CAC DPLB MCN: A DISTRIBUTED LOAD BALANCING SCHEME IN MULTIMEDIA MOBILE CELLULAR NETWORKS

Abhijit SHARMA*, Arvind SHAH[†], Monish CHATTERJEE[‡], Uma BHATTACHARYA[§]

Abstract. The problem of non-uniform traffic demand in different cells of a cellular network may lead to a gross imbalance in the system performance. Thus, the users in hot cells may suffer from low throughput. In this paper, an effective and simple load balancing scheme CAC_DPLB_MCN is proposed that can effectively reduce the overall call blocking. This model considers dealing with multi-media traffic as well as time-varying geographical traffic distribution. The proposed scheme uses the concept of cell-tiering thereby creating fractional frequency reuse environment. A message exchange based distributed scheme instead of centralized one is used which help the proposed scheme be implemented in a multiple hot cell environment also. Furthermore, concept of dynamic pricing is used to serve the best interest of the users as well as for the service providers. The performance of the proposed scheme is compared with two other existing schemes in terms of call blocking probability and bandwidth utilization. Simulation results show that the proposed scheme can reduce the call blocking significantly in highly congested cell with highest bandwidth utilization. Use of dynamic pricing also makes the scheme useful to increase revenue of the service providers in contrast with compared schemes.

Keywords: Call Carrying Capacity, Cellular Network, Channel Allocation, Frequency Reuse Partitioning, Load Balancing.

1. Introduction

Mobile cellular communication is one of the fastest growing and most demanding telecommunication applications. The fast growth of the cellular market necessitates high

---

* Department of Computer Applications, National Institute of Technology, Durgapur, Durgapur - 713209, West Bengal, India, abhijit.sharma@ca.nitdgp.ac.in

[†] Microsoft Global Delivery India Pvt. Ltd., Gachibowli, Hyderabad - 500032, Telangana, India, arsha@microsoft.com

[‡] Department of Computer Science and Engineering, Asansol Engineering College, Asansol, Burdwan, West Bengal, India, monish chatterjee@yahoo.com

[§] Department of Computer Science & Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah - 711103, West Bengal, India, ub@cs.iests.ac.in

capacity cellular systems. But, the frequency spectrum used in wireless communication is very limited. This leads to reuse the frequency as much as possible in order to support the many thousands of simultaneous calls that may arise in any typical mobile communication environment. Mobile cellular networks are evolving into versatile Internet Protocol (IP)-based wireless networks that can provide multimedia services to mobile users. Eventually, the next generation of cellular networks is expected to carry multimedia traffic - combinations of voice, video, images, or data. On one side, there exists the requirement for the traditional voice calls, which are considered as a real-time (RT) or delay sensitive service. On the other side, there exist data services which are adaptive and tolerant to delay. They can operate over a wide range of bandwidth as non-real-time (NRT) or delay tolerant services. Thus, the challenge in a multiservice system is that the radio resources should be properly distributed among multiple traffic classes so that the QoS requirements of each traffic class can be satisfied while the resources are utilized as efficiently as possible.

For enhancing capacity and coverage quality to meet the growing traffic demand, a network operator usually seeks to employ several approaches such as cell splitting and cell sectoring. However, cell splitting is very expensive when few numbers of hot cells exists in the cellular network, as it requires additional equipment such as power plant, radio transceiver, antennas etc. Also, the cell sectoring is economically not desirable since new equipment is needed to upgrade omni cell system to sectorized cell system. In addition, the increased capacity mostly remains unused during off peak hours. As the peak hours contribute only to a small portion of the day, other inexpensive alternatives should be employed to overcome such temporary congestions.

Channel Borrowing is the popularly used scheme to overcome such situation of temporary congestion during peak hours. In Channel Borrowing scheme, if a cell needs a channel in excess of the channels previously assigned to it, that cell may borrow a channel from one of its neighboring cells. The major problem with channel borrowing is that when a cell borrows a channel from a neighbor cell, other nearby cells are prohibited from using the borrowed channel because of co-channel interference. This can lead to increased call blocking over time [30].

The main objective of reuse partitioning is to provide an increase in system capacity over that which can be achieved with a single factor, without relaxing SIR (Signal to Interference Ratio) performance requirements. In two-zone reuse partitioning, each cell is divided into

Figure 1. A two tier seven cell cellular System

two concentric sub cells, tiers or zones, where the inner zone and outer zone, each corresponds to a different reuse factor. The idea behind reuse partitioning is that because the inner zone is closer to the base station, the power level required for a desired Carrier to Interference Ratio (CIR) in the inner zone is much lower than the outer zone. As a result, the frequency reuse factor of inner zones can be smaller than that of outer zones, which results in higher spectrum efficiency compared to the system without reuse partitioning.

In spite of the existing schemes, however, the more spectral efficient frequency reuse partitioning scheme still deserve to be studied, to improve the system performance further more. Also, based on our review work in section 2 it has been observed that none of the earlier papers on reuse partitioning have addressed the problem of congestion due to time-varying geographical traffic distributions, which led to the motivation of our work in this area. In this paper, we propose to use reuse partitioning scheme to cope with the problem of traffic overloads and thereby congestion in the cellular network during peak hours in multiple hot cell environment. Cell tiering is achieved by splitting cells into two tiers: Reduced Coverage Tier (RCT) and Normal Coverage Tier (NCT), as shown in Figure 1.

The remainder of this paper is organized as follows. In section 2 Review work is discussed. Section 3 defines the problem we are investigating, section 4 introduces system model used. Section 5 describes the architecture of a two-tier cellular network. Section 6 defines the various terminologies used in this paper. In section 7 key concepts used in the proposed scheme CAC_DPLB_MCN is described, whereas, the proposed call admission control scheme is described in section 8. Section 9 briefly discusses about the earlier works considered for the comparison. Section 10 discusses about simulation environment, results and performance comparisons. Conclusions are drawn in section 11.

2. Review Work

Radio network operators are always pursuing network planning scheme that is both performance and revenue-assuring. Time-varying geographical traffic distributions bring problem to that goal as many cells may become densely-populated with users in the case of events or emergencies. Also, few cells may experience temporary congestion during peak hours. This section presents a brief review of some of the existing load balancing approaches for mobile cellular network.

Earlier work on load balancing mostly assumed a centralized controller that governs the base stations and the mobile stations with access to all the necessary information [27], [36], [7]. However, centralized approaches, for load balancing, may require excessive computational complexity and message overhead, which increase exponentially with the size of the network. Distributed cell load balancing is also being considered as a basic requirement in upcoming standards. For example, IEEE 802.16m WiMAX2 recently included parameters such as cell load and cell type in the system information broadcast [19], [16].

A framework for user association problem in wireless networks is proposed in [18] which is specifically focused on distributed load balancing under spatially non-homogeneous traffic distributions and showed the optimality condition of cell coverage areas that minimizes generalized system performance function.

In [12], the authors suggested a decentralized algorithm for Load Balancing in the downlink. The algorithm is based on the solution of a mixed integer optimization problem

solved using Lagrangian approach [12]. The algorithm aims at optimally balancing the load, while at the same time guaranteeing low levels of Inter-Cell-Interference (ICI).

In [6], the authors proposed a novel channel management algorithm, called distributed load balancing with selective borrowing (D-LBSB), for cellular mobile networks. The scheme proposes to migrate unused channels from suitable cold cells to the hot ones through a distributed channel borrowing algorithm. The suitability of a cold cell as a lender is determined by an optimization function consisting of three cell parameters, namely, the degree of coldness, nearness and hot cell channel blockade.

For load balancing, a cell breathing technique was investigated in [7], [13]. It contracts (or expands) the coverage of congested (or under-loaded) cells by reducing (or rising) the power level, and therefore the load becomes more balanced.

Using smart base station antennas, a scheme is proposed [10] which dynamically changes cellular coverage size and shapes according to geographic traffic distribution.

An on-line algorithm which dynamically controls the associations of relay stations with base stations and associations of mobile units with relay stations and base stations is proposed in [33] where relay stations (RS) are used to transfer over-loaded traffic from hot cells to neighboring cooler cells.

A dynamic load balancing integrated cellular and ad hoc relay (iCAR) scheme employs ad hoc relay stations (ARS's) in the cellular network to balance traffic loads efficiently and to share channels between cells via primary and secondary relaying. Regardless of the number and position of hot cells, an iCAR based algorithm [34] for a k-tier system is developed. It is shown that perfect load balancing can be achieved and the call blocking probabilities of the hot cells can be decreased substantially.

A cell-cluster based traffic load balancing strategy to solve the problem of cell congestion in cooperative cellular networks using relays is proposed [32].

A self-organizing load balancing [25] provides self-optimizing load balancing policies to improve adaptation and robustness of Fixed Relay Station (FRS) based cellular networks. A novel Comprehensive Load Balancing Policy Stack (CLBPS) is also proposed to utilize merits of various load balancing policies.

In [3], the author proposed a scheme for borrowing bandwidth from the existing calls in order to admit new calls. They showed that their scheme outperforms the previous schemes in terms of call dropping probability and call blocking probability. However, they considered only two types of traffic and they did not include priority.

An energy efficient load balancing is proposed [24] for delay tolerant data services using the mobile relay nodes (vehicular relays) in cellular system. It has been shown that by trading off message delivery delay considerable savings in energy are observed while at the same time cell load imbalance can be mitigated.

Another load balancing approach [23] uses online bandwidth management algorithms for multimedia cellular networks which try to minimize the maximum available bandwidth among cells in an adaptive manner based on current network traffic conditions. This approach is able to resolve conflicting performance criteria while ensuring efficient network performance for QoS sensitive multimedia services.

A game-theoretic analysis for load balancing is presented in [4] which remarkably increase the capacity usage in the network even when the cells act in a non-cooperative way. Here the amount of load to accept or to offload is considered to be decided independently by each cell.

A hybrid dynamic load balancing algorithm in heterogeneous wireless networks for packet services [35] adopts the combination of TRU (two-dimensional resource unit) borrowing scheme and load transfer scheme among overlapping heterogeneous cells. The hot-spot cell can not only borrow TRUs from neighbor homogeneous light-loaded cell, but also can transfer loads into overlapping heterogeneous cell with light load.

This work in [14] is concerned with the dynamic load balancing problem in 3GPP LTE multi-cell networks with heterogeneous QoS requirements. It is formulated as a multi-objective optimization problem, the objectives of which are load balancing index of services with QoS requirements and network utility of other services.

In a cellular network with non-uniformly distributed traffic demand, a hot cell can borrow a channel from cold cell to fulfil its demand. But due to lending out these free channels the lender cell eventually become the hot cell and try to borrow channel from nearby cold cell. This process continues redundantly. To solve this redundant channel borrowing problem Two Thresholds [20] scheme is used. In this scheme cell states are divided into three categories – Hot cell, warm cell and cold cell. This scheme reduces redundant channel borrowing by lending out channels from warm cells.

The reuse partitioning scheme was proposed to focus on spectral efficiency more than inter-cell interference mitigation compared to predefined reuse scheme (PRS) [28], [1]. The soft frequency reuse partitioning scheme (SFRPS) was proposed to increase spectral efficiency compared to reuse partitioning scheme (RPS) [21], [37].

As discussed in this section, it is evident that none of the papers have addressed the problem of reducing congestion due to geographical and multi-media traffic distributions in multiple hot cell environment which motivates to do our research work in this direction.

3. Problem Definition

Given a multi service mobile cellular network consisting of multiple number of cells which are associated with a limited resource (bandwidth) the issue is to distribute bandwidth and balance load so as to improve bandwidth reuse and utilize the unused capacity of the neighbour cells to cope with the problem of traffic congestion with desired quality of service (QoS). The objective is also to reduce call blocking probability and improve bandwidth utilization with increasing system load.

4. System Model

In this section, the traffic model and the network model used throughout the rest of the paper are presented.

4.1. Traffic Model

In this paper, different traffic types (audio, video and data) and therefore different application groups are considered. The different types of traffic are characterized by different arrival

Table 1. Multimedia Traffic Types Used in Simulation

Traffic Type	Main Traffic Class	Application Group	Traffic Class	Bandwidth Requirement (min/ max)	Average Bandwidth Requirement	Average Connection Duration	Example
1	α	α – Voice User Group A	Real Time (CBR)	30 kbps	30 kbps	90 s	Voice calls (short calls)
2		α – Voice User Group B	Real Time (CBR)	30 kbps	30 kbps	180 s	Voice calls (medium calls)
3		α – Voice User Group C	Real Time (CBR)	30 kbps	30 kbps	380 s	Voice calls (long calls)
4		α - Other	Real Time (CBR)	256 kbps	256 kbps	300 s	Video phone and Video conference
5	γ	γ – Type1	Non Real Time (UBR)	10-20 kbps	15 kbps	30 s	Email, Paging and Fax
6		γ – Type2	Non Real Time (UBR)	64-512 kbps	256 kbps	180 s	Remote Login and Data on Demand
7		γ – Type3	Non Real Time (UBR)	1-10 Mbps	5 Mbps	120 s	File Transfer and Retrieval Service

rates, service rates, and bandwidth requirements. Many applications especially audio and video transmission can support a variable bit rate. For example, MPEG-4 supports very low bit rate coding with bandwidth requirements of 5-64 Kbps [15], while using silence detection and sophisticated coding techniques used in audio [5] result in encoding that supports variable bit rate.

In this paper, it is assumed that multi service cellular network traffic is classified into two main traffic classes as shown in Figure 2, which are referred to as delay-sensitive (α) and delay-tolerant (γ) ones. Class α calls have the highest priority followed by class γ calls.

It is assumed that the class α (α) calls belong to real time traffic with constant bit rate such as voice calls and video phone calls. Thus, the bandwidth required for real time voice and video phone calls are fixed. During congestion, on the basis of average call duration, the real time voice traffic is further categorized into three user groups (User Group A, User Group B and User Group C) as short, medium and long duration callers [Section 6.2].

Class γ belongs to non-real time traffic with variable or unspecified bit rates such as email, remote login and file transfer services. The bandwidth requirement for class γ varies from

Figure 2. Traffic Types

their minimum to their required (maximum) bandwidth. This permits to arbitrarily control the bandwidth according to network situation.

In this paper, it is assumed that every class has a maximum (requested) bandwidth, and a minimum bandwidth. The assumption is that this class of traffic can support degraded performance down to the minimum bandwidth. The network can, if needed, borrow some bandwidth from any active call with the condition of not violating its minimum bandwidth requirement. Constant bit rate sources are a special case where the minimum bandwidth is the same as the maximum bandwidth.

All traffic types are shown in Table 1 along with their duration and bandwidth requirements.

4.2. Network Model

The proposed scheme in this paper considers a cellular system consisting of N number of hexagonal cells. Non-uniform traffic demand is considered in the given cellular network. The results from the central hexagonal cell and its six surrounding cells are used to avoid the boundary effect. Radius of each cell is R and every cell is served by a base station. Base stations are connected together by using a wireless or wired network. Initially, the total available frequency is allocated according to fixed assignment scheme to a cluster of N cells. The central cell of the cluster is assumed here to be the most congested cell for simplicity though the proposed scheme is applicable for any random hot cell and multiple hot cells due to the distributive nature of the proposed scheme.

The coverage area of each cell in the cluster is divided into two tiers: Reduced Coverage Tier (RCT) and Normal Coverage Tier (NCT). For a cluster size of N , the radii of the RCT and NCT are considered as R' and R respectively where $R' = R/\sqrt{N}$ [29]. The selected value of R' ensures that bandwidth allocated at RCT of any cell in the cluster can be reused simultaneously at RCTs of all other cells in the cellular system without interference.

Both the tiers (RCT and NCT) in a cell share the same base station (or cell site). Bandwidth allocated to a particular cell can be used in both of its tiers but not simultaneously.

The mobile terminals in the RCT and NCT are at different distances from the base station of the cell. The base station uses two different transmitted power levels based on the radii of the respective tiers. RCT uses lower transmitted power level than that in NCT.

Different classes of calls with different priorities, arrival rates, service times, and bandwidth requirements are considered and shown in Table 1. The call blocking ratio, the bandwidth utilization (network utilization) and system delay are the main parameters of interest. In this paper the assumptions that are being considered for the proposed scheme CAC_DPLB_MCN are as given below.

4.2.1. Assumptions

1. Hand-off calls are not considered in this paper to maintain simplicity.
2. The scheme CAC_DPLB_MCN is independent of specific transmission technology such as TDM, FDM or CDMA.
3. The user can receive a variable bandwidth.
4. Arrival rate in hot cells is considered to be higher than that in cold cells.
5. Arrival rate and leaving rate of calls are equal among all NCTs/(RCTs).
6. Arrival rate and leaving rate of calls may differ in between RCT and NCT of a cell.
7. Arrival rate is Poisson distributed and Service time is exponentially distributed.
8. Call requests can wait for bandwidth allocation in a priority queue up to a threshold time.
9. Each cell contains three bandwidth pools namely Pre-allocated Bandwidth Pool, Shared Bandwidth Pool and Borrowed Bandwidth Pool. Description of these bandwidth pools is given in nut-shell in the next subsection 4.2.2.
10. Each cell contains two tables named as LENDER_TABLE and BORROWER_TABLE which are described in the section 7.1.1.

4.2.2. Bandwidth Pools

In the proposed work, following three types of bandwidth pools are maintained at each cell. Details of the bandwidth pools are as follows:

A. *Pre-allocated Bandwidth Pool*

Table 2. Structure of Pre-allocated Bandwidth Pool in a cell

Amount of (initially) allocated bandwidth	Amount of utilized bandwidth	Amount of bandwidth shared with neighbor cells
Pre-defined	Initially zero	Initially zero

Table 3. Structure of Shared Bandwidth Pool in a cell

Total Amount of bandwidth in the shared pool	Total Amount of utilized bandwidth within shared pool	Current Cell		Neighbor-1		Neighbor-2		...	Neighbor-6	
		Shared Bandwidth	Utilized bandwidth within shared	Shared Bandwidth	Utilized bandwidth within shared	Shared Bandwidth	Utilized bandwidth within shared		Shared Bandwidth	Utilized bandwidth within shared
Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	...	Initially zero	Initially zero

Table 4. Structure of Borrowed Bandwidth Pool in a cell

Amount of bandwidth in Borrowed bandwidth pool	Amount of utilized bandwidth within borrowed	Current Cell		Neighbor-1		Neighbor-2		..	Neighbor-6	
		Borrowed Bandwidth	Utilized bandwidth within borrowed	Borrowed Bandwidth	Utilized bandwidth within borrowed	Borrowed Bandwidth	Utilized bandwidth within borrowed		Borrowed Bandwidth	Utilized bandwidth within borrowed
Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	Initially zero	...	Initially zero	Initially zero

Pre-allocated Bandwidth Pool contains the relevant information about the initially allocated bandwidth at each cell. It also contains the amount of utilized bandwidth within the pre-allocated one and the total amount of bandwidth shared with neighbour cells i.e. the bandwidth transferred to shared bandwidth pool from the pre-allocated one at any moment during operational phase.

The structure of the Pre-allocated Bandwidth Pool is given in Table 2.

B. Shared Bandwidth Pool

Table 5. Format of UPDATE message

Sequence No./Timestamp	Message Type	Sender Cell ID	Amount of Bandwidth to be moved	Bandwidth Specific other information
M	UPDATE	Cell _i	30	As Required

Table 6. Format of the RSRB message

Sequence No./Timestamp	Message Type	Sender Cell ID	Receiver Cell ID	Amount of Bandwidth
M	RSRB	Cell _i	Cell _j	30 (say)

Table 7. Format of Is Shared Bandwidth Available (ISBA) message

Sequence No./Timestamp	Message Type	Sender Cell ID	Amount of Bandwidth	Bandwidth Specific other information
M	ISBA	Cell _i	30 (say)	As required

Table 8. Format of Bandwidth Return Request (BRR) message

Sequence No./Timestamp	Message Type	Sender Cell ID	Amount of Bandwidth	Bandwidth Specific other information
M	BRR	Cell _i	30 (say)	As required

Table 9. Format of BW_AVAILABLE_MSG message

Sequence No./Timestamp	Message Type	Sender Cell ID	Amount of Bandwidth	Cell Type
M	BW_AVAILABLE_MSG	Cell _x	30 (say)	Lender

Whenever a call request originates at RCT of a particular cell, the bandwidth to this call request is allocated from shared bandwidth pool of that cell. It contains the amount of shared bandwidth and the amount of utilized bandwidth within shared bandwidth for the cell itself and also for its all neighbour cells. The structure of the Shared Bandwidth Pool is given in Table 3.

C. Borrowed Bandwidth Pool

A cell can borrow unused bandwidth of its neighbours, if required. Accordingly, Borrowed Bandwidth Pool contains the relevant information about the bandwidth which is borrowed

Table 10. Format of BW_REQUIRED_MSG message

Sequence No./Timestamp	Message Type	Sender Cell ID	Amount of Bandwidth	Cell Type
M	BW_REQUIRED_MSG	Cell _y	30(say)	Borrower

Table 11. Format of LEND_BW message

Sequence No./Timestamp	Message Type	Lender Cell ID	Borrower Cell ID	Amount of Bandwidth	Bandwidth Specific other information
M	LEND_BW	Cell _x	Cell _y	30 (say)	As required

from neighbour cells. It also contains the utilized one among borrowed bandwidth for the cell and all its neighbour cells. Total amount of borrowed bandwidth and utilized bandwidth for the cell and all its neighbours are stored also within the structure. The structure of the Borrowed Bandwidth Pool is given in Table 4.

4.2.3. Messages Used in the Proposed Work

In the proposed cellular system, all the base stations are directly connected together by using a wireless or wired link and they can communicate among themselves without having any centralized controller. Message exchanges among neighbour cells take place to ensure distributed load balancing which are described below.

A. UPDATE MESSAGE

It has already been mentioned that the bandwidth allocated in RCT of a cell can be reused in RCT of any other cell. The required bandwidth needs to be transferred from Pre-allocated Bandwidth Pool to Shared Bandwidth Pool to facilitate this bandwidth reuse. In the process whenever a cell transfers its pre-allocated bandwidth to its Shared Bandwidth Pool the cell transmits an UPDATE message to all of its neighbour cells so that all of them update their Shared Bandwidth Pools. The sample format of the 'UPDATE' message sent by Cell_i is shown in Table 5 and its working is illustrated in Figure 3.

B. REQUEST TO SHARE RETURNED BANDWIDTH (RSRB) MESSAGE

Any cell borrowing bandwidth from another cell wants to return this borrowed bandwidth to the lender cell and also wants the lender cell to add this returned bandwidth into the lender's Shared Bandwidth Pool, the cell transmits a Request to Share Returned Bandwidth (RSRB)

message to the lender cell. The structure of the RSRB message is given in Table 6 and the working is depicted in Figure 4.

Figure 3. Working of UPDATE message

C. *IS SHARED BANDWIDTH AVAILABLE (ISBA) MESSAGE*

Any cell is allowed to take back its shared bandwidth from Shared Bandwidth Pool to its Pre-allocated Bandwidth Pool, if required. This will only be possible if all the neighbouring cells are not using that bandwidth from their respective Shared Bandwidth Pools to avoid interference. Thus, in order to know that whether the required shared bandwidth remains unutilized in all neighbour cells, the cell transmits an Is Shared Bandwidth Available (ISBA) message to all of its neighbour cells for confirmation. Format of the ISBA message is given in Table 7.

D. *BANDWIDTH RETURN REQUEST (BRR) MESSAGE*

As already mentioned above, in order to take back shared bandwidth from Shared Bandwidth Pool to its pre-allocated bandwidth pool Cell_i first sends an ISBA message to all of its neighbours to confirm whether the required bandwidth shared by Cell_i remains unutilized in the Shared Bandwidth Pools of all the respective neighbours. After validation, the Cell_i transmits a Bandwidth Return Request (BRR) message to its neighbour cells upon which the neighbours release the specified amount of shared bandwidth from their respective Shared

Bandwidth Pool and return it to the pre-allocated bandwidth pool of the Cell. Sample format of the BRR message is given in Table 8.

Figure 4. Working of RSRB message

E. *BW_AVAILABLE_MSG MESSAGE*

A cell transmits a *BW_AVAILABLE_MSG* message to its neighbour cells to announce its wish to lend specified amount of bandwidth. Format of the *BW_AVAILABLE_MSG* message is given in Table 9.

F. *BW_REQUIRED_MSG MESSAGE*

A cell transmits a *BW_REQUIRED_MSG* message to its neighbour cells to announce its need to borrow specified amount of bandwidth. Format of the *BW_REQUIRED_MSG* message is given in Table 10.

G. *LEND_BW MESSAGE*

The lender cell broadcasts a *LEND_BW* message to its neighbour cells to lend a specific amount of bandwidth to a borrower cell. Format of the *LEND_BW* message is given in Table 11. On receipt of this message the intended borrower cell update its Borrowed Bandwidth

Figure 5. Adjacent hexagonal cells

Pool while all other neighbour cells update their respective LENDER_TABLE and BORROWER_TABLE as discussed in section 7.1.1.

5. Architecture of a two-tier cellular network

The main objective of a two-tier cellular network is to provide an increase in system capacity over that which can be achieved with a single factor, without relaxing SIR (Signal to Interference Ratio) performance requirements. In two-tier cellular network, the power level required for a desired Carrier to Interference Ratio (CIR) in the inner zone is much lower than the outer zone since the inner zone is closer to the base station. Architecture of a two-tier cellular network requires (i) Radius computation for reduced coverage tier and (ii) Mobile Proximity and tier selection. Those are described in the following subsections.

5.1. Radius computation for Reduced Coverage Tier (RCT)

The key characteristic of a cellular network is the ability to re-use frequencies for increasing both coverage and capacity. Adjacent cells must use different frequencies, however there is no problem with two cells sufficiently far apart operating on the same frequency. The elements that determine frequency reuse are the reuse distance and the reuse factor.

The *reuse distance*, D is calculated as

$$D = R\sqrt{3N} \quad (1)$$

where R is the cell radius and N is the number of cells per cluster [31].

From equation (1), we get Co-channel reuse ratio as

$$Q = \frac{D}{R} = \sqrt{3N} \quad (1)$$

A small value of Q provides large capacity whereas a large value of Q improves the

transmission quality because of smaller level of co-channel interference.

In Figure 5(a) two adjacent hexagonal cells are shown which uses two different sets of frequencies (channels). The actual center-to-center distance between these two adjacent hexagonal cells is $2R\cos 30^\circ$ or $\sqrt{3}R$. Maintaining the co-channel reuse ratio, we can create two virtual cells at the centers of the adjacent large hexagonal cells as shown in Figure 5(b). Same set of frequencies can be allocated to these virtual cells if minimum co-channel distance, which equals $\sqrt{3}R$ is maintained between them.

Let r' is the cell radius and $d' (= \sqrt{3}R)$ is the reuse distance for these virtual cells. By using equation 2 r' can be calculated as

$$r' = \frac{R}{\sqrt{N}} \quad (3)$$

5.2. Mobile Proximity and Tier Selection

The channels are allocated depending on the user's location (Cell Tier) in the cell from which a channel is requested. Location of a user can be tracked by a base station. To determine the appropriate cell tier for channel allocation, predetermined Mobile Proximity Threshold (MPT) is decided on the basis of the size of RCT. The base station allocates channels based on the received signal strength indication (RSSI) from mobile terminals. The RSSI of a mobile terminal is compared with MPT to select appropriate tier for channel allocation to the mobile terminal.

IF $RSSI > MPT$

A channel is allocated to the mobile terminal in RCT.

ELSE

A channel in NCT is allocated.

6. Terminologies Used

$ReqBW_{RCT} / (ReqBW_{NCT})$: Sum of minimum required bandwidth to admit all waiting call requests in RCT/(NCT).

$BW_{minimum_k/(l)}$: The minimum required bandwidth to admit $k^{th} / (l^{th})$ call request from RCT/(NCT) waiting in queue.

$P/(Q)$: The total number of call requests waiting in the queue from RCT/(NCT).

$AvailableBW_i / (RequiredBW_i)$: Available / (Required) bandwidth for cell_i.

$BW_{(i, available\_pre-allocated)}$: Unused Bandwidth available in Pre-allocated Bandwidth Pool of cell_i.

$BW_{(i, available\_borrowed)}$: Unused Bandwidth available in Borrowed Bandwidth Pool of cell_i.

$BW_{(i, available\_shared)}$: Unused Bandwidth available in Shared Bandwidth Pool of cell_i.

BW_Use : The amount of bandwidth required to borrow from all active users of Application Group x belonging to class γ .

$BW\_Utilizatio$: The total amount of bandwidth utilized by all ongoing calls of Application Group x .

p : A variable representing any member of class γ Application Group.

$RequiredBW_i$: The amount of additional bandwidth that the cell_i requires.

n : the total number of ongoing calls of Application Group x .

7. Key Concepts used in the proposed CAC_DPLB_MCN scheme

Key concepts used in the proposed scheme are discussed below:

7.1. Resource (Bandwidth) handling

Resource (Bandwidth) handling is divided in three parts such as resource pre-allocation, resource allocation, and resource de-allocation which are described below.

7.1.1. Resource Pre-allocation

Resource Pre-allocation consists of the sub-schemes A. Prioritization of Call Requests in the waiting queue, B. Scanning priority queue, C. Check bandwidth availability in the cell and D. Bandwidth Accumulation which are described below.

A. Prioritization of Call Requests in the waiting queue

After the call request is added into the priority queue for waiting call requests, the call request is arranged at proper place in the queue along with other waiting call requests which is decided according to the predefined priorities of the traffic type as given in section 4.1. Equal priority call requests are sorted on the basis of their arrival time.

B. Scanning priority queue

At first the priority queue is scanned to find waiting call requests. If there exist waiting call requests in the priority queue, the sum of minimum required bandwidth to admit all waiting call requests in RCT ($ReqBW_{RCT}$) and in NCT ($ReqBW_{NCT}$) are computed as follows:

$$\text{ReqBW}_{RCT} = \sum_{k=1, k \in RCT}^P BW_{\text{minimum}_k} \quad (4)$$

$$\text{ReqBW}_{NCT} = \sum_{l=1, l \in NCT}^Q BW_{\text{minimum}_l} \quad (5)$$

C. Check bandwidth availability in the cell

The cell_i with the help of various bandwidth pools available in the cell computes to find whether it can share bandwidth to neighbor cells or it requires additional bandwidth. To check this, following computations are performed.

```

AvailableBWi = 0
RequiredBWi = 0
Req1i = BW(i, available_shared) – ReqBWRCT
IF Req1i >= 0
    Req2i = (BW(i, available_pre-allocated) + BW(i, available_borrowed)) – ReqBWNCT
    IF Req2i >= 0
        { AvailableBWi = Req2i
        Send BW_AVAILABLE_MSG to all neighbors to let them know that it can lend
        'AvailableBWi' bandwidth.
        }
    ELSE
        { RequiredBWi = |Req2i|
        }
    END
ELSE
    { Req3i = (BW(i, available_pre-allocated) + BW(i, available_borrowed)) – (|Req1i| + ReqBWNCT)
    IF Req3i >= 0
        { AvailableBWi = Req3i
        Send BW_AVAILABLE_MSG to all neighbors to let them know that it can lend
        'AvailableBWi' bandwidth
        }
    ELSE
        { RequiredBWi = |Req3i|
        Send BW_REQUIRED_MSG to all neighbors to let them know that it needs to borrow
        'RequiredBWi' bandwidth
        }
    END
END

```

D. Bandwidth Accumulation

Bandwidth-Accumulation facilitates the bandwidth borrowing from ongoing calls as well as from the neighbor cells. The sub-module 'Bandwidth Accumulation' is activated periodically and during each such activation tasks completed are described as follows.

- (i) Bandwidth degradation from ongoing calls

Table 12. LENDER_TABLE_i

Sl. No.	Neighbor Cell ID	AvailableBW from neighbor	Amount of bandwidth lent out from neighbor
	x		
...	...	...	...

Table 13. BORROWER_TABLE_i

Sl. No.	Neighbor Cell ID	Required BW from neighbor	Amount of bandwidth borrowed from neighbor
	y		
...	...	...	...

As already mentioned, the maximum and minimum required bandwidth of calls belonging to class α are same, bandwidths are degraded only from the active users of those delay-tolerant application groups that belongs to class γ . This step for bandwidth degradation from ongoing calls is applied to all application groups belonging to class γ . At this stage the amount of bandwidth needed to borrow from each active user x belonging to an application group of class γ is determined and then degraded from each ongoing call of that application group. The fraction/amount of bandwidth that would be borrowed /degraded from each active user of a particular Application Group x belonging to class γ is given in equation 6.

$$BW\_User_x = \left(\frac{\frac{BW\_Utilization_x}{\sum_{p=5}^7 BW\_Utilization_p} \text{Required} BW_i}{n} \right), \quad x \in \{5,6,7\} \quad (6)$$

Here, the objective is to fairly distribute the amount of bandwidth that would be borrowed from each user of a particular application group of class γ . This sub-module will borrow bandwidth from each user of a particular application group fairly in a way maintaining the constraint that the bandwidth of any call will not be reduced below its minimum required one thus ensuring QOS requirement of that call.

(i) *Bandwidth sharing among neighbor cells*

On the basis of the messages BW_AVAILABLE_MSG and BW_REQUIRED_MSG, each cell_i creates a table LENDER_TABLE_i sorted by AvailableBW from neighbor cells in descending order and another table BORROWER_TABLE_i sorted by RequiredBW from neighbor cells in descending order as given in Table 12 and Table 13 respectively.

First row of the LENDER_TABLE_i and BORROWER_TABLE_i indicates information about the least congested neighbor and most congested neighbor respectively. The first lender

neighbor cell $Cell_x$ initiates the lending of bandwidth to the first borrower cell $Cell_y$ whose $RequiredBW_y$ is not completely borrowed. The amount of bandwidth lent out by the lender cell $Cell_x$ to the borrower cell $Cell_y$ is $\text{Minimum}(\text{AvailableBW}_x, \text{RequiredBW}_y)$. $Cell_x$ initiates the lending process by broadcasting a LEND_BW message to its neighbors. The bandwidth lent in this way is thus added to the Borrowed Bandwidth Pool of $Cell_y$. However, all other neighbor cells (say k) update their respective LENDER_TABLE $_k$ and BORROWER_TABLE $_k$ with the information $\text{Minimum}(\text{AvailableBW}_x, \text{RequiredBW}_y)$.

Thereafter, if bandwidth remains in $Cell_x$ to lend out it broadcasts LEND_BW message and the next borrower cell in BORROWER_TABLE $_i$ responds and it goes on as described in last paragraph. If $Cell_x$ already lends out all of its additional bandwidth, the next lender cell is identified from LENDER_TABLE $_i$ and that lender cell broadcasts LEND_BW message in order to lend bandwidth to the first borrower cell $cell_m$ in the BORROWER_TABLE $_i$ whose $RequiredBW_m$ is not completely borrowed.

The process continues either till all borrower cells borrowed their required additional bandwidth or till all the lender cell lends out their additional available bandwidth. This system goes on periodically within the process of bandwidth allocation to all users.

7.1.2. Resource Allocation

Resource allocation for the call requests is described in section 8.

7.1.3. Resource De-allocation

Whenever an ongoing call terminates (for example due to completion of the call), the allocated resources (bandwidth) are released. Thus, the released bandwidth is added into the same bandwidth pool from where it was allocated.

7.2. Categorization of voice callers

The authors [17] found that 78% of total calls are with call duration less than 100 seconds which account for only 40% of traffic. Many times [17] such a short duration business call is blocked due to congestion i.e. all channels are allocated to other calls. A long-duration call reserves a channel for a longer period of time whereas a short-duration call for shorter period of time. Thus, in unit time period, a channel can support more number of short-duration calls than that of long-duration calls. During peak hours, a bandwidth allocation algorithm must pick the users that are having a high possibility of releasing bandwidth quickly than the others so as to allocate them to most of the call requests and the number of new call blocks would remain under required Grade-of-Service (GoS). This can be achieved by taking the advantage of traffic forecasting methodologies. Generally, cell operators maintain user call records in the database for the administration purpose and these can be easily utilized by the traffic forecasting algorithms. Our literature review in section 2 indicates that there has been some recent works [9], [26] on modeling and predicting future call duration of individual users. The proposed CAC_DPLB_MCN scheme categorizes voice callers into three groups (User groups A, B and C) based on their average call duration in a particular peak hour. With the

Figure 7. Flow Chart for bandwidth allocation at RCT

Fairness in call admittance after categorization of voice callers

During congestion when several call request waits in the priority queue, the proposed algorithm allocates bandwidth in the ascending order of average call duration of user groups

A, B and C. So, the highest priority is assigned to the user group A (the user group with shortest call durations). The call request queue is scanned continuously and at first bandwidth is allocated to the call requests from user group A. Then the remaining unallocated bandwidth (if any) is allocated to the call requests from user group B and after this allocation if unallocated bandwidth is available then it's allocated to user group C. This is to mention that this allocation is applicable only during peak hours or congestion period. Hence, the proposed algorithm will block few long duration calls to support many short duration calls and thereby reduces call blocking and increases overall call carrying capacity.

Although in this work we have classified users as user groups A, B and C on the basis of their average call duration during a particular peak hour, their actual call durations are not bounded at all. Thus, a user belonging to user group A is free to make a call for longer duration.

During a particular peak hour in normal situation, a user (belonging to user group A) having shorter average call duration in that particular peak hour is given preference in call admittance; hence, the user is getting incentive for keeping his call duration shorter during congestion periods. But, if the user makes a longer call, it will affect his average call duration for that particular peak hour, and in future request for call, he might lose preference in call admittance. So, most likely such users will not continue for longer duration calls unnecessarily. During congestion when bandwidth is unavailable in a cell, all new call requests would inevitably get blocked, whereas, by assigning priority in order to user groups A, B and C, the total number of blocked calls can be reduced. But, giving preference to user group A in call admission can lead to longer delays and call drops in extreme situation for user group B and user group C. There could be some situations when users belonging to user groups B and C need to make important business calls or calls with shorter duration. In these situations, such users can avail bandwidth for calls by opting for higher price by using dynamic pricing policy as described in section 7.3. Also, the users would not have to pay higher price if their call duration does not exceed a predetermined threshold. The threshold can be considered as the average duration of the user at that particular peak hour. Following this approach, the best interest of the users and as well as of service providers is being protected.

7.3. Dynamic Pricing

There exist some unsatisfied users who did not avail the services as per their expectation. For satisfying such user, dynamic pricing is integrated in the proposed call admission control scheme. The users who are in immediate need for the service have to accept higher price for the service and they will be served immediately or relatively sooner. Those users are termed as premium users. As for the rest of the users, they would find the dynamic price unacceptable and leave the system or put their call request in the waiting queue (those users are termed as conventional users) and wait for service.

For premium users if no bandwidth is available for this premium user the longest ongoing call is preempted to make way for the premium call. If the required bandwidth is not available yet, the latest multiple ongoing calls may be pre-empted for this purpose. While the conventional users are delayed by system for certain amount of time, then, they will be served with their expected price.

If dynamic pricing is deployed at the time of congestion, the users who will accept the dynamic price (higher than normal) are the ones who are in immediate need of the service. As for the rest of the users, they would find the dynamic price unacceptable and leave the system or put their call request in the waiting queue and wait for service. Some of them might try accessing the system again. Thus, deployment of dynamic pricing will shift some of the traffic load from peak hours to non-peak hours. As the system operates, the network will be less congested and possibly yield more revenue.

This module will return success in case of availability of bandwidth to serve the call request otherwise it will return failure.

8. Description of the proposed work

This section provides the detailed description of the proposed work. The flow chart of the proposed scheme is depicted in Figure 6.

When the system starts execution, each cell in the given cellular network takes the following steps.

- a) The three bandwidth pools (Pre-allocated Bandwidth Pool, Shared Bandwidth Pool and Borrowed Bandwidth Pool) [section 4.2.2] are initialized.
- b) A priority queue for waiting call request is created.
- c) A database USER_CALL_DURATION as described is created for maintaining user call register and thereby facilitating the computation of average call duration of the users.
- d) Two tables LENDER_TABLE & BORROWER_TABLE [section 7.1.1] are initialized.

Whenever a user requests for a new call, the cell collects the following information (hereafter, represented as **Collected_Information**) with respect to the requested call.

- a) The call request ID
- b) Time at which call request is generated
- c) Application Group (AG_x) of the call request
- d) Tier-Position of the user in the cell (whether user is located at RCT or NCT of the cell)
- e) The required bandwidth of the call request, $BW_{required}$, to accept the call.
- f) The minimum required bandwidth of the call request, $BW_{minimum}$ to accept the call.

Now the call request is added into the priority queue for waiting call request. The requested call has to be accepted within 'threshold' time, otherwise, the call request will be blocked. As mentioned in Section 7.3, a waiting user can opt to pay higher price to avail the service earlier. This option may be provided by the user, *either at the time the call request is added into the priority queue or before the call request is getting blocked*. The call requests from such users are designated as premium call requests and priority of their call requests is considered as superior to all other waiting call requests.

If the Application Group of the requested call is 'α voice', it is designated as either user group A, user group B or user group C on the basis of their average call duration with the help of the database USER_CALL_DURATION. User group A belongs to the shortest duration callers whereas user group C belongs to the longest duration callers. Depending upon the traffic type of the requested call (Table 1), the call request is positioned appropriately in the priority queue. Equal priority call requests are sorted on the basis of their arrival time.

The order of the priority in decreasing order is given below.

Premium Call Requests, α – Voice User Group A, α – Voice User Group B, α – Voice User Group C, α – Other, γ – Type1, γ – Type2, γ – Type3.

While call requests are waiting in the priority queue, the sub-module Bandwidth Accumulation [Section 7.1.1] is triggered periodically.

Hereafter, the highest priority call request from the priority queue is taken out for bandwidth allocation. Depending upon the location of the originated call request (RCT/NCT) and also traffic class of the call request (Class α / class γ) the process of bandwidth allocation is different.

The principle for allocating bandwidth to call requests from RCT & NCT is explained below in sub-section 8.1. Flow charts for bandwidth allocation at RCT and NCT are shown in Figure 7 and Figure 8 respectively.

8.1. Principle for allocating bandwidth to call requests from RCT & NCT

Cell-tiering concept has been used to increase bandwidth utilization by increasing frequency reuse and in turn call requests from RCT of the central cell and all neighboring cells can use the same bandwidth [section 5.1]. This common bandwidth is available in Shared Bandwidth Pool. Thus, call requests from RCT will always be served from Shared Bandwidth Pool. On the other hand, any two adjacent NCTs cannot share same bandwidth and also the bandwidth from Shared Bandwidth Pool because of interference. Consequently, emphasis is put over the scheme that call requests from RCT and NCT will be served from Shared Bandwidth Pool and Pre-allocated Bandwidth Pool respectively as long as the bandwidth remains available in the respective bandwidth pools.

In case of RCT, however, if sufficient bandwidth is not available in Shared Bandwidth Pool to accept a call request, Borrowed Bandwidth Pool is tested first for bandwidth availability and if available it will be transferred to Shared Bandwidth Pool. Thus, also in this case RCT is using the bandwidth from its neighbors. In the extreme situation, due to non-availability of bandwidth in Borrowed Bandwidth Pool, Pre-allocated Bandwidth Pool will be tested and if the required bandwidth to allocate the call request remains available, it will be transferred to Shared Bandwidth Pool and then it's allocated from Shared Bandwidth Pool.

Likewise, at first Pre-allocated Bandwidth Pool will be used for allocating bandwidth to a call request originating at NCT. If sufficient bandwidth is not available in Pre-allocated Bandwidth Pool the scheme will try to transfer cell's own shared bandwidth from Shared Bandwidth Pool to its Pre-allocated Bandwidth Pool. If now the requirement of bandwidth cannot be served from its own Shared Bandwidth Pool, in the extreme situation, Borrowed Bandwidth Pool will be tested for availability and if available it will be allocated to the call request. Whenever borrowed bandwidth is allocated, the same bandwidth is locked and cannot be allocated to call requests of other cells which are within the interference region.

Figure 8. Flow Chart for bandwidth allocation at NCT

9. Existing works used for comparison with the proposed work

The proposed adaptive call admission control scheme is evaluated by comparing it with the following two schemes.

9.1. First Come First Served (FCFS) Scheme

FCFS [2] is probably the simplest of all call admission protocols. In FCFS, it is assumed that all traffic types (classes) have the same priority. The entire bandwidth is allocated on First Come First Served basis. In this method, a new call is accepted if its average required bandwidth is available. If the condition fails, call will be rejected. This method has shown to produce a better utilization of the bandwidth [2] however it has been shown to be biased against calls with high bandwidth requirement.

9.2. Frequency Reuse Partitioning (FRP) based cellular systems (EX_CAC)

The author [22] proposed an analytical model to evaluate the performance of Frequency Reuse Partitioning based call admission control (EX_CAC) in cellular systems. The authors introduced a simple channel assignment procedure that selects a channel randomly from unassigned channels. The service area of a cellular system is divided into hexagonal cells of equal size, with the base stations at the center of each cell. Also, each cell is divided into two zones; inner zone using total system channel ($F = \{F_0, F_1, F_2\}; c$), and outer zone using one third of the total system channel ($F_i, i=1,2,3; c_o$, inner channel; $c_i = c - c_o$). When a call arrives, a MS requests a channel. If it is a MS in the inner zone (inner MS), one of the unassigned channels from among set F is randomly chosen and is assigned to the MS. Otherwise, if it is a MS in the outer zone (outer MS), one of the unassigned channels from among its outer channels (F_0, F_1, F_2) is assigned to the MS. In either case, if there is no channel available, the call is blocked and is cleared. The assigned channel is released and becomes available to other calls upon the completion of the call.

10. Simulation Environment, Results and Performance Comparisons

10.1. Simulation Environment

To assess the performance of the proposed CAC_DPLB_MCN scheme, a mobile cellular system with homogeneous cells and a fixed bandwidth which is permanently allocated to each cell is considered. Simulation is performed using MATLAB version 7.12.0.635 (R2011a). A cellular network consisting of 49 cells has been considered here for the simulation. The cell (base station) has a bandwidth capacity of 30 Mbps. The coverage area of each cell can be divided into Reduced Coverage Tier (RCT) and Normal Coverage Tier (NCT). Mobile Stations are assumed to be uniformly distributed across the coverage area. New call arrivals and call holding time are assumed to follow a Poisson distribution with

Figure 9. Overall call blocking probability vs. increasing system loads in Scenario-I

mean rate λ and an exponential distribution with mean $1/\mu$ respectively. Seven representative traffic types [11] shown in Table 1 are carefully chosen for the simulation. Table 1 shows the exact characteristics of the traffic used in the simulation. The arrival rate for the different call categories is adjusted such that the load on the network from the different types of traffic is equal, thus the different types are equally sharing the network bandwidth. Simulations are carried out as the traffic load of every class increases. During congestion, the probability that a user will opt for higher price [8] is considered as 0.3. In order to evaluate the proposed scheme, it is compared with a First Come First Served (FCFS) scheme [Section 9.1] and an EX_CAC scheme [Section 9.2].

10.2. Results and Performance Comparisons

Overall blocking probability, bandwidth utilization and system delay are the parameters of interest in this topic. Results of the simulation performed are described in the following subsections 10.2.1, 10.2.2 and 10.2.3 respectively.

10.2.1. Blocking Probability vs. System Load

Blocking probability is calculated as $\text{Blocking probability} = (\text{number of calls blocked} / \text{number of incoming calls})$.

Scenario I: Central cell is congested and is surrounded by uncongested neighbor cells

Overall call blocking

Figure 10. (a)Blocking Probability of FCFS in RCT vs. increasing system loads, (b) Blocking Probability of FCFS in NCT vs. increasing system loads and (c) Blocking Probability of EX-CAC in NCT vs. increasing system loads

Figure 9 depicts the overall call blocking of the three compared schemes. FCFS scheme shows the highest call blocks among all compared schemes. The performance of EX_CAC scheme was found better compared to the FCFS scheme. However, in all offered traffic loads the call blocks in CAC_DPLB_MCN is found zero, which shows the effectiveness of the CAC_DPLB_MCN scheme. CAC_DPLB_MCN scheme thus performed well both in RCT as well as in NCT, whereas EX_CAC is good only at RCT in terms of call blocks.

Figure 11. Overall call blocking probability vs. increasing system loads in Scenario-II

Blocking Probability in RCT

Figure 10(a) illustrates the variations of the call blocking probabilities of all traffic types with the increasing offered load for the compared schemes in RCT.

In case of RCT, call blocks are observed only in FCFS scheme amongst the three. This shows the effectiveness of EX_CAC and CAC_DPLB_MCN in reducing blocking by improving the frequency reuse.

Blocking Probability in NCT

Figure 10(b) illustrates the variations of the call blocking probabilities of all traffic types with the increasing offered load for the FCFS scheme in NCT of the cell while Figure 10(c) illustrates the variations of the call blocking probabilities of all traffic types with the increasing offered load for the EX_CAC scheme in NCT of the cell. The CAC_DPLB_MCN scheme shows its effectiveness in NCT too where also no call block occurs for all offered traffic load. It shows that the proposed CAC_DPLB_MCN scheme utilizes the borrowed bandwidth efficiently to reduce call blocks from NCT. Also, in CAC_DPLB_MCN scheme effective priority is assigned to the arrived call request and as the load increases, waiting users may opt for higher price and due to this the call blocks can further be decreased in even higher offered load. Hence, it is justified that the number of blocked calls is zero in CAC_DPLB_MCN scheme when compared with FCFS and EX_CAC. Even if it is absolutely evident that call block for the proposed scheme may show a very low value for very high load because of the underlying principle of the proposed scheme.

Scenario II: Central cell along with four congested neighbor cells

Overall call blocking

Figure 11 depicts the overall call blocking of the three compared schemes. FCFS scheme shows the highest call blocks among all compared schemes. The performance of EX_CAC scheme was found better compared to the FCFS scheme. However, in all offered traffic loads the call blocks in CAC_DPLB_MCN is found to be minimum among all compared schemes. Since, five out of seven neighbor cells were highly congested, the proposed CAC_DPLB_MCN scheme could not borrow sufficient bandwidth from neighbor cells to serve call requests. Yet the observed call blocks in CAC_DPLB_MCN scheme is significantly lower than the EX_CAC scheme due to fact that the CAC_DPLB_MCN scheme has efficiently utilized the borrowed bandwidth from two non-congested neighbor cells. Furthermore, the effective assignment of priority to the arrived call requests and the option to higher pricing facilitated the CAC_DPLB_MCN scheme to keep the call blocks minimum. CAC_DPLB_MCN scheme thus performed well both in RCT as well as in NCT, whereas EX_CAC is good only at RCT in terms of call blocks.

Blocking Probability in RCT

Figure 12(a) illustrates the variations of the call blocking probabilities of all traffic types with the increasing offered load for the compared schemes in RCT. In case of RCT, call blocks are observed only in FCFS scheme amongst the three. This shows the effectiveness of EX_CAC and CAC_DPLB_MCN in reducing blocking by improving the frequency reuse.

Blocking Probability in NCT

Figure 12(b) illustrates the variations of the call blocking probabilities of all traffic types with the increasing offered load for the FCFS scheme in NCT of the cell while Figure 12(c) illustrates the variations of the call blocking probabilities of all traffic types with the increasing offered load for the EX_CAC scheme in NCT of the cell. Figure 12(d) illustrates the variations of the call blocking probabilities of all traffic types with the increasing offered load for the CAC_DPLB_MCN scheme in NCT of the cell. Due to traffic prioritization and higher pricing total number of call blocks are much less compared to other two schemes. Call blocks observed in CAC_DPLB_MCN scheme are mostly belong to traffic type 6 and traffic type 7 due to their lowest priority among other traffic types.

10.2.2. Bandwidth Utilization vs. System Load

Bandwidth utilization is the ratio of the bandwidth utilized to the total bandwidth of the system. Figure 13 illustrates the overall bandwidth utilization of the three compared schemes in Scenario I (when only central cell is congested and is surrounded by uncongested neighbor cells). Whereas, Figure 14 demonstrates the overall bandwidth utilization of the three compared schemes in Scenario II (when central cell along with four neighbor cells are congested). FCFS scheme shows the lowest bandwidth utilization among all compared

Figure 12. (a)Blocking Probability of FCFS in RCT vs. increasing system loads, (b) Blocking Probability of FCFS in NCT vs. increasing system loads and (c) Blocking Probability of EX-CAC in NCT vs. increasing system loads (d) Blocking Probability of CAC_DPLB_MCN in NCT vs. increasing system loads

schemes. The performance of EX_CAC scheme was better when compared with FCFS because of frequency reuse implemented by EX_CAC. However, the proposed CAC_DPLB_MCN performed best when compared with the other two. This happens because of load balancing concept by considering time varying geographical traffic distribution in addition to frequency reuse by CAC_DPLB_MCN.

Figure 13. Overall Bandwidth Utilization vs. increasing system loads in Scenario I

Figure 14. Overall Bandwidth Utilization vs. increasing system loads in Scenario II

It is observed that the bandwidth utilization by CAC_DPLB_MCN scheme in Scenario-II is less than Scenario-I. This is due to the fact that in Scenario-II the central congested cell could borrow available bandwidth only from two neighbor cells as all other neighbor cells were congested.

10.2.3. System Delay vs. System Load

System delay is the time elapsed between arrival of call and allocation of bandwidth to that call. No system delay is observed in all the three compared schemes. The schemes FCFS and EX_CAC do not use any queue, and hence their delays are zero. In case of CAC_DPLB_MCN scheme due to effective prioritization of traffic types and efficient bandwidth utilization no delay was found for class α . Similarly, no delay was observed for class γ in case of CAC_DPLB_MCN as these call request types are blocked if bandwidth could not be allocated at the first attempt.

11. Conclusion

The proposed CAC_DPLB_MCN scheme considered concept of using the time-varying geographical multimedia traffic distributions over the cell-tiering concept in a cellular network. This newly proposed scheme utilized the pre-allocated bandwidth more efficiently and at the same time was also able to utilize the unused bandwidth of less congested cells to serve call requests of highly congested cell. CAC_DPLB_MCN scheme used concept of cell tiering thereby enabling bandwidth to be utilized simultaneously in multiple neighbor cells. Thus, the proposed scheme enhanced the bandwidth reuse in the given cluster and hence was effective in reducing the overall call blocks in the given cellular system. A message exchange based distributed scheme had been implemented so that the neighbor cells could communicate among themselves. This eradicated the need for a centralized controller which was required to keep the bandwidth usage information of various cells. This message exchange based distributed scheme also help the proposed scheme be implemented in multiple hot cell environment. Result analysis showed that under various traffic loads, the proposed scheme CAC_DPLB_MCN significantly balanced the load during congestion which is evident from the reduction of the blocking of calls to a large extent from the given cellular network. Furthermore, bandwidth utilization is most effective and system delay is found to be nil.

In addition, concept of dynamic pricing which is used in the proposed CAC_DPLB_MCN scheme serves the best interest of the users as well as of the service providers. The proposed scheme thus may be used for increasing revenue of the service providers in contrast to FCFS and EX_CAC.

References

- [1] 3GPP R1-082779. Proposal of Inter Cell Interference Coordination Scheme. Hitachi, Ltd.
- [2] Aboelaze M., A Call Admission Protocol for Cellular Networks That Supports Differentiated Fairness. In Proceedings of the IEEE 59th Vehicular Technology Conference (VTC 2004- Spring), May, 2004, Vol. 3, pp. 1616-1620.

- [3] Al-Sharaeh S H., Dynamic Rate-Based Borrowing Scheme for QoS Provisioning in High Speed Multimedia Wireless Cellular Networks. *Journal of Applied Mathematics and Computation*, 2006, 179(2): pp. 714-724.
- [4] Awada A, Wegmann B, Viering I, Klein A., A Game-Theoretic Approach to Load Balancing in Cellular Radio Networks. *IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2010)*, Istanbul, Turkey, September, 2010, pp. 1184-1189.
- [5] Crossman A., A Variable Bit Rate Audio Coder for Video Conferencing. In *Proceedings of the IEEE Workshop on Speech Coding for Telecommunications*, October, 1993, pp. 29-30.
- [6] Das S K, Sen S K, Jayaram R, Agrawal P., A Distributed Load Balancing Algorithm for the Hot Cell Problem in Cellular Mobile Networks. In *Proceedings of the Sixth IEEE International Symposium on High Performance Distributed Computing*, August, 1997, pp. 254-263.
- [7] Das S, Viswanathan H, Rittenhouse G., Dynamic Load Balancing Through Coordinated Scheduling in Packet Data Systems. In *Proceedings of the Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003)*, April 2003, Vol. 1, pp. 786-796.
- [8] Dhake T P, Patane R D, Jagtap S., Dynamic Call Management for Congestion Control in Cellular Networks. In *Proceedings of the International Conference and Workshop on Emerging Trends in Technology (ICWET 2011)*, Mumbai, India, February, 2011, pp. 946-947.
- [9] Dong Y, Tang J, Lou T, Wu B, Chawla N., How Long Will She Call Me? Distribution, Social Theory and Duration Prediction. *Machine Learning and Knowledge Discovery in Databases*, September, 2013, Vol. 8189, pp. 16-31.
- [10] Du L, Bigham J, Cuthbert L., Towards Intelligent Geographic Load Balancing for Mobile Cellular Networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, November 2003, 33(4): pp. 480-491.
- [11] Elnaggar A, Aboelaze M, Musleh M., A Call Admission Control Protocol for Multimedia Cellular Networks. *Journal of Networks*, 2008, 3(8): pp. 45-54.
- [12] Giovanidis A, Qi L, Stanczak S., A Distributed Interference-Aware Load Balancing Algorithm for LTE Multi-Cell Networks. *International ITG Workshop on Smart Antennas (WSA)*, Dresden, March, 2012, pp. 28-35.
- [13] Hanly S V., An Algorithm for Combined Cell-Site Selection and Power Control to Maximize Cellular Spread Spectrum Capacity. *IEEE Journal on Selected Areas in Communications*, September 1995, 13(7): pp. 1332-1340.
- [14] Hao W, Lianghui D, Ping W, Zhiwen P, Nan L, Xiaohu Y., Dynamic Load Balancing in 3GPP LTE Multi-Cell Networks with Heterogeneous Services. *5th International ICST Conference on Communications and Networking in China (CHINACOM)*, Beijing, August, 2010, pp. 1-5.
- [15] Hooi C C, Beg M S., Realizing MPEG-4 Video Transmission Over Wireless Bluetooth Link via HCI. *IEEE Transactions on Consumer Electronics*, 2004, 49(4): pp. 1028-1034.
- [16] IEEE Std 802.16m Part 16: Air Interface for Broadcast Wireless Access Systems: Advanced Air Interface. *IEEE Standard 802.16m*, 2010.

- [17] Junqiang G, Fasheng L, Zhiqiang Z., Estimate the Call Duration Distribution Parameters in GSM System Based on K-L Divergence Method. International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2007), Shanghai, September, 2007, pp. 2988-2991.
- [18] Kim H, Veciana G de, Yang X, Venkatachalam M., Distributed α -Optimal User Association and Cell Load Balancing in Wireless Networks. IEEE/ACM Transactions on Networking, February 2012, 20(1): pp. 177-190.
- [19] Kim H, Yang X, Venkatachalam M, Chen Y -S, Chou K, Fu I -K, Cheng P., Handover and Load Balancing Rules for 16m. IEEE 802.16 m-09/0136r1, January, 2009, pp. 1024-1037.
- [20] Kim J, Lee T, Hwang C S., A Dynamic Channel Assignment Scheme with Two Thresholds for Load Balancing in Cellular Networks. IEEE Radio and Wireless Conference (RAWCON 99), Denver, Colorado, August, 1999, pp. 141-145.
- [21] Kim K T, Oh S K., An Incremental Frequency Reuse Scheme for an OFDMA Cellular System and Its Performance. IEEE Vehicular Technology Conference (VTC Spring 2008), May, 2008, pp. 1504-1508.
- [22] Kim S Y, Ryu S, Cho C H, Lee H W., Performance Analysis of a Cellular Network Using Frequency Reuse Partitioning. Performance Evaluation, February, 2013, 70(2): pp. 77-89.
- [23] Kim S, Varshney P K., Adaptive Load Balancing with Preemption for Multimedia Cellular Networks. IEEE Wireless Communications and Networking (WCNC 2003), March, 2003, Vol. 3, pp. 1680-1684.
- [24] Kolios P, Friderikos V, Papadaki K., Load Balancing via Store-Carry and Forward Relaying in Cellular Networks. IEEE Global Telecommunications Conference (GLOBECOM 2010), Miami, FL, December, 2010, pp. 1-6.
- [25] Lexi X, Yue C, Yue G., Self-organizing Load Balancing for Relay Based Cellular Networks. 10th IEEE International Conference on Computer and Information Technology (CIT 2010), July, 2010, pp. 791-796.
- [26] Melo P O S V de, Akoglu L, Faloutsos C, Loureiro A A F., Surprising Patterns for the Call Duration Distribution of Mobile Phone Users. Machine Learning and Knowledge Discovery in Databases, September, 2010, Vol. 6323, pp. 354-369.
- [27] Navaie K, Yanikomeroglu H., Downlink Joint Base-Station Assignment and Packet Scheduling for Cellular CDMA/TDMA Networks. In Proceedings of the IEEE International Conference on Communications (ICC '06), June, 2006, Vol. 9, pp. 4339-4344.
- [28] Oh E, Cho M G, Han S, Woo C, Hong D., Performance Analysis of Dynamic Channel Allocation based on Reuse Partitioning in Multi-cell OFDMA Uplink Systems. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2006, E89-A(6): pp. 1566-1570.
- [29] Sharma A, Bhattacharya U., Load Balancing Scheme for Wireless Cellular Networks. 7th International Conference on Ubiquitous Information Management and Communication (ICUIMC), Kota Kinabalu, Malaysia, 2013, Article 43, pp. 1-6.
- [30] Sharma A, Sharma P, Chaki R, Bhattacharya U., Reducing call blocks in cellular network with non-uniform traffic conditions. Foundations of Computing and Decision Sciences, 2014, 39(4): 301-318.
- [31] Vijay K. Garg, "Wireless Communications & Networking", Elsevier, Morgan Kaufmann Publishers (Ebook), San Francisco, 2014, ISBN: 978-0-08-054907-1.

- [32] Wang X, Tian H, Jiang F, Li X, Hong X, Li T., Cell-Cluster Based Traffic Load Balancing in Cooperative Cellular Networks. 7th IEEE Consumer Communications and Networking Conference (CCNC), January, 2010, pp. 1-5.
- [33] Yang Z, Niu Z., A New Relay Based Dynamic Load Balancing Scheme in Cellular Networks. In Proceedings of the IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), Ottawa, Canada, September, 2010, pp. 1-5.
- [34] Yanmaz E, Tonguz O K, Mishra S, Wu H, Qiao C., Efficient Dynamic Load Balancing Algorithms Using ICAR Systems: A Generalized Framework. In Proceedings of the IEEE 56th Vehicular Technology Conference, (VTC 2002-Fall), September, 2002, Vol. 1, pp. 586-590.
- [35] Yuan-yuan C, Bing Q, Yi-ting L, Liang-rui T., A Hybrid Dynamic Load Balancing Algorithm in Heterogeneous Wireless Packet Networks. 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012), Chongqing, China, May, 2012, pp. 2052-2056.
- [36] Zemlianov A, Veciana G. de., Load Balancing of Best Effort Traffic in Wireless Systems Supporting End Nodes with Dual Mode Capabilities. In Proceedings of the Conference on Information Sciences and Systems, March 2005, pp. 1-6.
- [37] Zhang X, He C, Jiang L, Xu J., Inter-Cell Interference Coordination Based on Softer Frequency Reuse in OFDMA Cellular Systems. International Conference on Neural Networks and Signal Processing, June, 2008, pp. 270-275.

Received 02.11.2015, Accepted 03.10.2016