A Semi-Markov Decision Process (SMDP) for Active State Control of A Heterogeneous Network

Janghoon Yang[1]
[1]Department of New Media Contents, Seoul Media Institute of Technology
661 Deungchon-dong, Kangseo-gu, Seoul 157-030, Korea
[e-mail: jhyang@smit.ac.kr]
*Corresponding author: Janghoon Yang

Received December 21, 2015; revised May 5, 2016; accepted May 31, 2016; published July 31, 2016

Abstract

Due to growing demand on wireless data traffic, a large number of different types of base stations (BSs) have been installed. However, space-time dependent wireless data traffic densities can result in a significant number of idle BSs, which implies the waste of power resources. To deal with this problem, we propose an active state control algorithm based on semi-Markov decision process (SMDP) for a heterogeneous network. A MDP in discrete time domain is formulated from continuous domain with some approximation. Suboptimal on-line learning algorithm with a random policy is proposed to solve the problem. We explicitly include coverage constraint so that active cells can provide the same signal to noise ratio (SNR) coverage with a targeted outage rate. Simulation results verify that the proposed algorithm properly controls the active state depending on traffic densities without increasing the number of handovers excessively while providing average user perceived rate (UPR) in a more power efficient way than a conventional algorithm.

Keywords: MDP, active state control, load balancing, heterogeneous networks, handover

http://dx.doi.org/10.3837/tiis.2016.07.017 **ISSN : 1976-7277**

1. Introduction

As mobile data traffic data increases in a cellular network, the density of cells increases to provide satisfactory wireless service. In the past, micro cells or pico cells were installed to deal with this problem. Nowadays, including femto cells, the various types of cells are deployed, which constitutes a heterogeneous cellular network. As the cellular system evolves, variety of heterogeneities occurs. Especially internet of things (IOT) is likely to have great impact on the architecture of a next generation cellular system. In the perspective of cellular systems, more heterogeneous cells are likely to appear. Severe heterogeneity in cells may incur several critical issues such as load balancing, handover, and inter-cell interference. Thus, controlling cellular system parameters to optimize some performance for heterogeneous networks will be a very difficult task.

As a large number of cells are installed in a system, managing cells in an energy-efficient way also lays down some issues. The energy consumption incurred by cellular systems is known to take several tenths percentage of the world's energy consumption [1]. To address the energy-efficiency more systematically, energy efficiency evaluation framework (E3F) was developed [2]. According to the model of base station (BS) power consumption [3], power amplifier usually consumes 55-60% of the overall power at full load for macro cells while around 30% for small cells. Power control at BSs can be broadly classified into transmit power control for interference control to improve quality of service (QoS) and active power state control for energy efficiency (EE). The former may help to reduce power consumption marginally. Thus, it is expected that active state control with consideration of QoS need to be developed to deal with energy-efficient operation with respect to a given cell load condition.

Several researches on this problem have been conducted for the past few years. A joint active state control and cell association which was mapped to Knapsack-like problem was proposed to minimize total energy used in the network [27]. Since a brute-force optimal BS control has exponential complexity with the number of BSs, many of them take a form of a greedy algorithm [4][8][17][19]. In [8], an active set control was applied to a subset of BSs to support peak traffic while others being turned on always. Starting with all BSs being in active state, a dynamic cell reconfiguration algorithm to minimize the total power with constraint on cell load sequentially switches off BSs in a greedy way [19]. Discontinuous transmission schemes were also considered as an energy efficient scheme [7][18]. In [7], BSs were switched off to save energy in sacrifice of transmission delay without offloading the mobile traffics while optimal delay was sought with the cost function of average delay and average power in [18]. Some novel methods were also introduced [5][10]. Ecological self-organization based method for distributed inter-BS cooperation heuristically switches off a BS when cell load is below a certain threshold while it is turned on by the request of surrounding neighbor cells which request load sharing [10]. Tabu search was exploited to find an active BS set which provides the best BS-RS(relay station) association minimizing energy consumption under a QoS constraint [5]. Some simple methods such as choosing one from the predefined sets of active BSs based on mobile traffic density [6] and determining active state from using average distance between mobile stations (MSs) and a BS [11] were also developed.

Active state control can be formulated into a sequential decision problem. One of the most efficient methods for solving sequential decision problem is to exploit the framework of Markov decision process (MDP). There are several researches to apply MDP to wireless network optimization problems such as call admission control for multiple radio access

technologies (RATs) [12][14], and joint radio resource management [13]. Only a few are closely related to the network optimization for EE. Femto cell active control in a heterogeneous network using MDP for a very specific model tried to minimize total power from maximizing total reward [16]. Delay-optimal control of BS discontinues transmission was formulated into partially observable MDP (POMDP) with the cost function of average delay and average power rather than focusing on energy-efficient operation [18].

There are two important limitations in existing related research. First, many of them focus on homogeneous network in which the effect of different coverage is not considered properly. There are some researches for heterogeneous networks. However, most of them have been studied for controlling the active state of the single type of cells. A policy for controlling the sleep mode of small cells in a heterogeneous network based on stochastic geometry was developed to maximize energy efficiency (EE) in [23]. Algorithms for Switching on/off macro cells in a heterogeneous network have been developed to improve EE in [24] and to have a tradeoff between EE and the cost of mobile network operators (MNOs) with a combinational auction framework in [25]. Second, many existing algorithms do not consider broadcasting coverage for signaling channel when active state is controlled. One common assumption in existing research on energy-efficient optimization of a wireless network is that all MSs are in the coverage of some cells regardless of the number of active cells. However, this can be hardly true in a realistic network environment. This can be more problematic especially in a heterogeneous network environment where small cells and macro cells co-exist. In [8], the coverage of wireless network was explicitly considered in active state control. However, it is required to verify whether it satisfies the coverage condition whenever it switched off a BS, which could not be implemented in a realistic system due to the instability of service. An optimal density of micro BSs and macro BSs for energy-efficient operation in consideration of coverage constraint was also calculated numerically from using stochastic geometry theory [15]. However, a method for controlling the active state of each cell in consideration of cell coverage jas not been developed to the best of author's knowledge.

In this paper, we propose an ASC algorithm from the framework of semi-MDP (SMDP) in a heterogeneous network to have energy-efficient operation satisfying coverage constraint. It starts with learning channel environment from the measurement report of each MS to know the feasible set for active state control satisfying coverage condition. Then, it learns whether BS can be switched off or on depending on cell load.

The rest of the paper is organized as follows. A system model is given in section 2. In section 3, a problem is formulated in terms of power of each cell and coverage constraint. In section 4, the framework of SMDP for minimizing power with coverage constraint which approximately transforms a problem in the continuous domain into one in the discrete domain is developed with defining associated cost, transition probabilities, actions, and states. Since the transition probabilities are not available in a closed form, on-line reinforcement learning algorithm was introduced in section 5. In section 6, simulation setups are specified, and the performances of the proposed ASC algorithm are verified through numerical simulation. We make some concluding remarks in section 7.

2. System Model

We consider a heterogeneous downlink multi-cell system with frequency reuse-1 where the different types of cells exist in a system. For simplicity, the two types of cells, macro cells and small cells, are considered without loss of generality throughout this paper. There are N_{MC}

macro cells and N_{SC} small cells. Each BS may have a single sector or multi-sector depending on network design. With slight abuse of terminology, we call a sector as a cell throughout this paper. We make several assumptions to restrict the scope of this research and elucidate considered system setup.

Every BS has only one transmit antenna to be free from dependency on the types of multi-input multi-ouput (MIMO) techniques and to focus on the characterization of the ASC algorithm. We differentiate macro cell and small cell depending only on transmit power. Each BS is perfectly synchronized to common clock such as global positioning system (GPS) so that there may not be synchronization error in the received signal of MS. Likewise, each MS keeps perfect synchronization to its serving BS, and has perfect channel estimation. There exists a centralized processor which gathers perfect information on loading of each cell and controls its active state. There may be some uncertainty on this information due to delay and estimation error in practice. However, it may not have significant effect on the performance since active state control is likely to be done on long-term basis such as order of minutes or hours, and averaging out cell load over sufficient time interval may provide stable statistics.

One of important factors influencing cell load is a mobile traffic distribution. The arrival of mobile traffic is assumed to follow Poisson point process with arrival rate per unit area λ where the size of each traffic is exponentially distributed with mean $1/\mu$. Consequently, the traffic density χ can be calculated as λ/μ. For simplicity, we focus on homogeneous mobile traffic which can be readily extended to inhomogeneous case by making it dependent on location. A system load density associated with the cell c can be defined as follows [4][5].

$$\kappa_c(z, F_{ON}) = \frac{\chi}{r_c(z, F_{ON})} \tag{1}$$

where F_{ON} is a set of active cells, and $r_c(z, F_{ON})$ is the achievable transmission rate with the quality of radio frequency channel at location z served by the cell c. Exploiting (1), one can define loading of the cell c as follows.

$$L_c = \int_R \kappa_c(z, F_{ON}) q_c(z) dz \tag{2}$$

where $q_c(z)$ is the probability distribution of being associated with the cell c at location z, and R is the region of the system. $q_c(z)$ will be 0 if location z is out of coverage of the cell c.

Cell coverage usually depends on signal to noise ratio (SNR) or signal to interference plus noise ratio (SINR). For simplicity, we consider SNR as a measure for defining coverage throughout this paper. Let $E_C(z)$ be the event that maximum signal to noise ratio (SNR) over cells belonging to the set C is less than γ_T at location z when they are active. The probability of this event can be expressed as

$$\Pr(E_C(z)) = \Pr(\max_{c \in C} \gamma_c(z) \le \gamma_T) = \prod_{c \in C} \Pr(\gamma_c(z) \le \gamma_T) \tag{3}$$

where $\gamma_c(z)$ is the SNR of the signal received from the cell c. Exploiting this probability, one may define a feasible set collection F consisting of the sets of active cells which support outage level ζ over the region R in the following way.

$$F = \{F_i \mid \int_R \Pr(E_{F_i}(z)) f_Z(z) dz \le \zeta\} \tag{4}$$

where $f_Z(z)$ is a geometric traffic density. Active set control problem in terms of minimizing total power with outage constraint can be formulated as

$$\min_{F_i \in F} \sum_{c \in F_i} P_c \tag{5}$$

where P_c is the power consumption of the cell c. Even though the solution of this problem provides the minimization of total transmit power, quality of service (QoS) may not be good enough to be chosen as a practical solution. One may alternatively add loading constraint to balance the traffic and provide QoS.

$$\min_{F_i \in F} \sum_{c \in F_i} P_c + \sum_{c \in F_i} \lambda_c (L_c - L_T) \tag{6}$$

where λ_c is a nonnegative constant parameter, and L_T is the allowable maximum loading at each cell. It is noted that computing the optimal solution to (5) or (6) is a combinatorial problem which necessitates the exponential computational complexity with the number of cells. Throughout this paper, we will propose a suboptimal solution to tackle this problem with moderate complexity. We also provide the description of notations and summary of used acronyms and corresponding their meaning in **Table 1** and **Table 2** respectively.

Table 1. The description of Notations.

Notation	Description
N_{MC}	Number of macro cells
N_{SC}	Number of small cells
λ	arrival rate per unit area
$1/\mu.$	Average traffic size
χ	Traffic density
κ_c	System load density associated with a cell c
L_c	Cell load at a cell c
P_c	power consumption of the cell c.
β	a positive discount parameter for defining cost in a continuous time domain
l_{MC}	the average cell loads of macro cell type
l_{SC}	the average cell loads of small cell type
s_{MC}	the discrete cell load states of macro cell type
s_{SC}	the discrete cell load states of small cell type
T_H	thresholds for high cell load
T_L	thresholds for low cell load
ζ	Outage threshold

Table 2. Summary of used acronyms and corresponding their meaning.

Acronyms	Description
BS	Base Stations
MDP	Markov Decision Process
SMDP	Semi-Markov Decision Process
SNR	Signal to Noise Ratio
UPR	User Perceived Rate
ASC	Active State Control
MS	Mobile Station
SCM	Spatial Channel Model
3GPP	The 3rd Generation Partnership Project
HO	Hand Over
M-GOFF	Modified Greedy OFF

3. Problem Formulation

In this section, we formulate semi-Markov decision process for controlling active state of cells. One can easily find that the solution of (5) will be the same regardless of mobile traffic conditions. This means that it depends only on the geometry of system environment. This solution may not be a good working solution when there is space-time variation in the vloume of data traffic. The solution of (6) may provide a trade-off between the power minimization and cell load balancing. We will implicitly exploit this fact to formulate the problem in terms of semi-Markov decision process.

Let x_k t_k , a_k be state, continuous time, and action at time step k, respectively. A transition distribution can be expressed as

$$T_{i,j}(\tau,a)=P\{t_{k+1}-t_k\le\tau,x_{k+1}=j\mid x_k=i,a_k=a\) \tag{7}$$

With simple manipulation, this distribution can be expressed as

$$T_{i,j}(\tau,a)=P_{i,j}f_{i,j}(\tau\mid a) \tag{8}$$

where $f_{i,j}(\tau\mid a)=P\{t_{k+1}-t_k\le\tau\mid x_k=i,x_{k+1}=j,a_k=a\)$ and transition probability $P_{i,j}(\tau,a)\equiv$ $=P_{i,j}=\lim_{\tau\to\infty}T_{i,j}(\tau,a)$. One can define total discounted cost with policy $\pi=[a_0,a_1,\cdots]$ starting from state i as follows

$$C_\pi(i)=\lim_{N\to\infty}E\{\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}e^{-\beta t}g(x_k,a_k)dt\mid x_0=i\} \tag{9}$$

where $g(x_k,a_k)$ is cost per stage which is assumed to be constant for an interval $[t_k,t_{k+1}]$, and β is a positive discount parameter in continuous time domain. $C_\pi(i)$ can be decomposed into two parts, expected single stage cost $G(i,a_0(i))$ and expected cost to go from the next state [20].

$$C_\pi(i)=G(i,a_0(i))+E\{e^{-\beta\tau}C_{\pi_1}(j)\mid x_0=i,a_0(i)\} \tag{10}$$

where $\pi_1=[a_1,a_2,\cdots]$, $G(i,a_0(i))=\sum_{j=1}^{N_S}P_{ij}(a_0(i))\int_0^\infty(\int_0^\tau e^{-\beta t}g(i,a_0(i))dt)\frac{T_{ij}(d\tau,u)}{P_{ij}(u)}$, and N_S is the number of states. The second term in the right side of (10) can be further calculated by

$$E\{e^{-\beta\tau}C_{\pi_1}(j)\,|\,x_0=i,a_0(i)\}=E_j\{E\{e^{-\beta\tau}\,|\,i,a_0(i),j\}C_{\pi_1}(j)\,|\,i,a_0(i)\}$$
$$=\sum_{j=1}^{N_S}P_{ij}\hat{\beta}_{ij}(a_o(i))C_{\pi_1}(j) \tag{11}$$

where $\hat{\beta}_{ij}(a_o(i))=\int_0^\infty e^{-\beta\tau}df_{i,j}(\tau\,|\,a_o(i))$. (10) can be rearranged by using (11) as

$$C_\pi(i)=G(i,a_0(i))+\sum_{j=1}^{N_S}P_{ij}\hat{\beta}_{ij}(a_o(i))C_{\pi_1}(j) \tag{12}$$

Even though this equation appears to be same as the system equation of discrete MDP, it is not, since summing $P_{ij}\hat{\beta}_{ij}(a_o(i))$ over j is not equal to 1. Thus, we derive the system equation of Markov decision process with discounted cost from the upper bound of $P_{ij}\hat{\beta}_{ij}(a_o(i))$.

$$C_\pi(i)\approx G(i,a_0(i))+\overline{\beta}\sum_{j=1}^{N_S}P_{ij}C_{\pi_1}(j) \tag{13}$$

$$\overline{\beta}=\max_{\{(i,a,j)|i\in S,a\in A,j\in S\}}\hat{\beta}_{ij}(a) \tag{14}$$

For instance, if transition time follows exponential distribution with the transition rate of $b_{ij}(a)$, then $\overline{\beta}$ can be determined as $\max_{\{(i,a,j)|i\in S,a\in A,j\in S\}}1/(\beta+b_{ij}(a))$. Setting β as in (14), we can find the relationship between β and $\overline{\beta}$ with the following proposition.

Proposition-1 : $\overline{\beta}$ is monotonically non-increasing with β.

Proof : since $\hat{\beta}_{ij}(a_o(i))$ is monotonically non-increasing with β. The maximum of $\hat{\beta}_{ij}(a_o(i))$ over $\{(i,a,j)\,|\,i\in S,a\in A,j\in S\}$ is also monotonically non-increasing.

This proposition implies that larger $\overline{\beta}$ is equivalent to smaller β in continuous time domain. That is, setting $\overline{\beta}$ larger in discrete domain means to consider future cost more aggressively in continuous time domain.

An expected single stage cost $G(i,a_0(i))$ can be upper bounded with approximation in the following way.

$$G(i,a_0(i))=E\{\int_0^\tau e^{-\beta t}g(i,a_0(i))dt\}$$
$$=g(i,a_0(i))E_j\{E_\tau\{\frac{1-e^{-\beta\tau}}{\beta}\,|\,i,a_0(i),j\}\}$$
$$\le g(i,a_0(i))\sum_{j=1}^{n}P_{ij}(a_0(i))\frac{1-\exp^{-\beta E\{\tau|i,a_o(i),j\}}}{\beta}\approx g(i,a_0(i))\bar{\tau}_i(a_0(i)) \tag{15}$$

where $\bar{\tau}_i(a_0(i))=\sum_{j=1}^{n}P_{ij}(a_0(i))E\{\tau\,|\,i,a_o(i),j\}$. In (15), inequality and approximation come from Jensson's inequality and Taylor series expansion respectively. It is noted that the approximation is valid when $\beta E\{\tau\,|\,i,a_o(i),j\}$ is close to 0.

4. Markov Decision Process for Active State Control

In this section, we explicitly define the elements of the MDP for active state control in a heterogeneous network. A MDP is described as a tuple $< S, A, T, C >$ where S is a set of states, A is a set of actions, $T = \{P_{ij}(a) \mid i \in S, a \in A, j \in S\}$ is a set of transition probabilities, and $C : S \times A \to \Re$ is a cost function. In MDP, a next state and cost depend only on the current state and the chosen action which represents Markov property. An ASC problem can be formulated in many different ways with MDP by defining the elements of the MDP differently. It is known that MDP can have the curse of dimensionality with a large number of states. Thus, the proposed MDP will be constructed such that it can be implemented with moderate complexity.

State information is usually used for determining an action. One may determine the active state of each cell based on the active states of other cells, loading of each cell, the number of MSs in the system, and so forth. Theoretically considering every possible variable to determine a state may provide the best performance. However, it can lead to huge complexity and excessive convergence time if some parameters need to be learned. Thus, we define a state from a two-dimensional vector (l_{MC}, l_{SC}) where l_{MC} and l_{SC} represent the average cell loads of macro cells and small cells. Since the decision of the active state of each cell is closely related to average cell load, this information is taken as state information. Even though the state has only two dimensions, the number of state can be infinite if it takes a continuous value. Thus, we consider a discretized state space in the following way.

$$S = \{(s_{MC}, s_{SC}) \mid s_{MC}, s_{SC} \in \{H, M, L\}\} \tag{16}$$

where s_{MC} and s_{SC} are the discrete cell load states of macro cells and small cells, and "H", "M", and "L" represent three cell load states depending on average cell load. There is no strict rule to define the cell load state. It may depend on the goal of system operation and service management scheme. In this paper, this will be set by defining the interval of cell load for determining cell load state.

$$s_i = \begin{cases} H, & if\ l_i > T_H \\ L, & if\ l_i < T_L \\ M, & otherwise \end{cases} \tag{17}$$

where T_H and T_L are thresholds for high cell load and low cell load respectively, and $i \in \{MC, SC\}$.

For active state control, the action space may be simply defined as turning on and turning off. However, since we are interested in a heterogeneous system, we distinguish them depending on the types of cells. Consequently, the action space can be defined as

$$A = \{a \mid a \in \{M-On, S-On, M-Off, S-Off\}\} \tag{18}$$

where $M-On, S-On, M-Off,$ and $S-Off$ represents "Turn on a macro Cell", "Turn on a small cell", "Turn off a macro cell", and "Turn off a small cell" respectively. One may have more refined action space such that it can include a set of cells which need to change their active states. However, it will make MDP too complicated. Thus, the decision of this set will be done rather heuristically which will be defined in a subsequent section.

Taking an action for a current state incurs cost. Even though an objective function is given in (6), it cannot be directly used since it requires the parameter decision of λ_c. In addition, active state depends on cell load condition. There can be infinite number of ways to define this function while the systematic method of constructing this function is never known to the best

of authors' knowledge. Thus, from the observation that it is reasonable to control the active state of a cell based on cell loads, the single stage cost is proposed as follows.

$$G^{-0.5l_{MC}} = \{((H,*),M-ON)\},\quad G^{-0.5l_{SC}} = \{((*,H),S-ON)\}$$
$$G^{0.5(l_{MC}-1)} = \{((L,*),M-OFF)\},\quad G^{0.5(l_{SC}-1)} = \{((*,L),S-OFF)\} \tag{19}$$
$$G^{0.5} = \{((H,*),M-OFF),((*,H),S-OFF),((L,*),M-ON),((*,L),S-ON)\}$$

where $G^{\psi} \subset \{(\mathbf{s},a) \mid \mathbf{s} \in S, a \in A\}$ In (20), G^{ψ} denotes a set of state and action pairs of which single stage cost is ψ, and * means that any state in the state space can be allowed. Cost is designed such that it can have value between -0.5 and 0.5. Cost, 0.5 is assigned to the state-action pairs which are not desired to happen. Cost of turning on a cell in the case of high average cell load is designed to decrease linearly with average cell load, while cost of turning off a cell in the case of low average cell load linearly increases with average cell load. No cost is defined for the state (M,M), since it is considered as a terminal state..

For the approximate MDP defined in (13), a set of system equation for optimality can be defined as [20].

$$V^*(\mathbf{s}) = \min_{a\in A}[G(\mathbf{s},a) + \bar{\beta}\sum_{\mathbf{s}'} P_{\mathbf{s}',\mathbf{s}}(a)V^*(\mathbf{s}')] \quad for\ \forall \mathbf{s} \in S \tag{20}$$

where $V(\mathbf{s})$ is called as a value function. This is a Bellman equation for MDP. Corresponding optimal policy π^* can be expressed as

$$\pi^* = \arg\min_{a\in A}[G(\mathbf{s},a) + \bar{\beta}\sum_{\mathbf{s}'} P_{\mathbf{s}',\mathbf{s}}(a)V^*(\mathbf{s}')] \quad for\ \forall \mathbf{s} \in S \tag{21}$$

This equation is usually solved through value iteration or policy iteration [20]. To do so, knowledge on transition probability is required. However, this information is not available in the given problem formulation. When this information is not available, heuristic algorithm or learning algorithm can be an alternative.

5. Reinforcement learning for Active State Control

Some learning algorithms are known to converge to optimal solution for some conditions [21]. These learning algorithms are broadly classified into off-line and on-line algorithms. Since off-line algorithm requires precise modeling of physical environment, and its performance may degrade in the presence of model mismatch, we focus on on-line learning algorithm.

To derive on-line learning algorithm, first, we have to define a Q-factor $Q(\mathbf{s},a)$ which is also often called as an action-value function. Bellman equation for an optimal Q-factor $Q^*(\mathbf{s},a)$ can be expressed as

$$Q^*(\mathbf{s},a) = G(\mathbf{s},a) + \bar{\beta}\sum_{\mathbf{s}'} P[\mathbf{s}' \mid \mathbf{s},a]\min_{a'\in A} Q^*(\mathbf{s}',a') \quad for\ \forall \mathbf{s} \in S,\ and\ \forall a \in A \tag{22}$$

One of the most efficient algorithm to approximate $Q^*(\mathbf{s},a)$ is Sarsa algorithm[22]. This algorithm updates Q-value and policy at each step [22]. A updating equation for Sarsa algorithm at the k th step is given as

$$Q(\mathbf{s}_k,a_k) = Q(\mathbf{s}_k,a_k) + \delta(G(\mathbf{s},a) + \bar{\beta}Q(\mathbf{s}_{k+1},a_{k+1}) - Q(\mathbf{s}_k,a_k)) \tag{23}$$

where δ is a step size which determines averaging width. At each step, once the current state and action are given, associated single stage cost and next state can be determined. The action

of the next stage can be decided from a given policy. Policy may be updated at each step with updated Q-value if necessary. This procedure is repeated at every step. Convergence is guaranteed with ε -greedy policy if all state-action pairs are visited with asymptotically large number of times [21].

At each step, an action needs to be determined from a policy. The policy is a mapping from a state to an action. We adopt a heuristic random policy which combines a deterministic policy and a softmax policy.

$$\pi(\mathbf{s},a)=\frac{e^{-Q(\mathbf{s},a)}}{\sum_{\mathbf{s}'}e^{-Q(\mathbf{s}',a)}} \quad for\ \mathbf{s}\in\{(L,L),(L,H),(H,L),(H,H)\}$$
$$\pi((L,M),M-OFF)=1,\ \pi((M,L),S-OFF)=1$$
$$\pi((M,H),S-ON)=1,\ \pi((H,M),M-ON)=1 \tag{24}$$

The proposed policy exploits a softmax policy for the states consisting of the combination of L and H while it does a deterministic policy for other states. It is noted that the softmax policy is applied to the case that there is ambiguity in choosing a proper action for a given state. On the contrary the deterministic policy is applied to the states for which proper action is intuitively clear. For example, when $\mathbf{s}=(L,L)$ it is quite reasonable to turn off a macro cell or a small cell to lower average cell load. But it is unclear whether it will be better to turn on a macro cell or a small cell. Thus, a softmax policy probabilistically determines which action is going to be taken.

The action determined by the reinforcement algorithm does not specify the set of cells to be controlled at each stage. For simplicity, we set the maximum number of cells to change its active state as 1. For the purpose of load balancing, it will be advantageous to turn on the power of a cell of which neighbors are heavily loaded.

$$c_y=\arg_{i:\bar{P}_{i,y}=0}\max\frac{1}{|\{(j,y)\,|\,\bar{P}_{j,y}=1,(j,y)\in W_i\}|}\sum_{(j,y)\in W_i:\bar{P}_{j,y}=1}L_{j,y} \tag{25}$$

where $y\in\{MC,SC\}$, $\bar{P}_{j,y}$ represents the active state of the cell j of type y which takes the value 0 for being turned on and 1 for being turned off, and W_i is the set of neighbor cells of cell i which is defined more explicitly later. (25) selects a cell with the largest loading averaged over its neighbor cells. Similarly, the cell to be turned off will be determined as follows.

$$c_y=\arg_{i,\bar{Y}_i\neq\{\}}\min\frac{1}{|\{(j,y)\,|\,\bar{P}_{j,y}=1,(j,y)\in W_i\}|}\sum_{(j,y)\in W_i:\bar{P}_{j,y}=1}L_{j,y} \tag{26}$$

where $\bar{Y}_i=\{C_j\,|\prod_{(j',y)\in C_j}\bar{P}_{j',y}>0, C_j\in\hat{Y}_i\}$, and $\hat{Y}_i$ is the collection of the neighbor supporter sets which is defined in section 6. (26) simply selects a cell with the lowest loading averaged over its neighbor cells among cells which can be turned off safely with satisfying the coverage constraint.

The complexity of ASC algorithm depends on the reinforcement learning algorithm and the algorithm for selecting a cell to be turned on or off expressed in (25) and (26). The complexity of reinforcement learning algorithm itself is [28]. Since the number of neighbor cells and the number of active or inactive cells are proportional to the number of cells. The complexity due to (25) and (26) will be . Associated with this algorithm no extra signaling is required. However, the cell load of each cell needs to be collected at a centralized processor every period of load balancing algorithm, which is likely to be very minor overhead.

6. Simulation Results

6.1 Simulation Setup

We consider a conventional wrap-around hexagonal downlink network consisting of 57 macro cells where three cells are co-located as three sectors with nominal coverage of 120 degree and randomly located N_{SC} omni-directional small cells. Following conventional transmit power setup [3][26], the transmit powers of the macro cells and the small cells are set to be 20 watt and 1 watt respectively. Since we treat each sector as a cell, the total number of cells is $57+N_{SC}$. Each cell transmits signal with a single antenna. Each MS with two receiving antenna executes maximum ratio combining over received signal. Channel model follows the urban micro model of 3GPP SCM with non-line of sight. The number of MSs are 10 per macro cell at the start of simulation. They are geometrically randomly distributed with uniformly distributed velocity. That is, the m th MS has velocity $(100/570)m$. Each MS moves along a straight line with a random direction.

The average data traffic size for each call arrival was set to be 1k bits. Discount factor $\overline{\beta}$ was set as 0.9. T_H and T_L were 0.9 and 0.7 to make the cell load of each active cell rather high so that it can operate in a power-efficient manner. Simulation was executed over 50000 frames where frame length was 1 sec. For the first 10000 frames, feasible neighbor sets for supporting the coverage of each cell were estimated. Learning algorithm was executed over 40000 frames with the period of 10 frames. The practical active state control may be executed every several minutes or several tens of minutes. However, since it is important to include the effect of the short-term channel characteristics which is closely related with velocity of channel and the quality of service, we focus on the simulation on short-term scale rather than long-term scale. Even though learning period of 10 sec is rather too short, it is expected that setting control period short may not have great impact on characterizing the performance of the proposed algorithm. Adopting the conventionally used parameter values in a practical cellular system, we set simulation parameters. We summarize main simulation parameters in **Table 3**.

To simulate the proposed algorithm, a feasible set collection F needs to be estimated. Brute force search necessitates the exponential complexity with the number of cells which can not be feasible. To deal with this problem, we define a collection of neighbor supporter cells Y_c for cell c as follows

$$Y_c = \{S^{-c} \mid \int_{A_c} \Pr(E_{S^{-c}}(x)) f_Z(z \mid A_c) dz \le \zeta\} \tag{27}$$

where S^{-c} is a set of active cells with the power of the cell c being turned off, and A_c is the SNR coverage of the cell c which satisfies an inequality, $\int_{A_c} \Pr(E_{\{c\}}(x)) f_X(x \mid A_c) dx \le \zeta$.

Table 3. Simulation parameters.

parameter	value
Transmit Power of Macro cells	20Watt
Transmit power of Small cells	1Watt
Path loss exponent	4.05
Bandwidth	20MHz
Operating Frequency	1850MHz
Number of MSs	570
Minimum Inter-Macro BS distance	1000m

This inequality implies that SNR outage is less than ζ. Thus, Y_c is the collection of the sets consisting of neighbors providing the same SNR outage coverage of the cell c while the power of the cell c is turned off. To find Y_c, we define a set of candidate neighbor cells W_c as

$$W_c = \{a \mid \exists x_0 \; such that \; \gamma_a(x_0) \geq \gamma_T, \; a \neq c\} \tag{28}$$

For given W_c, there are $2^{|W_c|}-1$ possible candidate sets which can be included in Y_c. If $|W_c|$ is large, then it will add extra complexity. Thus, one can reduce the size of W_c in some sacrifice of performance which is expected to be negligible. Each cell counts the number of events that both SNRs of cell c, and cell a are greater than SNR threshold. The numbers of events are sorted with decreasing order. Cells corresponding to top N_W in the sorted list are selected as candidates to be used for finding Y_c approximately. At each frame, the event count of the i th combination of neighbor support cells can be counted as follows

$$V_{i,c} = V_{i,c} + \sum_{m \in M_c} (\oplus_{a \in C_i} I(\gamma_{a,m} \geq \gamma_T)) \tag{29}$$

where $C_{i,c}$ is the i th combination of neighbor supporter cells for cell c, M_c is the set of MSs served by the cell c, $\oplus$ is a logical or operation, and $I()$ is an indication function which has value 1 if the condition inside the bracket is true, otherwise 0. Exploiting this information, one can estimate Y_c as

$$\hat{Y}_c = \{C_{i,c} \mid (N_{T,c} - V_{i,c}) / N_{T,c} < \zeta, \; i = 1,\ldots 2^{|W_c|}-1\} \tag{30}$$

where $N_{T,c}$ is the total number of SNR report at the cell c. (30) implies that a collection of neighbor supporter cells is determined from the estimated outage probability of each set of neighbor supporter cells. In the simulation, W_c is estimated after the first 5000 frames. If the number of elements is greater than 10, then some elements of W_c are removed with the rule described above. Then, $\hat{Y}_c$ is determined from $V_{i,c}$ estimated for the second 5000 frames.

6.2. Simulation Results

In this subsection, we investigate the performance of the proposed ASC algorithm with simulation results. Three performance metrics, the percentage of active cells, the number of HOs, and average user perceived rate (UPR) are considered. The percentage of active cells is defined as the ratio between the number of active cells and the total number of cells. It will be given separately for macro cells and small cells to assess the power efficiency more accurately in relation with other metrics. The number of HOs is also an important parameter to be considered for controlling the active state of cells. HO often incurs delay and degradation in service reliability. HO performance in a practical system depends on HO algorithms. However, in this simulation HO is determined instantly with received signal strength for simplicity to be free from a specific HO algorithm and parameterization. Thus, the number of HOs will depend on the mobility of MSs and how often ASC algorithm switches on or off a cell. Depending on how many cells are turned off and which cells are off, service quality may be different. One of representative metrics for service quality at MS is average UPR. UPR is defined to be the ratio between the data traffic size and total time to finish its transmission. Thus, UPR represents the end user experience in having wireless service. In summary, the percentage of active cells, the number of HOs, and average UPR are chosen as representative performance metric to assess the power efficiency, service reliability, and end user service experience.

The performance of the proposed algorithm is likely to depend mostly on the threshold T_H and T_L which are used for deciding the active state depending on cell load. In **Fig. 1**, the effects of these thresholds on the active state were shown for two different traffic densities, 0.01, and 0.001. Two different types of plots were drawn together. First, the percentage of active cells was shown when T_L increased from 0.1 to 0.8 with the step of 0.1 while fixing $T_H = 0.9$. It can be observed that the percentage of active macro cells decreases as T_L increases. It is because there is more chance to turn off a cell when T_L is high. However, the percentage of active small cells is kept constant about 5%. High loading is required to turn on a cell for high T_H. However, small cells usually have low loading due to small coverage unless the wireless network is heavily crowded. Second, the percentage of active small cells was shown to decrease as T_H decreased from 0.9 to 0.2 with the step of 0.1 while fixing $T_L = 0.1$. Low $T_{:L}$ tends to make relatively large percentage of small cells active, since it activates the cells with small loading. However, the percentage of active macro cells does not change much with increasing T_H. As long as traffic density is not very low, and traffic is uniformly distributed geometrically, each macro cell may have some loading due to large coverage. Thus, every macro cell is active when $T_L = 0.1$, and traffic density is 0.01, while there is slight decrease in the percentage of active macro cells when $T_L = 0.1$, and traffic density is 0.001. These results confirm that high T_L leads to aggressive active state control while low T_H and low T_L lead to conservative active state control.

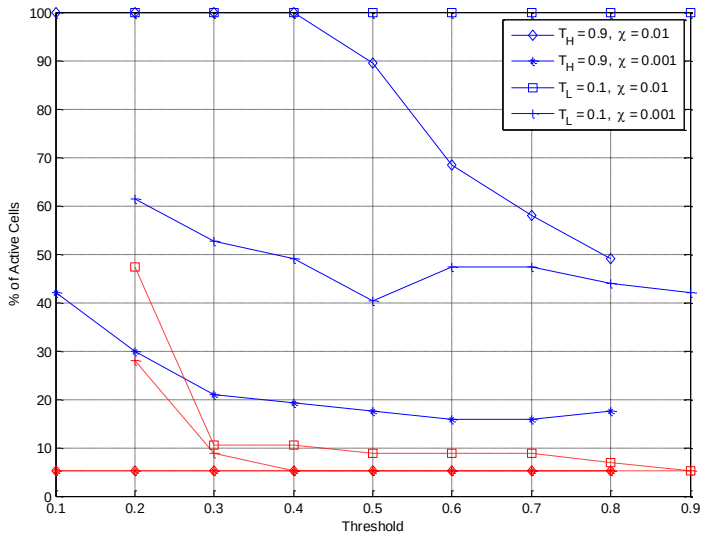


Fig. 1. Effects of load thresholds on active state (solid line : macro cell, dotted line : small cell).

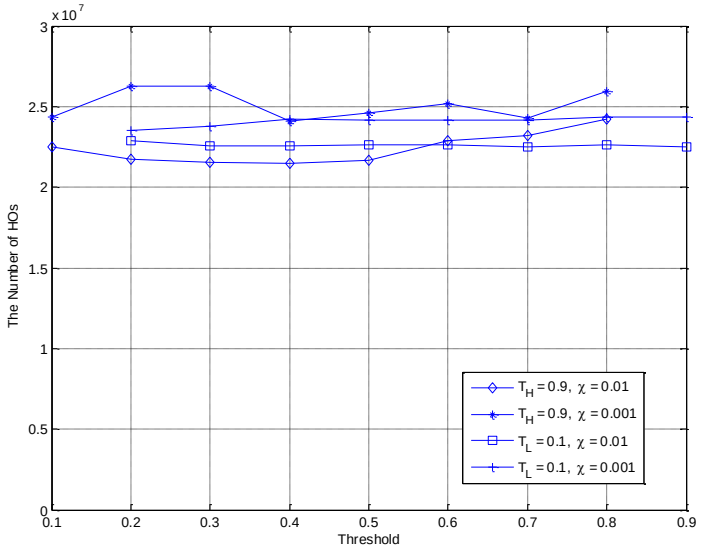


Fig. 2. Effects of load thresholds on the number of HOs.

The effect of load thresholds on the number of HOs is shown in **Fig. 2**. There is no significant change depending on load threshold and traffic density. The number of HOs with traffic density of 0.001 is slightly more than one with 0.01. Since the number of active cells is small when the traffic density is small, turning power on or off influences more MSs. When traffic density is fixed at 0.01, the number of HOs is found to increase slightly with the same reason as T_L increases while $T_H = 0.9$. Since HOs are incurred by both the mobility of MSs and the change in the active state of cells, the variation with load threshold depends on the period of active state control and observation time. Nonetheless, considering the short period of active state control, it is expected that the number of HOs is not very sensitive to the load thresholds.

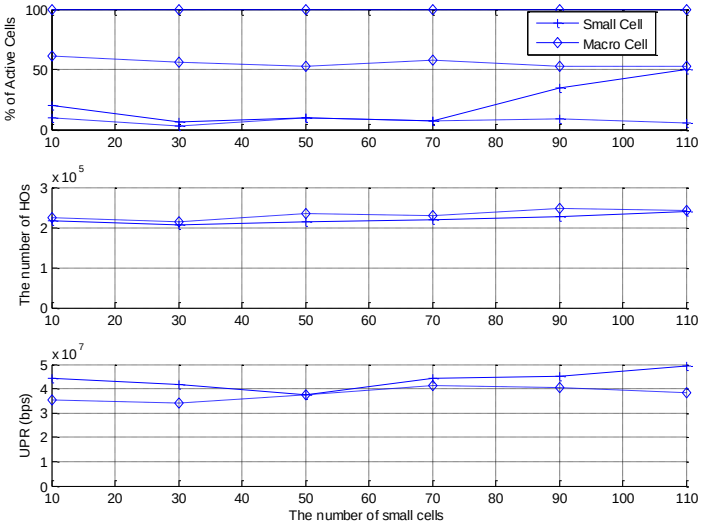


Fig. 3. The effect of the number of small cells on performances (solid line : traffic density of 0.1, dotted line : traffic density of 0.01).

The effect of the number of small cells on performances is shown in **Fig. 3**. It was investigated for two different traffic densities, 0.1, and 0.01. In the top figure, it can be observed that while there is no significant change in the number of active small cells for the traffic density of 0.01, it starts to increase around 70 for the traffic density of 0.1. It is conjectured that as the number of small cells increases, there can be more chance to have small cells with relatively large loading, which produces this result. It is interesting to note that increase in the number of small cells from 10 to 110 does not change the number of active small macro cells much. This result may not be generalized, since massively large number of active cells may have different effect. Due to the limited memory of simulator and simulation time, we could not evaluate the effect of the very large number of small cells which is left for future research with different simulation methodology. In the middle figure, the number of HOs was plotted with the increasing number of small cells. As in **Fig. 1**, the number of HOs with the traffic density of 0.01 is slightly more than one with the traffic density of 0.1, since the percentage of active macro cells is relatively smaller than one with the traffic density of 0.01. It is also observed that the number of HOs increases slightly as the number of small cells increases. However, it is not significant due to the small coverage of small cells. In the bottom figure, average UPRs were compared. In this paper, UPR is defined to be the ratio between the data traffic size and total time to finish its transmission. Thus, UPR represents the end user experience in having wireless service. Since active state is controlled such that the average loading of active cells can be maintained between the T_L and T_H. It is expected that the UPRs is not likely to be constant over different number of small cells. Even though there does not seem to be any particular relationship between UPR and the number of small cells, relative variation of UPR seems to be marginal for the different numbers of small cells.

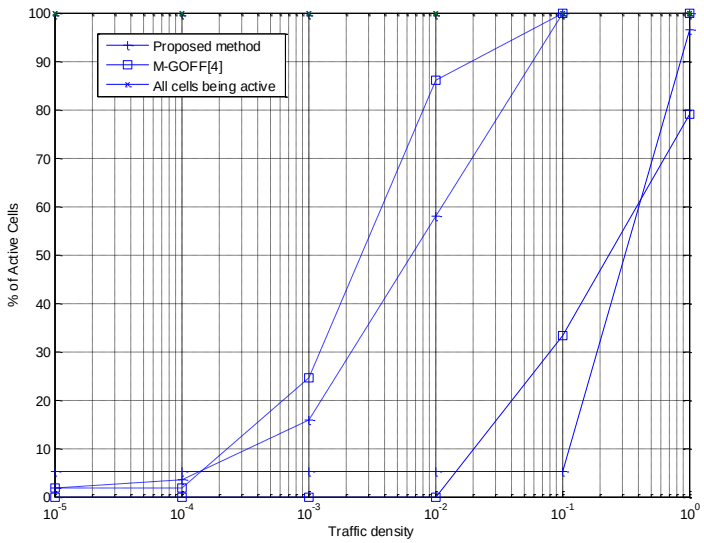


Fig. 4. The percentage of active cells for different data traffic densities (solid line : small cells, dotted line : macro cells).

The proposed algorithm is compared with an existing algorithm and the case of all cells being active in **Fig. 4**. There is no proper existing ASC algorithm for a heterogeneous network to be applicable to the system setup considered, to the best author's knowledge. Greedy OFF (GOFF) algorithm in [4] has good theoretical foundation with good performance verified with simulation results for a homogeneous network. However, GOFF algorithm stops when there is

no more cell to be turned off. Thus, its performance is likely to depend on how long averaging time is set and how often traffic is generated. At the same time, there is no chance to turn on a cell which was turned off. It means that it may not be robust to unstable statistics or dynamicity of data traffic. Thus GOFF was modified with exploiting algorithm structure developed by [17] which we call modified-GOFF (M-GOFF). M-GOFF starts with all cells being active. At each control stage, it compares the cell load of active cells L_c with θ_H and θ_L. If $L_c > \theta_H$ which means that it is over-loaded, the average cell load over its active neighbors is calculated for each inactive neighbor cell of the current over-loaded active cell. A cell among its neighbor inactive cells which has the largest average cell load over its active neighbors is turned on. If $L_c < \theta_L$, the corresponding cell is turned off. We set θ_H and θ_L as 0.9 and 0.1 respectively which seem to provide a reasonably good performance from trying several different parameterizations.

Total power consumption was compared in terms of the percentages of active macro cells and small cells separately in **Fig. 4**. Both the proposed ASC algorithm and M-GOFF are found to respond properly depending on traffic densities. The percentage of active macro cells is larger for both algorithms than that of active small cells in most cases, since macro cells provide larger coverage. Thus, considering the rolls of small cells as taking some portion of loading of macro cells, more refined method is called for as a future research direction. When traffic density is less than or equal to 0.0001, three small cells are still active with the proposed algorithm while their loading is much lower than . This implies that these small cells are active to satisfy the coverage constraint rather than loading condition. It can also be observed that the proposed ASC provides about 35% power gain for macro cells over M-GOFF at the traffic densities of 0.01 and 0.001. It also has about 85% power gain for small cells over M-GOFF at the traffic density of 0.1. For other cases, the proposed algorithm and M-GOFF has similar power gain over the case of all cells being active.

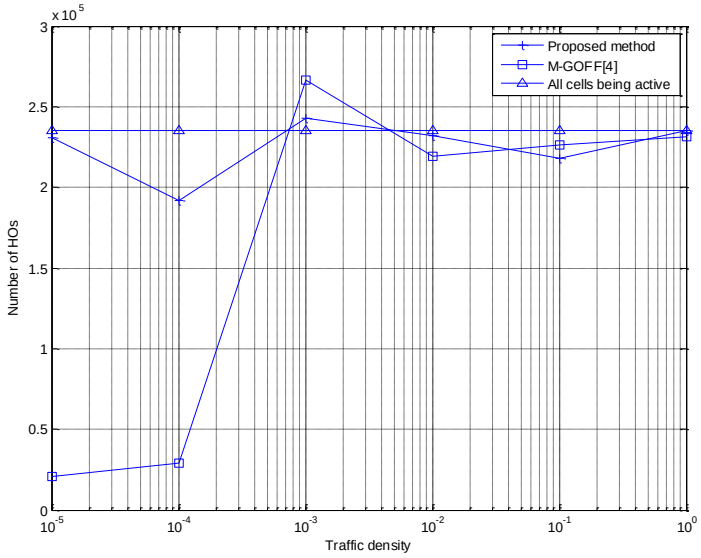


Fig. 5. The number of HOs for different data traffic densities.

The HO performances for different data traffic densities are shown in **Fig. 5**. When all cells are active, the number of HOs is the same regardless of traffic densities. This can be a good baseline to find the effect of the active state control on HOs. The number of HOs with the

proposed algorithm is comparable to the case of all cells being active. This may imply that the proposed ASC algorithm does not incur excessive HOs due to controlling the active states of cells. When traffic densities are greater than equal to 0.001, Difference in the number of HOs between the proposed ASC and M-GOFF is within 10%, which means that both algorithms provide similar HO characteristics as long as the traffic density is not too small. The marginal reduction of the number of HOs with the proposed algorithm at the traffic density of 10^{-4} can be attributed to the faster convergence and the small number of active cells. M-GOFF appears to reduce the number of HOs significantly for traffic densities, 10^{-5} and 10^{-4}, since the number of active cells is 1. Even though controlling this way may be good in terms of power efficiency and the number of HOs, it can result in significant degradation for some other performance measure such as average UPR as shown in **Table 4**.

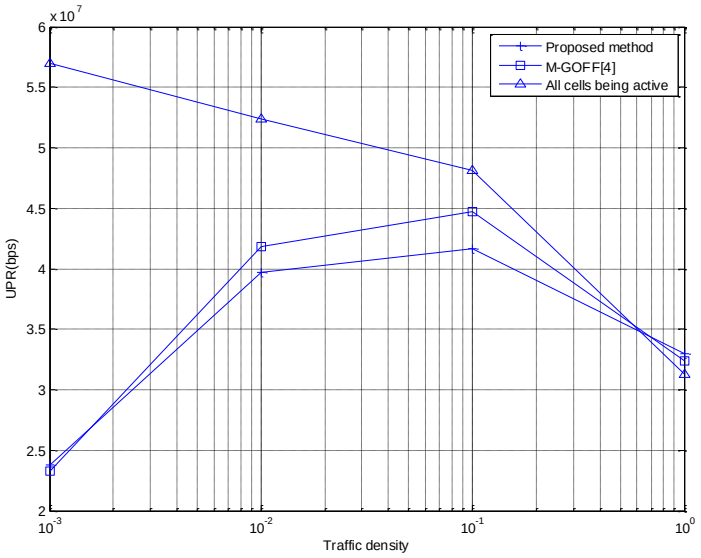


Fig. 6. The average UPR for different data traffic densities.

Fig. 6 shows average UPR for different data traffic densities. The average UPR with all cells being active is shown as baseline performance. As the traffic density decreases, there is more chance to be scheduled, which results in increase in average UPR. Increase in average UPR with decreasing traffic density is not significant due to channel quality degradation with the small number of active cells despite of increased scheduling opportunity with decreasing traffic density. Both the proposed algorithm and M-GOFF provide similar performance for the considered traffic densities. Average UPR influenced mainly by scheduling opportunity for traffic density of 1 while it is by channel condition due to smaller percentage of active cells for traffic density of 10^{-3}. However, the total power consumption of M-GOFF is larger than that of the proposed algorithm by 50 percent or so, which supports the efficiency of the proposed algorithm. Even though both algorithms basically depend on load condition, selecting a cell to be turned off is quite different. While M-GOFF turns off cells of which loading is below a threshold, the proposed algorithm does a cell with the smallest local average loading among cells turned on, which can be a reason why the proposed algorithm works in more power efficient way. When traffic density is less than or equal to 10^{-4}, average UPR may not be statistically stable enough. Thus, instead of average UPR, the numbers of completed data traffics are compared in **Table 4** for these traffic densities. The proposed method completed

the number of data traffics comparable to that of all cells being active while M- GOFF had significantly smaller number of traffics, since a single macro cell was turned on. This result supports that the proposed method provides a robust performance over various system conditions.

We also consider another comparing algorithm which takes a centralized approach with simulated annealing [29] to make sure that the proposed ASC provides better performance than existing algorithms. It tries to maximize the following utility function $U(F_{ON})$.

$$U(F_{ON}) = \alpha \frac{R_O(F_{ON})}{R_{REF}} + (1-\alpha)\frac{\sum_{c\in F_{ON}} P_c}{\sum_c P_c} \tag{31}$$

where α is a weight determining a tradeoff between user edge throughput gain and power gain, $R_O(F_{ON})$ is the user edge throughput which is defined as 5% percentile of UPR, and R_{REF} is a reference edge UPR. In simulated annealing procedure, there is a step for generating successor configuration which allows the change of active state over a single cell. Corresponding cell is randomly selected from cells which changes its active state from M-GOFF. In this regard, this algorithm can be thought as a refined version of M-GOFF, which we call simulated annealing M-GOFF (SA-M-GOFF).

In **Table 5**, we compare the performance of the proposed algorithm with exiting algorithms when traffic density is 0.05 and all MSs move with the same velocity of 3km/h or 60km/h. It can be observed that the proposed ASC algorithm provides similar average UPR to that of M-GOFF while SA-M-GOFF has lower average UPR. 95% confidence intervals of average UPR verify that the difference in the average UPR of the proposed ASC and M-GOFF is not statistically significant while it is marginally lower than the case of all cells being active. The difference in the number of HOs among the proposed ASC, M-GOFF and the case of all cells being active is within 10% of the number of HOs for the case of all cells being active. SA-M-GOFF incurs relatively larger number of HOs due to trying many different active states in simulated annealing process. The proposed algorithm shows that its turns on small cells and turns off macro cells with relatively larger proportion, which provides more power efficient operation that comparing algorithms. SA-M-GOFF seems to perform worse for this simulation condition, which may be attributed to the sensitivity to parameterization of simulated annealing and a cost function. In summary, regardless of the velocity of the MSs, the proposed algorithm shows the consistent performance trends as the case of MS movement uniformly distributed from 0km/h to 100km/h. That is, it provides power efficiency while the number of HOs is comparable to that of the case of all cells being active and the average UPR is similar to that of M-GOFF.

Table 4. Comparison of the number of completed traffics when traffic density is 10^{-4} or 10^{-5}.

Traffic Density	Proposed Method	M-GOFF	NOPC
10^{-4}	103	27	113
10^{-5}	9	2	15

Table 5. Performance of the proposed ASC algorithm when all MSs moves with the same velocity.

MS Movement	Algorithms	% of Active Small Cells	% of Active Macro Cells	No. of HOs	Average UPR (Mpbs)	Lower Bound of 95% Confidence Interval of Average UPR (Mbps)	Upper Bound of 95% Confidence Interval of Average UPR (Mbps)
3km/h for all MSs	Proposed ASC	8.77	59.65	55191	30.15	28.34	31.95
	M-GOFF	0	65.91	58434	31.23	29.22	33.24
	SA-M-GOFF	1.75	91.23	70433	21.43	19.68	23.17
	No ASC	100	100	59574	35.01	33.41	36.61
60km/h for all MSs	Proposed ASC	1.75	36.84	263697	31.79	29.20	34.38
	M-GOFF	0	66.67	248643	33.50	31.07	35.93
	SA-M-GOFF	0	68.42	250798	30.59	28.34	32.83
	No ASC	100	100	264125	47.45	44.58	50.31

7. Conclusions

In this paper, an active state control algorithm for a heterogeneous network was formulated into approximate Markov decision process. A learning algorithm with a random policy with coverage constraint was proposed to solve the problem. Simulation results verify that the proposed algorithm can properly controls the number of active cells depending on traffic density while providing consistent UPR for a wide range of traffic densities.

There are some limitations on this research, since we focused on the problem formulation and the characterization of the proposed algorithm. The proposed algorithm implicitly considers the coverage constraint only. To provide better user experience, it will be required for a system to provide consistent QoS. Controlling the active state of each cell can have significant influence on QoS. Thus, additional constraints such as minimum UPR or minimum delay need to be added to the active state control problem. However it may require the joint optimization of HO, power control, and load balancing, which is a very difficulty problem. This research also can be further extended to the case of a heterogeneous network with device to device (D2D) communication. Since D2D can be considered as an extension of a cell coverage, it may call for a totally different approach to control the active state of each cell.

References

[1] E. Oh, X. Liu, Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, Vol.49 , No.6 , pp. 56 - 61, June, 2011. Article (CrossRef Link)

[2] L.M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, L. Van der Perre, "Challenges and Enabling Technologies for energy aware mobile radio networks," *IEEE Communications Magazine*, Vol.48, No.11, pp. 66-72, Nov. 2011. Article (CrossRef Link)

[3] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume and A. Fehske, "How much energy is needed to run a wireless network?," *IEEE Wireless Commun.*, Vol. 18, No. 5, pp. 40-49, Oct. 2011. Article (CrossRef Link)

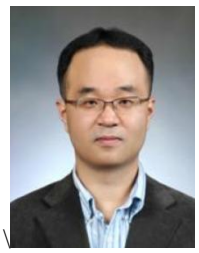
[4] K. Son, H. Kim, Y. Yi and B. Krishnamachari, "Base Station Operation and User Association Mechanisms for Energy-Delay Tradeoffs in Green Cellular Networks," *IEEE J. Select. Areas Commun.*, Vol. 29, No. 8, pp. 1525-1536, Aug. 2011. Article (CrossRef Link)

[5] Z. Yang and Z. Niu, "Energy Saving in Cellular Networks by Dynamic RS-BS Association and BS Switching," *IEEE Trans. Veh. Technol.*, Vol.62, No.9, pp. 4602–4614, Nov. 2013. Article (CrossRef Link)

[6] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal Energy Savings in Cellular Access Networks," in *Proc. of IEEE International Conference on Communications Workshops*, Dresden, June 2009. Article (CrossRef Link)
[7] R. Gupta and E. C. Strinati, "Base-Station Duty-Cycling and Traffic Buffering as a Means to Achieve Green Communications," in *Proc. of IEEE Vehicular Technology Conference (VTC Fall)*, Quebec City, Sept. 2012. Article (CrossRef Link)
[8] L. Chiaraviglio, D. Ciullo, G. Koutitas, M. Meo, and L. Tassiulas, "Energy-efficient planning and management of cellular networks," in *Proc. of 9th Annual Conference on Wireless On-Demand Network Systems and Services (WONS)*, 2012. Article (CrossRef Link)
[9] W.-C. Liao, M. Hong, and Z.-Q. Luo, "Base station activation and linear transceiver design for utility maximization in Heterogeneous networks," in *Proc. of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, BC, May 2013. Article (CrossRef Link)
[10] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "Distributed Inter-BS Cooperation Aided Energy Efficient Load Balancing for Cellular Networks," *IEEE Trans. Wireless Commun.*, Vol.12, No.11, pp. 5929–5939., Nov. 2013. Article (CrossRef Link)
[11] A. Bousia, E. Kartsakli, L. Alonso, and C. Verikoukis, "Dynamic energy efficient distance-aware Base Station switch on/off scheme for LTE-advanced," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, Dec. 2012. Article (CrossRef Link)
[12] G. H. Carvalho, I. Woungang, A. Anpalagan, R. W. Coutinho, and J. C. Costa, "A semi-Markov decision process-based joint call admission control for inter-RAT cell re-selection in next generation wireless networks," *Computer Networks*, Vol. 57, No.17, pp. 3545–3562., Dec. 2013. Article (CrossRef Link)
[13] M. Coupechoux, J.-M. Kelif, and P. Godlewski, "SMDP approach for JRRM analysis in heterogeneous networks," in *Proc. of 14th European Wireless Conference*, Prague, June 2008. Article (CrossRef Link)
[14] E. Stevens-Navarro, Y. Lin, and V. Wong, "An MDP-Based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks," *IEEE Trans. Veh. Technol.*, pp. 1243–1254, Mar. 2008. Article (CrossRef Link)
[15] D. Cao, S. Zhou, and Z. Niu, "Optimal Combination of Base Station Densities for Energy-Efficient Two-Tier Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications IEEE Trans. Wireless Commun.*, Vol.57, No.2, pp. 4350–4362, Mar. 2008. Article (CrossRef Link)
[16] L. Saker, S. Elayoubi, R. Combes, and T. Chahed, "Optimal Control of Wake Up Mechanisms of Femtocells in Heterogeneous Networks," *IEEE J. Select. Areas Commun. IEEE Journal on Selected Areas in Communications*, Vol.30, No.3, pp. 664–672. Mar. 2012. Article (CrossRef Link)
[17] S. Kim, S. Choi, and B. G. Lee, "A Joint Algorithm for Base Station Operation and User Association in Heterogeneous Networks," *IEEE Commun. Lett.*, Vol. 17, No.8, pp. 1552–1555, Aug 2013. Article (CrossRef Link)
[18] Y. Cui, V. K. N. Lau, and Y. Wu, "Delay-Aware BS Discontinuous Transmission Control and User Scheduling for Energy Harvesting Downlink Coordinated MIMO Systems," *IEEE Trans. Signal Process.*, Vol.60, No.7, pp. 3786–3795., July 2011. Article (CrossRef Link)
[19] K. Son, S. Nagaraj, M. Sarkar, and S. Dey, "QoS-aware dynamic cell reconfiguration for energy conservation in cellular networks," in *Proc. of 2013 IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, Apr. 2013. Article (CrossRef Link)
[20] D. P. Bertsekas, Dynamic programming and optimal control, 3rd ed. Belmont, Mass.: Athena Scientific, 2005.
[21] 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)
[22] R. S. Sutton, and A. G. Barto, Reinforcement Learning, Cambridge, Mass. The MIT Press, 1998.
[23] C. Liu, B. Natarajan, and H. Xia, "Small Cell Base Station Sleep Strategies for Energy Efficiecy," *IEEE Trans. Vehi. Tech*. Vol. 65, No. 3, pp. 1652-1661, Mar. 2016. Article (CrossRef Link)
[24] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, "Energy Efficient Heterogeneous Cellular

Networks," IEEE Journal on Select. in Commun. Vol. 31, No. 5, pp. 840-850, May 2013. Article (CrossRef Link)

[25] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Multiobjective Auction-based Switching Off Scheme in Heterogeneous Networks," *IEEE Trans. on Vehi. Tech*. Article (CrossRef Link)
[26] J. Pang, J. Wang, D. Wang, G. Shen, Q. Jiang, and J. Liu, "Optimized time-domain resource partitioning for enhanced inter-cell interference coordination in heterogeneous networks," in *Proc. of WCNC*, Shanghai, China, 2012. Article (CrossRef Link)
[27] E. Chavarria-Reyes, I. F. Akyildiz, and E. Fadel, "Energy Consumption Analysis and Minimization in Multi-Layer Heterogeneous Wireless Systems," *IEEE Trans. Mobile Comp.* Vol. 14, No. 12, pp.2474-2487, Dec. 2015. Article (CrossRef Link)
[28] M. Wiering, and J. Schmidhuber, "Fast Online Q(λ)," *Machine Learning*, 33, pp.105-115, 1998. Article (CrossRef Link)
[29] G. P. Koudouridis, G.Hui, and P. Legg, "A Centralized Approach to Power On-Off Optimization for Heterogeneous Networks," in *Proc. of IEEE VTC Fall*, pp.1-5, Quebec City, Sept. 2012. 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 new media contents, Seoul Media 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 Seoul Media Institute of Technology, Seoul, since 2010. He has published numerous papers in the area of multi-antenna transmission, signal processing, and control. His research interest includes wireless system and network, artificial intelligence, control theory, neuroscience, and affective computing.