المجلة الإلكترونية الشاملة متعددة التخصصات
العدد الخامس و الأربعون (شهر 3) 2022

An enhanced protocol for load balancing of cooperative relay networks

Maha Alsayyari and Mohammed Arafah

College of Computer and Information Science, King Saud University

Riyadh, Saudi Arabia

E-mail: malsayyari@ksu.edu.sa

Abstract
In this paper, we propose an Adaptive Bandwidth Allocation (ABA) algorithm that balances the load of the network cell and dynamically supports different levels of data rates. The algorithm guarantees Quality of Service (QoS) using the Call Admission Control (CAC) concept by restricting the access to network resources. It will accept a new admission request provided that there are adequate free resources to meet the QoS requirements of the new mobile without violating the committed QoS of already accepted mobiles. The algorithm can guarantee a level of fairness between the network users. However, there is a tradeoff between the QoS level perceived by the user and the utilization of scarce wireless resources.

Keywords
Relay station, cooperative communication, load balancing, multi-hop relay network, call admission control, adaptive bandwidth allocation, bandwidth utilization.

1. Introduction

Currently the overall look of the wireless broadband market shows that mobile traffic of data will double every year, and 80% of global mobile traffic will be driven by mobile broadband handsets and laptops (Soldani, 2009). Users are expected to grow rapidly, thus creating a need for greater scale and reliable wireless network. In wireless network, signals suffer from fading and that may cause unpredictable packet losses. This is a big problem from the communication point of view. An early effort to cope with this problem was to use multiple-input, multiple- output (MIMO) antenna array. This solution was effective, however, we have to consider dealing with mobility issue and in addition most portable wireless devices can carry one or two antennas only. To meet the growing demand for throughput, capacity, coverage, reliability and low cost radio network deployment, for both short-range and wide-area coverage, we need, in addition to advanced transmission techniques and antenna technologies, some major modifications in the wireless network architecture itself that will enable the effective distribution and collection of signals to and from wireless users. The integration of multi-hop capability into conventional wireless networks is perhaps the

most promising architectural upgrade. This introduces a special kind of nodes called Relay Stations (RSs) or Relay Nodes (RNs) to relay traffic for Mobile Stations (MSs). Relay nodes are currently one of the most promising solutions. Relay nodes are devices equipped with antennas that can intersect a neighboring transmitter's signal and relay it to the designated receiver. Relay stations deployment will help overcome the dependency on wired backbones, which will reduce the cost, offer flexible placement, and enhance the coverage area. Such cost-effective network that will provide higher average revenue per user (ARPU) will be well suited for many scenarios, for example when the traffic density is low and the users are distributed, or to relay signals to shadowed areas (Baldini, Karanasios, Allen and Vergari, 2014). In a relay chain, each receiver along the chain takes a copy of only the data that has been sent to it, while discarding signals from other transmitters. The idea of Cooperative Communication (CC) came from exploiting the broadcast nature of wireless communication, and the relaying ability of the nodes, which will help to achieve diversity through independent channels. That has been shown to have the potential to increase the capacity of wireless networks (Yang, Fang, Xue and Tang, 2010).

1.1 Aim

In this paper, we address the problem of distributing the network traffic among different access networks so that the desired user and network performance objectives can be achieved.

1.2 Methodology

To accomplish the research objectives, the thesis research methodology will address the following:

- Studying the concept of cooperative relaying networks, the load balancing and CAC techniques
- Developing an enhanced protocol for load balancing for clustered cooperative relay networks.
- Conducting computer simulations to study the performance and compare the results.

2. Related Work

In the past few years, cooperative communication has been a very active research field. Extensive research has focused on the physical layer aspects of cooperative communications (Laneman, Tse and Wornell 2004). After that it worked also with the upper layers. In a traditional cellular network having only powerful BSs, stations association and load balancing have been well studied in. Normally, an MS is connected to the BS with the strongest received downlink signal power or Signal to Interference plus Noise Ratio (SINR) to get the best service. However, this scheme will not work efficiently in a multi-hop relay network due to the following:

• For a MS associated with a RS, both its downlink and uplink signals have to traverse multiple wireless hops. The best RN-MS link quality does not guarantee a good composite link quality over multiple hops.

• It must be considered that the MS associated with a RS will consume resources from both donor BS and the RS.

While the MS directly connected to the BS will only consume resources from that BS. So there is a tradeoff between link quality and total resource consumption in order to maximize the global efficiency.

- In this scheme, the majority of the MSs will select the BSs as the serving nodes due to their higher transmission power; this may lead to inefficient RS resource utilization.

In Yu, Hu, Bontu and Cai (2011), the authors take in consideration the resource utilization in order to achieve the load balancing. The actual resource consumption by mobiles connected to the BS is donated. Also hypothetical resource consumption indexes are defined, they are given different weights to allow flexibility on load balancing. The values of these weights can be selected as a trade-off among signal quality, spectrum efficiency and load balancing needs.

Sadek,Su and Liu (2005), discussed the Mobility Load Balance (MLB), which is done by shifting the traffic from heavy-loaded cell to light-loaded cell by handover, and this done by changing and adjusting handover parameters. Under this scheme, the call blocking probability can be effectively reduced because the unused resource is allocated to the users handed-over from the heavy-loaded cell. However, the shifted users suffer from Signal to Noise Ratio (SNR) degradation. This will cause a reduction in the date rate for shifted cell edge users.

But an important point must be discussed; the traditional traffic shifting schemes concentrates on the load reduction of the heavy-loaded cell and ignores the load increment on the load in the partner cells. By shifting, some light-loaded cells upgrade to heavy-load, and then new imbalanced states occur, which may introduce the ripple load balancing and that reduces the load balancing efficiency.

Cuthbert, Xu, Chen and Gao (2011) proposed a self- optimizing load balancing scheme in fixed relay based cellular networks. In the scheme, heavy-loaded cell and partner cells determine the decision of traffic shifting cooperatively. As an important part of the scheme, the response mechanism can reduce the probability of ripple load balancing. The shifting process is done via cell-to-cell communication.

Some recent work on relay selection criteria includes physical layer parameters (such as shortest distance, minimum path-loss, maximum receiving power, maximum Signal to Interference Ratio (SIR)). It aimed to obtain the maximum achievable rate for each cooperative user, but actually, in the relay based cellular networks, the throughput of each MS also relates to the total number of users served by the BS and RN. It is also related to the resource allocation algorithm and the scheduling schemes.

Jiang and Wang (2011) derived a distributed load balancing based relay selection (LB-RS) algorithm. When adopting two hop transmissions, the proposed scheme will select the optimal RN for each MS, according to the current Channel State Information (CSI) as well as the number of users that the RN serves. The goal is to elevate the asymmetric traffic loading and proactively avoid congested RNs for individual MSs and locate a good service site.

Xu, Chen and Gao (2010) proposed a self-organizing load balancing framework in Fixed Relay Station (FRS) based cellular networks. The concept discussed is Self-organizing Cooperative Partner Cluster (SCPC) in which the partner selection and updating process are distributed controlled rather than centrally control.

In wireless systems, supporting QoS requirements of different traffic types is a more challenging problem. The following researches are related to CAC and ABA algorithms: An appropriate bandwidth allocation scheme in the IEEE 802.16 multi-hop relay network is expected in order to guarantee QoS transmission. The issue of mobile QoS or (MQoS), has been addressed in the literature. The typical strategy for MQoS is

to reserve required bandwidth at neighboring node before the mobile user handoff to the new node, which inevitably results in low bandwidth utilization. Yang, Mai, Lin and Chen (2014) explained why this traditional mechanism inappropriate for MQoS support in the multi- hop relay network because the medium is managed by the base station in a centralized control manner, which provides the feasibility of more sophisticated bandwidth management in the network.

Chowdhury, Jang and Haas (2013) proposed an efficient CAC algorithm that relies on multi-level ABA scheme for non-real time calls. The scheme allows reduction of the call dropping probability, along with an increase in the bandwidth utilization. The scheme also results in higher priority for the handover calls over the new calls. While the proposed scheme blocks more new calls instead of dropping handover calls, the scheme also reduces the number of handovers and the average call duration.

Chou and Shin (2002) developed a model for cellular networks with a combined ABA and traffic restriction CAC in multi-level QoS system where users or applications can tolerate a certain of QoS degradation. Instead of focusing only on the usual call level parameters, they derive two user-perceived QoS metrics that helps decide the operation point under different traffic conditions: Degradation Ratio (DR), which is the fraction of time a user, receives degraded QoS. Upgrade/Degrade Frequency (UDF), which is the frequency of changing the QoS level an admitted user receives.

These metrics has been used by Chou and Shin (2004), the authors showed the importance of UDF to QoS provisioning, especially because of its strong dependency on user mobility and trade off with user fairness. Fair admission and bandwidth allocation algorithms provided such that low DR and UDF can be achieved with only a slight increase in the blocking probability of new users. Also they showed how to exploit ABA to increase system utilization (for the system administrator) with controlled QoS degradation (for the users).

Also in multi-level QoS system Lu and Bigham (2005) proposed a utility-based ABA scheme for multi-class traffic (Elastic, Adaptive and real time) in wireless networks, bandwidth adaptation is decomposed into two processes bandwidth upgrades and bandwidth degrades to cope with the network resource fluctuation. The scheme allocates bandwidth to ongoing calls according to their traffic utility functions so that the total achieved utility is maximized. Also the algorithm has been designed with low computational complexity.

Chen and Shen (2012) proposed an ABA algorithm in heterogeneous wireless networks with multi-level QoS of multi-services based on multi-thresholds reservation (or as called safe guard) mechanism. With the bandwidth reservation mechanism by setting multi-thresholds in each network for each traffic, the ABA scheme can be formulated as an optimal problem with the constraints of the bandwidth allocation matrix for each traffic and all users based on the multi-homing technology. But the defined optimized objective function is only maximizing the real- time network throughput according to the variations of resource availability under bandwidth reserving thresholds and network capacity constraints.

Unlike the previous work, the proposed CAC that relies on ABA can deal with all QoS acceptable levels to utilize the bandwidth. The generic algorithm works in heterogeneous network where all kind of traffic is acceptable. The degradation algorithm proposed is fair; Instead of decreasing one random mobile like previous works by Chou and Shin (2004), or decreasing one kind of traffic (Chowdhury et al, 2013), the algorithm is going to decease all the mobiles in the cluster by a certain ratio at each step, while making sure that the Service-Level Agreement (SLA) is not violated. The handoff traffic can be given a higher priority without any kind of resource reservation. The scheme results in bandwidth utilization and decreases in the call blocking probability, all at the expense of an increase in the user's delay in the system.

3. System Model and Assumptions

We consider a Centralized Wireless Relay Network (CWRN) with three types of nodes: BS, RS and MS. Among the network nodes, only the BS has backhaul access to the Internet and also the BS is the central coordinator controlling resources. Only the downlink direction is considered in this study. MS of different kinds of traffic is arriving according to Poisson arrival process. The BS knows all locations and takes charge of both admission control and channel assignment for mobiles in its cell. As mentioned before, channels could be frequencies, time slots or codes depending on the radio technology used. In order to avoid the self-interference of RSs, the transmission and reception of a RS should take place on different frequency or different time slot, so transmission and reception of a RS are either separated in the frequency-domain, e.g., by using different carrier or subcarrier frequencies, or in the time-domain. This study is based on using different frequencies so three frequency division are possible, as shown in Table 1. To avoid early system bottleneck explained by Yu et al. (2011), we are assuming that no MS can directly be connected to the BS, so that will narrow the frequency division to one case. We denote the downstream frequency band of the BS by F1, and the downstream frequency band of the RS by F2. Therefore, concurrent transmissions on F1 and F2 are allowed.

Table 1: Possible frequency division relaying cases.

	BS-MS link	RS-MS link	BS-RS link
Case A	F1	F1	F2
Case B	F1	F2	F1
Case C	F1	F2	F3

According to Lu and Liao (2012), to achieve high sum rate and maintain fairness among users, there is two sub- problems: **(1)** The Cooperative RS Clustering Problem is to decide which RSs should cooperate together. A cluster can be either static or dynamic. In the static approach, the size and shape of the cluster is the same regardless of the network situation. In the dynamic approach however, shape and/or size of the cluster change according to the congestion level and traffic characteristics. Typically, dynamic clusters have a better performance at the expense of increased complexity (see Table 2). Here we are assuming that static clusters are already formulated.

Table 2: Cluster type vs. CAC performance.

Cluster type	CAC efficiency	CAC complexity
Static	Moderate	Moderate
Dynamic	High	High

(2) The MS Assignment Problem is to decide which MSs should be scheduled to receive data from which cluster, which we are going to focus on. Without loss of generality, we consider only two-hop relaying. That is, each RS is connected to a specific BS (called the serving BS) and the RS must relay data packets for that BS to downstream MSs. Our CWRN model is depicted in Figure 1.
In this example, cluster A, B and C contains different numbers of RSs cooperating, resulting in different recourses assigned to each cluster, which means each cluster have different capacity. The BS transmits to

the clusters on F1. While the three clusters can also transmit to the downstream MSs on F2 at the same time since there are no interferences toward the MSs. After that, deciding which MSs to receive data from a cluster or mobile admission needs to be discussed. We are going to use the CAC concepts to deal with the mobile admission. If the coverage areas of the two clusters overlap, then the MS can only receive from one cluster and that is decided based on load balancing criteria discussed later. In general, dropping a call in progress is considered to have a more negative impact from the user's perspective than blocking a newly requested call. That's why we assume that once the MS is admitted it cannot be dropped. And if a handed off MS arrived it will be given a priority.

Figure 1. CWRN network model

In order to achieve the goals, which are more users, load balancing, utilization, fairness and QoS, we are going to work on two points. First, Admission control where CAC concepts are applied on the MS trying to enter and receive data from each cluster, specifically; ABA is going to be used. Second, Priority queue where MS with higher priority should be scheduled first. Figure 2 illustrates the architecture of the proposed system.

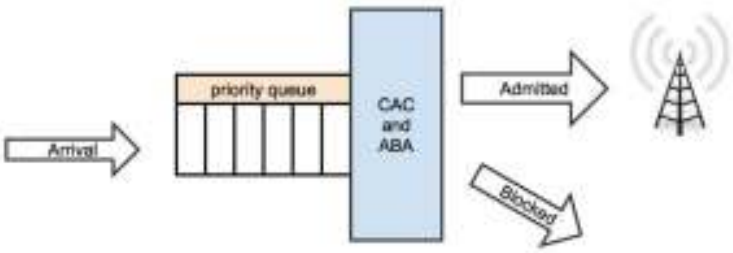


Figure 2. Architecture of the proposed system.

4. Proposed Algorithm

With global information at the BS, the cooperation behavior of RSs can be coordinated. The admission control module determines how radio resources are allocated. We assume that the BS uses Fixed Channel Assignment (FCA). It statically distributes the recourses between the clusters based on the previous knowledge of the usual traffic distribution. However, adaptive bandwidth allocation is being used inside each cluster. The available bandwidth in each cluster is channelized and cooperatively focuses on call-level QoS measures. Increasing the aggregate throughput of RS- MS links can increase the effective network throughput accordingly (Wang, Chen, Guo, Cai and Shen, 2008), and that is what we are aiming for.

4.1 Admission Control and ABA Algorithm

Due to the diversity of applications and QoS requirements for MS and the dynamic nature of wireless channel quality, developing an enhanced ABA algorithm would be necessary to improve utilization of the resources. Therefore, CAC strategies should be designed taking this into account. With the proposed ABA algorithm, when the network becomes congested, the amount of bandwidth allocated to some of the ongoing calls will be revoked to accommodate more incoming calls so that the call blocking probability can be minimized as possible, taking into consideration that the SLA is not violated for all mobiles.

Table 3: List of symbols and notations.

Notation	Meaning
C_j	Cluster j.
B_j	Bandwidth of cluster j.
CAB_j	Current Available Bandwidth in cluster j.
R_j	Coverage area of cluster j.
$maxMS_i$	Maximum data rate required by MS_i.
$minMS_i$	Minimum data rate required by MS_i.
$aveMS_i$	Average data rate required by MS_i.
CAD_i	Current assigned data rate of MS_i.
SLA	Service Level Agreement.
t_i	Amount of data traffic (in bits) queued at the BS destined for MS_i.
W	Step size, $0 < W < 1$.
PF	Priority factor.

The Generic Algorithm: Mobile requests are randomly added to a cell, and after the completion of their transmission or call according to their requirements, they go to idle mode or leave the cell. In our proposed algorithm, a cell has one Base Station, M Mobile Stations, K clusters Where
$1 \leq i \leq M$ and $1 \leq j \leq K$.
Table 3 lists the set of notations used in the proposed algorithm.

Mobile Assignment: This section presents the mobile assignments in case of a mobile under coverage of 1 cluster or several clusters. The CAC decision will follow the algorithm illustrated in the next two cases.

Case 1: M Si Under the Coverage of One Cluster (Cj). Initially, a mobile station will try to be admitted with the maximum data rate possible under that cluster. If there is no enough bandwidth to

assign to the mobile station with the maximum data rate, and however, the available bandwidth is more than the minimum required by a mobile station, then the available data rate is assigned to the mobile station. Otherwise, the data rate degradation algorithm starts working. Algorithm 1 illustrates the proposed admission control when a MS is under coverage of one cluster.

Algorithm 1: The proposed admission control when MS is under one cluster.

Input: CAB_j, $maxM\ S_i$, $minM\ S_i$, C_j.
Output: $M\ S_i$ is assigned to cluster C_j or not.

begin
  if $CAB_j \geq maxM\ S_i$ **then**
    Assign $M\ S_i$ to C_j
    $CAD_i \leftarrow maxM\ S_i$
  else if $CAB_j \geq minM\ S_i$ **then**
    Assign $M\ S_i$ to C_j
    $CAD_i \leftarrow CAB_j$
  else
    Execute data rate degradation algorithm (for cluster C_j)
  if $M\ S_i$ *is assigned to* C_j **then**
    return "Assigned"
  else
    return "Not Assigned"
end

Case 2: M Si under the coverage of N clusters where $1 \leq n \leq N$. Initially, a mobile station will try to be admitted with the maximum data rate possible under one of the clusters.
The choice of a cluster leads to dynamic load distribution by shifting the traffic to the most suitable cluster where it can be admitted under his SLA while trying to leave the clusters with big CABs for other new mobiles that might have higher data rate requirements. If the mobile was not admitted using the above process, the data rate degradation algorithm begins working, starting with the cluster with the biggest CAB first. This descending order is to save the rest of the clusters from going to the degradation algorithm. The cluster with big CAB has a better chance to host the new mobile with fewer rounds of the degradation algorithm since it already has some available bandwidth. The proposed admission control when a MS is under N clusters is described by Algorithm 2.

Algorithm 2: The proposed admission control when MS is under N clusters.

Input: CAB, $maxM\ S_i$, $minM\ S_i$, $(C_1, C_2, \ldots. C_N)$.
Output: $M\ S_i$ is assigned to cluster Cn or not.

begin
  if *All or some of* N *clusters have* $CAB \geq maxMS_i$ **then**
    Assign $M\ S_i$ to the cluster with minimum $CAB \geq maxMS_i$
    $CAD_i \leftarrow maxM\ S_i$
  else Sort clusters $(C_1, C_2, \ldots, C_N)$ in descending order based on CAB.
  if $CAB_1 \geq minM\ S_i$ **then**

```
7          Assign MS_i to C_1
8          CAD_i ← CAB_1
9      else
10       Execute data rate degradation algorithm (for cluster C_n)
11         if MS_i is assigned to C_n then
12               return "Assigned"
13         else if n = N then
14               return "Reject"
15         else
16               n ← n + 1
17               go to line 10
18     End
```

4.2 Data Rate Degradation Algorithm:

The data rate degradation algorithm is a kind of dynamic resource allocation. It is performed by borrowing the recourses from neighbor mobiles to the point where the new mobile can be admitted. Instead of decreasing one random mobile like previous works (Chou and Shin, 2004), or decreasing one kind of traffic (Chowdhury et al, 2013), the algorithm is going to decrease all the mobiles in the cluster by a certain ratio at each step, while making sure that the SLA is not violated. For the fairness sake, the amount of the decreasing is proportional to the current given data rate CAD of each mobile. This amount is decided by the step size (W). The algorithm will keep going through steps; degrading the CAD of the admitted mobiles until one of the following cases happen:

1. The mobiles reached their minimum data rate stated in the SLA.
2. The new mobile can fit in the cluster.
3. Number of steps reaches a certain threshold that is relative to the step size.

Number of steps threshold = 1/W. This threshold limits the complexity of the algorithm. Algorithm 3 describes the steps of data rate degradation.

Algorithm 3: The data rate degradation algorithm.

Input: Number of mobiles in each cluster (Q), $(CAD_1, \ldots, CAD_Q)$, CAB_j, B_j, MS_i, Cj, W.

Output: Accept or Reject assigning MS_i to cluster C_j.

```
1  begin
2        S ← 0
3        q ← 1
4        if (S ≤ 1/W) then   sum of CAD of all MS in C_j <= B_j
5           if (CAD_q − W * CAD_q ≥ minDR_q) then
6              CAD_q ← CAD_q − W * CAD_q
7              CAB_j ← CAB_j + W * CAD_q
8                 if (minMS_i ≤ CAB_j) then
9                   return "Accept" // Assign MS_i to C_j such that CAD_i = CAB_j.
10                  else
11                   go to line 13
12             else
```

```
13                  if (q = Q) then
14                    S ← S + 1
15                  q ← 1
16                  go to line 4
17                   else
18                    q ← q + 1
19                  go to line 5
20          else
21              return "Reject" // MS_i cannot be assigned to cluster_j.
22.     end
```

4.3 The Priority Queue

The priority queue proposed will sort the MSs in an ascending order based on the following metric:

$$(t_i / \frac{(maxMSi + minMSi)}{2}) \frac{1}{PF} \quad (1)$$

where, ti is amount of data traffic (in bits) queued at the BS destined for M Si, maxM Si and minM Si are the maximum and minimum data rate required by MSi, and PF is the priority factor. Priority Factor is introduced to give more priority to the handoff MS. The value of PF will be set according to the following equation:

$$PF = \begin{cases} 1 & if\ MS\ is\ new \\ > 1 & if\ MS\ is\ handoff \end{cases} \quad (2)$$

Making the rationale behind the priority is:

- MS that have a high mean data rate required, and have less data queued at the BS comes first.
- MS at the beginning of the queue have a better chance to be admitted to the cell.
- In case there is a handoff MS arrived it should be given a higher priority.
- Having the shortest job first will make the network accept more users.

Also, this factor can be used to give the priority to any kind of traffic, e.g., real time.

5. Simulation and Numerical Results

The performance of the proposed algorithm has been evaluated using a real time simulator, which was implemented using MATLAB. We present the results obtained from the simulation of the proposed scheme in different numerical scenarios and compared with some previous work in this field. We consider a CWRN of one cell, as shown in Figure 1. The cell contains three clusters that may overlap in their coverage area. New MSs requesting connection are arriving according to the Poisson arrival process. The distribution of the MSs places is random. Table 4 presents the parameters used in the simulator.

Table 4: Parameter list.

Parameter	Setting / Argument
Number of clusters	3
Cluster bandwidth (Mbps)	60–80–100
Arrival distribution model	Poisson
Arrival rate (λ)	2, 4, 6, 8, 10, 12, 14
Bit rate degradation step (W)	0.1
Data traffic queued at the BS destined for MS (Mb)	40

We should mention that all arriving mobiles are requesting a service and will release the channel and leave the cell after call completion so “call”, “user” and “mobile” will be used to refer to the same thing.

5.1 Performance Metrics

In each scenario, some or all of the call level performance measurements listed below were taken:
1. Blocking probability, which is the probability that the mobile requesting the service will not get it. It can be represented mathematically as the proportion of number of calls blocked to number of calls arrived.
2. Bandwidth utilization efficiency, which is the total data rates that are delivered to all mobiles in a network to the total bandwidth offered. It can be represented as $\frac{\sum_{q=1}^{Q} B_q - \sum_{q=1}^{Q} CAB_q}{\sum_{q=1}^{Q} B_q}$.
3. Average number of mobiles being served in the cell.
4. Mobile average sojourn time, which is the service time from entering the cell and starting the call until the completion of the call and leaving the cell.

5.2 Scenarios

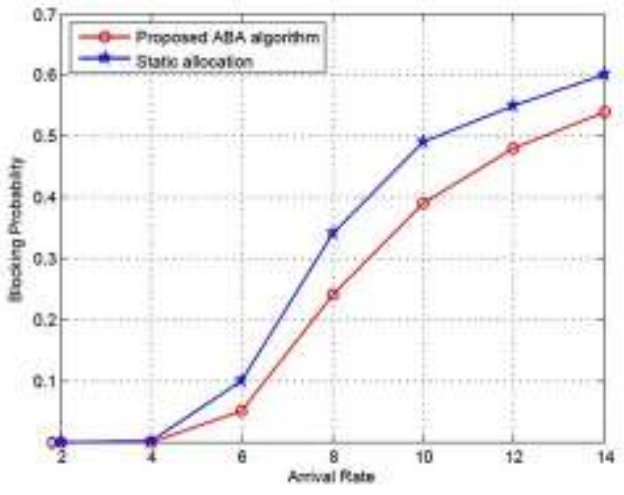
The proposed algorithm has been evaluated using the following numerical scenarios:
1. Proposed ABA vs. static allocation
2. Proposed ABA in two kinds of traffic.
3. The effect of the step size.
4. Priority queue efficiency.
5. With handoff arrival.

Scenario 1: To proof the efficiency of the proposed ABA, the performance of the proposed ABA algorithm has been compared with the static bandwidth allocation. We are going to consider one kind of traffic where the maximum, minimum, and required bit rate is shown in Table 5. The static allocation is based on the required bit rate, and if the mobile is under coverage of more than one cluster, the clusters will be sorted based on CAB in a descending order.

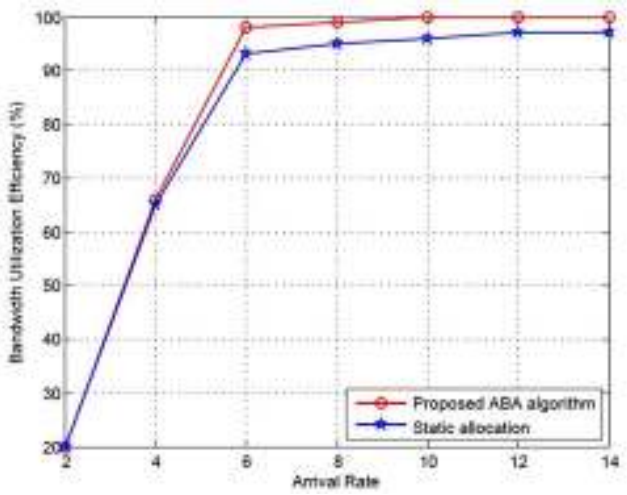
Table 5: Scenario 1 parameter list (Chen and Shen 2012).

Parameter	Setting / Argument
Traffic 1 maximum bit rate (kbps) (ABA Algorithm)	1024
Traffic 1 minimum bit rate (kbps) (ABA Algorithm)	256
Traffic 1 Required bit rate (kbps) (Static Algorithm)	512

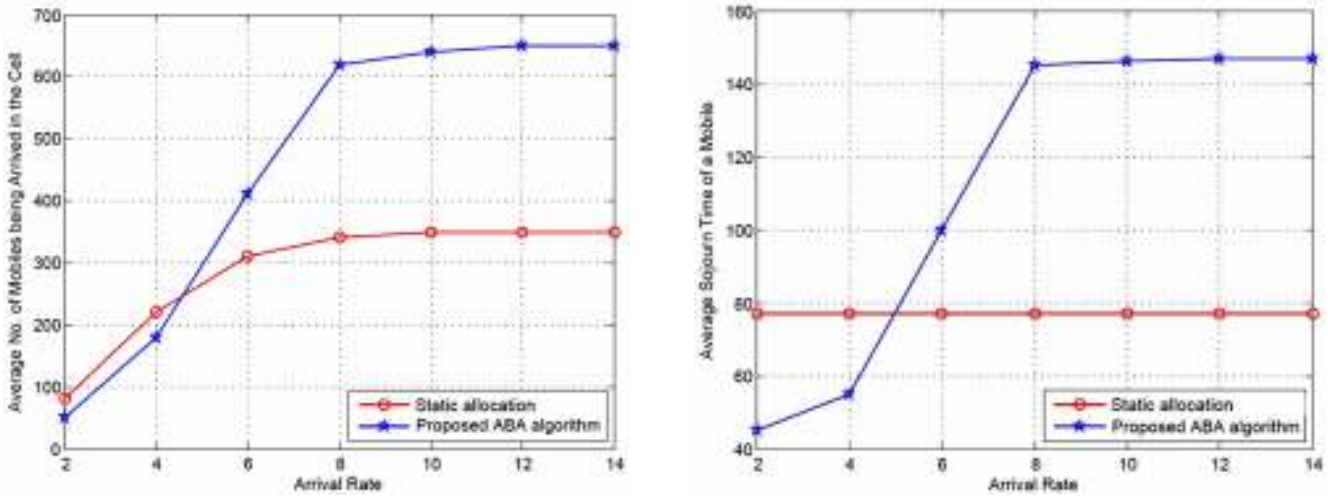
The blocking probability for both kind of bandwidth allocation is illustrated in Figure 3(a). It is shown that the call blocking probability is enhanced with more than 7% in the ABA algorithm compared with the static scheme. This result can be justified since the ABA algorithm uses the bit rate degradation algorithm according to the cell status in order to provide more bandwidth to admit more mobiles.
The bandwidth utilization efficiency with the proposed ABA algorithm is shown in Figure 3(b). As can be seen that the bandwidth utilization of the cell with the proposed ABA is more than 99% at the higher arrival rates. It is clear that the bandwidth utilization efficiency is enhanced with 2% in the ABA algorithm compared with the static scheme. This can be explained that more mobiles can be admitted into the cell using the proposed ABA algorithm to reduce the call blocking probability and the entire cell capacity can be utilized efficiently to maximize the cell throughput.
Figure 3(c) shows the average number of users in the cell at a certain moment. With light arrival rate, the average number of users in the cell using the static allocation is more than the users using the proposed ABA. This is due to the proposed algorithm utilization and load balancing mechanism, where in the light load cell the users are given the maximum bit rate possible and that will cause the reduction in the average sojourn time as shown in Figure 3(d). As the arrival rate increases and the cell gets congested, the bit rate degradation algorithm starts working. Eventually the bit rate of the mobiles decreases to the minimum acceptable limit, making the average sojourn time increases without SLA violation.



a. Blocking Probability



b. Bandwidth Utilization Efficiency



c. Average No. of Mobiles

d. Average Sojourn Time

Figure 3. The proposed ABA algorithm vs. static allocation on scenario 1.

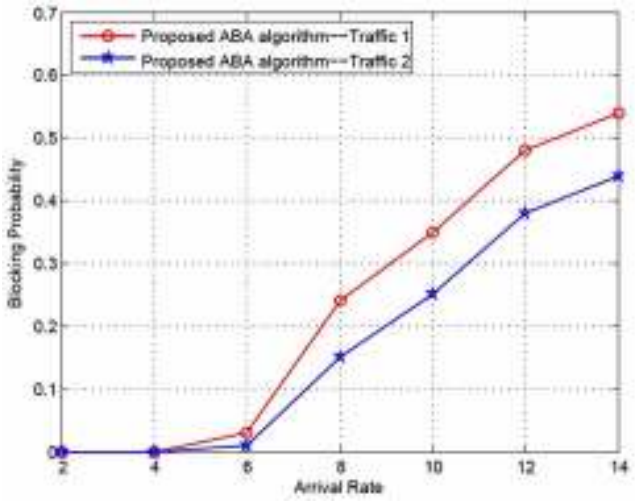
Scenario 2: In this scenario, the proposed ABA has been tested under two kinds of traffic. These traffics has different bit rate requirement as shown in Table 6.

Table 6: Scenario 2 parameter list (Chen and Shen 2012).

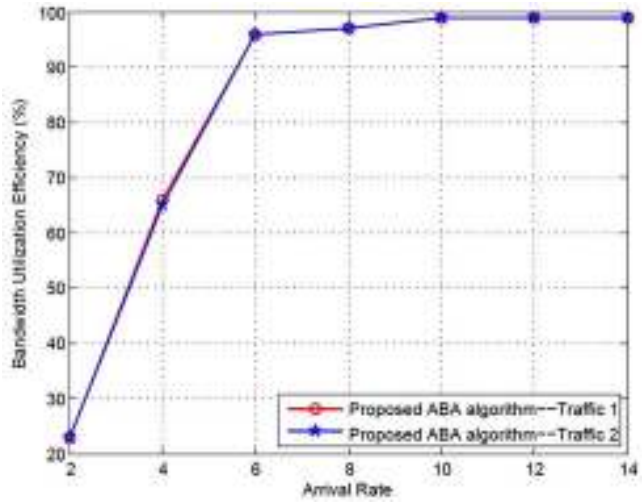
Parameter	Setting / Argument
Traffic 1 maximum bit rate (kbps)	1024
Traffic 1 minimum bit rate (kbps)	256
Traffic 2 maximum bit rate (kbps)	1024
Traffic 2 minimum bit rate (kbps)	128

The blocking probability for each traffic with the proposed ABA is shown in Figure 4(a). It can be seen that the blocking probability for Traffic 2 is lower than that of Traffic 1 by 10%. This can be explained that the Traffic 2 can reach lower transmission rate levels with the minimum transmission rate 128kbps, so Traffic 2 is characterized by larger adaptation for new calls compared with the Traffic 1. Since the proposed ABA is adaptive to different kinds of traffics, the bandwidth utilization efficiency is the same in both traffics as illustrated in Figure 4(b). Since Traffic 2 is characterized by larger adaptation for new calls compared with the Traffic 1, it is shown in Figure 4(c) that as the arrival rate increases and the bit rate degradation algorithm starts working, the average number of mobiles currently in the network is enhanced in Traffic 2 compared with Traffic 1. Figure 4(d) Illustrates the average sojourn time for the proposed ABA using different kinds of traffic. Initially, with arrival rate is 2, all mobiles from both traffics have the same performance because they have the same bit rate upper bound and the cell is not yet congested. As the arrival rate increases and the network becomes more congested, the bit rate of the mobiles eventually decreases to the minimum limits using the bit rate degradation algorithm. Since Traffic 2 mobiles have

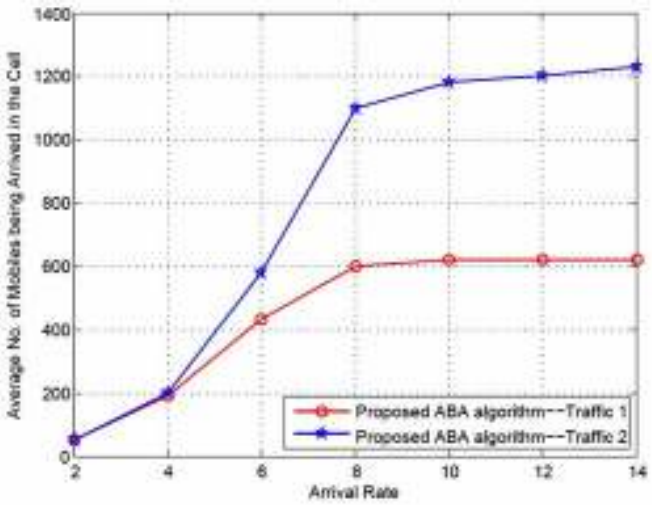
lower limit, they will free up more recourses, and thus accommodate more users and decrease the rejection rate. All that is at the expense of the sojourn time.



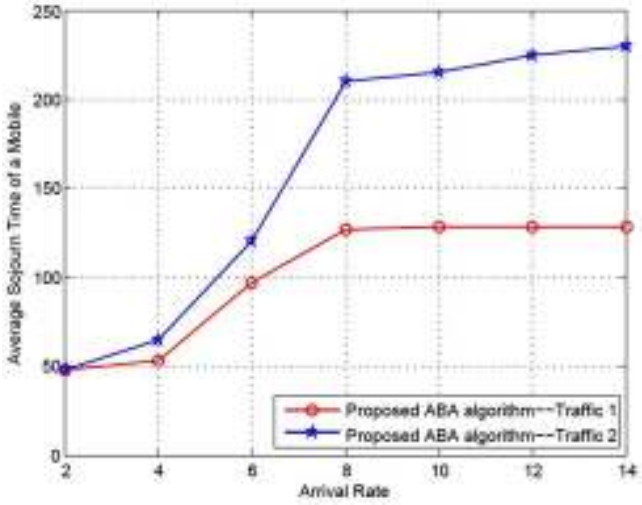
a. Blocking Probability



b. Bandwidth Utilization Efficiency



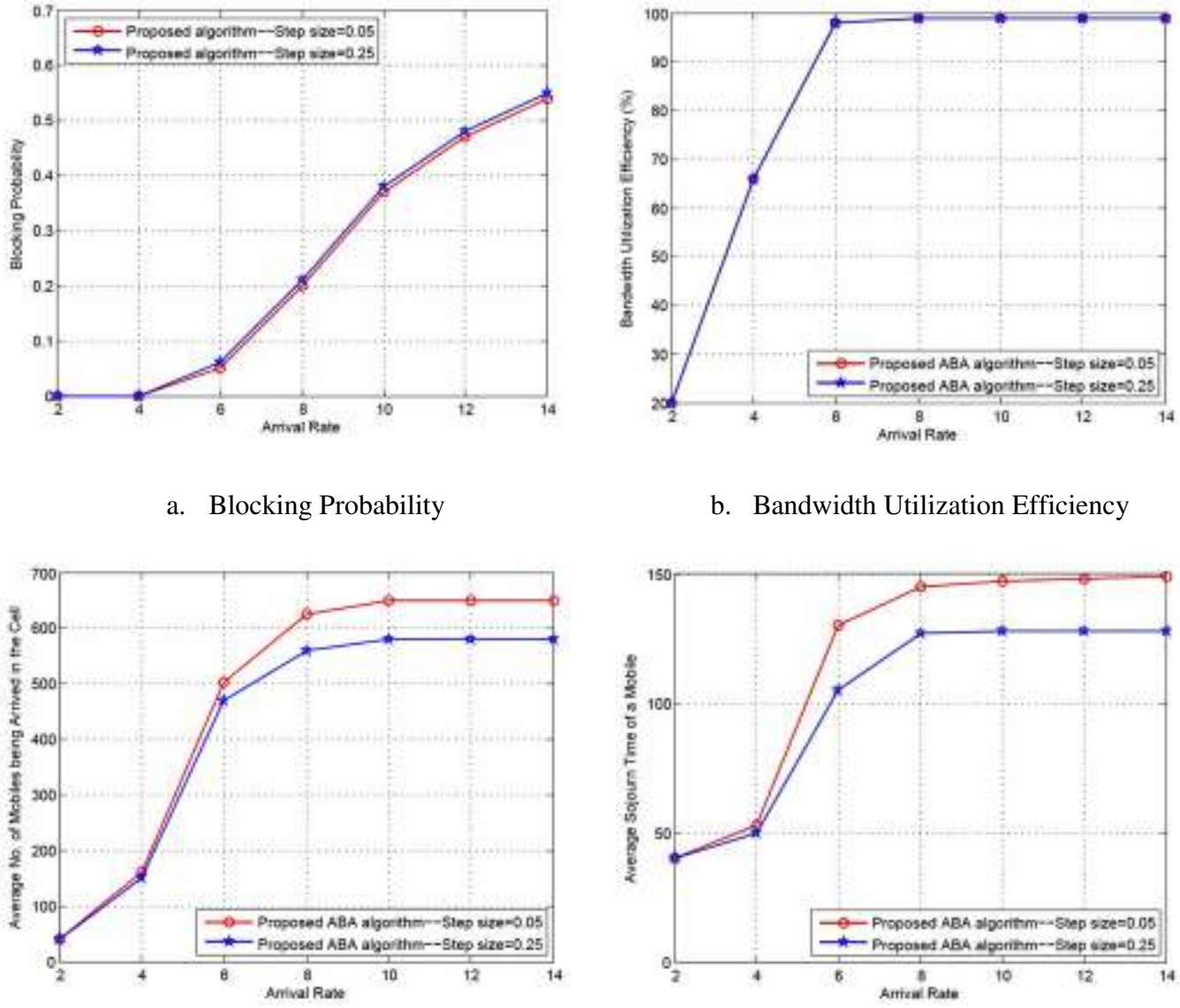
c. Average No. of Mobiles



d. Average Sojourn Time

Figure 4. The proposed ABA algorithm using different kinds of traffic--Scenario 2.

Scenario 3: In this scenario, the performance of the proposed ABA is compared when the bit rate degradation algorithm is performed using different step sizes. One kind of traffic is considered. The parameter list is shown in Table 7.



a. Blocking Probability

b. Bandwidth Utilization Efficiency

c. Average No. of Mobiles

d. Average Sojourn Time

Figure 5. The proposed ABA algorithm using different step sizes—Scenario 3.

Table 7: Scenario 3 parameter list.

Parameter	Setting / Argument
Bit rate degradation step (W)	0.05, 0.25
Traffic 1 maximum bit rate (kbps)	1024
Traffic 1 minimum bit rate (kbps)	256

Figure 5(a) illustrates the blocking probability for the proposed ABA algorithm using different step sizes. It can be concluded that the blocking probability of the cell is enhanced by only 1% using a small step size of 0.05 compared with 0.25 step size. This is due to better load distribution. Comparing with 0.25 step size, using 0.05 step size, the mobiles can support more transmission rate levels using the bit rate degradation algorithm. The improvement will get better if there was no threshold to limit the number of steps performed. Figure 5(b) illustrates the bandwidth utilization efficiency for the proposed ABA using different step sizes. It is clear that the bandwidth utilization efficiency of both step sizes is similar because the same proposed ABA algorithm is used. The only difference when using different

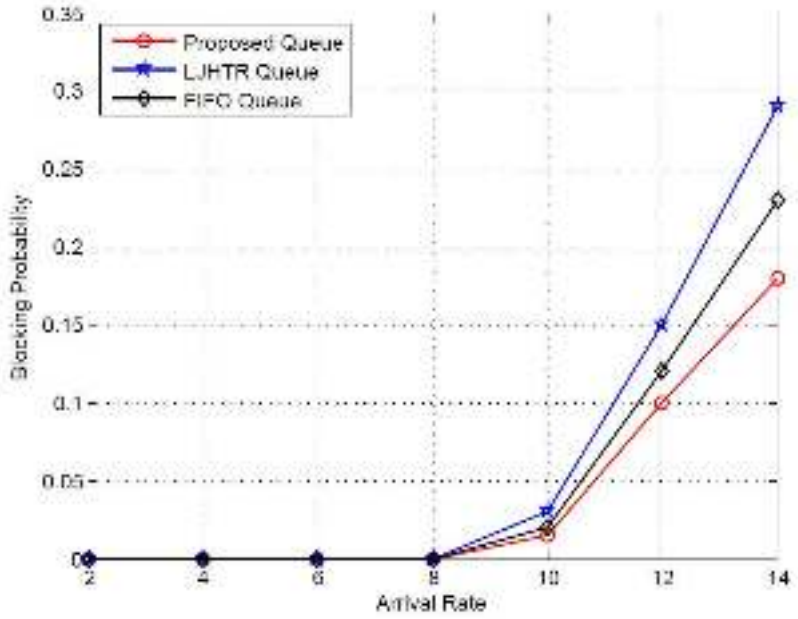


Figure 6. Blocking probability for the proposed ABA under different kinds of queues.

step sizes is in the bit rate distribution among the mobiles. In general, smaller step size will lead to a better performance in terms of the bit rate distribution and blocking probability. Figure 5(c) illustrates the average number of mobiles for the proposed ABA algorithm using different step sizes. It is shown that the average number of mobiles in the network is enhanced using 0.05 step size compared to the number of mobiles using 0.25 step size. Since the smaller step size will support more transmission rate levels, this will lead to freeing up more resources to accommodate more users. On the other hand, the average time for each mobile to complete its transmission is worsen using 0.05 step size compared to the average time using 0.25 step size

as shown in Figure 5(d). This is because that the bit rate dedicated to each mobile is less. The differences between these two cases increase as the arrival rate increases and the network becomes more congested.

Table 8: Scenario 4 parameter list.

Parameter	Setting / Argument
Traffic maximum bit rate (kbps)	1024
Traffic minimum bit rate (kbps)	(128–256)
Data traffic queued at the BS destined for MS (Mb)	2-40

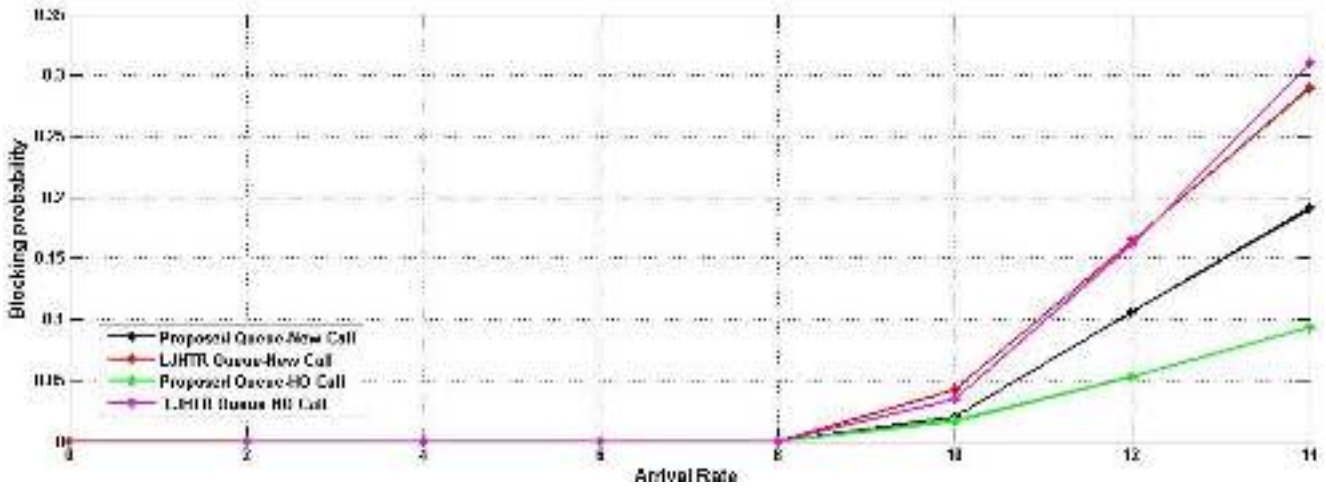
Scenario 4: In this scenario, the efficiency of the different scheduling queues has been investigated. The performance of our proposed queue is compared with the performance of two other scheduling queues which are: First In First Out (FIFO) and the Large Job High Transmission Rate (LJHTR) (Lu and Liao 2012).

To proof the adaptability of the proposed scheme, a mixed traffic is considered. Also, the amount of data that is designated to be downloaded for each mobile will be a random variable between 2 and 40 Mb as shown in Table 8. The same proposed ABA is going to be used with all queues. Figure 6 illustrates the blocking probability for the proposed ABA under different kinds of queues. It can be concluded that the blocking probability of the proposed queue is the lowest. This is due to the given priority to the shortest jobs with highest transmission rate. Having less blocking probability means high number of admitted mobiles. The bandwidth utilization efficiency is similar for all queues since we are using the same proposed ABA algorithm.

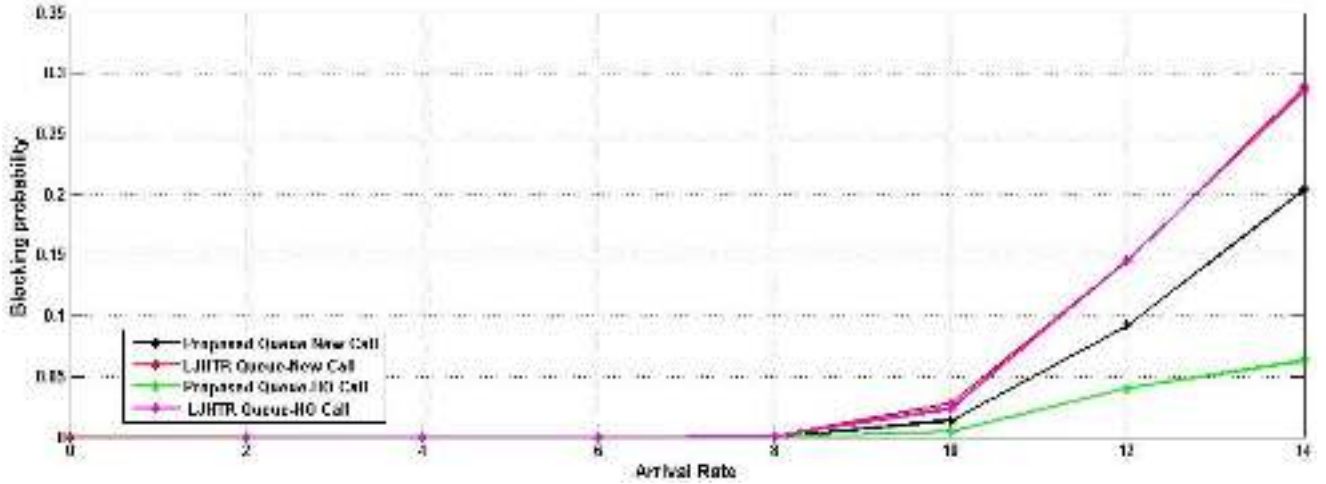
Scenario 5: In this scenario, the handoff mobiles are given a higher priority in the queue such that the hand off blocking probability will be decreased. This priority is controlled by the handoff priority factor. Parameters from Table 8 are being used. In addition, it is assumed that 20% of the arrival rate is a handoff traffic while 80% of the arrival rate is a new arrival.

Figure 7(a) illustrates the blocking probability of both new and handoff arrivals when the handoff priority factor is 2. It is shown that when setting the handoff priority factor to 2, the handoff blocking probability can reach up to 0.09, while the new call blocking probability = 0.19. That makes the handoff blocking probability less than the new call blocking probability by 10%. That is due to the priority given to the handoff mobiles. The blocking probability of LJHTR is the same for both handoff and new calls because there is no priority given over.

Figure 7(b) shows the case when the handoff priority factor is increased to 4. The handoff blocking probability will decrease to 0.06 while the new call blocking probability will be increased to 0.21. That makes the handoff blocking probability less than the new call blocking probability by 15%. Due to the higher priority factor given in this case, the handoff blocking probability will enhance, causing the new call blocking probability to increase. In conclusion, if the handoff priority factors increase, the blocking probability of the handoff calls will improve at the expense of the new call blocking probability.



a. When handoff priority factor =2



b. When handoff priority factor =4

Figure 7. Blocking probability of both new and handoff arrivals--Scenario 5.

6. Conclusion

In this paper, we proposed an algorithm for load balancing in CWRN cell that contains number of clusters. The algorithm uses CAC to prevent the network congestion before happening, CAC restrict the access to network resources unless the QoS parameter for all users is guaranteed. Moreover, we proposed an ABA algorithm that uses a data rate degradation algorithm. In addition to the high level of resource utilization

provided, the network can accommodate more users and reduce the call blocking probability. We have studied how the performance of the proposed algorithm was affected by varying degradation step size and traffic kind, we have seen an improvement in the rejection rate. Also the proposed scheme contains some kind of priority mechanism, which also will improve the handoff call dropping probability.

As future work, many opportunities have been observed, such as working on a data rate upgrading algorithm that will improve the ongoing calls if the situation of the congestion relief happened. Also the idea of hybrid channel assignment might be considered, where each cluster has a static set of channels and can dynamically borrow additional channels.

References

Baldini, G., Karanasios, S., Allen, D. and Vergari, F. (2014). Survey of Wireless Communication Technologies for Public Safety. IEEE Communications Surveys & Tutorials, 16(2), pp. 619–641.

Chen, G. and Shen, L. (2012) An Adaptive Bandwidth Allocation Algorithm Based on the Multi - thresholds in Heterogeneous Wireless Network. IEEE International Conference on Consumer Electronics - (ICCE-Berlin), pp. 190-194.

Chou, C,T. and Shin, K, G. (2002) Analysis of Combined Adaptive Bandwidth Allocation and Admission Control in Wireless Networks. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, (2), pp. 676-684.

Chou C, T. and Shin, K, G. (2004) Analysis of Adaptive Bandwidth Allocation in Wireless Networks with Multilevel Degradable Quality of Service. IEEE Trans. Mob. Comput, 3(1), pp. 5-17.

Chowdhury, M, Z., Jang, Y, M. and Haas, Z, J. (2013) Call Admission Control Based on Adaptive Bandwidth Allocation for Multiclass Services in Wireless Networks. J. Commun. Networks, 15(1), pp. 15-24.

Cuthbert, L., Xu, L., Chen, Y. and Gao, Y. (2011) A self-optimizing load balancing scheme for fixed relay cellular networks. IET Int. Conf. Commun. Technol. (ICCTA), pp. 306-311.

Jiang, F. and Wang, B. (2011) A Load Balancing Relay Selection Algorithm for Relay Based Cellular Networks. 7th Int. Conf. Wirel. Commun. Netw. Mob. (Comput), pp. 1-5.

Laneman J N., Tse, D, N, C, and Wornell, G, W. (2004). Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior. IEEE Trans. Inf. Theory, 50(12), pp. 3062-3080.

Lu, N. and Bigham, J. (2005) Utility-based Adaptive Bandwidth Allocation for Multi-Class Traffic in Wireless Networks. The 19th International Teletraffic Congress (ITC 19), pp. 879-888.

Lu, H. and Liao, W. (2012) Cooperative Strategies in Wireless Relay Networks. IEEE J. Sel. Areas Commun, 30(2), pp. 323-330.

Sadek, K,. Su, W. and Liu, K, J, R. (2005) Clustered Cooperative Communications inWireless Networks. GLOBECOM - IEEE Global Telecommunications Conference, 3.

Soldani, D. (2009). Multi-Hop Relay Networks. In Upena D Dalal and Y P Kosta (Ed). WIMAX New Developments (pp. 409-428). Retrieved from http://www.intechopen.com/books/wimax-new-developments/multi-hop-relay-networks

Wang, W., Chen, C., Guo, Z., Cai, J. and Shen, X, S. (2008) Isolation band based frequency reuse scheme for IEEE 802.16j wireless relay networks. 5th International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness.

Xu, L., Chen, Y. and Gao, Y. (2010) Self-organizing load balancing for relay based cellular networks. 10th IEEE International Conference on Computer and Information Technology (CIT), pp. 791-796.

Yang, C., Mai, Y., Lin, I. and Chen, J, Y. (2014) Adaptive Zone-based Bandwidth Management. The IEEE 802.16j Multi-hop Relay Network, J. Internet Technol., 15(2), pp. 163-173.

Yang D., Fang X., Xue G., and Tang, J. (2010) Relay station placement for cooperative communications in WiMAX networks. GLOBECOM - IEEE Global elecommunications Conference I, pp. 1-5.

Yu, Y., Hu, R, Q,. Bontu, C, S, and Cai, Z. (2011) Mobile association and load balancing in a cooperative relay cellular network. IEEE Commun. Mag, 49(5), pp. 83-89.