Transfer Reinforcement Learning Framework for Energy Saving in Next Generation Wireless Networks

Student Name: Shreyata Sharma

IIIT-D-MTech-ECE
July 6, 2016

Indraprastha Institute of Information Technology
New Delhi

<u>Thesis Committee</u>
Dr. Anand Srivastava (Chair)
Dr. Sumit Darak (Chair)
Dr. Anubha Gupta
Dr. Keith Blow

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Electronics and Communication Engineering

Keywords: Reinforcement learning, transfer learning, actor-critic algorithm, energy efficient access networks, moderate traffic, heterogeneous networks.

Certificate

This is to certify that the thesis titled **"Transfer Reinforcement Learning Framework for Energy Saving in Next Generation Wireless Networks"** submitted by **Shreyata Sharma** for the partial fulfilment of the requirements for the degree of *Master of Technology* in *Electronics and Communication & Engineering* is a record of the bonafide work carried out by her under my guidance and supervision at Indraprastha Institute of Information Technology, Delhi. This work has not been submitted anywhere else for the reward of any other degree.

Dr. Anand Srivastava and Dr. Sumit Darak
Indraprastha Institute of Information Technology, New Delhi

Abstract

Recent upsurge in data intensive applications over wireless communication networks is stimulating rapid expansion of such networks and thus presenting new research challenges pertaining to their efficient deployment. In the present communication networks, the increased traffic load entails network operators to expand their networks by the deployment of a large number of base stations (BSs) and access points (APs). Studies have reported that a major portion of energy consumption occurs at the access network entities. This means that the massive data traffic is being served at the expense of increased carbon footprint and huge energy consumption. Therefore, energy saving has emerged as one of the major aspects in such data intensive and high traffic communication networks. Considering this, the energy efficient operation of BSs and APs has become a major research problem and it is well taken up in this thesis for the case of wireless networks.

In this work, the research aim of energy saving has been considered for both, cellular BSs and Wi-Fi APs to cover the major part of the wireless communication networks. An actor-critic (AC) reinforcement learning (RL) framework is used to enable traffic based ON/OFF switching of BSs and APs. Furthermore, previously estimated traffic statistics is exploited through the process of transfer learning for further improvement in energy savings and speeding up the learning process. Herein, this novel approach is used for three cases: realization of a transfer learning framework for Wi-Fi networks, implementation of a three state RL based BS switching scheme for existing cellular networks and application of RL in heterogeneous networks (HetNets) consisting of macro and femto BSs. The use of practical scenario and real time data collected from institute's Wi-Fi network to validate the adopted scheme is an important feature of this study. The superiority of the proposed framework is depicted through simulations and relevant mathematical analysis.

Acknowledgments

Knowledge is a noble blessing of God. Thus, to be in this environment and become a part of new creations is like a fortunate blessing. I therefore first humbly acknowledge this blessing that came to me through God, my teachers and parents. My research efforts in the present M.Tech thesis and thereby contribution in technology and societal development is inspired by their gracious presence. I thus express my deepest gratitude to my thesis supervisors, Dr. Anand Shrivastava and Dr. Sumit Darak who introduced me to an interesting research challenge and cultivated my interest in grasping prior findings and imagine the new scope for their further extension. Their constant support, patience, encouragement and belief in my abilities is the sole factor behind the successful completion of this work.

Inspiration and motivation needs a suitable platform to be converted into a meaningful work, I would thus like to thank Dr. Pankaj Jalote, Director of the Institute and my teachers at IIIT Delhi for providing such an inspirational and supportive environment to work, specially so conducive for higher learning and research. Further, I sincerely thank the related lab, academic, library and administrative staff of the institution for their support in carrying out this research.

At this important juncture of my quest to live for my interest in learning and research, I would like to express my gratitude to my parents and my younger sister, Saloni, whose support and affection kept my moral high in all difficult and challenging situations.

I specially thank my friends who supported me in many ways. Their cheering presence gave me emotional and spiritual strength to think beyond myself and face tough challenges so lightly.

Contents

List of Figures

List of Tables

List of abbrevation

ICT	information and communication technology
OpEx	operational expenses
CapEx	capital expenses
BS	base station
AP	access point
RL	reinforcement learning
AC	actor-critic
HetNet	hetrogenous network
MDP	Markov decision process
LTE	long term evolution

Chapter 1

Introduction

1.1 Background

Over the past few years, the information and communication technology (ICT) sector has seen major technological advancements and there has been an explosive growth of data intensive applications over the communication networks. This massive data traffic is being served at the expense of increased carbon footprint and huge energy consumption. It has been reported that ICT industries contribute to 2-10% of worlds overall total energy consumption [1]. Furthermore, higher energy consumption would mean higher capital expenses (CapEx) and operational expenses (OpEx) for the network operators. Therefore, reducing the energy consumption in ICT operations is a major research challenge and it is extremely crucial to handle this from both ecological and economical perspectives.

Studies have shown that more than 55% of energy consumption in the current communication systems take place at the access network. To serve high traffic and high data rates, there is a need for deployment of large number of base stations (BSs) and access points (APs) in the access networks leading to high energy consumption [1]. Therefore, in order to reduce energy consumption in such network, ensuring energy efficient operation of BSs and APs is extremely important. In the current deployment of access networks, the BSs and APs are more or less active all the time with the capacity to serve peak load. The aspect of variation in the traffic load which is a practical scenario is generally not taken into account. This kind of deployment leads to inefficient usage of access network resources when the traffic load is low and causes high energy consumption. Therefore, there is a need to develop an optimal switching scheme such that the BSs/APs are switched ON/OFF according to the traffic load. Furthermore, the existing networks do not fully exploit the past usage statistics for optimal operation of BSs/APs. It is seen that there is a reasonable correlation between the current data traffic and the data traffic in the past. In the case of Wi-Fi networks, this could be attributed to a regular schedule of activities in the concerned organization or area where these are deployed. For instance, in an academic institution, a similar schedule is followed on the same days of the week. In case of cellular networks, there is a typical day-night behavior of users and daily movements of users carrying their mobile devices from residential to office areas and back [1]. Hence, there is a high possibility that the traffic variations at a given location follow a similar pattern at same time instants on different days. Therefore, the past data statistics is indeed significant and could be used effectively for devising present energy saving scheme. Considering these factors, there is an ample scope of new research initiative to address this vital aspect of design of energy efficient communication networks.

1.2 Motivation and Objectives

Recently, there has been increased research on traffic based BS switching in cellular networks. However, application of these schemes on Wi-Fi network APs and on real time data have not been studied extensively. This calls for significant research initiatives in this direction due to the deployment of new data-intensive services ranging from high-speed data to multimedia and operator's interest in cellular data off-loading on Wi-Fi networks. Same way, in the case of cellular networks, there are some bottlenecks in the current research direction for BS switching. In most of the research works, two state BS switching is considered wherein BSs are switched between the active mode and sleep mode according to the traffic load. In the active state, the BSs operate in tri-sectorized mode in which for each active sector there is an active power amplifier which consumes a large amount of energy. However, when the traffic load is moderate the BSs can be configured to operate in omnidirectional mode. This could lead to a reduction in system energy consumption owing to lower energy consumption of BSs operating in the omnidirectional mode as compared to those in tri-sectorized mode. Apart from this, to make the cellular networks more power efficient and sustain high speed data traffic, the propagation distance between nodes can be decreased to reduce the transmission power. Hence, next generation communication systems are marked by significant deployment of small cells. This calls for sincere research efforts for devising energy saving schemes for herogeneous networks (HetNets).

In the backdrop of above discussed status of technology and research, the objective of this work is to make efforts towards realization of energy efficient access networks for next generation communication systems. Broadly, the main objectives of this work are:

- Realization of a transfer learning framework for energy efficient Wi-Fi networks and performance analysis using real data.
- Implementation of a three state reinforcement learning (RL) based BS switching scheme.
- Application of RL for energy saving in HetNets consisting of macro and femto BSs.

1.3 Related Work

Traffic load based dynamic BS switching has been identified as a promising technique for energy efficient operation of wireless access networks. This technique is studied in [2–4] assuming prior information about the traffic load. In [2], a theoretical framework for BS energy saving is developed. A cost minimization problem is formulated and its solution is obtained using greedy ON-OFF algorithm. In [3], an optimization problem is formulated which is aimed at minimization of BS energy consumption with constraint that blocking probability is less than a threshold value. The traffic profile is modelled as a periodic sinusoidal profile and the key factors affecting energy saving are analysed. It is deduced that the energy saving is a function of traffic parameters and the number of neighbouring BSs. In [4], a centralized greedy algorithm is provided which requires the complete traffic and channel information followed by a decentralized approach, which requires only the local load information. In the centralized approach, the energy saving problem is formulated as an optimization problem with constraint on the bandwidth occupancy of the user associated with a certain BS. In decentralized approach, information requirement by the centralized algorithm is relaxed by triggering user-specific BS association by designing a BS selection preference function. In [5], RL is applied for optimal BS switching assuming traffic load to follow a Poisson distribution. Furthermore, the concept of transfer learning discussed in [6–9] is effectively applied to BS switching problem. Transfer learning has emerged as an effective way to exploit knowledge from a previous task (source task) to more efficiently learn and solve a new related task (target task).

The advent of long term evolution (LTE) is marked by the growth of small cell deployments. This is primarily because of the boost in capacity and quality provided by small cells which is extremely critical, given the tremendous surge in data traffic [10]. Moreover, due to shorter distances between the transmitter and receiver, same quality of service (QoS) can be achieved using lower transmit power, increasing energy efficiency of the communication network. This trend has given rise to expanding research interest in energy efficient HetNet deployment techniques including dynamic switching of BSs in HetNets. Sleep mode techniques for small cells are broadly introduced in [11]. In [12] optimal sleep/wake-up mechanism is derived using MDP (Markov decision process). The solution to the MDP is obtained through dynamic programming based Blackwell optimality conditions. However, for implementation simplicity a deterministic policy is used. A simple fixed time femto BS sleeping scheme is presented in [13]. Optimization problem formulation is used to derive ON/OFF scheme for two tier networks in [14–16].

The growth of research activities pertaining to dynamic BS/AP switching is evident from the above discussion. However, there are a few bottlenecks in the state of art dynamic switching schemes:

- There is a limited research on the application of these schemes in Wi-Fi networks and on real time data. Although, the concept of Wi-Fi AP sleep mode is introduced in [17] but the APs are switched ON and OFF periodically without learning the traffic pattern.

- In case of cellular networks, the current research mainly focuses on two state BS switching schemes. The BSs are switched to an active state at high traffic and sleep state at low traffic. The fact that moderate traffic can be served with lower transmit power and has a lower capacity requirement have not been exploited.

- In a practical scenario, when certain BSs are switched off the users would associate with different set of BSs thereby changing the traffic loads on these BSs. Hence, consecutive BS switching operations are correlated. The current BS switching operation would affect overall energy consumption in the future. Learning based BS switching schemes can provide foresighted BS switching strategy which would increase system energy efficiency in the long run [5]. Therefore, significant research efforts are needed in this direction.

- As discussed earlier, there could be a reasonable similarity in the current and future traffic pattern at a given location. Therefore, the past data statistics could be effectively used for devising present energy saving scheme. There is a limited research to exploit this concept of knowledge transfer.

  Effective application of RL along with transfer learning to achieve energy saving in cellular networks discussed in [5] provides an incentive to use this concept in various related research domains. In the current work, this concept is effectively used to:

  i) Achieve energy saving in localized Wi-Fi networks through an effective ON/OFF switching scheme exploiting transfer learning for Wi-Fi APs.

  ii) Realize a three state BS switching scheme for cellular networks to increase energy efficiency at moderate traffic.

  iii) Derive an optimal BS switching scheme for HetNets consisting of macro BSs and femto BSs.

1.4 Thesis Outline

The research problem formulation and findings of this work are organized as follows:

Chapter 2 describes the design of an actor-critic (AC) RL framework to enable traffic based ON/OFF switching of APs in Wi-Fi network. Furthermore, it discusses the use of knowledge transfer and validation of the proposed framework using real data.

Chapter 3 discusses the use of AC learning framework and transfer learning concept for dynamic sectorization based BS switching scheme. Herein, a three state BS switching scheme is studied in which apart from an active state at high traffic and sleep state at low traffic, the BSs are switched to an omnidirectional state at moderate traffic leading to a more efficient energy saving scheme. The performance of the proposed scheme is analysed through relevant mathematical formulations and simulations.

Chapter 4 discusses the application of RL for energy saving in HetNets consisting of macro and femto BSs. The performance of the proposed scheme is evaluated through simulations and a brief analysis on trade-off between the energy consumption and QoS is presented.

Chapter 5 concludes the thesis and suggests future research directions.

Chapter 2

Transfer Learning Framework for Energy Saving in Wi-Fi Network

2.1 Introduction

Rapid growth of business and institutional entities and the need for cellular data off-loading has led to a phenomenal increase in localized Wi-Fi network deployment. High data rate offered by Wi-Fi networks is being increasingly used to accommodate the recent upsurge in data intensive applications. Growing concern about the energy efficient operation of Wi-Fi networks is a natural consequence of their increased deployment. In such networks, a major portion of energy consumption occurs at the access network entities making energy efficient operation of Wi-Fi APs extremely crucial. The previous research works in this domain [2–5] mainly focus on traffic based switching of BSs in cellular networks. Moreover, in these works, switching schemes are applied on assumed traffic load or considering the traffic load to follow a given distribution. Switching scheme for Wi-Fi APs is previously discussed in [17]. However, APs are switched periodically without learning the traffic pattern. Considering the aforementioned factors, in this work a novel approach is discussed in which AC learning algorithm is used for switching the APs according to the traffic load variations. Furthermore, the previously learned data statistics is well exploited by using transfer learning approach in which data from appropriate period from the past is utilized to make decision on the AP switching at the present. This scheme indeed leads to further improvement in performance and speeds up the learning process. Moreover, to have a well founded analysis, the algorithm is applied on real data collected from distinct APs of a W-Fi network deployed in an academic institution.

2.2 System Model

For the present study, Wi-Fi network of an academic institution (IIIT-Delhi campus) has been taken as a model. The algorithm is applied on a defined area of the campus as indicated in Figure 2.1. The data is in five minutes average format, i.e a single data point corresponds to traffic load averaged over five minutes interval. The traffic data is collected from 20 distinct APs as indicated in Figure 2.1 and a Markov decision process (MDP) is formulated using the traffic variations. To ascertain reliability of the adopted scheme, the algorithm is applied on two set of APs. Set-1 APs are the subset of APs serving lecture blocks of the institution and Set-2 APs serve other areas like cafeteria and faculty cabins. An attempt has been made to cover all the APs in the academic block. Lecture halls are not present on the third, fourth and fifth floor, therefore lecture block APs are not present on these floors.

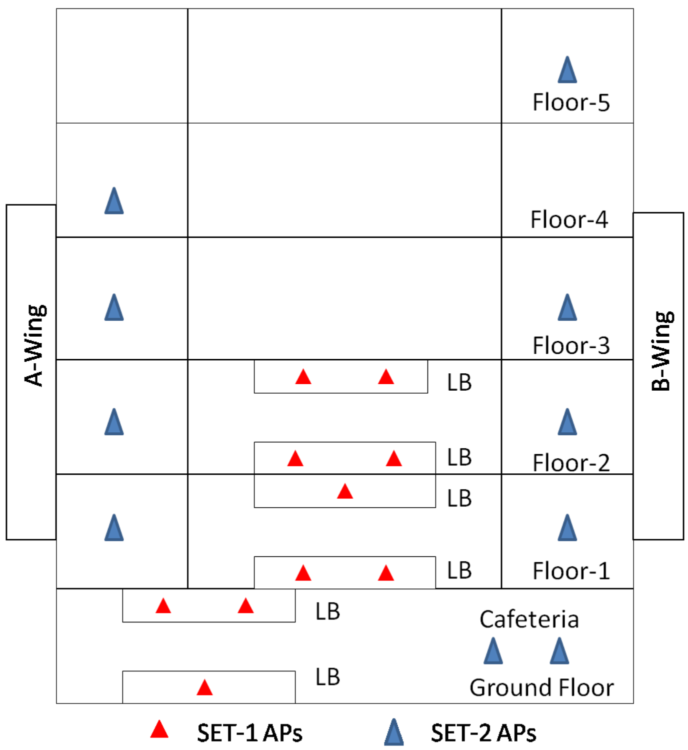


Figure 2.1: Floor plan of academic building with 20 APs distributed over 5 floors.(LB:Lecture Block)

An MDP is characterized by the tuple $M = <S, A, P, C>$, where S is the state space, A is the action space, P is the state transition probability, C is the cost function. At stage k, traffic load is at state $\mathbf{s}^{(k)}= \{s_1^{(k)}, s_2^{(k)}, ...\}$, where $s_i^{(k)}$ represent state of i^{th} AP at stage k. When an action $\mathbf{a}^{(k)}= \{a_1^{(k)}, a_2^{(k)}, ...\}$ is taken, the i^{th} AP is switched OFF if $a_i^{(k)} = 0$. Otherwise, if $a_i^{(k)} = 1$, the AP remains ON. If n APs are considered, there would be 2^n possible states and actions.

The system cost, C in this case is power consumption of the system. Power consumption of an AP consist of two parts: static power consumption which is independent of traffic load and dynamic power consumption which varies proportional to the traffic load. This can be expressed as follows:

$$C = \sum_{i \in A'} [(1 - q_i)\rho_i P_i + q_i P_i] \tag{2.1}$$

where q_i is fraction of static power consumption of i^{th} AP, ρ_i is traffic load density, P_i is total power consumption and A' is the set of active APs.

2.3 AC Algorithm based ON/OFF Switching Scheme for APs

As discussed in section 2.2, the traffic load variations are modelled as an MDP and an optimal AP switching scheme can be established by finding its solution. In this work, the solution to the formulated MDP is obtained through AC learning algorithm through a series of relevant steps discussed here in the subsections.

AC learning algorithm is a subclass of RL algorithms. In general, RL framework consists of an agent and an environment. There is a continuous interaction between the agent and the environment. At each time step, the agent implements a mapping from states to action, which is called agent's policy. There is a reward (or cost) associated with each action. The goal of RL is to maximize the reward received in the long run. If a problem is modelled in such a way that each action has a cost associated with it, the goal of RL is to minimize the cost incurred in the long run. In precise terms, if the sequence of rewards received after time step t is denoted by $r_{t+1}, r_{t+2}, r_{t+3}....$ then the goal of RL is to maximize the expected value of the return. The return, R_t can be numerically expressed as:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ... = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{2.2}$$

where, γ is the discount factor having value between 0 and 1. It can be seen that the value of γ^k decreases as k increases. This term is included to incorporate the fact that the worth of immediate reward is greater than the later rewards. In the present case, the goal of is to minimize the expected value of the discounted return which is termed as state value function. The state value function is given by,

$$V^\pi(s) = E_\pi \left[\sum_{k=0}^{\infty} \gamma^k C(s^k, \pi(s^k)) | s^{(0)} = s \right] \tag{2.3}$$

where, E is the expectation operator, $C(s^k, \pi(s^k))$ represents system cost at stage k which depends on state s^k and action $\pi(s^k)$ [18]. In the current work, the problem is modelled in such a way that there is a cost associated with each action. Here, the cost is the system energy consumption given in Eq. (2.1). Hence, the return used in the computation of state value function is in terms of cost instead of reward.

In AC algorithm, the policy structure is called the 'Actor' as it selects the action and the value function acts as a 'Critic' as its value determines how good is the action taken and consequently decides the future course of action.

In the present context, the objective of AC algorithm is to find an optimal strategy π which maps every state 's' to an action $\pi(s^{(k)})$ such that system cost, C, is minimized. As the learning proceeds, the policy structure tends towards optimal value and at each state optimal action is taken such that the energy consumption of the system is minimized.

The AC algorithm for optimal AP switching is applied to the system through a sequence of steps summarized in Figure 2.2. These steps are action selection, state value function update, policy update and knowledge transfer which are elaborated in this section. Prior to the application of these steps the policy and the state value function are initialized. Consequently, their values get updated at each stage of learning. Further, if n APs are considered the state value function is initialized as an $2^n \times 1$ vector as there is a state value associated with each of the 2^n possible states. Similarly, the policy would be a $2^n \times 2^n$ matrix as there is a policy value corresponding to each state-action pair.

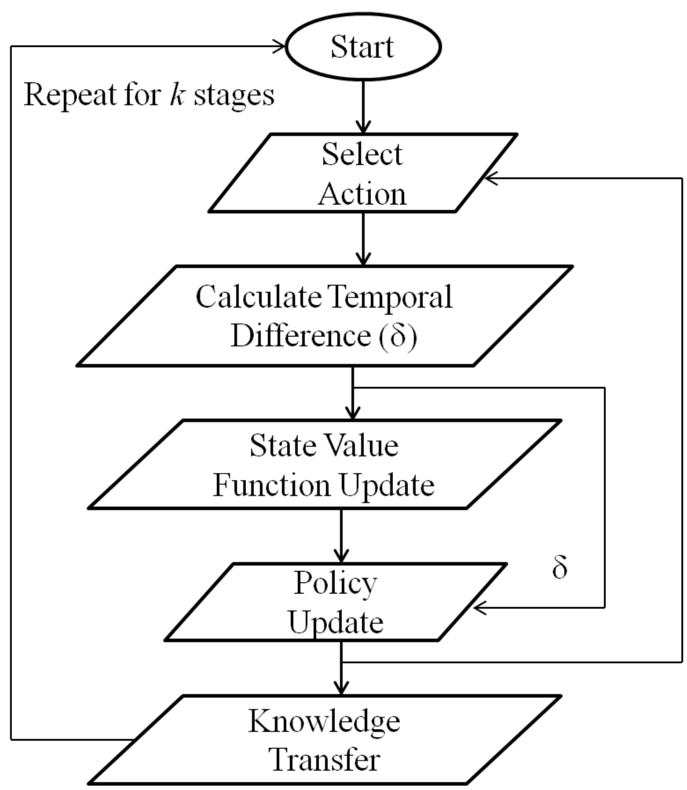


Figure 2.2: Flowchart of AC algorithm.

2.3.1 Action Selection

Action selection is done according to the policy structure. The selected action determines which APs to be switched ON and OFF when the system is at a given state. In deterministic terms, if the traffic load corresponding to an AP is high, it should be switched ON, otherwise should be switched OFF. However, a deterministic policy would inhibit the exploration factor in learning and may not guarantee convergence for various traffic models. Therefore, in this work a probabilistic policy is used in which action selection is done according to Boltzmann distribution as given in Eq. (2.4) [18]:

$$\pi^{(k)}(s^{(k)}, a) = \frac{exp\{p(s^{(k)}, a)/\tau\}}{\sum_{a' \in A} exp\{p(s^{(k)}, a')/\tau\}} \tag{2.4}$$

where, $\pi^{(k)}(s^{(k)}, a)$ is the probability with which an action is taken at a given stage, τ is a positive constant and $p(s^{(k)}, a)$ specify the tendency to select an action a in state $s^{(k)}$.

2.3.2 State Value Function Update

The data transmission occurs after the APs are switched according to the selected action. Consequently, load at active APs change which in turn changes the state of the system. This requires a corresponding change in the value function discussed in Eq. (2.3). To update the value function, the temporal difference (δ) is calculated as follows:

$$\delta(s^{(k)}, a^{(k)}) = C^{(k)}(s^{(k)}, a^{(k)}) + \gamma V^{(k)}(s^{(k+1)}) - V^{(k)}(s^{(k)}) \tag{2.5}$$

Hence, $\delta(s^{(k)}, a^{(k)})$ is the difference between estimated state value function at the preceding stage, $V^{(k)}(s^{(k)})$ and the foresighted state value function at stage k if action $a^{(k)}$ is taken at state $s^{(k)}$ i.e., $E_\pi[C(s^k, a^k) + \gamma^k C(s^{k+1}, a^{k+1}) + \gamma^{k+1} C(s^{k+2}, a^{k+2}) ....] = E_\pi[C(s^k, a^k) + \gamma(C(s^{k+1}, a^{k+1}) + \gamma^k C(s^{k+2}, a^{k+2}) ....)] = C^{(k)}(s^{(k)}, a^{(k)}) + \gamma V^{(k)}(s^{(k+1)})$.

A negative value for $\delta(s^{(k)}, a^{(k)})$ suggest that the foresighted cost of taking the corresponding action at that state is less than the estimated system cost. Therefore, the value function corresponding to the next state after taking action $a^{(k)}$ should be decreased. Hence, value function is updated as:

$$V^{(k+1)}(s^{(k)}) = V^{(k)}(s^{(k)}) + \alpha(\nu_1(s^{(k)}, k))\delta^{(k)}(s^{(k)}, a^{(k)}) \tag{2.6}$$

where, $\nu_1(s, k)$ represents number of times state s occurs in k stages and $\alpha(.)$ is the positive step size parameter. From the update rule it is clear that the value function decreases if δ is negative and increases otherwise [5].

2.3.3 Policy Update

State value function update is followed by policy update so as to increase the probability of taking a favourable action at a given state and decrease its probability if the action leads to an increase in the system cost. The policy update is done according to,

$$p^{(k+1)}(s^{(k)}, a^{(k)}) = p^{(k)}(s^{(k)}, a^{(k)}) - \beta(\nu_2(s^{(k)}, a^{(k)}, k))\delta^{(k)}(s^{(k)}, a^{(k)}) \tag{2.7}$$

where, $\nu_2(s, k)$ represents number of times action $a^{(k)}$ is taken at state $s^{(k)}$ in k stages and $\beta(.)$ is the positive step size parameter. Certainly, a negative δ indicates a favourable action as discussed earlier. Hence, when δ is negative $p^{(k+1)}(s^{(k)}, a^{(k)})$ is increased, as a result, the probability of taking the favourable action increases according to Eq. (2.4) [5].

2.3.4 Knowledge Transfer

Transfer learning is a vital part of present study. In this work, the concept of transfer learning has been evaluated by its application on real time data collected from the APs of a localized Wi-Fi network chosen here. As indicated in the literature review, some transfer learning based studies have been reported but these have not covered Wi-Fi networks specifically. The present study is thus a step to add more knowledge in this domain while considering a practical scenario and use of real time data. As discussed previously, there is a reasonable similarity in traffic variation at a given location at the same instants in the past and the present. Hence, it is quite worthwhile to exploit past data statistics in the learning process to improve the performance and to eliminate the need for learning from scratch. For transferring the knowledge gained from the past, the overall policy is divided into two parts 'native policy' and 'exotic policy'.

$$p_{overall} = (1 - \zeta(k))p_{native} + \zeta(k)p_{exotic} \tag{2.8}$$

where $\zeta(k) = \theta^k$ is transfer rate that determines fraction of exotic policy that contributes to overall policy, $\theta \in (0, 1)$ is the transfer rate factor, p_{native} is the policy which is continuously updated as the learning proceeds and p_{exotic} is the previously learned policy which is transferred to the current task. Hence, p_{native} helps to explore new optimal values pertinent to the current scenario and p_{exotic} helps to exploit the optimal values learned previously. When transfer learning is used, $p(s^k, a)$ in Eq.(2.4) is the overall policy, $p_{overall}$. Too much dependence on exotic policy may have negative impact on learning because the native policy becomes more and more optimal with each stage of learning process. Therefore, the knowledge transfer rate should decrease as the learning proceeds. Further, as discussed in [6] the selection of source task from which knowledge is being transferred should be given due consideration to avoid negative transfer and ensure reasonable improvement in performance. Selection of source task should be such that there is sufficient similarity between the source task and the current task [6]. For the present study, the choice of source task is instinctive. This is due to reasonable similarity between the last week's and last day's schedule with the current day's schedule. However, when there is irregularity in system's traffic pattern more robust approaches discussed in [8] can be used for source task selection.

2.4 Results and Discussions

Considering a practical scenario, the AC algorithm is applied on two sets of APs as shown in Figure 2.1. The algorithm is applied on real time data on 6 different dates, viz., November 18, 2015, November 24, 2015, November 25, 2015, January 13, 2016, January 21, 2016 and January 22, 2016 . The performance is measured in terms of a metric termed as 'Energy Consumption Ratio (ECR)' which is the ratio of energy consumption of the system on a particular instant of learning process to the energy consumption of the system when there is no learning and all APs are ON at that instant. The learning process proceeds in k stages and each stage is termed as an Episode.

In this work, model of an academic institution is taken where a weekly academic schedule is followed. Hence, there is certain correlation between the traffic variations at same days of the week. Similarly, there are factors like weather conditions, a nearby festival or an extra-curricular college activity which could lead to similar traffic patterns on consecutive days. Hence, while transferring the knowledge from the past the exotic policy, p_{exotic} is taken as the combination of the last day's policy and the last week's policy i.e.,

$$p_{exotic} = \frac{1}{2}[p_{lastday} + p_{lastweek}] \tag{2.9}$$

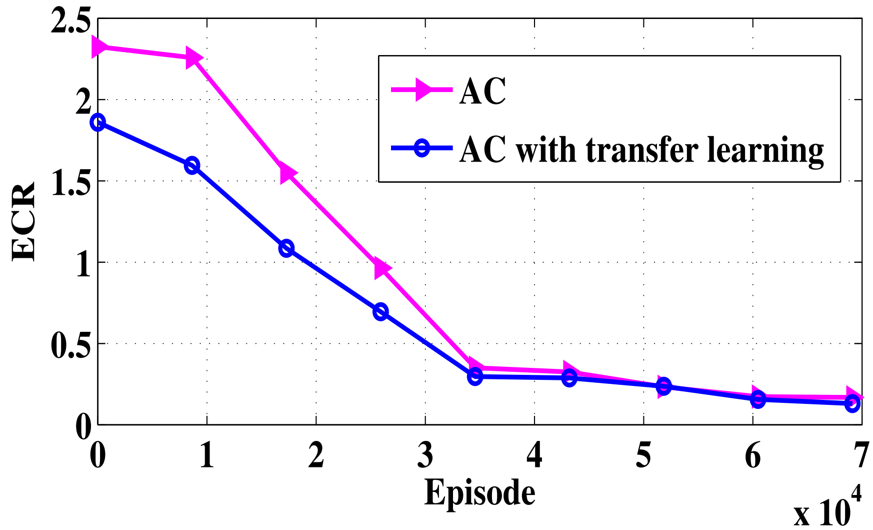
Figures 2.3 and 2.4 represents the simulation results when the algorithm is applied on Set-1 APs for Nov. 2015 and Jan. 2016 data respectively. Simulation results depict that as the learning proceeds, optimal value of policy is attained and the system energy consumption reduces in-turn reducing the ECR. Figures 2.5 and 2.6 depict the simulation results when the algorithm is applied on Set-2 APs. Furthermore, it can be inferred from Figures 2.3-2.6, that there is a significant improvement in performance after application of transfer learning.

The transfer rate factor, θ can be varied between 0 and 1 which consequently varies ζ (transfer rate). The simulation results shown in Figures 2.3-2.6 are with $\theta = 0.2$. Conceptually, lesser value of θ would mean lesser dependence on exotic policy and a greater exploration factor in learning. Figure 2.7. shows variation of mean ECR with θ for four different dates viz. Nov. 24, 2015, Nov 25, 2015, Jan. 13, 2016 and Jan. 22, 2016. It can be seen that minimum value of mean ECR and hence the best performance is obtained for $\theta = 0.2$. A similar trend is observed for the other dates as well. Therefore, for the present system model, the value of θ is selected as 0.2.

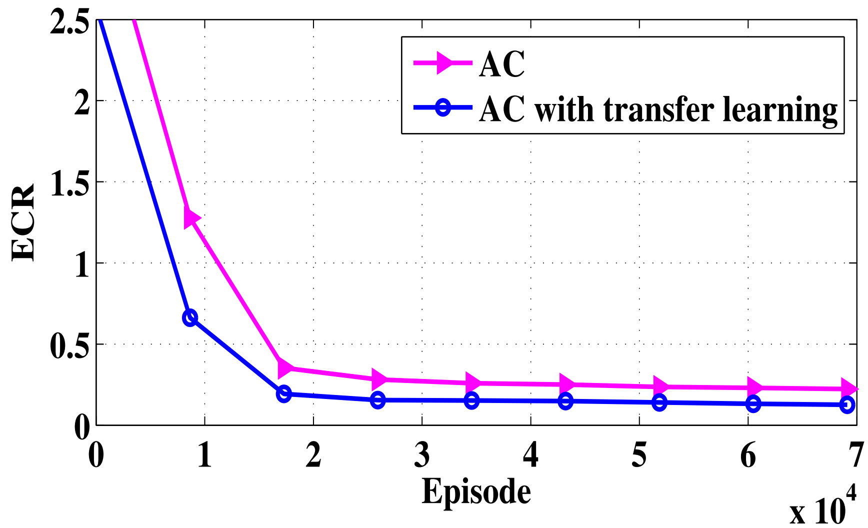
The percentage reduction in energy consumption as compared to the case when AP switching is not used is tabulated in Table 2.1. The maximum possible improvement refers to the improvement when genie-aided policy is used where there is prior knowledge of traffic data. It can be seen from the numerical results in Table 2.1 that by using AC algorithm along with transfer learning a significant amount of energy saving can be achieved. There is consistency in reduction of energy consumption on all chosen days, however the different level of energy reduction on different days is due to specific traffic pattern.

Table 2.1: Percentage energy saving with the proposed scheme.

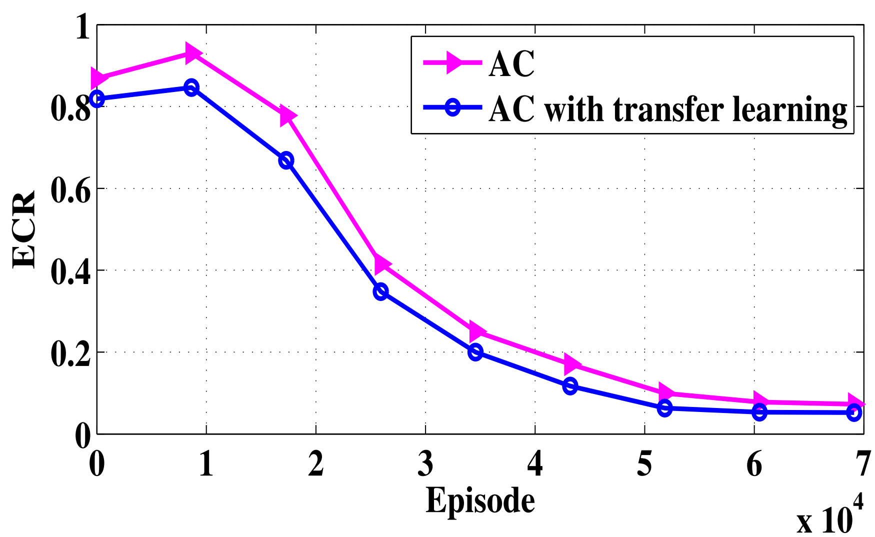
	Date	Mean % improvement with AC algorithm	Mean % improvement with AC algorithm and knowledge transfer	Maximum possible Improvement
Set-1	18/11/2015	47.5	69.5	96.4
	24/11/2015	60.9	78.6	96.9
	25/11/2015	73.6	79.1	95
	13/1/2016	81.36	86.9	96.09
	21/1/2016	81.49	88.09	97.18
	22/1/2016	80.46	85.33	95.2
Set-2	18/11/2015	74.49	83.2	85
	24/11/2015	75.4	81	91
	25/11/2015	81.8	87	90
	13/1/2016	63	67.3	94.12
	21/1/2016	94.7	96.16	98.5
	22/1/2016	94.74	96.4	98.6



(a)

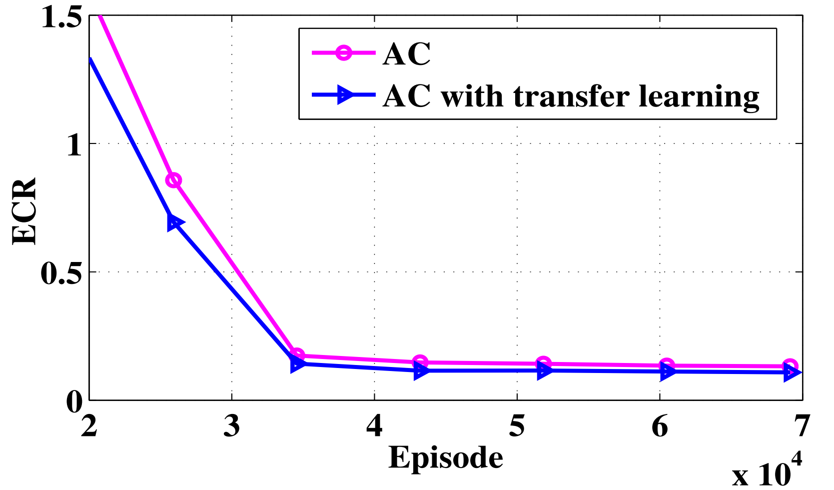


(b)

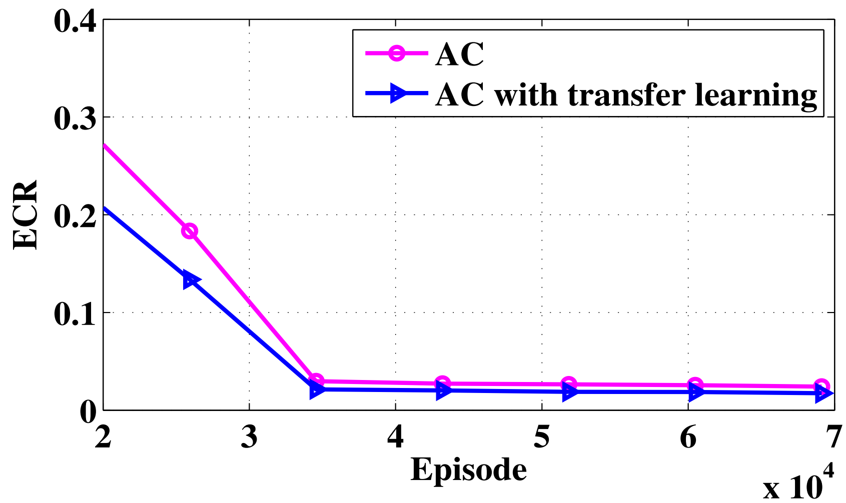


(c)

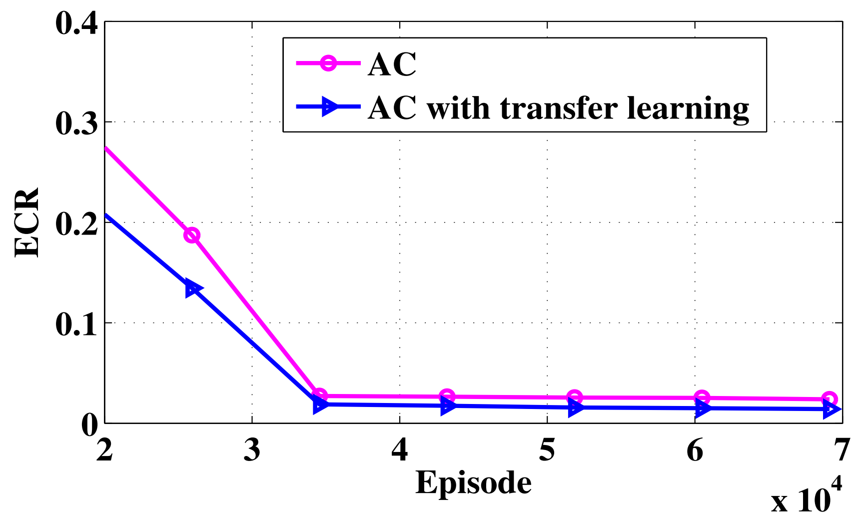
Figure 2.3: ECR curve for Set-1 APs (a) Combined policy of Nov. 11 and Nov. 17, 2015 as applied on Nov. 18, 2015 (b) Combined policy of Nov. 17 and Nov 23, 2015 as applied on Nov. 24, 2015 (c) Combined policy of Nov. 18 and Nov. 24, 2015 on Nov. 25, 2015



(a)



(b)



(c)

Figure 2.4: ECR curve for Set-1 APs, (a) Combined policy of Jan. 6 and Jan. 12 , 2016 as applied on Jan. 13, 2016. (b) Combined policy of Jan. 14 and Jan. 20, 2015 as applied on Jan 21, 2016. (c) Combined policy of Jan. 15 and Jan. 21, 2016 as applied on Jan. 22, 2016.

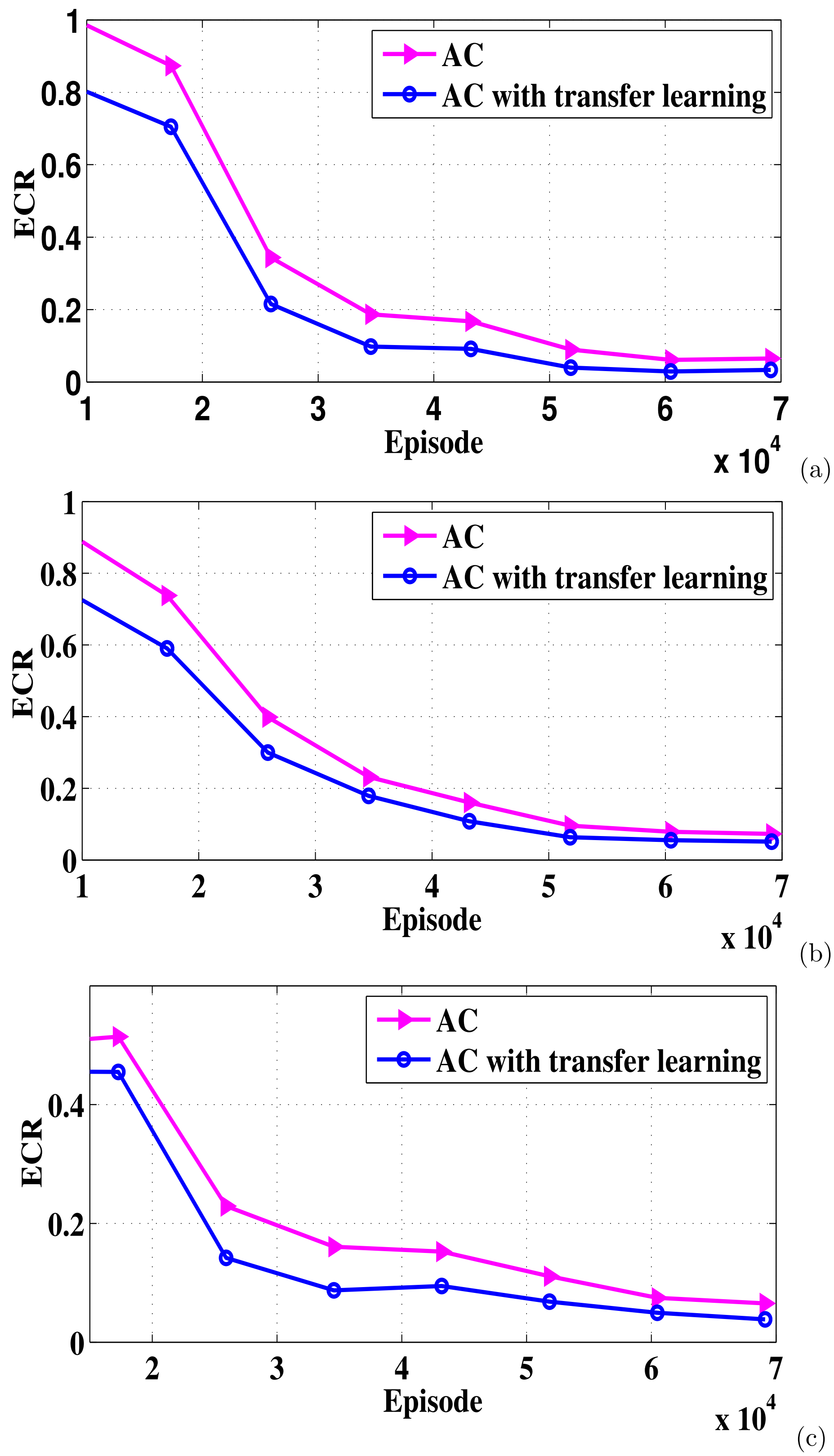
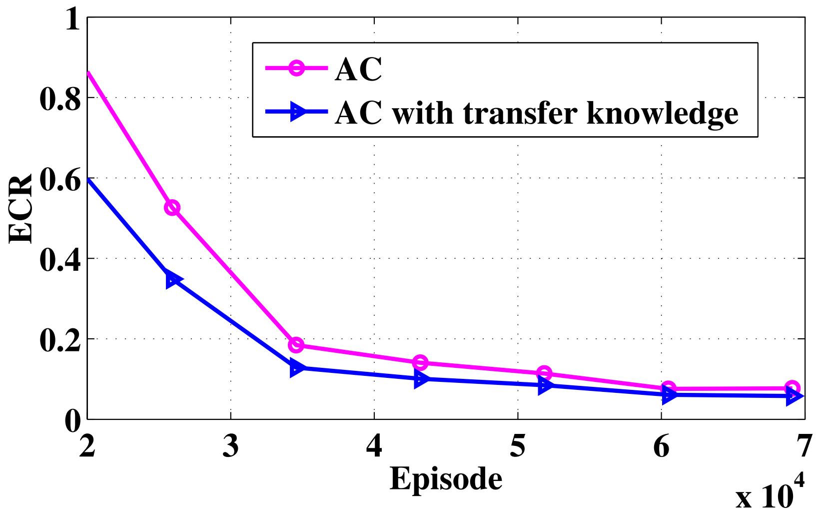
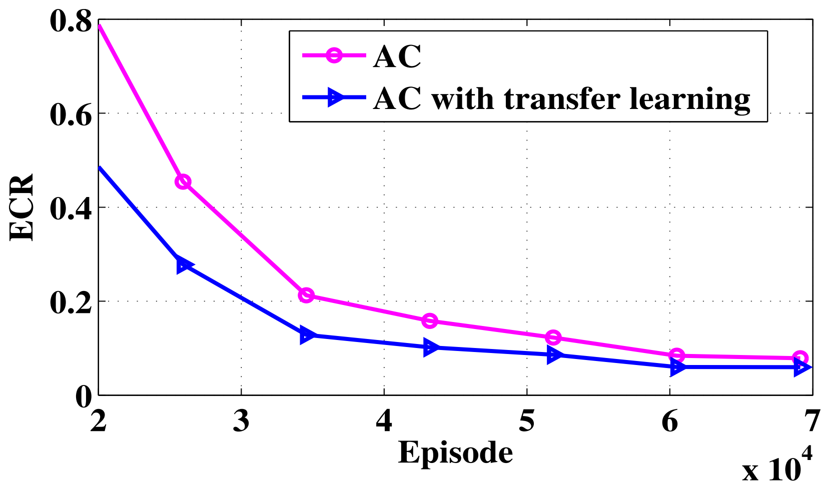


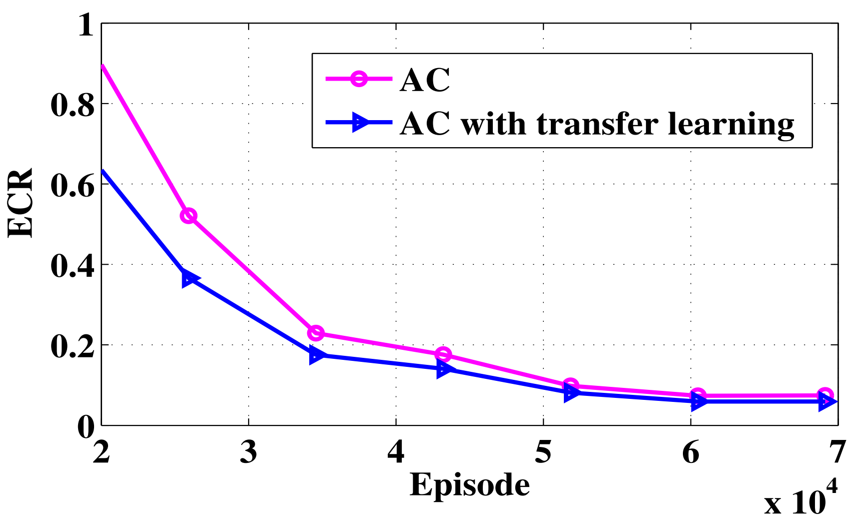
Figure 2.5: ECR curve for Set-2 APs, (a) Combined policy of Nov. 11 and Nov. 17 , 2015 as applied on Nov. 18, 2015. (b) Combined policy of Nov. 17 and Nov. 23, 2015 as applied on Nov 24, 2015. (c) Combined policy of Nov. 18 and Nov. 24, 2015 as applied on Nov. 25, 2015.



(a)



(b)



(c)

Figure 2.6: ECR curve for Set-2 APs, (a) Combined policy of Jan. 6 and Jan. 12 , 2016 as applied on Jan. 13, 2016. (b) Combined policy of Jan. 14 and Jan. 20, 2015 as applied on Jan 21, 2016. (c) Combined policy of Jan. 15 and Jan. 21, 2016 as applied on Jan. 22, 2016.

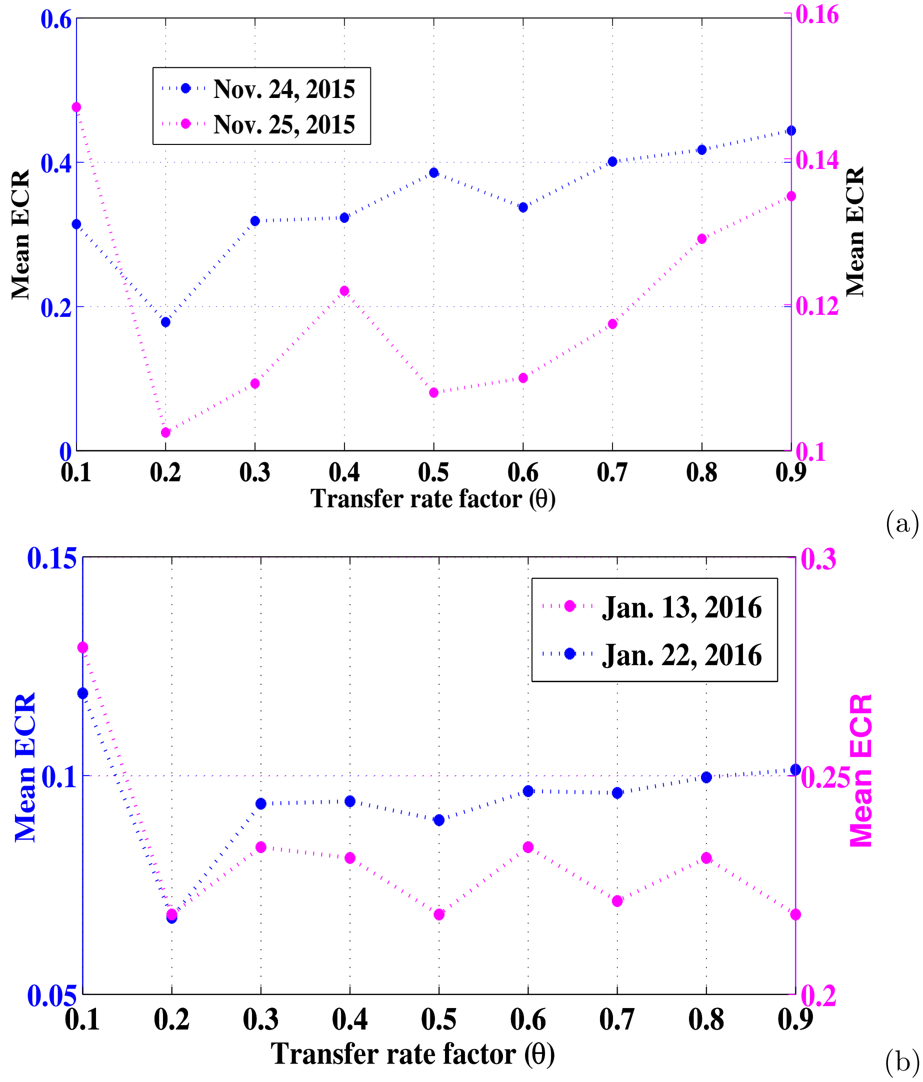


Figure 2.7: (a)Variation of mean ECR with θ for Nov. 24, 2015 and Nov 25, 2015. (b) Variation of mean ECR with θ for Jan. 13, 2016 and Jan 22, 2016.

2.5 Summary

The study on energy saving in Wi-Fi networks using real time data, presented in this chapter is well conforming to the first objective outlined for the research work. This is significant as energy efficiency in Wi-Fi networks is a vital factor, considering increased deployment of such networks in organisations worldwide. Further, it addresses the requirement of add-on applications in present and future communication networks with low consumption of energy. As an innovative way, RL is used to device an optimal AP switching scheme which results in significant reduction in energy consumption of the system. A further improvement in performance is achieved through transfer learning process. On an average, the scheme presented in this chapter can lead to around 75% saving in energy consumption as compared to the case when AP switching is not used. In both the cases, the results are derived considering a model of limited area Wi-Fi communication network in an academic institution and using real time traffic data of the network. Importantly, these results are applicable to similar networks in other organizations as well and the study can be extended for large size Wi-Fi networks. In nutshell, the innovative approaches of RL and transfer learning adopted here for a Wi-Fi network have yielded satisfactory result in terms of energy saving.

Chapter 3

RL Framework for Three State BS Switching

3.1 Introduction

As discussed in section 1.1, the energy consumed by the BSs constitutes a major portion of energy consumption of the overall communication system. Figure 3.1 depicts the energy consumption by different components of a cellular BS. It is evident from Figure 3.1 that the maximum energy consumption occurs at the radio heads, specifically within the power amplifiers. In the present deployment, when a BS is active it operates in tri-sectorized mode wherin for each active sector there is an active power amplifier leading to a high energy consumption. To achieve energy saving in BSs, several BS switching schemes have been proposed in [2–5, 11–17]. In these schemes BSs are switched between active state and sleep state depending on the traffic load. However, as a different approach when the traffic load is moderate the BSs can be configured to operate in omnidirectional mode. This could lead to a reduction in system energy consumption owing to lower energy consumption of BSs operating in the omnidirectional mode as compared to those in tri-sectorized mode. This chapter discusses a novel scheme in which the previous work on RL based BS switching [5] has been extended for three state BS switching. In the proposed scheme, apart from an active state at high traffic and sleep state at low traffic, the BSs are switched to an omnidirectional state at moderate traffic leading to a more efficient energy saving scheme. It is quite intuitive that with such a scheme, the reduction in energy consumption would be maximum in case of moderate traffic load. This is because at moderate traffic a larger fraction of BSs would go into low power omnidirectional mode. In this chapter, in addition to simulation, a suitable mathematical formulation is developed to represent the new approach and reduced energy consumption.

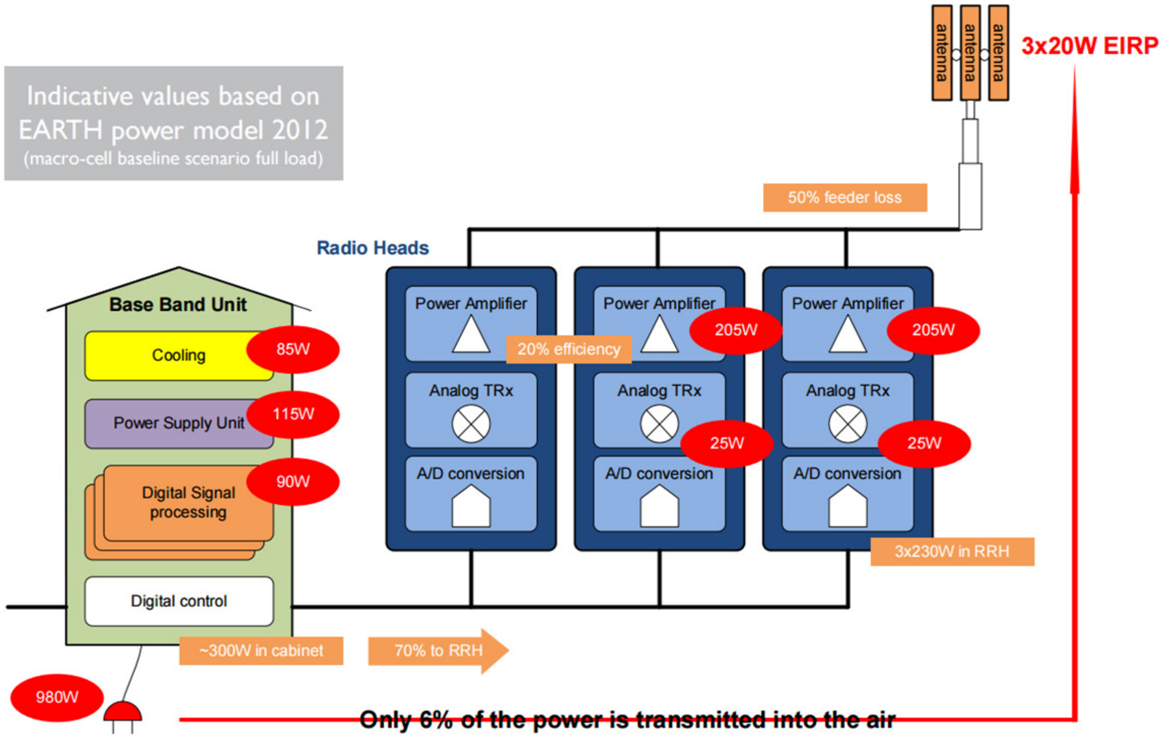


Figure 3.1: Power consumption of different BS components [25].

3.2 System Model

To analyse three state BS switching, the algorithm is applied on a system containing uniformly distributed users in a region served by a set of overlapping BSs. The traffic load at a given location is considered to follow a Poisson distribution with arrival rate λ [19–21], then an MDP is formulated using the traffic load variations. When an action $\mathbf{a}^{(k)} = \{a_1^{(k)}, a_2^{(k)}, ...\}$ is taken if $a_i^k = 0$ the i^{th} BS is switched OFF else it is switched ON if $a_i^k = 1$. When a BS is ON , if the corresponding traffic load is greater than a threshold, it is made to operate in tri-sectorized mode otherwise it is switched to omnidirectional mode.

The BSs are switched in a similar way as the Wi-Fi APs following the AC algorithm steps described in chapter 2. In this case, there is an additional step after action selection i.e user association and rest of the steps remain the same. After action selection, when certain BSs are switched OFF, users associate themselves with the BSs belonging to the set of ON BSs according to the following metric:

$$i^*(x) = arg\,max_j \frac{c(x,j)}{P_j} \tag{3.1}$$

where, $c(x, j)$ is the upper-bound on the capacity of the link between user located at x and j^{th} BS calculated according to Shannon's theorem and P_j is the power consumed by the j^{th} BS. According to Eq. (3.1), a user located at position x chooses to be served by a

BS j if the link between them provide maximum capacity and the BS consumes minimum power.

3.3 Mathematical Analysis

The energy saving scheme presented here indeed works better than conventional two state schemes. This argument can be supported through underlying mathematical formulation developed in this study. Let S_i be a random variable which represent the state of i^{th} BS in the set of available BSs. The random variable S_i is a function of BS traffic load which in-turn is a Poisson random variable. S_i is either 0 or 1 depending on whether the BS is ON or OFF and can be described as:

$$S_i = \begin{cases} 0, & \text{when } BS_{traffic} < threshold \\ 1, & \text{when } BS_{traffic} \geq threshold \end{cases}$$

Therefore, in a two state system, expected value of power consumption of a BS can be given as,

$$E = f(S_i = 0)P(S_i = 0) + f(S_i = 1)P(S_i = 1) \tag{3.2}$$

where $f(S_i = s_i)$ represents the probability that i^{th} BS is at state s_i and $P(S_i = s_i)$ represents power consumed by the i^{th} BS at state s_i. Now, $f(S_i = 0) = f(BS_{traffic} < threshold)$ and $f(S_i = 1) = f(BS_{traffic} \geq threshold)$ therefore,

$$E = f(BS_{traffic} < threshold)P(S_i = 0) + f(BS_{traffic} \geq threshold)P(S_i = 1) \tag{3.3}$$

Substituting $P(S_i = 0) = 0$ and $P(S_i = 1) = P_{sectorized}$,

$$E = f(BS_{traffic} < threshold) \times 0 + f(BS_{traffic} \geq threshold) \times P_{sectorized} \tag{3.4}$$

Traffic load at a given BS is the sum of traffic due to all users associated with it. Traffic from each user follow Poisson distribution i.e. $F(X = x) = e^{-\lambda}\lambda^x/x!$ $(Poisson(\lambda))$. From the property of Poisson random variable, sum of independent $Poisson(\lambda i)$ distribution is a $Poisson(\sum \lambda_i)$ distribution. Let $\sum \lambda_i = \lambda$ and $BS_{traffic} = x$

$$f(BS_{traffic} \geq threshold) = \sum_{x=threshold}^{\infty} e^{-\lambda}\lambda^x/x! \tag{3.5}$$

Therefore,

$$E = P_{sectorized} \times [\frac{e^{-\lambda}\lambda^{threshold}}{threshold!} + \frac{e^{-\lambda}\lambda^{threshold+1}}{(threshold+1)!} \cdots] \tag{3.6}$$

In the three state model presented in this work, at a particular instance, a given BS can be either in active mode, sleep mode or omnidirectional mode. Hence, in this case the random variable S_i can be defined as,

$$S_i = \begin{cases} 0, & \text{when } BS_{traffic} < threshold \\ omni, & \text{when } threshold \leq BS_{traffic} \leq n \times threshold \\ 1, & \text{when } BS_{traffic} \geq n \times threshold \end{cases}$$

where, n is a scalar such that $n > 1$. In this case, the expected value of power consumption of a BS would be,

$$E_{omni} = f(S_i = 0)P(S_i = 0) + f(S_i = 1)P(S_i = 1) + f(S_i = omni)P(S_i = omni) \quad (3.7)$$

$$\begin{aligned} E_{omni} = f(BS_{traffic} < threshold)P(S_i = 0) + f(threshold \leq BS_{traffic} \leq n \times threshold)P(S_i = omni) \\ + f(BS_{traffic} > n \times threshold)P(S_i = 1) \end{aligned} \quad (3.8)$$

$$\begin{aligned} E_{omni} = P_{omni} \times \left[\frac{e^{-\lambda}\lambda^{threshold}}{threshold!} + \frac{e^{-\lambda}\lambda^{threshold+1}}{(threshold+1)!} \cdots + \frac{e^{-\lambda}\lambda^{n \times threshold}}{n \times threshold!} \right] + \\ P_{sectorized} \times \left[\frac{e^{-\lambda}\lambda^{(n+1) \times threshold}}{(n+1) \times threshold!} + \cdots \right] \end{aligned} \quad (3.9)$$

In Eq.(3.6) entire range of values are multiplied by $P_{sectorized}$ while in Eq.(3.9) values for which $BS_{Traffic} > n \times threshold$ are multiplied by $P_{sectorized}$ and rest by P_{omni}. As $P_{omni} < P_{sectorized}$, therefore $E_{omni} < E$. Hence, the expected value of power consumption for omnidirectional mode is always less than the sectorized mode.

3.4 Results and Discussions

To evaluate the performance improvement through the proposed scheme, the AC algorithm is applied on the system under two scenarios: a) the BSs are switched between active mode and sleep mode according to traffic load b) BSs are switched between active, omnidirectional and sleep modes based on traffic load. As the traffic load is considered to follow a $Poisson(\lambda)$ distribution, higher the value of λ, higher would be the traffic load. Simulations are performed for varied traffic load. Figure 3.2 depicts the variation of mean differential improvement with λ i.e. traffic load. Mean differential improvement is the average of difference between energy consumption in a two state model and the proposed three state model. The improvement is highest when the load is moderate as the fraction of BSs going into low power omnidirectional mode would be highest in this case. The proposed scheme gives better performance for wide range of traffic models. At moderate traffic, the performance gain is high and performance is equal to the existing scheme at very low and very high traffic. Further improvement in performance can be achieved by exploiting the past data statistics and using the concept of transfer learning as discussed in section 2.3.4. This improvement in performance is depicted in Figure 3.3 for moderate load. In this case also, the performance is measured in terms of ECR vs Episode curve. It is observed that the proposed three state BS switching using AC learning framework leads to 15% additional reduction than the two state switching scheme. Furthermore, the application of transfer learning to this scheme leads to 40% reduction which is a quite significant amount.

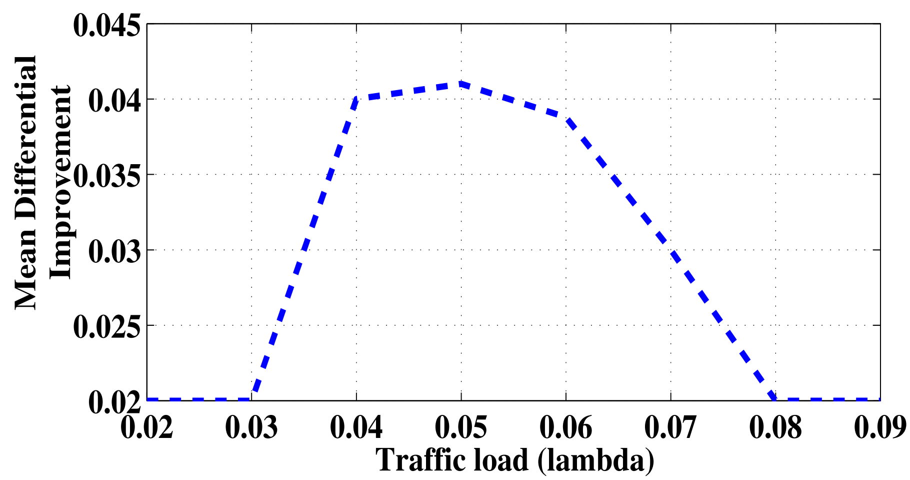


Figure 3.2: Variation of Mean differential improvement with respect to traffic load.

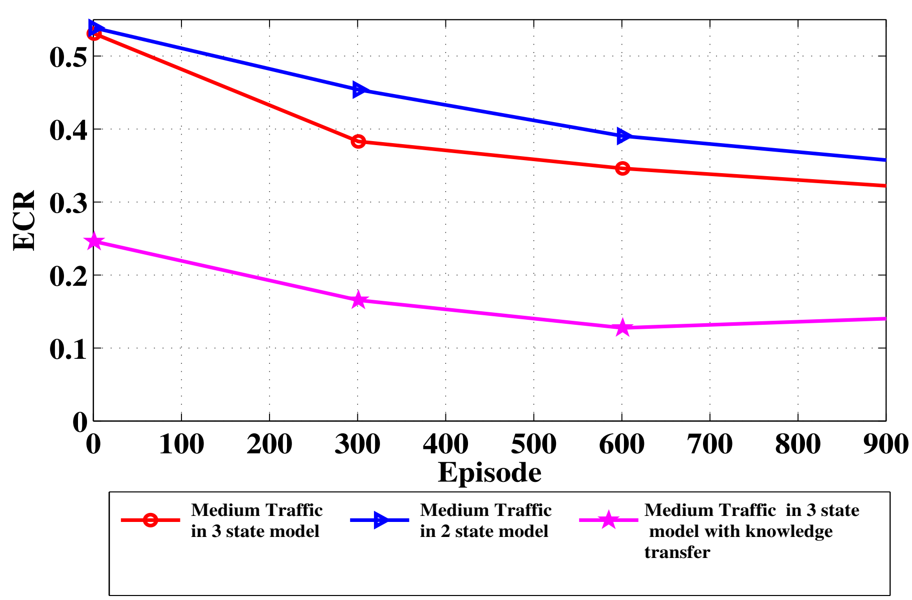


Figure 3.3: Reduction in system energy consumption at moderate traffic depicted through ECR curve.

3.5 Summary

The analysis presented in this chapter addresses the requirement of a suitable energy reduction approach in cellular networks. The three state base station switching is well explained through related mathematical formulation and precise simulation results. A significant drop in energy consumption level is observed in case of moderate traffic load owing to switching of BSs to omnidirectional state. It is further dropped with the application of transfer learning approach.

Chapter 4

RL framework for energy saving in HetNets consisting of macro and femto BSs

4.1 Introduction

In the previous chapters, RL based energy saving schemes are discussed for single tier systems. This chapter presents the use of RL framework to develop an energy saving scheme for a two-tier HetNet consisting of macro and femto BSs. Small cell deployment has emerged as a promising solution to serve ever increasing data traffic in next generation communication systems. Growing attraction towards small cells is due to their ability to extend the coverage and boost the network capacity by reducing the propagation distance between nodes and offloading the macro cell traffic. However, the inclusion of additional BSs to serve greater traffic demands and data intensive services has become a major challenge towards the energy efficient deployment of cellular networks. Considering this trend, there have been many recent research efforts in this direction [11–16]. This chapter discusses the use of AC learning algorithm and application of transfer learning concept in the context of energy efficient BS switching in HetNets comprising of macro cells and femto cells. Further, the trade-off between system delay and energy saving is analysed. Relevantly, various sleep mode techniques for small cells are discussed in [11]. Sleep mode techniques for small cells can be broadly classified into three categories, viz., small cell controlled sleep mode, core network controlled sleep mode. In this work, core network controlled sleep mode is considered to avoid the requirement of active user detection capability within the small cells and reduce implementation complexity as compared to user equipment controlled sleep mode technique.

4.2 System Model and Application of AC Algorithm Based BS Switching Scheme on the System

To analyse the performance of AC learning based energy saving scheme, the algorithm is applied on a hypothetical system which is discussed in this section. As a simple case, a small region is considered consisting of two macro BSs separated by a distance of around 1 km and 20 femto BSs at fixed location within this region. Figure 4.1 shows the schematic representation of the system.

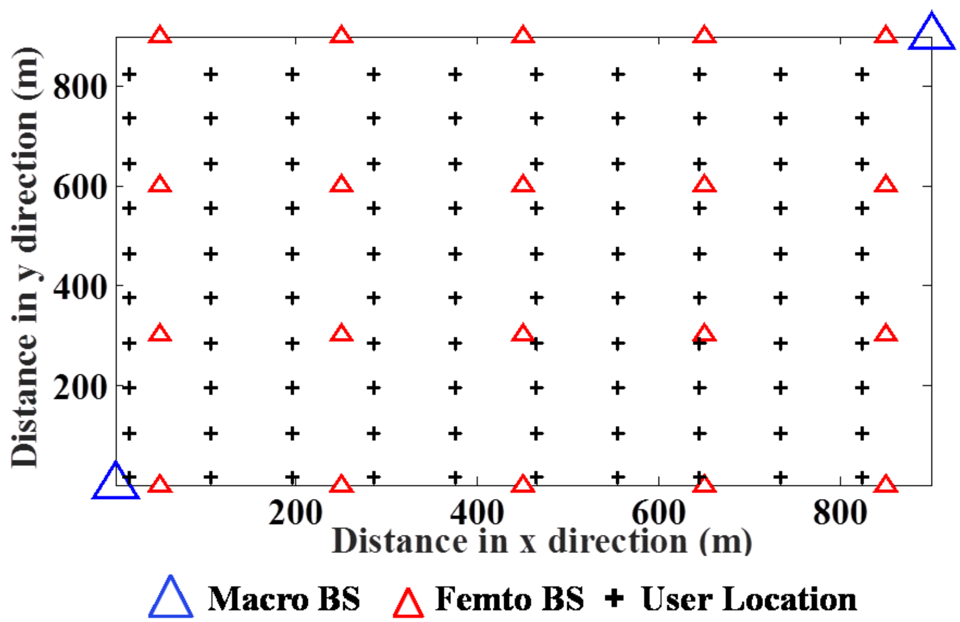


Figure 4.1: Schematic representation of the system.

Traffic from each user follows Poisson distribution as discussed in chapter 3. The present framework is modelled considering core network controlled sleep mode [11]. In LTE, the control functionality is embedded within the macro BSs [22]. Hence, macro BSs are responsible for providing coverage and control functionality in the HetNets. Considering this, in the present scheme the macro BSs are held in ON state and the femto BSs are switched based on AC learning algorithm as described in section 2.3.

In this case also, action selection step is followed by user association. Based on the description of core network controlled sleep mode given in [11], the user association occurs in two steps:

- Firstly, a connection between the macro BS and user is established based on the macro cell coverage.
- Then, the user is associated with the femto BS lying in that macro cell region depending on the metric given by Eq.(3.1).

The user association step is followed by state value function update and policy update as described in section 2.3. Finally, transfer learning is used for further improvement in performance. The steps involved in the application of the algorithm to the system are summarized in Figure 4.2.

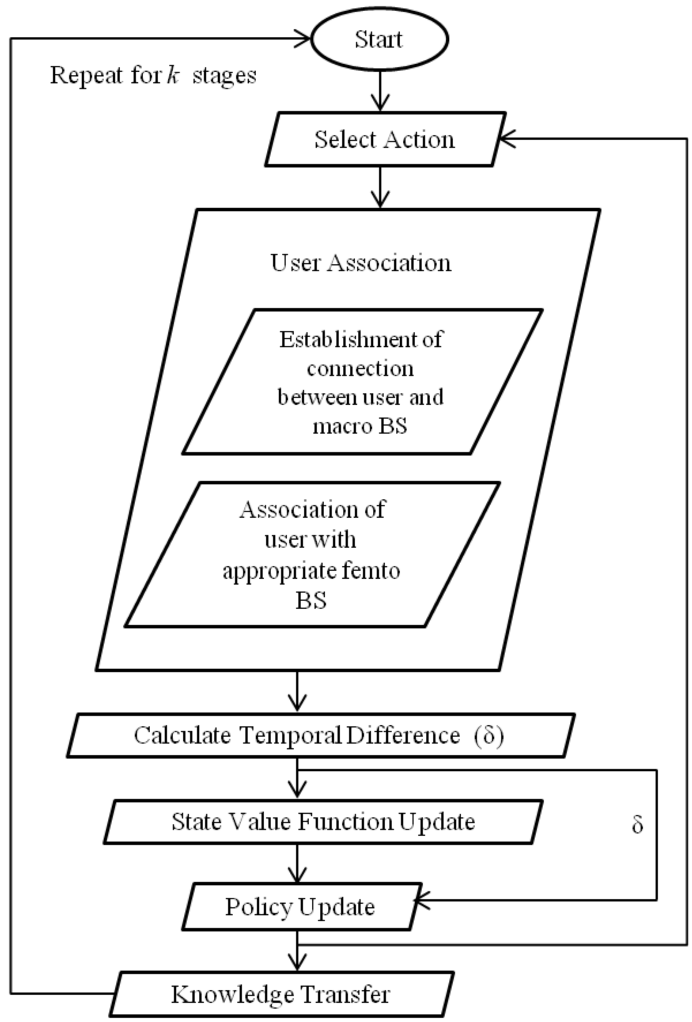


Figure 4.2: Flowchart of the algorithm.

4.3 Results and Discussions

The energy saving obtained by the proposed scheme is validated through simulations and the results are discussed in this section. For link capacity calculation for user association step, COST-231 modified Hata model [23] is used for macro BS link and ITU-R specified path loss model [24] is used for femto BS link. Figure 4.3 shows the ECR vs Episode curve when the proposed scheme is applied on the system. It can be seen that as the learning proceeds, the energy consumption of the system is reduced. Numerically, mean improvement achieved on the application of AC learning algorithm is 78% and when transfer learning is applied, around 82% of improvement is achieved as compared to the case when BS switching is not used.

To analyse the trade-off between system delay and energy saving, the system cost described in Eq.(2.1) is modified as,

$$C = \sum_{i \in A'} [(1 - q_i)\rho_i P_i + q_i P_i] + \varsigma C_d \tag{4.1}$$

where, C_d is the delay equivalent cost given by $C_d = \sum_{i \in A'} \frac{\rho_i}{(1-\rho_i)}$. As discussed in [20], minimizing C_d is equivalent to minimizing the average delay. A' is the set of active BSs and ς is a scalar that determines the weightage which is to be given to the delay equivalent cost reduction [5].

The simulations are performed by varying the value of ς between 500 and 3500. For each value of ς, the mean ECR and delay are computed and are plotted against each other. Figure 4.4 indicates that a higher value of ς corresponds to a lower delay and greater energy consumption. This is due to the fact that higher value of ς amounts to a greater importance to delay equivalent cost reduction. If there is a lower tolerance to delay, lesser number of BSs would be turned OFF and hence the energy consumption would be greater which is apparent from Figure 4.4. Therefore, to ensure the required quality of service it is necessary to take care of the trade-off between the system energy consumption and delay. Further, it is observed that when the value of ς is increased from 500 to 1500 a small increase in energy consumption is observed. After $\varsigma = 1500$ a steep increase in energy consumption is observed, suggesting $\varsigma = 1500$ to be optimal value for the present model.

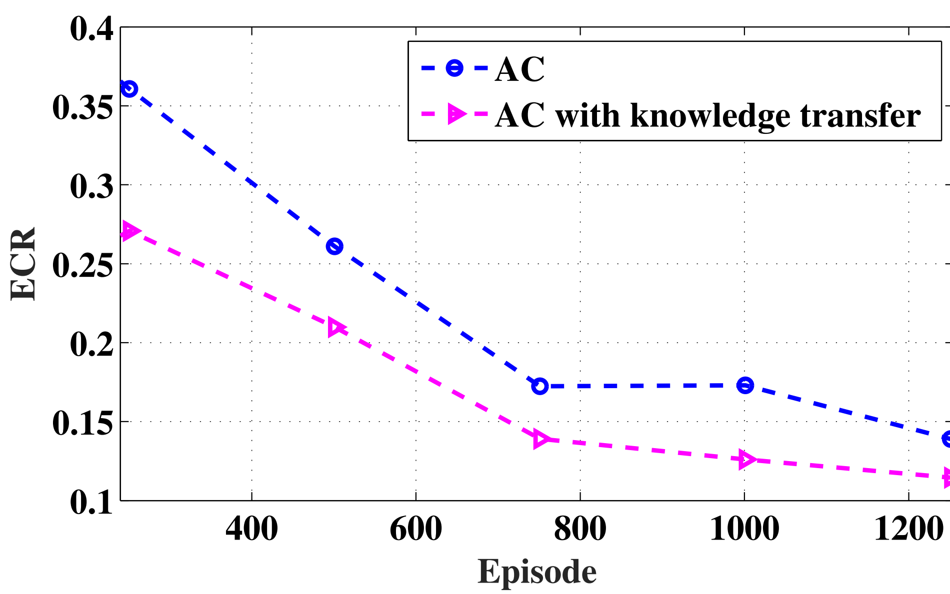


Figure 4.3: Reduction in system energy consumption depicted through ECR curve

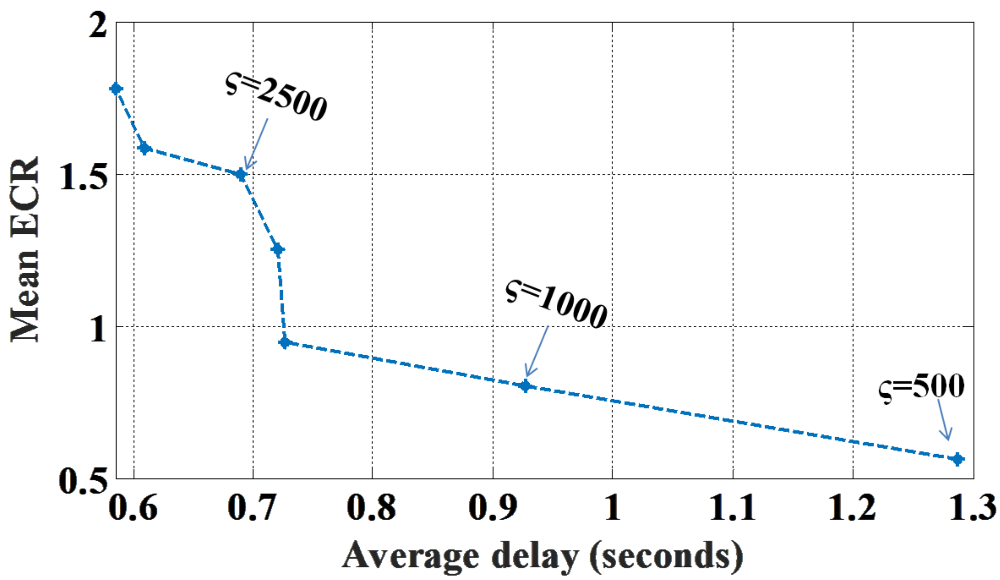


Figure 4.4: Variation of mean ECR with delay importance parameter (ς)

4.4 Summary

In the preceding chapters, the objective of reducing energy consumption in wireless networks has been well attempted for localized Wi-Fi network and cellular networks. This research is broadened by considering the case of HetNets consisting of macro and femto BSs. The AC algorithm is relevantly modified to analyse performance of chosen scheme of BSs deployment. Same way, the effect of transfer learning is analyzed for the HetNet configuration. One more important aspect related to the trade-off between system delay and energy saving is analysed. The tolerable limits of system delay with respect to desired level of energy reduction is a point of intelligent trade-off and the discussed results are undoubtedly a significant outcome of the research efforts here.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

ICT is one of the important areas which is dominating technological growth in the present era. Rapid technological growth presents new research challenges to keep it sustainable and in broad perspective, to conform with societal, economical and ecological aspects. This is true with communication systems technology as well. Making communication systems energy efficient is an interesting research motivation from technical, economical as well as ecological perspective. The research work presented in this thesis well addresses this issue and contributes in terms of presenting some novel approaches and their validation through a significant reduction in energy consumption in key segments of wireless communication networks i.e. cellular and Wi-Fi networks. Various research initiatives for development of energy efficient wireless communication networks are reported in the literature. However, there is ample of scope for improvement in order to address the requirement of next generation communication systems which are heavily loaded with surging data applications and traffic.

As such, the research work presented in this thesis relevantly attempts to address the requirement of energy efficiency in next generation communication networks. Following are the main contributions of the present research work:

Firstly, the innovative approaches of RL and transfer learning are applied to a Wi-Fi network, this has yielded a satisfactory result in terms of energy saving. Importantly, results are validated with real time data of a selected Wi-Fi network of an academic institution.

Secondly, for the case of cellular networks, in addition to usual two state base station switching scheme of active and sleep modes, a three state scheme comprising of an additional omnidirectional mode is analyzed with related modification in AC algorithm, relevant mathematical formulation and simulations. A significant drop in energy consumption level is observed in the case of moderate traffic load owing to switching of BSs to omnidirectional state. It is further dropped with the application of transfer learning approach.

Thirdly, the energy efficiency aspect is analyzed for the case of HetNets consisting of macro and femto BSs. The AC algorithm is relevantly modified and transfer learning approach is applied to the system. A significant drop in energy consumption is observed in this case as well. Importantly, the aspect related to the trade-off between system delay and energy saving is also analyzed here.

5.2 Future Work

– The concept of RL and Transfer learning for energy efficient access network deployment is applied to a limited case here, this can be extended to different deployment scenarios. For example, this scheme could be applied to larger size Wi-Fi networks. Furthermore, the study can be extended by the inclusion of data offloading from cellular BSs to WiFi APs for high data rate applications.

– Preliminary attempt has been made here to analyse the trade-off between system delay and energy saving for HetNets. In addition to system delay parameter, the trade-off analysis could be extended to other compromising parameters when evaluating limits of energy efficiency with respect to over all efficiency of the network. For instance trade-off between energy efficiency and spectral efficiency could be analysed. Further, interference among different cells could be taken into account.

Bibliography

[1] J. Lorincz, A. Capone, and M. Bogarelli, "Energy savings in wireless access networks through optimized network management," *5th IEEE International Symposium on Wireless Pervasive Computing (ISWPC)*, pp. 449-454, Modena, Italy, May 2010.

[2] 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 Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1525-1536, Sept. 2011.

[3] E. Oh and B. Krishnamachari, "Energy Savings through Dynamic Base Station Switching in Cellular Wireless Access Networks," *IEEE Global Telecommunications Conference (GLOBECOM 2010)*, pp. 1-5, Miami, FL, Dec. 2010.

[4] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," *ACM MobiCom*, vol. 9, no. 262, pp. 10-12., Sept. 2009.

[5] R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, "Tact: a transfer actor-critic learning framework for energy saving in cellular radio access networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 2000-2011, April 2014.

[6] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *The Journal of Machine Learning Research*, vol. 10, pp. 16331685, 2009.

[7] D. W. Aha, M. Molineaux, and G. Sukthankar, "Case-based reasoning in transfer learning," In *Case-Based Reasoning Research and Development*, pp. 29-44. Springer Berlin Heidelberg, 2009.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345 1359, Oct. 2010.

[9] L. A. Celiberto, J. P. Matsuura, D. M'antaras, R. Lopez, and R. A. Bianchi,"Using transfer learning to speed-up reinforcement learning: a cased-based approach," *Robotics Symposium and Intelligent Robotic Meeting (LARS),* pp. 5560, Oct. 2010.

[10] J. Wannstrom and K. Mallinson. (2014). *Heterogeneous Networks in LTE* [Online]. Available: http://www.3gpp.org/technologies/keywords-acronyms/1576-hetnet

[11] I. Ashraf, F. Boccardi and L. Ho, "SLEEP mode techniques for small cell deployments," In *IEEE Communications Magazine,* vol. 49, no. 8, pp. 72-79, August 2011.

[12] L. Saker, S. E. Elayoubi, R. Combes and T. Chahed, "Optimal Control of Wake Up Mechanisms of Femtocells in Heterogeneous Networks," In *IEEE Journal on Selected Areas in Communications,* vol. 30, no. 3, pp. 664-672, April 2012.

[13] Y. Li, H. Celebi, M. Daneshmand, C. Wang and W. Zhao, "Energy-efficient femtocell networks: challenges and opportunities," In *IEEE Wireless Communications,* vol. 20, no. 6, pp. 99-105, December 2013.

[14] J. Kim, W. S. Jeon and D. G. Jeong, "Base-Station Sleep Management in Open-Access Femtocell Networks," In *IEEE Transactions on Vehicular Technology,* vol. 65, no. 5, pp. 3786-3791, May 2016.

[15] Yuan Gao et al., "A novel energy aware dynamic on-off control of base stations in wireless networks," *IEEE 16th International Conference on Communication Technology (ICCT),* Hangzhou, pp. 132-137, Oct. 2015.

[16] D. Sinha, V. Kavitha and A. Karandikar, "Load dependent optimal ON-OFF policies in cellular heterogeneous networks," *12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt),* Hammamet, pp. 159-166, May 2014.

[17] L. Haratcherev, M. Fiorito, and C. Balageas, "Low-power sleep mode and out-of-band wake-up for indoor access points," *IEEE GLOBCOM Workshops,* pp. 16, Nov. 2009.

[18] R. S. and A. G. Barto, Reinforcement learning: An introduction, vol. 1, no. 1.

[19] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," *IEEE INFOCOM,* vol. 1, pp. 786796, March 2003.

[20] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "alpha-optimal user association and cell load balancing in wireless networks," *IEEE INFOCOM,* pp. 15, March 2010.

[21] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," Wireless Networks, vol. 14, no. 1, pp. 103120, Feb. 2008.

[22] M. Nohrborg (2014). *LTE Overview* [Online]. Available: http://www.3gpp.org/technologies/keywords-acronyms/98-lte

[23] IEEE 802. 16 Boradband Wireless Access Working Group (July 2008). *IEEE 802. 16m evaluation methodology document (EMD)* [Online]. Available: http://ieee802.org/16

[24] Saunders, Simon R., et al., Femtocells: opportunities and challenges for business and technology, John Wiley Sons, 2009.

[25] B. Debaillie, C. Desset (2014). *Power Modeling of Base Stations* [Online]. Available: https://wireless.kth.se/5green/wp-content/uploads/sites/19/2014/08/BDebaille.pdf