wong, "Hot-spot traffic relief in cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 6, pp. 934-940, Aug. 1993.
- [6] B. Eklundh, "Channel Utilization and Blocking Probability in a Cellular Mobile Telephone System with Directed Retry," *IEEE Trans. Commun.*, vol. 34, no. 4, pp. 329-337, Apr. 1986.
- [7] H. Jiang and S.S. Rappaport, "Hybrid channel borrowing and directed retry in highway cellular communications," in *Proc. Vehic. Tech. Conf (VTC'96)*, vol. 2, pp. 716-720, May 1996.
- [8] D. Everitt and D. Manfield, "Performance Analysis of Cellular Mobile Communication Systems with Dynamic Channel Assignment," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 8, pp. 1172-1180, Oct. 1989.
- [9] M. Alanyali and B. Hajek, "On Simple Algorithms for Dynamic Load Balancing," in *Proc. IEEE INFOCOM'95*, vol. 1, pp. 230-238, Apr. 1995.
- [10] O. K. Tonguz and E. Yanmaz, "On the Theory of Dynamic Load Balancing," in *Proc. IEEE Global Telecom. Conf. (GLOBECOM'03)*, vol. 7, pp. 3626-3630, Dec. 2003.