<u>Auto-Tuning In LTE Networks Using Joint RRM Optimization</u>

By

Lt Cdr Syed Hassan Raza (PN)

Submitted to the Faculty of Electrical Engineering Department,
Military College of Signals, National University of Sciences and Technology,
Islamabad in partial fulfillment for the requirements of an M.S. Degree in
Electrical (Telecom) Engineering.

JULY 2014

ABSTRACT

Auto-tuning or Self-optimization in mobile networks is under continuous evolution in current era. Self optimization means the process where UE (user equipment) and eNB (enhanced Node B) performance measurements are used to auto-tune the network Radio Resource Management (RRM) parameters (in order to achieve optimal network performance. With the growth in the size and complexity of the mobile wireless networks self-optimization is important, in order to reduce the operational cost of the network.

Research on auto-tuning has already been reported for GSM networks. Many contributions on RRM auto-tuning in 3G networks have been reported. Although there were important industrial and academic efforts, auto-tuning was unable to get its place in the UMTS standard. The ambitious targets for the LTE system have given Self Organizing Networks (SON) functionalities a place in the standard including self-configuration, self-optimization and self healing. Different RRM parameters have been used for auto tuning such as Inter-Cell Interference Coordination (ICIC) and load balancing using handover parameter optimization.

In this research, we will specially investigate the joint optimization of more than one RRM parameters i.e. Mobility and Inter-Cell Interference Co-ordination (ICIC) to achieve the network optimization objective. It was expected that joint optimization of two RRM parameters will yield better network performance in terms of KPIs (Key Performance Indicators) as compared to the case when single RRM parameters is optimized.

The performance of proposed auto tuning algorithm has been evaluated. Simulation results have shown better network performance in terms of KPIs with joint optimization.

ACKNOWLEDGEMENTS

In the beginning I would like to thank Almighty Allah whose guidance showed me the light and gave me courage so that I can play my part in this world and contribute, praise be to Hazrat Muhammad (Peace be upon Him) who is a beacon of knowledge for all humanity. Special thanks to Hazrat Ali a.s who is an ideal personality for me and whose following being the door of knowledge has lead me to this day where I have accomplished this feat.

Words are lacking to express my humblest obligations for my supervisor Dr. Moazzam Islam Tiwana, Assistant Professor, COMSATS, Rawalpindi, who was kind enough to spare his valuable time in helping me to complete my thesis. It was because of his patience in attitude and interest in this field that I have been able to fulfill this task.

I would also like to thank Lt Col. Dr. Adnan Rashdi, Lt. Col. Dr. Adil Masood, and Lt. Col. Dr. Imran Tauqir who were the members of my Guidance and Examination Committee (GEC) and were kind enough to share their knowledge for my guidance.

Last but not the least I owe this to my parents, brothers and sister, my wife and beautiful daughter whose love, prayers and support was the reason that I accomplished this task.

Lastly, I thank my course mates, friends, job mates, lab attendants and library staff who were patient enough to bear me out and remained helpful throughout my period of research.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

Acronym	Meaning
HOM	Hand Over Margin
ICIC	Inter Cell Interference Coordination
RL	Reinforcement Learning
FLC	Fuzzy LogicController
FQL	Fuzzy-Q Learning
eNB	Evolved Node B
UE	User Equipment
SFR	Soft Frequency Reuse
HO	Hand Over
RSRP	Fuzzy-Q Learning
SON	Self Optimization Network
UMTS	Universal Mobile Telecommunication System
OFDMA	Orthogonal Frequency Division Multiple Access
SCFDMA	Single Carrier Frequency Division Multiple Access
GSM	Global System for Mobile Communication
GPRS	General Packet Radio Service
EDGE	Enhanced Data for Global Evolution
WCDMA	Wideband Code Division Multiple Access
HSDPA	High Speed Downlink Packet Access
LTE	Long Term Evolution
QoS	Quality of Service
BCR	Block Call Rate
ABR	Average Bit Rate

MFTT .... Mean File Transfer Time

MSINR.. Mean Signal to Interference Noise Ratio

RSRP.. Reference Signal Receive Power

PBGT .... Power Budget

PRB .... Physical Resource Block

Chapter 1

INTRODUCTION

1.1 Chapter Overview

This chapter explains, in brief, the long term evolution (LTE) and self organizing networks (SONs), their background and their importance. Then it presents the objectives of this work, the problem statement and author's contribution in this field. In the later part, organization of this thesis is discussed.

1.2 Background

In the recent past, with the wireless mobile communication boom, auto-tuning and self-optimization of network parameters are more than ever key issues to provide high-quality services for the end-user and to decrease the operational expenditure of the network operation. The special attention drawn to the self-optimization of radio resource management (RRM) parameters is motivated by the user need for ubiquitous communication and by the increasing complexity of networks resulting from the cooperation of radio access technologies. Formerly, RRM has been based on some algorithms (admission control, resource allocation, handover …) governed by a set of fixed thresholds. Today, RRM procedures have undergone a considerable change and the paradigm will shift towards a completely automatic network management. Optimal auto-tuning mechanisms, can considerably improve network management functions with respect to traditional RRM algorithms with fixed parameters.

Though there are research papers available on the auto-tuning of the 2G and 3G networks, the auto-tuning of LTE networks is still an evolving field. Furthermore, there is less work available on the joint optimization of more than one RRM parameters. This research will involve joint optimization using reinforcement learning algorithm and its validation will be done using network level simulation.

1.2.1 Long Term Evolution (LTE)

A project was started to define the long-term evolution (LTE) of Universal Mobile Telecommunications System (UMTS) cellular technology in Nov. 2004. It is a major step in cellular communication and is designed in such a way to meet requirements for high-speed data rate and is equipped to meet the challenges of next-generation mobile networks. It offers bandwidth from less than 5MHz to 20MHz. It employs Orthogonal Frequency Division Multiple Access (OFDMA) and Single Carrier FDMA (SC-FDMA) for downlink and uplink transmission respectively.

LTE is a step forward from 2G and 3G networks to '4G' mobile systems. Technically it is based upon 3GPP system that encompasses GSM, GPRS, EDGE, WCDMA and now HSPA (High Speed Packet Access). LTE mainly focuses upon higher and less erroneous data rates by making more efficient use of finite bandwidth and thus making mobile service faster and highly reliable.

1.2.2 Self Organizing Networks (SONs)

Wireless communication is expanding nowadays by leaps and bounds especially because of internet access demand. Hence from business perspective huge benefit can be gained by providing users highly reliable and fast data rates in minimum cost. This can be done by making optimum use of scarce resource which in our case is bandwidth. High data rates can be achieved by using better techniques in network but reliability comes with better operation and maintenance (O & M). The Self Organizing Network (SON) as part of LTE is playing a key role in this regard [1]. Its primary aim is to reduce the cost in installation and management of networks through self-configuration, self-healing and self-optimization.

Self-optimizing and self-healing mechanisms are meant for improving quality of service (QoS) through reduction in malfunctions which come from inaccuracy in planning or equipment failure. Self healing detects and cures the fault as early as possible and self optimization uses certain techniques to optimize the network parameters under interference and overload conditions.

1.2.3 Radio Resource Management

Since bandwidth is a scarce resource we can use resources available to us like radio resources and manage them in such a way to have better QoS. This management includes following three types.

- Transmission power
- Mobility
- Scheduling of radio resources management.

The foundation of LTE is based upon how intelligently this management is done [2] so that it can meet our future needs and can sustain more loads keeping high speeds of data rates. With increase in use of wireless technology it is necessary that available resources are scheduled in such a manner that they remain available to users even while they are moving and there should not be any collision among scheduled resources. Thus SINR (signal to interference noise ratio) of the network must be enhanced through careful management so that users receive the signals in good quality.

1.3 Statement of Problem

Joint optimization of more than one RRM parameters i.e., Mobility and Inter-Cell Interference Co-ordination (ICIC) will be investigated to achieve the network optimization objective. It is expected that joint optimization of two RRM parameters will yield better network performance in terms of KPIs (Key Performance Indicators) as compared to the case when single RRM parameters is optimized.

1.4 Objectives

The following are the objectives of this research work.

- Detailed study of LTE technology
- State of the art study on auto-tuning and self optimization
- Propositions of the new auto-tuning algorithms.
- Network level simulations to validate the auto-tuning concepts.
- The thorough investigation of network level simulations for LTE in MATLAB
- Establishment of a platform for the testing of new ideas related to LTE network
- Investigate ways to reduce mobile networks operation cost by self-optimization

1.5 Author's Contribution

The author has proposed joint optimization of two RRM parameters to achieve better network performance in terms of KPIs as compared to single RRM parameter optimization

The performance of proposed optimization has been evaluated graphically on basis of results obtained from simulator.

1.6 Thesis Organization

Organization of remaining portion of thesis is as follows. Chapter 2 presents review of literature for proposed work. Chapter 3 and 4 discuss fuzzy systems and reinforcement learning for rule optimization of those systems respectively. Chapter 5 explains the proposed work. Chapter 6 gives the comparison of obtained results. Chapter 7 gives the conclusion of thesis and proposes future directions.

Chapter 2

LITERATURE REVIEW

2.1 Chapter Overview

This chapter presents a brief review of the two RRM parameters being optimized, benefits of using fuzzy logic and achieving optimum results through reinforcement learning.

2.2 RRM Parameters

LTE comprises of all IP core networks known as the Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) and Evolved Packet Core (EPC) [3]. Multiple techniques in E UTRAN, such as load balancing and ICIC, which are the radio resource management parameters, aim to resolve the bandwidth bottleneck and interference caused in this part of the network.

Both these RRM parameters will be used in joint optimization of LTE network.

2.2.1 Hand-Over Margin

Hand over margin (HOM) is an important feature of SON as it balances the load thus increasing network capacity [4]. Implementation of load balancing is carried out by tuning hand over margins between adjacent cells thus decreasing block call rate (BCR). HO process is the one responsible for transfer of an ongoing call between two adjacent ells. Adjustment in HO parameters settings results in the modification of the service area of a cell by sending users to adjacent cells.

Thus, there is a reduction in area of the over loaded cell thereby increasing the area of neighboring cell which is less loaded by transferring users from edge of congested cell. This arrangement results in more uniform traffic distribution, thus reducing BCR in over loaded ell.

A lot of effort has been put in to study the literature on balancing of load in adjacent cells in a network. Rida Nasri and Zwi Altman [5] presented a load balancing technique using simulator has been shown to reduce cell congestion. A. Tolli and P. Hakalin [6], showed that the amount of unnecessary HO attempts and failures can be reduced to a minimum by tuning the load based HO thresholds in a GSM/WCDMA scenario, where we suppose that a considerable amount of users are handed over to less loaded cells when the load exceeds the threshold. A. Pillekeit, F. Derakhshan, E. Jugl, and A. Mitschele-Thiel [7] presented the decision of triggering inter-system HOs for load balancing between UMTS and GSM is dependent on the load in the source and the target cell, the quality of service, HO overhead and the time taken since the last HO. It also proves the fact that both the capacity and the provided quality of service can be significantly improved in overload scenarios. Min Sheng, Chungang Yang, Yan Zhang, Jiandong Li [8] showed a game theoretic approach for load balancing has been discussed for LTE which is zone based. Pengfei Li Xiaohui Chen Weidong Wang [9] discusses an efficient algorithm for mobile load balancing in LTE which uses cell specific offset to distribute loads in a better way. Pablo Mu˜noz, Raquel Barco and Isabel de la Bandera [10] involves reinforcement learning for load balancing in adjacent cells.

In LTE, the network is the one which has to decide when to perform hand over in order to maintain call quality. This work involves algorithm based upon power budget hand over technique.

2.2.2 Inter-Cell Interference Coordination

Inter-Cell Interference Coordination (ICIC) is another important radio resource management parameter for improving efficiency of LTE networks [11]. Orthogonal

Frequency Division Multiple Access (OFDMA) is being used by LTE in downlink. The reason for OFDMA selection is provision of high spectral efficiency and efficient performance in high mobility scenarios and fading environments. LTE is using frequency reuse-1 system in order to make full use of available bandwidth but it suffers the disadvantage of inter cell interference. In case of absence of inert cell interference coordination (ICI) mechanism this interference becomes critical on cell edges. One of those mechanisms is soft reuse scheme which uses different portions of frequency bandwidth are allocated with different amount of power for the mobile users. The power allocation pattern is denoted as the power mask of the cell. This thesis presents a solution for the ICIC based on adaptive soft frequency reuse scheme. Following figure shows the frequency and power distribution model in a mobile network.

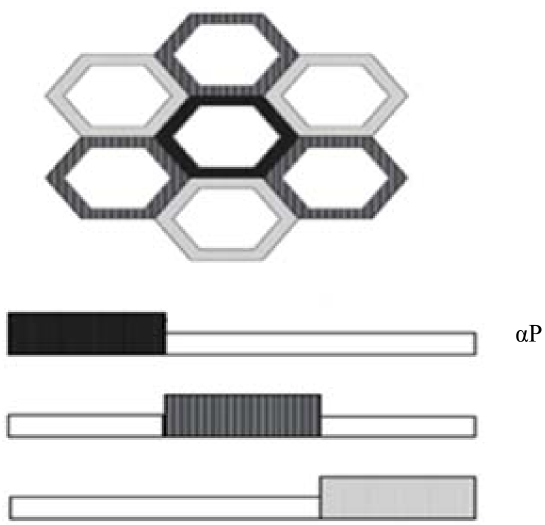

Figure 2.1 Inter Cell Interference Coordination

Different techniques to mitigate interference have been studied including frequency reuse schemes such as soft reuse and fractional reuse schemes [12]. Ahmed TRIKI, Loutfi NUAYMI [13] discusses changes in some algorithms performing ICIC have been discussed. Also an exchange based interference coordination algorithm has been proposed. An analytic qualification approach for interference mitigation by using the technique of fractional frequency reuse has been discussed in [14] [Marko Porjazoski and Borislav Popovski]. A comparison of soft frequency reuse and fractional frequency reuse has been discussed in [15] [Manli Qian]. SFR uses grouping in carriers as well as power allocation optimization jointly to mitigate interference. Marko Porjazoski and Borislav Popovski [16] have proposed another FFR technique for ICIC in which different parts of frequency spectrum are allocated to edge users of adjacent bands while users close to base station have all same chunk of frequency bandwidth. Mariana Dirani and Zwi Altman [17] have used reinforcement learning for interference mitigation.

2.3 Fuzzy-Q Learning

Nowadays application of Fuzzy logic is vast having scope in engineering, artificial intelligence, and computer science. Moreover, the ability of fuzzy logic that it can be used in combination with several learning techniques, such as reinforcement learning, neural networks and genetic algorithms makes it helpful in improving performance of various systems. Fuzzy logic converts human experience into a set of rules and thus helps in solving complex non linear problems like wireless network.

But the rule base of Fuzzy Logic Controller (FLC) should be designed in such a manner, so as to take into account, both the current network state and the network operator policy. That is where Q-Learning is so handy. [18], load sharing problem in GERAN is described. FLC is used which adapts the HO margins to minimize the call blocking rate in the network. All of the above references show that FLCs are quite handy for self optimization. Fuzzy Logic has the edge because of their capability to translate human knowledge into a set of rules.

Fuzzy logic controllers require a set of rules which interpret from input the required output in linguistic terms. These rules are based upon human expertise. In mobile communication operator experience comes in handy in order to establish these expertise. However, most of the times these expertise are either unavailable or not that accurate so different methodologies have been sorted out to polish these rules, e.g. evolutionary computing, neural networking or reinforcement learning (Q Learning in our case). Q Learning is a Reinforcement Learning (RL) method for learning from interaction, when it is impossible to ascertain best possible outcome for all the situations in which the controller has to get involved [19].

FUZZY INFERENCE SYSTEMS

3.1 Fuzzy Logic

In fuzzy logic, there is a huge difference with traditional crisp logic because its foundation block is uncertainty. Fuzzy logic is based upon fuzzy set theory, in which the concept of uncertainty is there which allows partial membership of membership function. Moreover with fuzzy logic, random complex environments can be modeled.

3.2 Fuzzy Set

A set without a hard, clearly sharp defined boundary is known as fuzzy set [20]. If ***X*** is the complex random environment (the input space variable) and its elements are given by *x*, then a fuzzy set *A* is defined as:

$$A = \{x, \mu_A(x)\ | x \in X, 0 \leq \mu_A(x) \leq 1\} \tag{3.1}$$

where μ denotes the membership function. The membership function relates each element of ***X*** to a crisp value between 0 and 1. Membership function has many types e.g. triangular, trapezoidal etc. The general trapezoidal curve has been shown in Fig. 3.1:

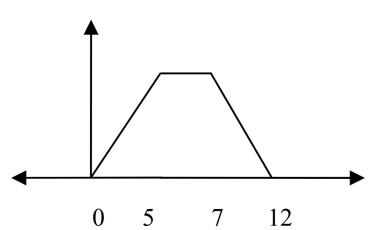

Fig. 3.1 General trapezoidal MF

3.3 Linguistic Variable

X_f is the variable whose values are represented by words and not by crisp integers. It is because we want to show the quality of the variable and not the quantity. This variable is known as linguistic variable (fuzzy variable). Under the umbrella of linguistic variable there can be many linguistic labels (L_i), and there is a fuzzy set associated to each one of them:

$$X_f = L_i \mid 1 \leq i \leq n \tag{3.3}$$

$$L = \{x, \mu_{L_i}(x) \mid x \in X, 0 \leq \mu_{L_i}(x) \leq 1\} \mid 1 \leq i \leq n \tag{3.4}$$

Any non precise words in human language can serve as linguistic labels for translation of different variables. For instance to describe the speed it can be slow, medium or fast and so on. In a complex environment any number of linguistic terms can be defined which will segregate the whole model of the fuzzy variable known as fuzzy partitioning. It has two types:

- Weak partition: This type dictates that there is no compulsion on construction of membership functions (labels) of the variable.
- Strong partition: this type of partitioning is based upon following equation:

$$\sum_{i=1}^{n} \mu_{L_i}(x) = 1 \quad \forall x \in X \tag{3.5}$$

which shows that for all possible x at least one fuzzy set will have some membership degree and their sum must be equal to 1. This equivalence to unity makes it more useful as it covers the complete variable.

3.4 Rule base:

Main aspect of fuzzy logic is its rule base. It is formulated through some expert knowledge or through some learning technique. It translates input to desired output. If χ is the input and y is the output and A and B are the membership functions of input and output respectively then if x is 10% of A and the rule says that if x is A then y is B then this employs that (y is 10% of B).

3.5 Fuzzy Inference System (FIS)

FIS is an expert system governing the output and is based upon fuzzy rules. It controls the mapping of an output from a given input using the given rule base. This mapping provides the foundation for making future decisions. Its functionality depends upon following four functions and they are discussed later in this chapter.

- Fuzzify the input vector
- Evaluate the rules
- Deriving truthness of triggered rules
- Defuzzify into ouput

This system has two types:

- Mamdani type (M-FIS)
- Takagi-Sugeno type (TS-FIS).

Their main difference is in determination of output for these two systems. In M-FIS the output is fuzzy in nature while T-FIS has linear or constant output and its processing is more simple and efficient. In proposed work TS-FIS method is described.

3.6 Takagi-Sugeno FIS (TS-FIS)

It has two types.

- Zero-order

  In this type all output membership functions are singleton spikes.

- First-order

  In this type the output membership functions are linear combination of input variables

Hence consequent part of the rules creates the difference.

In general the rule-base in a FIS consists of arbitrary number of different form of rules constructed out of AND and/or OR operators.

Only some of the rules from rule base will be activated. These are the rules whose all antecedent membership values are not equal to zero when X is applied and A(X) is the set containing all such rules.

Until this point the input vector has been fuzzified and membership functions have been calculated. Now antecedent membership values of activated rules are multiplied to get degree of truthness of each activated rule. Its calculation is carried out as per following equation:

$$\alpha_{Ri}(X) = \prod_{j=1}^{n} \mu_{L_j^i}\left(x_j\right), \qquad \forall R_i \in A(X) \tag{3.6}$$

Where $\mu_{L_j^i}$ is the MF associated to L^i_j.

This truth value dictates the magnitude of single spike generated at output.

Now , $Y_o(X)$, which is the total output of the system is given by:

$$Yo\,(X) = \frac{\sum_{R_i \in A(X)} \alpha_{R_i}(X) o^i}{\sum_{R_i \in A(X)} \alpha_{R_i}(X)} \tag{3.7}$$

which means that total output is average of magnitude of triggered rules output.

The controllers discussed in this proposed work are MISO (Multi Input Single Output) in nature. They are preferred over MIMO (Multi Input Multi Output) because they are less complex in nature.

3.7 FIS Architecture

The architecture makes the functionality of system very evident. As the functions are 4 so architecture is 4 layered.

- Input layer
- Fuzzification layer

- Rule evaluation layer
- Output layer, which keeps all weights of triggered rules equal to 1

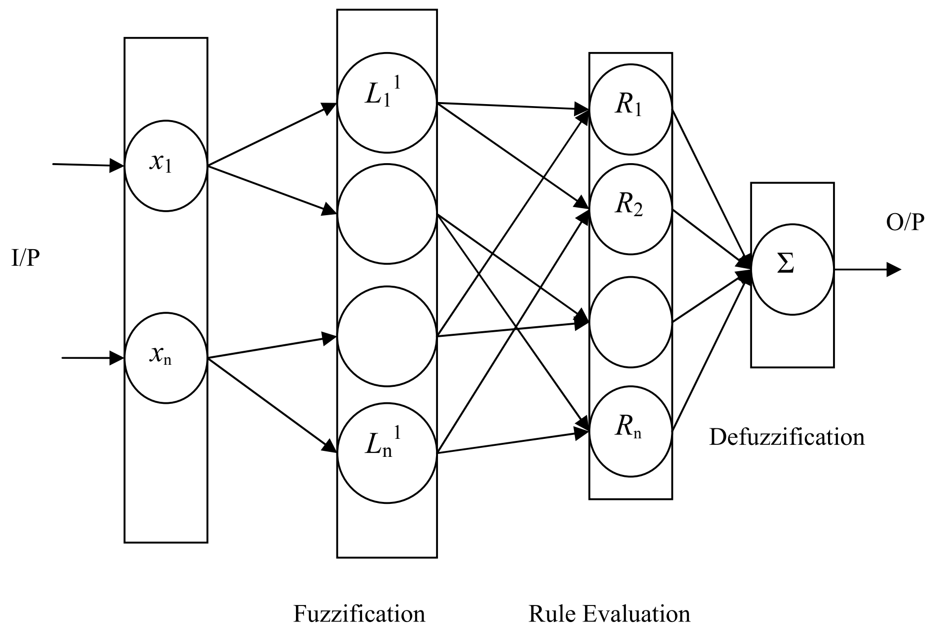

Fig. 3.2 FIS Architecture

The above figure relates to MISO technique and can be converted to a MIMO system by having more outputs in final layer.

C h a p t e r 4

REINFORCEMNT LEARNING

4.1 Introduction

In complex environments the initial data is insufficient to make an efficient rule base for the fuzzy controller. Moreover in many scenarios we cannot predict the input vector to the controller. Solution to these problems lies in reinforcement learning (RL). The use of RL with FLC makes the controller very efficient in handling complex environments.

RL is a self learning technique with which efficient actions can be chosen while being in an environment.

4.2 Reinforcement Learning Model

This model consists of sensors which take an input vector i_t from environment at time t then performs an action a_t to take the system to a next state and also gets feedback r_t which is the reinforcement signal to declare the reward if the action is towards the desired goal. So the main objective is to find the best action out of the ones previously tried which will result in maximum summation of these signals resulting in maximum reward. Thus the standard RL model can be shown in Fig 5.1. The input vector is taken from the current state of the environment and ***R*** is the reward which takes towards the final goal and is known. Now RL, can be defined as:

"Reinforcement learning is learning what to do, how to map situations to actions, so as to maximize a numerical reward signal.

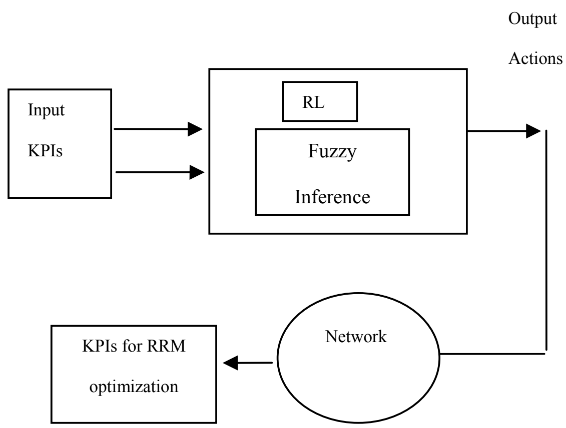

Fig. 4.1 RL model

Now for RL to choose the best action it has to keep on exploring new actions whose reward is yet unknown. This is done on the basis of an exploration and exploitation (EEP) policy.

4.3 Reinforcement Learning Components

RL is based upon five components.

- Agent: It means the learner
- Environment: Complex media which is required to be modeled.
- Policy π: Set of rules which determine present as well as future actions.
- Reinforcement signal R: It is the reward function which maps state action pair to a number.
- Value function: Starting from a particular state the total expected reward is known as value function of that state. It can be calculated in different ways and is the main difference between different RL techniques.

4.4 Reinforcement Learning Challenges

- Exploration/Exploitation trade-off: Exploitation is the process in which best actions with maximum rewards are chosen to take the system from one state to the other. But this can only be done if sufficient number of actions have been tried already and out of those the best one is chosen. The process of trying different possible actions is called exploration. There exists a tradeoff between how much to explore and when to exploit and is known as exploration/ exploitation dilemma. In this work a hybrid approach is used i.e., ε -Greedy method, which will be discussed in next chapter.
- Curse of dimensionality: In modeling complex environments especially wireless communication networks the possible input vector dimension is very large. Thus memory requirements become a critical issue. in order to handle large data required for saving and tracking possible transitions and value functions etc large look up tables with high memory requirements are necessary. This is known as curse of dimensionality and will be dealt with more advanced RL techniques.

4.5 Markov Decision Process

4.5.1 Frame Work

Markov Decision Process (MDP) is the one which satisfies the relation between an agent and environment with all functions related to RL. In this stochastic system, various states can be defined with transition probabilities to next states. Following figure shows an MDP in which an action is taken to move from one state to the other.

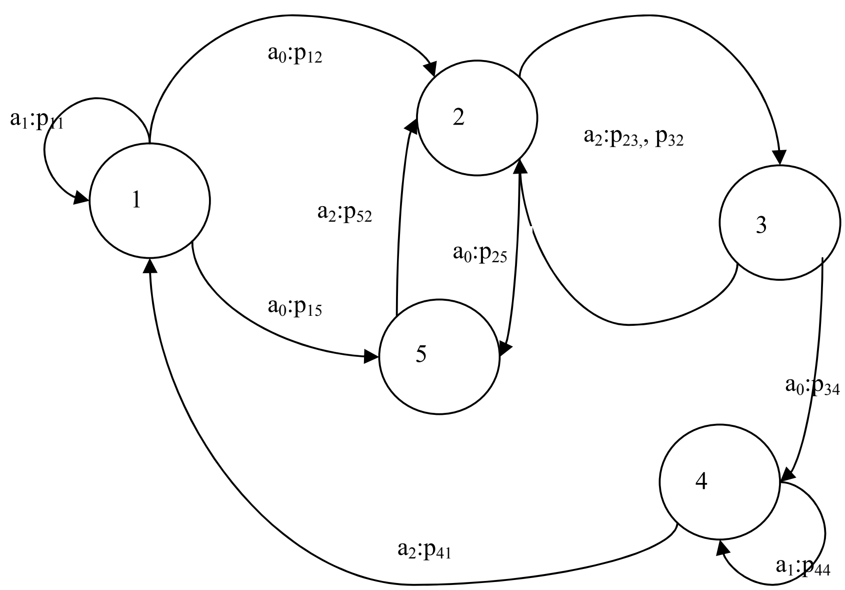

Fig. 4.2 Markov Decision Process

In Fig. 4.2 an agent can take an action a_n from nth state . From each state the agent can reach the next state with a probability p_{ab} from state a to state b. The model given above must satisfy following conditions of probability:

$$p_{ab}\,(a_n\,)\in[0,1], \qquad \forall i, j, k.$$

$$\sum_j p_{ab}\,(a_n\,)=1, \quad \forall i, j, k \tag{4.1}$$

Modeling the environment with this property implies that one only needs current state information for making the decision of which action to take. Current state carries enough knowledge comprising of probability of moving to next state and reward associated with it to make the decision and is the same even if the information about whole environment including all the data of all states actions and rewards is known. This feature is known as markov property and is very important in RL learning. This means that p_{ab} is entirely dependent upon state a in which agent is present and a_n which is the action agent has to take.

Above discussion employs that by markov property: if random variable S_t denotes the state of the environment at time step t, A_t the action taken at this time (equal to one of available possible actions), and s' and r as the resulting state and reward respectively, then by means of Markov property we have:

$$\begin{aligned} p_{s_t s_{t+1}}(A_t) &= \Pr(S_{t+1} = s', r_{t+1} = r | S_t, A_t) \\ &= \Pr(S_{t+1} = s', r_{t+1} \\ &= r | S_t, A_t, r_t, S_{t-1}, A_{t-1}, r_{t-1}, \dots \dots, r_1, S_0, A_0) \end{aligned} \tag{4.2}$$

$\forall\, S_t, r, s', A_t$ and past state action reward triples.

where agent is in state S_t at time t, A_t is the action taken by agent at this time, $s'$is the next state, r is the resulting reward and r_t is the reinforcement signal received when agent reached S_t . since $\boldsymbol{S}$ denotes the set containing states, and $\boldsymbol{A}$ the set containing action then:

$$\text{Reward function } \boldsymbol{R}\text{:}\ \boldsymbol{S} \times \boldsymbol{A} \tag{4.3}$$

$$\text{Transition probability function } p\text{:}\ \boldsymbol{S} \times \boldsymbol{A} \times \boldsymbol{S} \to [0, 1]. \tag{4.4}$$

4.6 Dynamic Programming (DP)

Problems related to environments which are modeled by MDP can be efficiently solved by using DP algorithms. By means of these algorithms one can find optimal policy and corresponding optimal value function for solution of problem. π is the optimal policy which has the highest value function for all states of the environment:

$$V^*(s) = \max_\pi E_\pi \lfloor \textstyle\sum_{t=0}^{\infty} \gamma^t r_t \rfloor \qquad \forall s \in S, 0 \leq \gamma \leq 1. \tag{4.5}$$

Where V^* is corresponding optimal value function.

Classical DP methods have their own drawbacks. Curse of dimensionality as discussed earlier is one of them in which optimal policy can only be chosen if data of all states is stored which in real world scenario is much difficult.

However in complex environments such as wireless system, apriori knowledge of all features of MDP modeling cannot be known in the beginning so as to apply RL algorithm. In such scenarios certain advanced alterations are required to be done in the algorithm. Such advanced techniques are known as adaptive DP and they resolve such complexities efficiently. Basic RL techniques require all features of MDP model to be known such as all states and transitional probabilities etc while adaptive DP techniques work without having complete MDP model picture. They are of two types:

- Model-free

  As the name indicates this type of technique is completely independent of knowing system model in advance. These direct techniques learn the policy and value functions when they interact with the system and thus make complex systems such as wireless to be efficiently optimized. These techniques are more generic in nature and one such technique is Q learning.

- Model-based. These are also known as indirect techniques because they firstly learn the model of environment and then utilise that model to establish an efficient policy which is optimal in nature and makes the controller efficient.

4.7 Q-learning

Q-learning is simple computationally efficient and model free adaptive DP technique Watkins was its pioneer and its working principle is based upon determination sate-action qualities. It differs from common DP techniques which are based upon state-action values. Assume that expected discounted reward, Q^* (s, a), results when an agent takes an action a from state s. Then like before an estimation of optimal state-action value function, Q^* (s, a), denoted by $Q^{\sim}(s, a)$ will be:

$$Q^{\sim}_{t+1}(s,a) = Q^{\sim}_{t}(s,a) + \alpha\left[r_{t+1} + \gamma \max_{a'} Q^{\sim}_{t}(s',a') - Q^{\sim}_{t}(s,a)\right]. \quad 0 < \alpha < 1 \quad (4.6)$$

where α is the learning rate.

In Eq. 4.6 a' is the next action to be taken from state s'. Now in Q-learning this action a' must be chosen among all the available actions with the highest Q value. This means that the policy being evaluated is closer to the optimal policy for the current Q-function. Q-learning is also independent of exploration/exploitation policy and will finally achieve convergence to optimal values.

4.8 Fuzzy Q-learning

4.8.1 Framework

FQL is basically a combination of fuzzy logic with Q learning used for optimization of different problems.

FQL structure consists of a fuzzy controller with a rule base governed by Q learning. Fuzzy logic controller is based upon certain rules which come from human expertise or operator knowledge. Dealing with complex systems where human knowledge is insufficient to generate a strong rule base Q learning algorithm becomes handy in optimization of rule base.

4.8.2 FQL Algorithm

The complete FQL learning process during one time step can be summarized as follows; assume that the current time stamp is $t+1$, the agent has already performed the action $U_t\,(X_t\,)$ at previous time step t, and has even received the reward r_{t+1} , then:

- Initialize q table
- Calculate the degree of truthness for all rules
- For each activated rule select an action with the EEP policy
- Calculate the inferred action corresponding to all activated rules
- Calculate the corresponding quality of inferred action
- Execute the action and observe new state and reinforcement

- Calculate the value function of the new state
- Calculate the variation of the quality ΔQ
- Update the elementary quality table for each activated rule
- $t \leftarrow t + 1$

If convergence is attained then stop learning.

4.9 Chapter Summary

In this chapter, a brief taxonomy of fuzzy logic theory and related work in each category has been presented. Reinforcement learning when combined with fuzzy logic helps in finding optimal solutions to complex problems in a time efficient manner.

Chapter 5

PROPOSED WORK

5.1 Chapter Overview

This chapter presents, in detail, the proposed joint optimization algorithm. First, it briefly reviews the two controllers being employed and their working. Then a detailed description of the joint working of both controllers for the optimization task has been discussed. Finally, the proposed strategy for joint optimization employing fuzzy q learning algorithm has been given.

5.2 Parallel Operation Of Controllers

Idea of using two parallel controllers is not new but has been discussed in [21] and reduces complexity of rules e.g.; consider the following IF-THEN statement containing an OR operation:

"IF x_1 is X_1 AND x_2 is $X2$ OR x_3 is X_3 AND x_4 is X_4 THEN y is Y."

By convention, this is understood as

"(IF x_1 is X_1 AND x_2 is X_2) OR (IF x_3 is X_3 AND x_4 is X_4) THEN (y is Y)."

This means that the above mentioned rule is equivalent to the combination of the following two IF-THEN statements:

(1) "IF x_1 is X_1 AND x_2 is X_2 THEN y is Y."

(2) "IF x_3 is X_3 AND x_4 is X_4 THEN y is Y."

Hence, the fuzzy logic OR operation is not necessary to use: it may shorten IF-THEN statement, but it maximizes the complexity of the rules. Same idea is being used in this proposed work.

5.3 Controllers

In this work I have used two controllers for joint optimization of HO and ICIC. Both use fuzzy logic and rule optimization is being handled by Q learning algorithm. The rule base of controllers in linguistic terms is:

- Controller 1: If BCR is imbalanced from cell i to cell j and handover value between them is low then output will be the one having maximum reward
- Controller 2: If BCR of a cell is high and average BCR of neighboring cells is low then output will be the one having maximum reward

The optimum results in both controllers depend upon the reward function which has been selected so as to achieve our goal in terms of better KPIs for improved performance of network. Q learning algorithm makes use of this reward function in order to optimize the consequent part of fuzzy logic rules.

Since controller 1 gives output in terms of change in handover margin which impacts directly on KPIs of system such as BCR, SINR, ABR and FTT etc, their improvement will further enhance if controller 2 works in combination with controller 1 which is indeed the case. The less congested cell may be suffering interference problems so when users will be handed over to that cell their KPIs will degrade. Proposed joint optimization technique will enhance the KPIs by ensuring interference mitigation in those cells. Controller 2 has the output in terms of change in alpha (α) of the cell for power allocation αP to the subbands thus reducing interference. Thus joint optimization will improve overall performance of network.

5.3.1 Controller Design

Both fuzzy logic controllers are designed making use of fuzzy logic toolbox. The initial data for rules of controllers has been provided so as to intialise them.

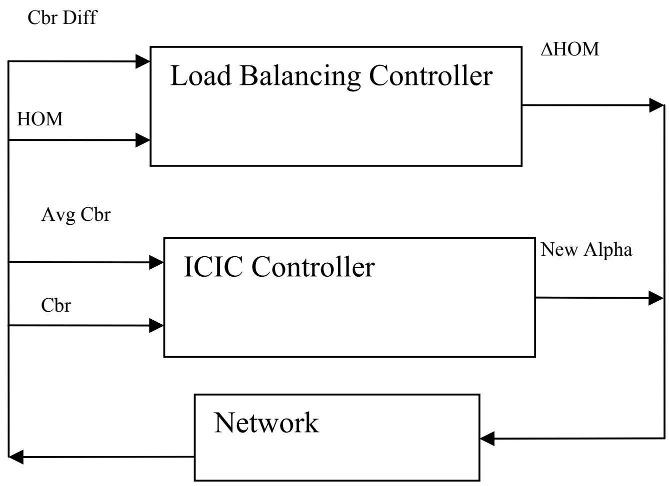

Fig 5.1 Joint Optimization

5.3.2 Design of Controller 1:

For controller meant for load balancing its impact of sharing traffic between cells is that the call blocking is reduced, especially in highly loaded cells. Call Blocking Ratio (CBR) or Block Call Rate (BCR) is defined as:

$$CBR = \frac{N_{blocked}}{N_{offered}} = \frac{N_{blocked}}{N_{blocked}+N_{accepted}} \quad (5.1)$$

where $N_{accepted}$ and $N_{blocked}$ and are the number of accepted and blocked calls by the admission control respectively and $N_{offered}$ is the number of calls offered.

In LTE, the network is responsible for deciding when performing a HO to maintain the connection quality. The network also determines how often each mobile terminal has to send signal measurements back to their serving eNBs, as well as the interval time to perform each measurement.

One of the most widely used algorithms for the HO-triggering decision is the Power Budget (PBGT) HO. This algorithm triggers the execution of the HO procedure if the following condition is fulfilled for a specific time period determined by the Time-To-Trigger (TTT) parameter:

$$\overline{RSRP_j} > \overline{RSRP_i} + HOM_{i\to j} \tag{5.2}$$

where $\overline{RSRP_i}$ (Reference Signal Receive Power) and $\overline{RSRP_j}$ are the averaged values of RSRP measured for serving cell i and target cell j respectively, and $HOM_{i\to j}$ is the HO margin defined between cell i and cell j.

The two inputs for this controller are the BCR difference between the two cells and the previous value of handover which requires updation.

For BCR difference a 45x45 matrix has been calculated which is continuously updated after each iteration. Similarly a 45x45 matrix of hand over margin values is intialised with initial value put to 6. The output is in the form of change in handover margin which balances the load between the cells

5.3.3 Design of Controller 2:

Controller meant for ICIC utilises adaptive soft reuse 1 technique which is that all the bandwidth is available for all the cells in a network to reuse but the power being transmitted is controlled for interference management.

Bandwidth is divided into three disjoint subbands. One of them is allocated to mobiles in poor signal quality and is known as an edge band. Edge of the cell is where users normally experience poor signal quality, but radio conditions may even get worse near the base station because of deep shadow fading. The rest of the two frequency subbands are known as centre bands.

We use notation c and e for centre and edge respectively. The interference in the system comes from transmissions on the centre band which interferes with adjacent cell edge band. Let us denote the maximum transmitted power per subcarrier by P. When a base station strongly interferes with its adjacent cells, the ICIC designed process reduces the transmitted power per subcarrier in the centre band to αP. PRBs allocation is performed based on a priority scheme for accessing the edge subbands. Let s denote the serving base station of the user u. A quality metric h_u is calculated using the pilot channel signal strengths

$$h_u = \frac{Pr_{su}}{\sum_{s' \neq s} Pr_{s'u} + \sigma_n^2} \tag{5.3}$$

where Pr_{su} and $Pr_{s'u}$denote the mean pilot power received by the user u of the signals transmitted by base stations s and s′ respectively, and σ_n^2 is the thermal noise power corresponding to the pilot channel. h_u is slightly different from SINR in that the power of data channels used to calculate the SINR is being controlled.

Users in poor radio conditions with bad h_u are allocated resources from the protected band and get maximum power of base station. When the protected subband is full, the centre band starts contributing. The channel model used to calculate the SINR is adapted for implementation in a semi-dynamic network simulator that can assess performance of large network size with tens to hundreds of base stations. Within each subband b, namely center (b = c) or edge (b = e) subbands of base station s, the power allocated to the subcarriers is identical, i.e. $P_{sc} = \alpha_s P$ or $P_{se} = P$ respectively. The allocation of a mobile u to a given subband depends on the quality metric.

Consider a mobile u attached to base station s. The average interference perceived by u and produced by eNBs s′, s′ ≠ s is given by

$$I_{ub} = \sum_{s' \neq s} \Lambda(s, s') \eta_{s'b} \frac{P_{s'b} G(s', u)}{L(s', u)} \tag{5.4}$$

Λ (s, s′) equals one if eNBs s and s′ use the same frequency bandwidth and zero otherwise. $P_{s'b}$ is the transmitted power per subcarrier belonging to frequency subband b, b ∈ {c, e}. $\eta_{s'b}$ represents the load of subband b of base station s′ and is defined as the ratio between the number of PRBs allocated in subband b. $N_{s'b}^{allocated}$, and the total number of PRBs available in this subband, $N_{s'b}^{available}$:

$$\eta_{s'b} = \frac{N_{s'b}^{allocated}}{N_{s'b}^{available}} \tag{5.5}$$

The load coefficient (4) expresses the fact that the average interference on a given sub channel belonging to the frequency subband b is proportional to the portion of time the sub channel is used. Hence $\eta_{s'b}$ equals the probability of interference produced by s′. G (s′, u) is the antenna gain of base station s′ in the direction of the mobile u. The channel loss L (s′, u) at a distance d between s′ and mobile u is given by

$$L(s',u) = A_{\chi}(s',u)\left(\frac{1}{d(s',u)}\right)^{v} \tag{5.6}$$

where A is measured path-loss at a reference distance and ν is the path-loss exponent which depends upon environment. Now SINR per subcarrier for the mobile u is:

$$SINR_{ub} = \frac{P_{sb}G(s,u)}{L(s,u)(I_{ub}+\sigma_z^2)} \tag{5.7}$$

where σ_z^2 is the thermal noise per subcarrier and I_{ub} is interference perceived by mobile u in subband b.

The two inputs for this controller are the CBR of the centre cell and average CBR of the neighbouring cells. If CBR of central cell is high as compared to neighbours output will be increase in alpha of that cell. Thus KPIs of central cell improve. For average CBR the positioning of cells in the simulator has been observed and on that basis the inputs of central and neighbouring cells are given to controller.

Output of controller is used to increase or decrease power in the subbands of the centre cell.

5.3.4 Membership functions

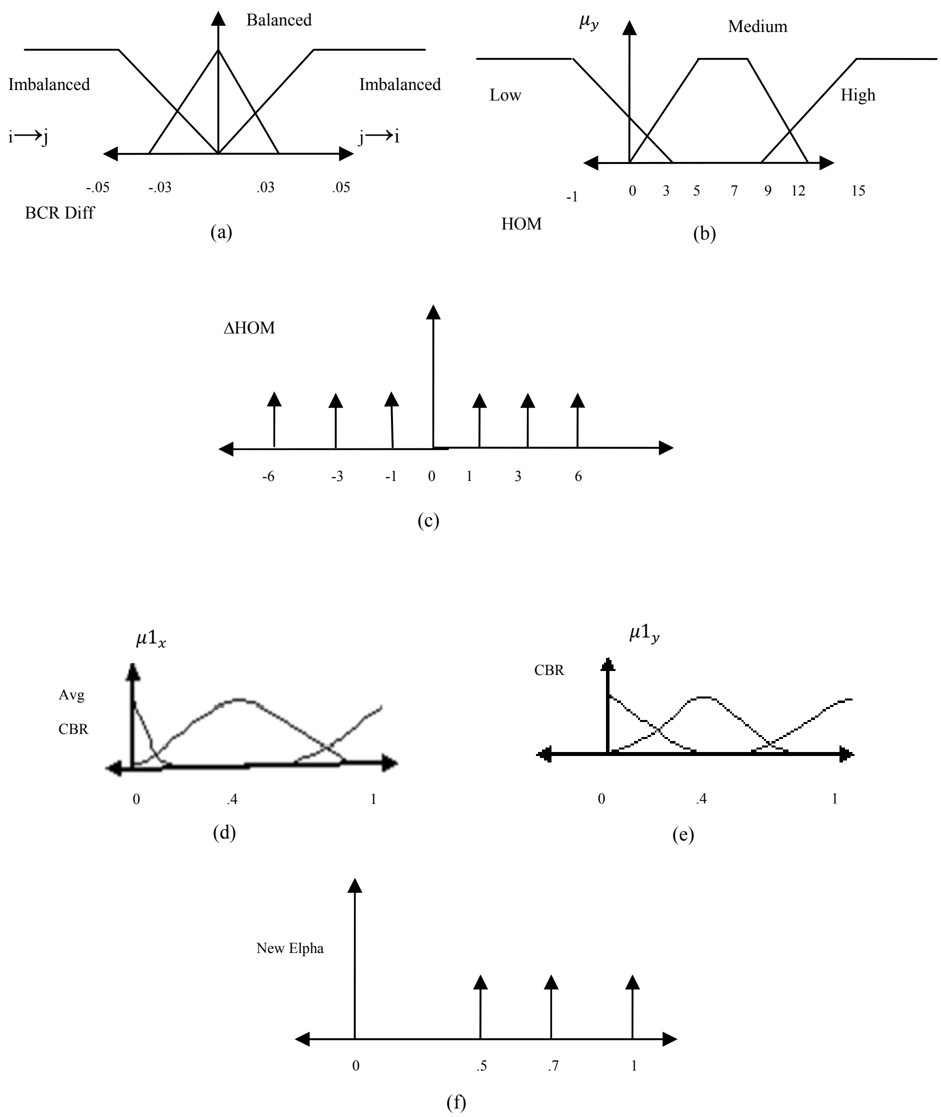

Fig 5.2(a)Block Call Rate Difference For Controller 1 (b)Hand Over Margin (c)Change In Hand Over Margin (d)Block Call Rate For Controller 2 (e)Average Block Call Rate (f)New α For Power Masking Of Subbands

5.3.5 Modes Of Controllers

- Mode 11 is the mode in which results are gathered without any optimization
- Mode 12 is the mode in which only handover margin is optimized for comparison
- Mode 13 is the mode in which only ICIC is optimized
- Mode 14 is the mode in which joint optimization is carried out. The optimal actions stored at the end of execution of mode 15 are applied in this mode.
- Mode 15 is the mode in which EEP policy is followed. Both controllers are initialized first then on basis of EEP policy the optimal actions are chosen based on the reward function. At the end the actions having maximum reward function are chosen which will yield best results in terms of improved KPIs

5.4 Learning algorithm:

Once both controllers are initialized they contain certain set of rules with some random actions. For taking optimal actions Q learning is applied which explores the actions having maximum reward. That action is stored and if later on any action is calculated to have reward value more than the previous one then it is stored overwriting the previous one. Exploration of optimal actions is a complex and time consuming task so it is limited by EEP policy. The algorithm in this regard is shown below. η is learning rate which is used for maintaining the balance between exploitation and exploration. (e.g., when $\eta = 0$ it shows that 100% exploitation is being used, that is, always choose the best action):

- Initialize q values in look up table
  $q[i,j] = 0, 1 \leq i \leq N \; and \; 1 \leq j \leq J \; for \; HO$

  $q1[i,j] = 0, 1 \leq i \leq N \; and \; 1 \leq j \leq J \; for \; ICIC$

- Select an action for each activated rule with EEP policy ε
  $a_i = \arg max_k \, q[i,k] \; with \; probability \; 1-\varepsilon$

  $a_i = random\{a_k, k = 1,2, \dots J\} \; with \; probability \; \varepsilon$

  $a1_i = \arg max_k \, q1[i,k] \; with \; probability \; 1-\varepsilon$

  $a1_i = random\{a1_k, k = 1,2, \dots J\} \; with \; probability \; \varepsilon$

- Calculate global actions inferred by FLC

$$a = \sum_{i=1}^{N} \alpha_i(s) \times a_i$$

$$a1 = \sum_{i=1}^{N} \alpha 1_i(s) \times a1_i$$

  Where $\alpha_i(s)$ is degree of truth for HO and $\alpha 1_i(s)$ is for ICIC for rule i

- Approximate Q function from current q values and degree of truthness

$$Q(s(t), a) = \sum_{i=1}^{N} \alpha_i(s) \times q[i, a_i]$$

$$Q1(s(t), a1) = \sum_{i=1}^{N} \alpha 1_i(s) \times q1[i, a1_i]$$

  Where $Q(s(t), a)$ is he value of Q function for state s at iteration t and action a

  for HO while $Q1(s(t), a1)$ is for action a1 in the same state for ICIC

- Leave the system to evolve to next state s(t+1)
- Observe the reinforcement signals r(t+1),r1(t+1) and compute the value of new state denoted by V and V1

$$V_t(s(t+1)) = \sum_{i=1}^{N} \alpha_i(s(t+1)) \,.\, max_k q[i, a_k]$$

$$V1_t(s(t+1)) = \sum_{i=1}^{N} \alpha 1_i(s(t+1)) \,.\, max_k q1[i, a1_k]$$

- Calculate difference ΔQ:

$$\Delta Q = r(t+1) + \gamma \times V_t s(t+1) - Q(s(t), a)$$

$$\Delta Q1 = r1(t+1) + \gamma \times V1_t s(t+1) - Q1(s(t), a)$$

Where γ is discount factor

- Update q values by ordinary gradient descent method
  $q[i, a_i] = q[i, a_i] + \eta . \Delta Q . \alpha_i(s(t))$

  $q1[i, a_i] = q1[i, a_i] + \eta . \Delta Q1 . \alpha 1_i(s(t))$

  Where η is the learning rate

- Repeat the above described process starting from step 2 for the new current state until the convergence is achieved

5.5 Chapter Summary

This chapter presented the proposed technique of joint optimization and explained the design of both controllers which are being employed for joint RRM optimization. It also presents the fuzzy q learning algorithm and membership functions being used in both controllers.

C h a p t e r 6

RESULTS AND ANALYSIS

6.1 Chapter Overview

This chapter describes the platform specifications, the characteristics of the simulator, compares the different modes of autotuning and evaluates the results graphically. a comparison of cumulative distribution function of all modes is also presented in order to weigh the improvement in performance.

6.2 Simulator characteristics:

The results presented in this section have been obtained using a semi-dynamic network simulator. The simulator performs correlated snapshots to account for the time evolution of the network with a time resolution of the order of a second.

FTP-type data traffic is considered. During a time interval between two consecutive snapshots, the following operations are performed: users' arrivals and departures and update of base station loads; admission control algorithm is executed for each new arrival; the user arrival follows a Poisson process, and the data volume for download is fixed to 5 Megabits. After download completion the user leaves the network and the QoS is calculated for the terminating call. At the end of each time interval, the simulator computes the new positions, radio conditions and handovers of the users.

A LTE network comprising 45 base stations positioned on a non-regular grid is considered. The inter-site distance varies from 1.5 to 2 km. Each base station has a capacity of 9 PRBs per subband, namely a total of 27 PRBs per base station. Users requesting to initiate a file transfer are allocated 1 to 3 PRBs according to resource availability. Users' speed of 13.88 m/sec is chosen. The noise power spectral density is taken as −173 dBm/Hz. The Okumura-Hata propagation model is used. The path loss at a reference distance of 1 km and the path loss exponent are chosen as −128 dB and 3.76 respectively. The standard deviation of the shadowing process is 6 dB.

The periodicity of the FQLC is of 40 seconds. The KPIs serving as inputs to the FQL controller are filtered using an averaging filter over the same duration of 40 seconds. The learning rate ε is set to 0.1 and the discount factor γ - to 0.95.

6.3 Experimentation and Results

In this section, comparison has been shown between auto tuning using non optimization, single optimization and joint optimization of RRM parameters. The KPIs being compared are BCR, Mean SINR, ABR and MFTT. Cumulative distribution function (CDF) of users' file transfer time for the three systems has also been compared. The proposed work solution achieves lower values for the file transfer time, when compared even with the system implementing the load balancing and ICIC. This does not come as a surprise since joint optimization achieves joint improvement for all users. In other words, improvement of users with bad quality does not come at the expense of users with good quality.

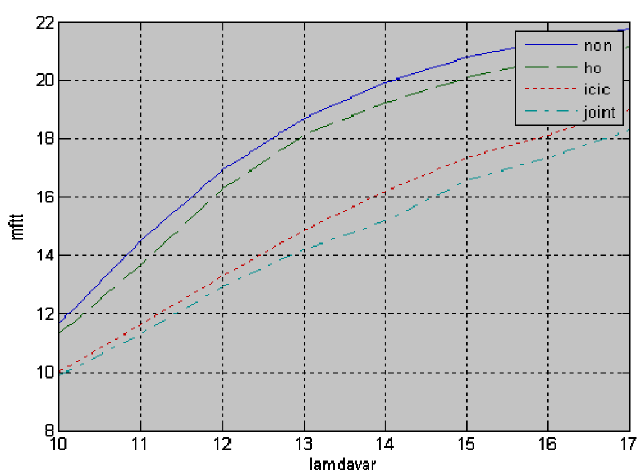

Figure. 6.1 Mean File Transfer Time

Fig. 6.1 compares the mean file transfer time of the three systems. Significant improvement with respect to joint optimization is obtained. For traffic intensity up to 14 arrivals/sec, joint optimization outperforms the rest of the schemes, and then no further improvement as the cells become overloaded too much with incoming calls.

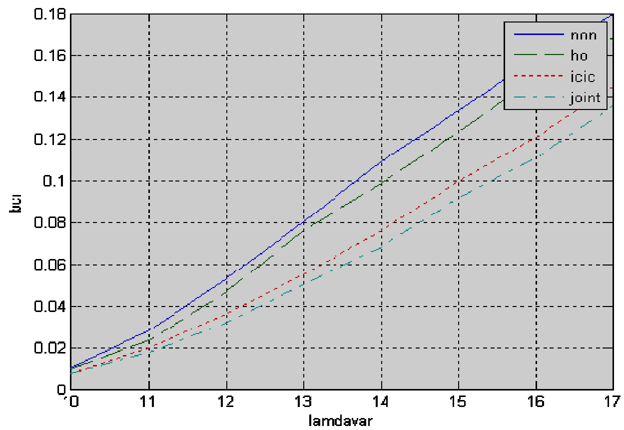

Fig 6.2 Block Call Rate

Fig. 6.2 depicts the blocking rates for the three systems as a function of traffic intensity. The blocking rate provides an indicator of capacity, namely the traffic intensity that can be served by the network for a given blocking rate. Joint optimization approach offers a better system capacity with respect to the other systems being compared.

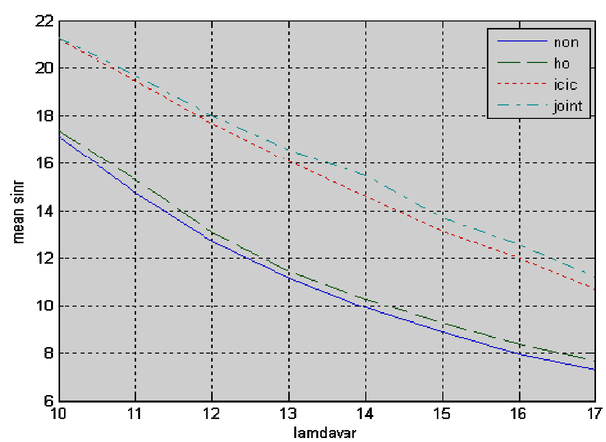

Fig 6.3 Mean Signal To Interference Noise Ratio

Fig. 6.3 presents comparison for the mean SINR perceived by all users. Joint optimization is better than other schemes for moderate traffic intensity. The tendency decreases for higher traffic intensity. Systems having no optimization and with HO as optimization technique suffer from low average SINR due to inter-cell interference.

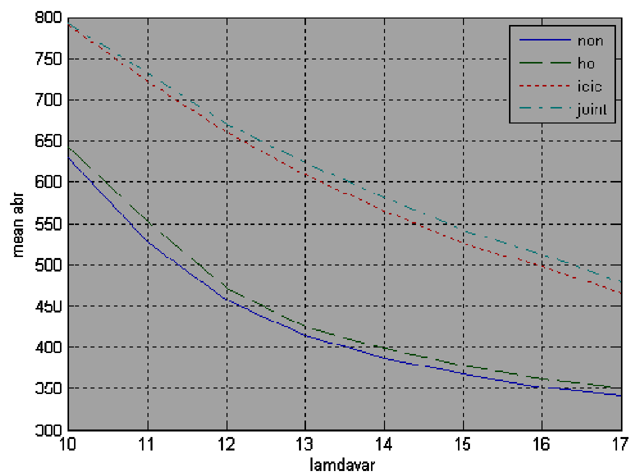

Fig 6.4 Average Bit Rate

Fig 6.4 compares the average bit rates achieved by the users in the network. Comparison shows that joint optimization technique achieves highest data rates among all. This is because both RRM parameters are being optimized simultaneously.

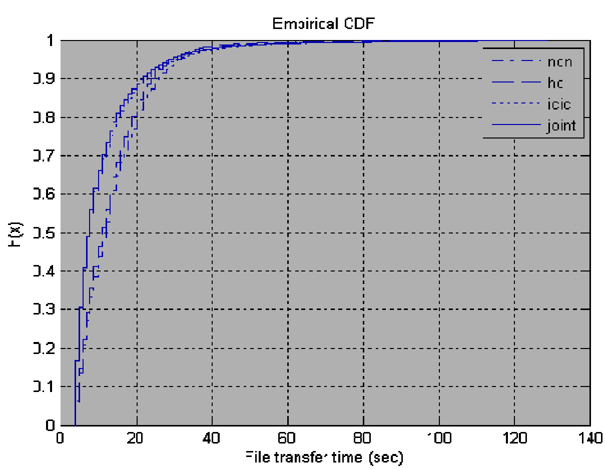

Fig 6.5 Cumulative Distribution Function

Fig. 6.5 compares the cumulative distribution function (cdf) of users' file transfer time for the three systems. Joint optimization achieves lower values for the file transfer time, This does not come as a surprise since joint optimization enforces harmonic mean fairness and achieves joint improvement for all users. In other words, improvement of users with bad quality does not come at the expense of users with good quality.

Chapter 7

CONCLUSION AND FUTURE DIRECTIONS

7.1 Chapter Overview

This chapter provides overview of the carried research, objectives achieved, limitations of proposed solution and future directions.

7.2 Research Overview

In this thesis, auto tuning in LTE using joint RRM optimization has been proposed. The proposed algorithm resolves a number of issues that exist with networks with optimization of single RRM parameter. There is a significant decrease in call blocking and file transfer time while the mean SINR of network increases. A brief introduction of LTE, SONs and RRM parameters has been presented in Chapter 1. Chapter 2 presented the material which hs been gone through during this research. It includes concepts of load balancing interference mitigation fuzzy logic and reinforcement learning. Chapter 3 presented the details of fuzzy inference structure with study of basic fuzzy logic and concept of its architecture. Chapter 4 presented the concept of reinforcement learning and markov system model upon which it is applicable. Chapter 5 presented the proposed joint optimization technique alongwith the learning algorithm. In chapter 6 results have been evaluated.

7.3 Objectives Achieved

This research produced a simulator for auto tuning of LTE network using joint optimization of RRM parameters namely HO and ICIC. Optimization is achieved by using a fuzzy logic controller. The rule base used in controller employs Q learning technique in order to maximize the reward which is set to be the throughput of the network. Output KPIs are monitored and results are compared with auto tuning techniques involving optimization of single RRM parameter which reveal that joint optimization yields better results in terms of KPIs namely BCR, MFTT and SINR.

7.4 Future Directions

As demand for wireless access to the Internet and Internet-based services is expanding, competitive advantages in the mobile business can be gained by offering enhanced user experience through cost effective broadband mobile access. A promising approach is to maximize total performance of networks, i.e., provide not only wireless access with higher performance but also more efficient operation and maintenance (O&M). The Self Organizing Network (SON) introduced as part of the 3GPP Long Term Evolution (LTE) is a key driver for improving O&M. It aims at reducing the cost of installation and management by simplifying operational tasks through automated mechanisms such as self-configuration and self-optimization.

Mobility Robustness Optimization (MRO) encompasses the automated optimization of parameters affecting active mode and idle mode handovers to ensure good end-user quality and performance, while considering possible competing interactions with other SON features such as, automatic neighbor relation and load balancing.

The objective of MRO is to dynamically improve the network performance of HO in order to provide improved end-user experience as well as increased network capacity. This is done by automatically adapting cell parameters to adjust handover boundaries based on feedback of performance indicators and by the optimization of time to trigger (TTT) for HO. Typically, the objective is to eliminate Radio Link Failures and reduce unnecessary handovers. Automation of MRO minimizes human intervention in the network management and optimization tasks.

REFERENCES

[1] M. Dottling and I. Viering, “Challenges in mobile network operation: Towards self-optimizing networks,” in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009

[2] Technical Solutions for the 3G Long-Term Evolution by H Ekstrom 2006

[3] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (EUTRAN); Overall description; Stage 2, version 9.1.0 2009

[4] S. Luna-Ram´ırez, M. Toril, M. Fern´andez-Navarro, and V. Wille, “Optimal traffic sharing in GERAN,” *Wireless Personal Communications*, vol. 57, no. 4, pp. 553 –574, April 2010

[5] Ridha Nasri , Zwi Altman "Handover Adaptation For Dynamic Load Balancing In 3gpp Long Term Evolution Systems" 2013

[6] A. T¨olli and P. Hakalin, “Adaptive Load Balancing between Multiple Cell Layers,” in *Proc. of IEEE 56th Vehicular Technology Conference (VTC), Fall*, 2002.

[7] A. Pillekeit, F. Derakhshan, E. Jugl, and A. Mitschele-Thiel, “Force-based Load Balancing in Co-located UMTS/GSM Networks,” in *Proc. of IEEE 60th Vehicular Technology Conference (VTC), Fall*, 2004.

[8] Min Sheng, Chungang Yang, Yan Zhang, Jiandong Li " Zone-based Load Balancing in LTE Self-Optimizing Networks" A Game Theoretic Approach Transactions on Vehicular Technology IEEE, 2013

[9] Pengfei Li Xiaohui Chen Weidong Wang "A High-efficient Algorithm Of Mobile Load Balancing in LTE System" Ying Yang University of Science

and Technology of China Hefei,China , 2012

[10] Pablo Mu˜noz, Raquel Barco and Isabel de la Bandera "Optimization of Load Balancing using Fuzzy Q-Learning for Next Generation Wireless Networks" 2013

[11] 3GPP TR 36.913 v8.0.0, ”Requirements for Further Advancements for E-UTRA (LTE-Advanced)”, (Release 8), June 2008.

[12] Manli Qian, Wibowo Hardjawana, Yonghui Li, Branka Vucetic, Jinglin Shi, Xuezhi Yang, "Inter-cell Interference Coordination Through Adaptive Soft Frequency Reuse in LTE Networks", IEEE 2012

[13] Ahmed TRIKI, Loutfi NUAYMI "InterCell Interference Coordination Algorithms in OFDMA wireless systems" Institute Telecom; Telecom Bretagne 2, rue de la Chˆataigneraie CS 17607 35576 Cesson S´evign´e Cedex, France, 2011

[14] Marko Porjazoski and Borislav Popovski "Analysis of Intercell Interference Coordination by Fractional Frequency Reuse in LTE" University “Sts Cyril and Methodius” Faculty of Electrical Engineering and information technologies Skopje, Republic of Macedonia, 2010

[15] Manli Qian "Inter-cell Interference Coordination Through Adaptive Soft Frequency Reuse in LTE Networks" 2012

[16] Marko Porjazoski and Borislav Popovski "Contribution To Analysis Of Intercell Interference Coordination In Lte: A Fractional Frequency Reuse Case" Faculty of electrical engineering and information technologies University “Sts Cyril and Methodius” – Skopje Karpos 2 b.b, 1000, Skopje Republic of Macedonia, 2010

[17] Mariana Dirani and Zwi Altman "A Cooperative Reinforcement Learning Approach for Inter-Cell Interference Coordination in OFDMA Cellular Networks" Altran, 58 Boulevard Gouvion St Cyr - 75017 Paris, France †Orange Labs, 38/40 rue du General Leclerc,92794 Issy-les-Moulineaux 2010

[18] M. Toril and V. Wille, "Optimization of Handover Parameters for Traffic Sharing in GERAN," *Wireless Personal Communications*, vol. 47, no. 3, pp. 315 –336, 2008.

[19] R. S. Sutton and A. G. Barto, *Reinforcement Learning: an Introduction*. MIT Press, 1998.

[20] Guanrong Chen Trung Tat Pham University of Houston, Clear Lake Houston, Texas "Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems, 2001

[21] Tom M. Mitchell "Machine Learning" 1997

[22] Fuzzy logic toolbox Tutorials