A Markov Decision Process (MDP) based Load Balancing Algorithm for Multi-cell Networks with Multi-carriers

Janghoon Yang[1]

[1]Department of Newmedia, Korean German Institute of Technology
661 Deungchon-dong, Kangseo-gu, Seoul 157-030, Korea
[e-mail: jhyang@kgit.ac.kr]

*Corresponding author: Janghoon Yang

Received November 18, 2013; revised June 23, 2014; accepted July 29, 2014; published October 31, 2014

Abstract

Conventional mobile state (MS) and base station (BS) association based on average signal strength often results in imbalance of cell load which may require more powerful processor at BSs and degrades the perceived transmission rate of MSs. To deal with this problem, a Markov decision process (MDP) for load balancing in a multi-cell system with multi-carriers is formulated. To solve the problem, exploiting Sarsa algorithm of on-line learning type [12], α -controllable load balancing algorithm is proposed. It is designed to control tradeoff between the cell load deviation of BSs and the perceived transmission rates of MSs. We also propose an ε -differential soft greedy policy for on-line learning which is proven to be asymptotically convergent to the optimal greedy policy under some condition. Simulation results verify that the α -controllable load balancing algorithm controls the behavior of the algorithm depending on the choice of α . It is shown to be very efficient in balancing cell loads of BSs with low α .

Keywords: MDP, dynamic programming, load balancing, multi-carriers, multi-cells

http://dx.doi.org/10.3837/tiis.2014.10.006

1. Introduction

With surge of smartphones, wireless data traffic has been increasing explosively. Network operators install more base stations(BSs) to increase network capacity and satisfy quality of service (QoS) for different types of data traffic. An alternative way to deal with this problem is to use more bandwidth. Multi-carrier operation or carrier aggregation for long term evolution advanced (LTE-A) is introduced to support efficient wideband usage [13]. As the number of BSs and the number of used frequency bands increase, load balancing takes a critical role to use wireless resource efficiently [14].

Load balancing (LB) algorithms have been developed for various types of wireless network with multiple carriers. For a code division multiple access (CDMA) system, load balancing algorithms were developed for distributing voice traffic over multiple carriers in the same frequency band [1] and in frequency bands of 800MHz and 1900MHz [2]. Joint scheduling and load balancing was also proposed for the efficient transmission of data traffic over multiple carriers in a CDMA network [3]. Son and et. al. proposed a simple dynamic load balancing algorithm which jointly optimized fractional frequency reuse and load balancing [4]. Classifying access states as unloaded/balanced/overloaded, a cell association algorithm for wireless local area network (WLAN) was experimentally shown to increase the total wireless network throughput [5]. Joint scheduling and load balancing for the LTE-A system with carrier aggregation was developed to use radio resources distributed in frequency and space [6].

One of efficient methods to derive load balancing algorithm is to use dynamic programming (DP) which is a set of system equations defined over a problem composed of states, actions, and rewards [15]. A dynamic load balancing self-stabilizing distributed pseudo-tree optimization procedure (DLB-SDPOP) to balance the load of WLAN which exploits multiagent constraint optimization based on dynamic programming was shown to provide robust performance in a dynamically changing environment [8]. DP-based load balancing algorithm can be found to be applied to several different contexts outside of wireless network. DP-based power control for internet servers and data centers which implicitly balances loads from observing the system load and thermal status was formulated to capture the power-delay tradeoff [9]. Similarly, DP-based algorithm for turning on and off geometrically distributed servers for content delivery networks (CDNs) was developed to maximize energy reduction while minimizing the impact on client-perceived service availability [10].

There are also a few researches which are implicitely related to load balancing based on DP. For a heterogeneous network consisting of WLAN and 3G network, DP for markov decision process (MDP) formulated from a user-network association was solved using value iteration [11]. Even though it was characterized to particular scenario with an explicit model, it showed the potential of DP approach for load balancing in a cellular network. Similarly MDP-based vertical handoff algorithm was proposed with the objective function of maximizing the expected reward which is composited by bandwidth, delay and signaling cost [7]. However, it focused on the maximization of total reward rather than load balancing itself.

In this paper, we propose an load balancing alogirthm for multi-cells with multi-carriers called "α-controllable load balancing algoithm" which explicitely allows the network to control tradeoff between the cell load deviation of BSs and the perceieved transmission rates of MSs with choice of α. To this end, we define MDP for load balancing and formulate DP. To solve the DP without an explicit model, Sarsa learning algorithm [12] of on-line type is exploited. We also propose an ε- differential soft greedy policy to make it robust to dynamicity of networks which is proved to be optimal in the sense that it converges to greedy policy under some conditions. Simulation results verify the efficiency of the proposed algorithm.

The following notations are introduced for the rest of the paper. $E[\cdot]$ denote the expectation. $\min(\cdot)$ and $\max(\cdot)$ are the minimization and maximization of the function in the parenthesis respectively. Boldface lowercase letters imply vectors. Finally, $|A|$ denotes the number of elements in the set A, while $|a|$ is the absolute value of a.

The remainder of this paper is organized as follows. The basic assumptions and system models are made in section II. In section III, MDP for load balancing is defined and corresponding DP is formulated. To solve the DP for load balancing without an explicit model, α-controllable LB algorithm with ε- differential soft greedy policy which exploits Sarsa learning algorithm is proposed, and its convergence is analyzed in section IV. To evaluate the performance of the proposed algorithm, simulation results are provided in section V. Finally, we make conclusions and comments on remaining issues for future research in section VI.

2. System Model

We consider a downlink multi-cell system with frequency reuse-1 where it uses multi-carriers to increase capacity with larger bandwidth. Different carriers can be adjacent or located at totally different bands. Even though the system model itself is transparent to how carriers are allocated to frequency bands, this research focuses more on the multi-carriers over totally different bands in the simulation. For the ease of analysis without losing generality much, we make several assumptions which limit the scope of this research and interpretation of research results.

First, MSs have initial association based on signal strength with a single BS and a single carrier, which is prevalent in a conventional cellular network. Second, all BSs are perfectly synchronized for all carrier frequencies, and all MSs have perfect synchronization with BSs. In a real synchronization system, all MSs and BSs work with tolerable synchronization error which scarcely degrades performance. Third, there exists a centralized processor which gathers perfect information on SINRs of all BS and all carriers, and the number of MSs served by each BS and each carrier, and executes load balancing algorithm based on it. In practice, even though it may not be possible to have perfect information, a centralized processor may have estimated and delayed information. Since the proposed algorithm use information depending on the average characteristics of channels and systems, delayed information may not have significant effect on performance. In addition, SINR estimation error may not be significant as long as the large number of samples are averaged out.

For a mutli-cell downlink system with multi-carriers, the SINR $\xi_b^f(k)$ of MS k in association with BS b and carrier f can be expressed as

$$\xi_b^f(k) = \frac{g_b^f(k)}{\sum_{b'=1, b'\neq b}^{B} g_{b'}^f(k) + \sigma_n^2} \tag{1}$$

where $g_b^f(k)$ is the received signal power at MS k from BS b with carrier f, B is the number of BSs for each carrier, and σ_n^2 is thermal noise power at each BS. In a multi-carrier network where there are large frequency gaps between different bands, MSs may have more frequent association with the carriers of lower frequency which have less propagation loss. In addition, the number of BSs supporting each carrier may not be the same for all carriers. In these contexts, there can be imbalance of the number of MSs among BSs. Thus reassociating MSs over BSs properply can reduce the imbalance of the load of BSs and improve the MS perceived transmssion rates of MSs.

3. Problem Formulation

In this section, we formulate the MDP for load balancing and define DP from MDP to find a proper load balancing which provides a tradeoff between the minimum perceived transmission rate of MSs and the cell load deviation of BSs depending on a parameter.

3.1. Formulation of Markov Decision Process for Load Balancing

MDP formulation is usually described by states, actions, transition probability and associated rewards. At each predefined period, a centralized processor executes a LB algorithm. To make problem simple, it is assumed that it can execute a forced hand off (HO) for at most one MS. That is, it chooses a candidate MS for the forced HO and decides whether it will execute the forced HO or not.

The central processor decides an action from an action space A. The elements of A in this problem are associated with forced handoff decision. Decision of action depends on a current state from a state space S. The state information for load balancing can be identification number for each BS, perceived transmission rate of all MSs, number of MSs for each cell, and etc. States describe the system environment, which has effect on deciding action for each decision stage. Since forced HO can change loads of BSs and minimum perceived transmission rates of MSs, state $\mathbf{s}_{k+1}$ at the decision stage $k+1$may depend on the current state $\mathbf{s}_k$ and action a_k taken at stage k. The state transition probability $P[\mathbf{s}_{k+1} \mid \mathbf{s}_k, a_k]$ completely describes interaction among current state, currently chosen action, and next state.

The transition from the current state with a selected action incurs reward $g(\mathbf{s}_k, a_k)$. Depending how this reward is defined, the objective of the problem can be different. For the load balancing problem, it can be related to transmission rate, the number of MSs served by each BS, signaling cost, degree of congestion of wired connection with core network, or degree of occupying processors. The actual rewards depend on how to choose an action for a given state. A sequence of controls u_k which map states $\mathbf{s}_k$ into $a_k = u_k(\mathbf{s}_k)$ is called a policy which is denoted by $\pi = \{u_0, u_1, \cdots, u_{N-1}\}$. For a given policy, one can define the expected total reward of the policy as

$$J^\pi(\mathbf{s}_0) = \lim_{N\to\infty} E\{\sum_{k=0}^{N-1} \gamma^k g(\mathbf{s}_k, a_k) \mid \mathbf{s}_0\} \tag{2}$$

where $\mathbf{s}_0$ is an initial state, and γ is a discount factor with $0 < \gamma < 1$ which controls the contribution of future rewards to the current reward. Depending on the transition probability and the policy, $J^{\pi}(\mathbf{s}_0)$ is determined. One can determine the optimal policy for the expected reward as follows

$$\pi^* = \arg_{\pi} \min J^{\pi}(\mathbf{s}_0) \ \ for \, \forall \mathbf{s}_0 \tag{3}$$

When the it is a stationary process, optimal policy can be determined from the corresponding optimal control u^* as $\pi^* = \{u^*, u^*, \cdots, u^*\}$

3.2. Dynamic Programming for Load Balancing

In this subsection, we explicitly define the elements of the MDP for load balancing in a multi-cell network with multi-carriers. States, actions, and reward are defined in more detail. DP for the embodied MDP is also formulated to derive an algorithm from it.

We consider two different types of states. The state space of the first type S_8 is defined as

$$S_8 = \{(s_1, s_2, \cdots s_8) \mid s_1, s_5 \in \{1,2,\cdots B\}, s_2, s_6 \in \{1,2,\cdots F\}, s_3, s_7 \in W, s_4, s_8 \in H\} \tag{4}$$

where F is the number of carriers in a cell, s_1 and s_2 are the indices of serving BS and serving carrier respectively, s_3 and s_4 are states associated with the minimum perceived transmission rate of MS and the cell load associated with the serving BS, s_5, s_6, s_7 , and s_8 are correspondingly defined for the target BS. W and H have N_w and N_h elements which are associated with quantized minimum perceived transmission rate of MS and cell load of BS respectively. One can express the states of the minimum perceived transmission rate and the cell load in numerous ways. Rather than setting the states as values themselves, we make them represent a normalized form in the following way.

$$w = \frac{\overline{m}_A}{(A_{b(t)}^{f(t)} + \overline{m}_A)}, h = \frac{N_{b(t)}^{f(t)}}{(N_{b(t)}^{f(t)} + \overline{m}_N)} \tag{5}$$

where $t = 0$ and $t = 1$ represent serving BS and target BS respectively, $b(t)$ and $f(t)$ are the indices of BS and carrier for a given t, $A_{b(t)}^{f(t)}$ is the minimum perceived transmission rate of MS in the BS $b(t)$ with carrier $f(t)$, and $\overline{m}_A$, and $\overline{m}_N$ are the averages of minimum perceived transmission rates and cell loads. One can easily verify that these quantities range from 0 to 1, which makes it possible to define quantized states more simply. With these quantities, the state space of the second type S_2 can be defined as follows.

$$S_2 = \{(s_1, s_2) \mid s_1, s_2 \in R\} \tag{6}$$

where the elements of the set R is the discrete levels of the composite value r which is defined to be

$$r = \alpha w + (1-\alpha)h \tag{7}$$

Since both w and h range from 0 to 1, r also has value within the same range. It is noted that the number of feasible states for S_8 is significantly larger than S_2 as long as N_r, the number of elements in R has similar value as N_w or N_h. Efficiency of simplified state space S_2 will be addressed further with simulation results.

When the load balancing algorithm operates, it changes MS/BS association or remains unchanged depending on the action and the current state. Since it is assumed that there can be at most one forced HO at each decision stage for simplicity, the control space, A can be defined as

$$A = \{YES, NO\} \tag{8}$$

where *YES*, and *NO* indicate whether the algorithm executes the forced HO at the decision stage or not.

The action chosen under current state incurs reward. Defining the reward structure for DP usually depends on the object of control. One of most important factors defining load will be the number of MSs served by each BS. When MSs are evenly associated with each BS, it can be considered to be a perfect state for load balancing in the perspective of cell load. While this scheme may evenly balance the load among BSs, balancing load may degrade the perceived transmission rate of MSs. It will be particularly more problematic when MSs are not uniformly located over service region, and radio channel environments are different among BSs. Thus it will be advantageous to have a load balancing algorithm such that it can provide a tradeoff between the cell load deviation of BSs and the perceived transmission rates of MSs with a parameterization. To realize this characteristic, reward is defined as follows.

$$g(\mathbf{s}_k, a_k) = \alpha \frac{\min_{b,f} A_b^f(k)}{\overline{m}_A} + (1-\alpha)\frac{\min_{b,f} N_b^f}{\max_{b,f} N_b^f} \tag{9}$$

Controlling α can make the reward provide a tradeoff between the cell load deviation of BSs and the perceived transmission rates of MSs. When $\alpha = 1$, the object of DP is to maximize the minimum perceived transmission rate, while it is to equalize the loads of BSs with $\alpha = 0$.

For the optimization problem of (2) formulated from MDP, a set of system equation for optimality can be developed as [15]

$$J^*(\mathbf{s}) = \max_{a\in A}[E\{g(\mathbf{s},a) + \gamma \sum_{\mathbf{s}'} P[\mathbf{s}' | \mathbf{s}, a] J^*(\mathbf{s}')\}] \quad for\, \forall \mathbf{s} \in S \tag{10}$$

This is a Bellman equation for MDP, which is often called as "Dynamic Programming". This set of system equations are usually solved with value iteration or policy iteration [15]. However it requires the knowledge of state transition probability. This probability can be analytically calculated for some specific simplified model [11]. However, it is very difficult to find one for a general cellular system. Thus, rather than finding the explicit model of state transition probability, a model free approach will be used throughout this paper.

4. α - Controllable Load Balancing Algorithm

When information on the model of MDP is not available, learning algorithms are often used to learn the model from observation [12]. In this section, we develop Sarsa-learning based LB algorithm to be applicable to any cellular environment with on-line learning. Since the reward function in (9) depends on the parameter α, we call it as α-controllable LB algorithm throughout this paper.

4.1. α - Controllable LB Algorithm

While value in (2) is defined for each state, the value of action in association with states is known as the Q-factor which is defined as

$$Q^*(\mathbf{s},a)=g(\mathbf{s},a)+\gamma\sum_{\mathbf{s}'}P[\mathbf{s}'|\,\mathbf{s},a]\min_{a'}Q^*(\mathbf{s}',a')] \tag{11}$$

By definition, it requires knowledge on the state transition probability. One of well known Q -factor approximation algorithms is Sarsa algorithm which updates Q value and policy continuously [12]. The Sarsa algorithm for (11) can be expressed as [12]

$$Q(\mathbf{s}_k,a_k)=(1-\beta_k)Q(\mathbf{s}_k,a_k)+\beta_k(g(\mathbf{s}_k,a_k,\mathbf{s}_{k+1})+\gamma Q(\mathbf{s}_{k+1},a_{k+1}) \tag{12}$$

where β_k is a step size which needs to be properly selected to converge to 0 with a proper rate. When implementing Sarsa algorithm, it observes reward associated with the current action and the subsequence stat. Then, it decides next action based on the current policy. The policy is usually designed so that it can provide tradeoff between exploration and exploitation. ε -greedy, or soft max policy [16] are often used for Sarsa algorithm. However, the ε -greedy policy does not take into account on Q value while the soft max policy may take random action with relatively large probability when maximum Q value is not significantly larger than others. To overcome issues in the existing policy, we propose ε -differential soft greedy policy of which ε varies with control stage k as follows

$$d_k=(1-\tau)d_{k-1}+\tau\left|\frac{Q_{new}-Q_{old}}{Q_{new}}\right| \tag{13}$$

$$\varepsilon_k=1-e^{-ad_k} \tag{14}$$

where τ is a time constant for averaging the normalized difference of Q value, Q_{old} is the Q value of the state and the action visited at the control stage k while Q_{new} is the update of Q_{old} with equation (12), and a is a parameter with positive value which controls the effect of the normalized difference of Q on ε. As a gets close to 0, it takes exploitation more often, while it does exploration more often with large a. Whatever a may be, as Q value gets stable, ε will get smaller. In a cellular network, it is usually non-stationary environment since some new MSs may appear and some existing MSs may close connection. This may necessitate changes in Q value in response to change in system environment. In this case, the ε -differential soft greedy policy is likely to increase chance of exploration accordingly.

Finally, we need to define how to determine a candidate MS for forced HO, and target BS and carrier. r in (7) has relatively high value when the minimum perceived transmission rate of MSs is low, and the number of MSs served by BS is larger. Since r can be calculated for each BS, we determine the BS and the carrier serving the candidate MS for the forced HO first, which has the largest r. Then, the candidate MS is determined to be one which has the minimum perceived transmission among MSs served by the BS and the carrier. Similarly the target BS and the carrier is determined such that they have the smallest r.

It is noted that the proposed algorithm is transparent to how many MSs are selected for forced HO at each stage, even though only one MS was considered for the implementation in this paper. For the practical implementation of this algorithm, the execution period of the algorithm need to be set properly. It will depend on the length of transmission slot, the computational capability of a system, backhaul bandwidth, dynamicity of data traffic, and etc. Alternatively one may change the number of MSs for forced HO at each stage. Thus, the execution period of the proposed algorithm is likely to be implementation and environment specific. Since, we focus on the characterization of the proposed algorithm, the optimized determination of this period is left for future research.

4.2. Convergence of ε-differential soft greedy policy

Since the proposed α-controllable LB algorithm is Sarsa type learning algorithm, Q value can converge to optimal value if the learning policy optimally chooses an action as the number of decision stages approaches to infinity. The Sarsa type algorithm is known to converge if the learning policy is greedy in the limit with infinite exploration (GLIE) [12]. [12] proved that Bolzmann policy and ε-greedy policy are GLIE for some specific conditions. Exploiting GLIE, the convergence of the α-controllable LB algorithm with ε-differential soft greedy policy can be expressed in the following way.

Proposition 1*: ε-differential soft greedy policy is GLIE when there exists l' such that* $d_k \geq -\frac{1}{a}\log(1-\frac{c}{\min_{s,a} n(\mathbf{s},a)})$ *for* $\min_{s,a} n(\mathbf{s},a) \geq l'$.

Proof: Let $n(\mathbf{s},a)$ be the number of visits to the state $\mathbf{s}$ with choice of action a. The probability of choosing non-optimal action at control stage k can be expressed as $(1-e^{-ad_k})/|A|$. It is well known that the sum of sequence c/l where c is an arbitrary positive constant approaches to infinite as the number of summation terms goes to infinite. Exploiting this property, the ε-differential soft greedy policy will be GLIE when there exist l' such that $\frac{1-e^{-ad_k}}{|A|} \geq \frac{c}{n(\mathbf{s},a)} \geq \frac{c}{l'}$ for all $(\mathbf{s},a)$ pairs. It can be readily calculated that it is the case when there exists l' such that $d_k \geq -\frac{1}{a}\log(1-\frac{c}{\min_{s,a} n(\mathbf{s},a)})$ for $\min_{s,a} n(\mathbf{s},a) \geq l'$. ■

Even though the proposition-1 proves that the ε-differential soft greedy policy is GLIE in some condition, one can not tell analytically whether the ε-differential soft greedy policy is GLIE or not, since there is no explicit way to analyze decaying rate of average normalized differential Q value. Alternatively one may set parameter a large enough to make the probability of choosing non-optimal action non-negligible for a long time.

5. Simulation Results

In this section, we evaluate the performance of the proposed algorithm through simulations. Two different simulations are considered. First, we characterize the basic properties of the α-controllable LB algorithm applied to a very simple case where there are two cells and two carriers. The effects of related parameters, and convergence properties will be addressed with this simulation. Second, the α -controllable LB algorithm will be executed over conventional 2 tier 19 hexagonal cells with 3 sectors per cell to evaluate its performance in a more realistic cellular environment where two carriers in different frequency bands are available.

5.1 Two Cells

Two cells with two carriers are considered to characterize the behavior of the proposed algorithm. Two BSs are located at (-1,0) and (1,0) over two dimensional space, and MSs are uniformly distributed within unit circle with center at origin. It is assumed that one carrier frequency is two times higher than the other carrier frequency. Single input single output (SISO) flat Rayleigh fading channels with log normal shadowing with standard deviation of 7dB and the path loss exponent of 3.5 are generated for each link between a MS and a BS. For each drop, 20 MSs are generated over 5000 frames with independent fading channels. We also average out the performance over 500 drops. Since we are more interested in the characterization of the algorithm, α -controllable LB algorithm are executed at every frame with assumption that perfect average SINR is available at the central processor. Since we also execute the algorithm at each frame, rather than calculating the average perceived transmission rate of MSs from each sample, we treat $\log_2(1+E\{\xi_{\lambda_B(k)}^{\lambda_F(k)}(k)\})/N_{\lambda_B(k)}^{\lambda_F(k)}$ as a nominal perceived transmission rate where $\lambda_F(k)$ and $\lambda_B(k)$ are the indices of the carrier and the BS serving the MS k .

In the formulation of MDP for load balancing, we defined the two different types of state spaces, S_2 and S_8 . If the numbers of possible states for w , h , and r are defined to be the same, the total number of states for S_8 is likely to be much greater than S_2 . In **Fig. 1** and **Fig. 2**, we compare the performances of the proposed algorithm for MDPs with differently defined state spaces, S_2 and S_8 . In this particular case s_3, s_4, s_7, s_8 in S_8 and s_1, s_2 in S_2 are set to have 10 different states composed of {0.05,0.15,...,0.95}, which results in 1600 times larger number of states for the MDP with S_8 . We also set the exponent a for the ε -differential soft greedy policy as 1000 so that it can have learning period for a long time which is required for embedded Sarsa algorithm to converge with this policy. It can be observed that when α is small, load deviation which is defined as $\sqrt{(1/(BF))\sum_{b=1}^{B}\sum_{f=1}^{F}(N_b^f - \overline{m}_N)^2}$ is small. However, spectral efficiency at 5% percentile of the perceived transmission rates of MSs is much smaller than one without load balancing, since the proposed algorithm tries to balance the cell load without considering spectral efficiency at 5% percentile of perceived transmission rates of MSs. It is also noted that whatever α may be, the load deviation gets smaller than one without load balancing. It is clear that α controls the tradeoff between the perceived transmission rate of MSs located at cell edge and cell load deviation of BSs. The MDP with S_8 uses information on minimum perceived transmission rate of MSs and cell load for serving and target BSs and carriers. Thus, the optimal action may be different for each cell and carrier. On the contrary, the MDP with S_2 uses composite state information constructed similarly to reward. Thus it is expected that if the MDP with S_2 are better designed in harmony with reward, it may provide comparable performance to the performance of one with S_8 despite of significantly smaller number of states. **Fig. 1** and **Fig. 2** confirm that this is the case with the α -controllable LB algorithm constructed with S_2 . They show the similar performance of perceived transmission rate and cell load. This may imply that knowledge on serving and target BSs and carriers with the predefined rule of selecting candidate MS for the forced HO may not help much to improve the performance of load balancing. Based on these results, we focus on evaluating the α -controllable LB algorithm of the MDP with S_2 for the subsequent simulations.

The proposition-1 tells that exponent a need to be large enough for the Q values of the α -controllable LB algorithm with ε -differential soft greedy policy to converge. **Fig. 3** and **Fig. 4** compare the performance of the α -controllable LB algorithm for different exponents. It can be observed that even though the algorithm with $a = 10$ has slight performance degradation, there is no discernable performance difference among different exponents of 100, 1000, and 10000. To investigate the convergence characteristics of the ε -differential soft greedy policy for different exponents, ε values for increasing control stages are shown in **Fig. 7** where d_0 is 0.5 and α is 0. The ε -differential soft greedy policy converges faster to a greedy policy with smaller a . It can be observed that as a increases every time by 10 folds, the graph of ε values appears to be shifted with several hundreds of control stages. This implies that the average normalized difference of Q is reduced by order of one tenth with several hundreds control stages. To show how fast Q value converges when $a = 1000$, Q values for which the algorithm makes the most frequent visits for each case is plotted in **Fig. 8**. As expected from the convergence of ε , Q value converges fast.

The performance of the proposed algorithm may have some dependency on how finely states are discretized. **Fig. 7** and **Fig. 8** compares the performance for the different discretizations of states. It can be observed that it does not have much dependency on how finely states are discretized. Since the predefined rule for choosing a candidate MS for forced HO which is embedded in the α -controllable LB algorithm does not consider serving and target BSs and carriers, the effect of the discretization of states may not be significant.

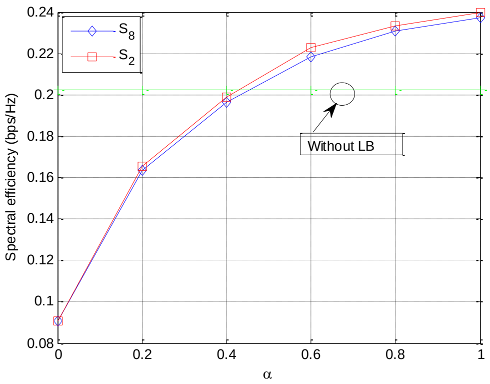

Fig. 1. Spectral efficiency at 5% percentile of perceived transmission rate with α -controllable LB algorithm

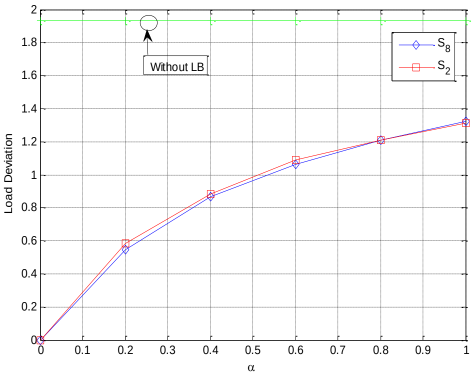

Fig. 2. Load deviation of the α -controllable LB the algorithm

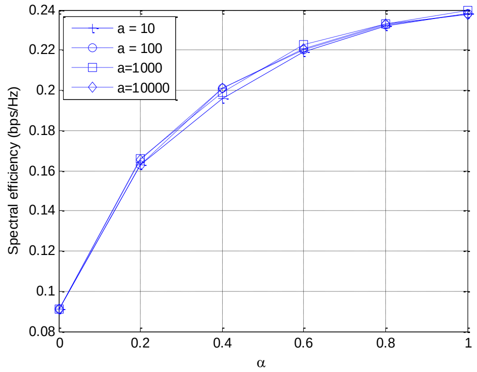

Fig. 3. Spectral efficiency at 5% percentile of perceived transmission rate of MSs with the α -controllable LB algorithm for different exponent a of ε - differential soft greedy policy

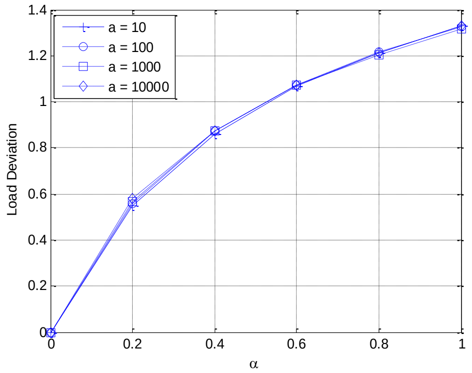

Fig. 4. Load deviation of the α -controllable LB algorithm with different exponent a of ε - differential soft greedy policy

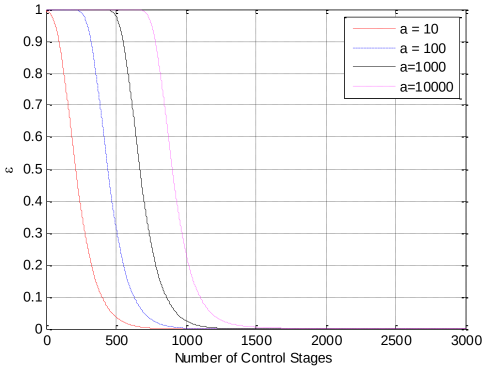

Fig. 5. ε of ε - differential soft greedy policy for different exponent a

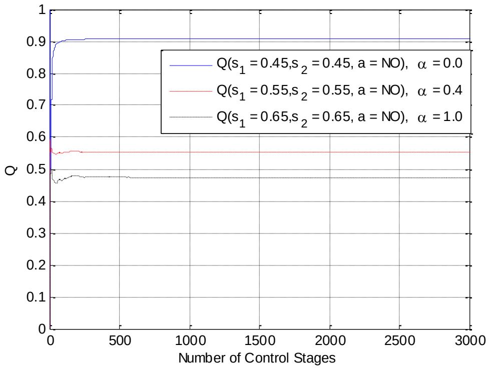

Fig. 6. Convergence characteristics of Q values

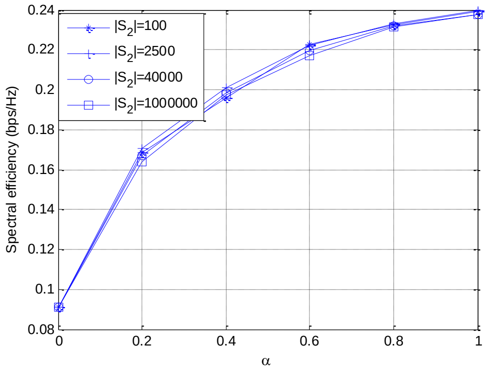

Fig. 7. Spectral efficiency at 5% percentile of perceived transmission rate of MSs with the α -controllable LB algorithm for different discretization of states

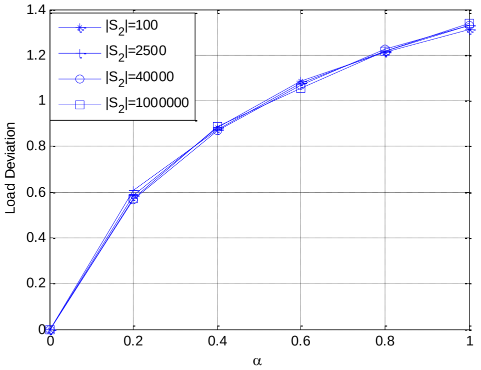

Fig. 8. Load deviation of the α -controllable LB algorithm with the two different discretization of states

5.2. 2 tier 19 cells with 3 sectors

To evaluate the performance of the proposed algorithm in a more realistic environment, we consider a 2 tier cellular wrapped around system with 19 cells. Each cell has 3 sector and every BS and MS is assumed to have one antenna. 10 MSs per sector are geometrically uniformly distributed. The non-line of sight (NLOS) urban micro channel of 3GPP SCM model [17] with velocity of 30km/h is used for simulation. To mimic the multi-carrier operation of LTE system, two carriers with operating frequencies of 850MHz and 1850MHz are considered. Bandwidth is assumed to be 20MHz. Proportional fair scheduling is exploited with scheduling period of 1 msec. Even though load balancing may be performed with period of order of every several minutes or so, due to extensive simulation time, it is performed every 10 msec with assumption that information required for load balancing is perfectly available. Even though this assumption may be rather unrealistic, it may not have much effect on the characterization of the proposed algorithm. We also set exponent a as 1000, and s_1 and s_2 in S_2 are set to have 10 different states composed of {0.05,0.15,...,0.95}. For each given α, the proposed algorithm was tested for 5 drops with 10000 frames. We consider two different cases. The first one is that the number of BSs supporting each carrier is same. In this case, imbalance in cell loads often results from frequency dependent path loss and shadowing. In the second case, 9 randomly chosen cells do not support the carrier of higher frequency. The imbalance of cell loads is likely to be due to interaction among carrier frequency, shadowing, and available number of carriers, which incurs the state of more severely imbalanced cell loads. As a baseline performance, the load balancing scheme of [4] which was designed for maximizing the perceived transmission rate was considered. The on-line implementation of [4] for multi-carrier operation can be described as follows.

$$\hat{r}_b^f(k) = \frac{G(N_b^f)E\{\log_2(1+\xi_b^f(k))\}}{N_b^f} \approx \frac{G(N_b^f)\log_2(1+E(\xi_b^f(k))}{N_b^f} \qquad (15)$$

$$k_b^f = \arg\max{}_k \max{}_{b',f'} \frac{\hat{r}_{b'}^{f'}(k)}{\hat{r}_b^f(k)} \qquad (16)$$

where $G(\cdot)$ is a function representing multi-user diversity gain, and k_b^f is a user selected for forced HO at the cell *b* using carrier *f*.

Fig. 9 shows the load deviation with different α. Similarly to the characteristics in the case of two cells, the load deviation increases as α increases while it is zero when $\alpha = 0$ for the first case. The reason not to have zero load deviation for the second case is due to the indivisible number of MSs for the given number of BSs and carriers rather than severely unbalanced initial loads. The number of MSs served by every BS is found to be 6 or 7 for the second case when $\alpha = 0$. The load deviation with $\alpha = 1$ is marginally smaller than one with $\alpha = 0.8$. It is conjectured that maximizing minimum perceived transmission rate may help to reduce load deviation by moving MSs located in cell edge to less crowded cells more often. This trend is found to be similar for both cases. However, in the second case of some BSs supporting only one carrier, relatively higher load deviation than the first case for the same α can be observed. It is also noted that the proposed scheme outperforms [4] for cell load deviation when α is small while [4] provides the slight better performance when α is larger than 0.6.

The perceived transmission rates of MSs for the first case are compared for different α and different percentiles in **Fig. 10**. The perceived transmission rate at 5% and 10% percentile degrades significantly as α gets close to zero, since the α-controllable algorithm forcefully associates MSs with BSs without considering much about the perceived transmission rates of MSs to reduce the cell load deviation. As α increases, the perceived transmission rates at 5% and 10% percentile larger than that of [4] by 3~10%. Similar characteristics can be found for the second case in **Fig. 11**. It is also noted that there is considerably large improvement on the perceived transmission rate at 90% percentile especially for the second case when α is small. There can be a tradeoff between scheduling opportunity and multi-user diversity (MUD) gain depending on the number of MSs served by BS. Moving one MS from a crowded cell to other cell with small number of MSs can benefit MSs in the crowded cell, in terms of the perceived transmission rates. Even though the perceived transmission rates of the MSs in the target cell may degrades due to increase in the number of MSs served by the cell, the number of MSs located closely to the target cell is likely to be smaller than one in the crowded cell. Thus, these two facts may explain the noticeable increase of the perceived transmission rate at 90% percentile for the second case. Observing **Fig. 9**, **10**, and **11** reveals that the proposed scheme has better performance both in the cell load deviation and perceived transmission rate when α is around 0.45. In addition, the proposed scheme provide a good control of tradeoff between the cell load deviation and perceived transmission rate compared to [4].

Simulation results verify that the α-controllable LB algorithm can have tradeoff between the perceived transmission rates of MSs located at cell edge and cell load deviation of BSs from choosing α. While the proposed algorithm exhibits remarkable performance in equalizing cell loads, the improvement on the perceived transmission rates of MSs located in cell edge is very limited regardless of choice of α. Similar results can be found in [4] when LB is applied to homogeneous multi-systems with full buffer traffic model and uniform MS distribution. In this case, LB algorithms have to seek for gain from the tradeoff between more scheduling opportunities and channel quality degradation, which is found to be very limited. However, it is noted that the α-controllable LB algorithm can still provide slight better perceived transmission rate with reduced imbalance in cell load by properly choosing α.

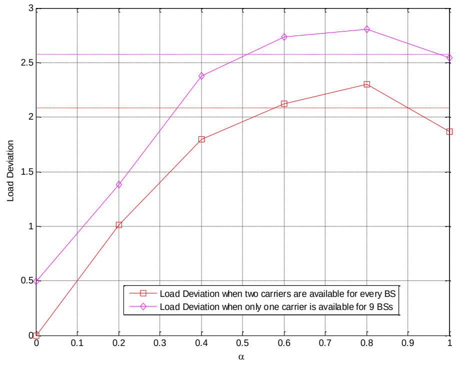

Fig. 9. Load deviation of the α-controllable algorithm in a 2 tier multi-cell system (dash-dot lines are the performance of [4])

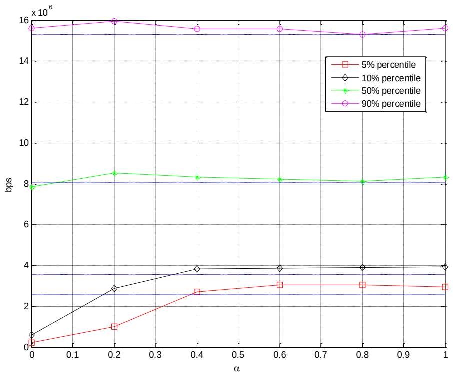

Fig. 10. Perceived transmission rate with the α-controllable algorithm in a 2 tier multi-cell system when two carriers are available to every cell (dash-dot lines are the performance of [4])

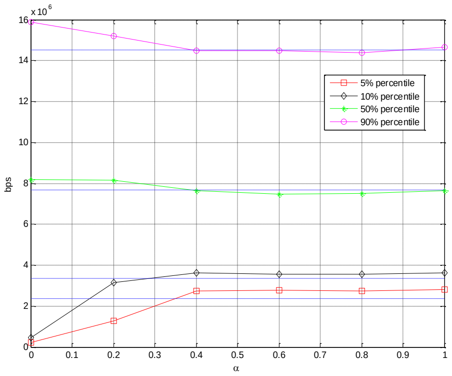

Fig. 11. Perceived transmission rate with the α -controllable algorithm in a 2 tier multi-cell system when only one carrier is available to 9 cells (dash-dotlines are the performance of [4])

6. Conclusions

In this paper, we formulated MDP for load balancing in a multi-cell system with multi carriers. To solve the MDP, we proposed the α -controllable algorithm exploiting Sarsa learning. To have efficient tradeoff between exploration and exploitation in the learning process of the α -controllable LB algorithm, ε -differential soft greedy policy was proposed and proved to be an asymptotically optimal policy under some condition. The simulation results verified that the α -controllable algorithm with ε -differential soft greedy policy could have tradeoff between cell load deviation and perceived transmission rate depending on the choice of α . It was also found to be particularly efficient in balancing cell loads.

There are still remaining issues associated with this research which calls for further future research. The implementation of the proposed algorithm requires a central processor. Even though computational complexity and communication overhead is not so large, it may be more efficient and robust if it is implemented in a distributed way. The load balancing algorithm can be more tailored to the specific system environment. Recently, small cell technologies are developed for beyond 4G system such as carrier aggregation and cell range expansion [14]. Algorithms considering the embedded technologies may improve the performance of load balancing further in the future wireless cellular system.

References

[1] N. D. Tripathi, and S. Sharma, "Dynamic Load Balancing in a CDMA system with Multiple Carriers," *VTC 2001 Fall*, 2001. Article (CrossRef Link)

[2] P. Guturu, and A. Lachtar, "An Efficacious Method for Dual Band Multi-carrier Traffic Allocation in CDMA Wireless Systems," *Globecomm 2003*, 2003. Article (CrossRef Link)

[3] R. Attar, D. Ghosh, C. Lott, M. Fan, P. Balck, R. Rezaiifar, and P. Agashe, "Evolution of cdma2000 cellular networks: Multi-carrier EV-DO," *IEEE Comm. Mag.*, Feb. 2006. Article (CrossRef Link)

[4] K. Son, S. Chong, and G. D. Veciann, "Dynamic Association for Load Balancing and Interference Avoidance in Multi-Cell Networks," *IEEE Trans. Wireless Comm.*, July. 2009. Article (CrossRef Link)

[5] H. Velayos, V. Aleo, and G. Karlsson, "Load Balancing in Overlapping Wireless LAN Cells," *ICC 2004*, 2004. Article (CrossRef Link)

[6] Y. Wang, K. I. Pederson, P. E. Mogensen, and T. B. Sorensen, "Resource Allocation

Considerations for Multi-Carrier LTE-Advanced Systems Operating in Backward Compatible Mode," *PIMRC 2009*. Article (CrossRef Link)

[7] E. S-Navarro, Y. Lin, and V. W. S. Wong, "An MDP-based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks," *IEEE Trans. Veh. Tech*, Vol. 57, No. 2, pp.1243-1254, Mar. 2008. Article (CrossRef Link)

[8] S. Cheng, A. Raja, J. Xie, and I. Howitt, "DLB-SDPOP : A Multiagent Pseudo-tree Repair Algorithm for Load Balancing in WLANs," in *Proc. of IEEE/WIC/ACM Web Intelligence and Intelligent Agent Technology (WI-IAT)*, vol.2., pp. 311-318, Toronto, Canada, Sept. 2010. Article (CrossRef Link)

[9] L. Mastroleon, N. Bambos, C. Kozyrakis and D. Economou, "Autonomic Power Management Schemes for Internet Servers and Data Centers," in *Proc. of GLOBECOM '05. IEEE*, Dec., 2005. Article (CrossRef Link)

[10] V. Mathew, R. K. Sitaraman, and P. Shenoy, "Energy-Aware Load Balancing in Content Delivery Networks," in *Proc of INFOCOM*, pp.954-962., 2012. Article (CrossRef Link)

[11] D. Kumar, E. Altman, and J.-M. Kelif, "Globally Optimal User-Network Association in an 802.11 WLAN & 3G UMTS Hybrid Cell," *Managing Traffic Performance in Converged Networks LNCS,* Vol. 4516, pp 1173-1187, 2007. Article (CrossRef Link)

[12] S. Singh, T. Jaakkola, M. L. Littman,and C. Szepesvari, "Convergence Results for Single-Step On-Policy Reinforcement-Learning Algorithms," *Machine Learning*, Vol. 39, pp. 287-308, 2000. Article (CrossRef Link)

[13] D. Kumar, E. Altman, and J.-M. Kelif, "Globally Optimal User-Network Association in an 802.11 WLAN & 3G UMTS Hybrid Cell," *Managing Traffic Performance in Converged Networks LNCS*, Vol. 4516, pp 1173-1187, 2007. Article (CrossRef Link)

[14] A. Ghosh, N. Mangalvedhe, R. Ratasuk, and B. Mondal, "Heterogeneous cellular networks: From theory to practice," *IEEE Comm. Mag.*, Vol.50, No. 6, pp.54-64, June 2012. Article (CrossRef Link)

[15] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, Vol.18 , No.3, pp.10-21, June 2011. Article (CrossRef Link)

[16] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd Ed. Vol.1 Atheana Scientific, Belmont, Massachusetts. Article (CrossRef Link)

[17] R. S. Sutton, and A. G. Barto, *Reinforcement Learning*, The MIT Press, Cambridge, Massachusetts. Article (CrossRef Link)

[18] 3GPP, "Spatial channel model for multiple input multiple output (MIMO) simulations", *3GPP TR 25.996 V6.1.0*, Sep. 2003. Article (CrossRef Link)

Janghoon Yang received his Ph.D. in Electrical Engineering from University of Southern California, Los Angeles, USA, in 2001. He is currently an associstate professor at the department of newmedia, Korean German Institute of Technology, Seoul, Korea. From 2001 to 2006, he was with communication R&D center, Samsung Electronics. From 2006 to 2010, he was a Research Assistant Professor at the Department of Electrical and Electronic Engineering, Yonsei University. He has been with the Department of Newmedia, Korean German Institute of Technology, Seoul, since 2010. He has published numerous papers in the area of multi-antenna transmission and signal processing. His research interest includes wireless system and network, artificial intelligence, neuroscience, and affective computing.