TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO.: 072/MSI/616

An Effective Handover Scheme in Heterogeneous Networks using Multi Armed Bandit Based Learning Approach

BY

SUBASH CHANDRA PAKHRIN

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULLFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN INFORMATION AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

NOVEMBER, 2017

An Effective Handover Scheme in Heterogeneous Networks using Multi Armed Bandit Based Learning Approach

BY

SUBASH CHANDRA PAKHRIN

072MSI616

SUPERVISED BY:

Dr. Dibakar Raj Pant

A THESIS SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENIGNEERING IN PARTIAL FULFULLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

INSTITUTE OF ENGINEERING, PULCHOWK CAMPUS

TRIBHUVAN UNIVERSITY

LALITPUR, NEPAL

NOVEMBER, 2017

COPYRIGHT©

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering for acceptance, a thesis entitled "**An Effective Handover Scheme in Heterogeneous Networks using Multi Armed Bandit Based Learning Approach**", submitted by Subash Chandra Pakhrin in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Information and Communication Engineering**".

_____

Dr. Dibakar Raj Pant
Supervisor,
Department of Electronics and Computer Engineering,
Pulchowk Campus, Institute of Engineering

_____

Om Bikram Thapa
External Examiner,
Vianet Communication Pvt. Ltd.

_____

Dr. Dibakar Raj Pant
Committee Chairperson,
Department of Electronics and Computer Engineering,
Pulchowk Campus, Institute of Engineering

Date: 28th November, 2017

Departmental Acceptance

The thesis entitled "**An Effective Handover Scheme in Heterogeneous Networks using Multi Armed Bandit Based Learning Approach**", submitted by **Subash Chandra Pakhrin** in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Information and Communication Engineering**" has been accepted as a bona fide record of work independently carried out by him in the department.

_____

Dr. Dibakar Raj Pant

Head of the Department

Department of Electronics and Computer Engineering,

Pulchowk Campus,

Institute of Engineering,

Tribhuvan University,

Nepal

ACKNOWLEDGEMENT

I would like to express my gratitude to my thesis supervisor, Dr. Dibakar Raj Pant for his continuous encouragement, advice, help and invaluable suggestions to my study and my life. He is such a nice, generous, helpful and kindhearted person. During my study at Institute of Engineering, Pulchowk Campus, he builds a relaxing, comfortable and active environment during study. I owe my research achievements to his experienced supervision.

I sincerely thank to Prof. Dr. Dinesh Kumar Sharma, Prof. Dr. Shashidhar Ram Joshi, Prof. Dr. Subarna Shakya, Dr. Surendra Shrestha, Dr. Ram Krishna Maharjan, Dr. Sanjeev Pandey, Dr. Nanda Bikram Adhikari, Mr. Daya Sagar Baral and the whole Department of Electronics and Computer Engineering for giving this opportunity and inspiring me all the time.

My special thanks to our program coordinator, Dr. Basanta Joshi for extraordinary coordination and support.

I also would like to thank my wife, Nilam Lama and my daughter Serena Pakhrin for understanding, assistance and company during my study. I also thank my parents for the support. This thesis could not have been completed without their supports.

Last but not least, special thanks to all of my friends, who continuously motivated me to carry out this research.

ABSTRACT

Deploying pico cell and femto cell nodes within a macro cell layout is known as heterogeneous networks. It is a promising solution to enhance overall system performance, cell-edges coverage. Indeed this type of deployment leads to an improvement of spectral efficiency and achieves load balance by offloading macro cell traffic to low power nodes. Heterogeneous networks deployment incurs new technical challenges related to handover performance of user equipment, which will be impacted especially when high velocity user equipment's traverse pico cells. To tackle this problem, reinforcement learning techniques; Multi Armed Bandit and Bayesian Multi Armed Bandit has been proposed. User equipment's learn the best cell based on the posterior distribution of reward and continuous optimal cell range expansion value is predicted through linear regression. These equipment's are scheduled based on their velocity and previous rates (exchange among tiers). Information entropy is also used to evict the user equipment from overcrowded cell to the cell that has relatively less traffic, better throughput and higher signal to interference noise ratio. The potential reward on each base stations channel is calculated; then the channel with the maximum accumulated rewards is formally chosen. The proposed learning based approach with entropy measures for load balancing out performs the Multi Armed Bandit based mobility management in terms of user equipment throughput. In average, a gain of up to 86 % is achieved for user equipment throughput, while the handover failure probability is reduced to a factor of two by the proposed reinforcement based mobility management approaches. Simulation value of user equipment's throughput validates the proposed scheme is better over the classical RSRP handover scheme.

Index Terms – Cell range expansion, Heterogeneous Networks, mobility management, reinforcement learning, multi armed bandits.

Table of Contents

Copyright iii

Recommendation iv

Department Acceptance v

Acknowledgement vi

Abstract vii

Table of Contents viii

List of Figures x

List of Tables xi

List of Abbreviations xii

1. INTRODUCTION 1

1.1 Background and Motivation 1

1.2 Problem Statement 4

1.3 Objective 4

2. LITERATURE REVIEW 5

2.1 Literature Review 5

3. RELATED THEORY 10

3.1 Reinforcement Learning 10

3.1.1 The Agent-Environment Interface 10

3.2 Multi Armed Bandit Algorithm 12

3.2.1 $\in$-greedy 14

3.2.2 Upper Confidence Bound (UCB) 14

3.2.3 Thompson Sampling 15

3.2.4 Bayesian Bandit 16

3.3 Exploration – Exploitation Tradeoff 18

3.4 Entropy 18

3.5 Heterogeneous Network Concepts 19

3.6 Heterogeneous Network Deployment Challenges 20

3.6.1 Frequency Allocation 21

3.6.2 Backhauling 21

3.6.3 Handover 21

3.6.4 Self – organizing 21

3.6.5 Interference 22
3.7 Cell Range Expansion 23
3.8 Shadowing (Large-scale Fading) 26
3.9 Small-scale Fading (Fast Fading) 26
4. METHODOLOGY 27
4.1 Proposed Framework 27
4.2 Model network 27
4.3 Data Collection 29
4.4 The Beta Distribution 30
4.5 Linear Regression to Predict REB 31
4.6 Signal to Interference plus Noise Ratio (SINR) 32
4.7 Flowchart 33
4.8 Short-term solution: a context-aware scheduler 36
4.9 Multi-Armed Bandit Based Learning Approach 37
4.10 Bayesian Bandit 39
4.11Tools Used 39
5. Result and Analysis 40
5.1 Path Loss Models 40
5.2 Exploration and Exploitation Dilemma 40
5.3 Application of dataset to UCB algorithm 41
5.4 Posterior updating through different pulls 43
5.5 Linear regression to predict REB and entropy for load balancing 44
5.6 Simulation of HetNet in NS3 45
5.7 UE Throughput and sum-rate 46
5.8 Segmentation of UE in HetNets 47
5.9 HOF and PP probability 48
5.10 Validation of Result 51
6. Conclusion 53
7. Limitations and Future Works 54
REFERENCES 55

LIST OF FIGURES

Fig. No.	Title	Page No.
1	Example of the decay of the power profile from the Macro Base Station and Pico Base Station as the UE moves away from the Macro Base Station towards the Pico Base Station.	2
2	Heterogeneous Cellular Network	3
3	Interaction between agent and environment.	11
4	Greedy Octopus pulling N slot machine once in a random order	13
5	Macro - Femto Interference Scenarios.	23
6	Range Expansion Concept.	24
7	Handover in HetNets with RE.	25
8	Proposed framework for handover using reinforcement learning approach.	28
9	Model Network based on velocity calculation and history based scheduling	28
10	Beta distribution for different values of α and β	31
11	2D SINR map without shadowing and without wrap-around, with shadowing and wrap around	33
12	Flow chart of the proposed learning-based MM approaches.	35
13	Different path loss model	40
14	Exploration – Exploitation dilemma	41
15	Uniform distribution of eNBs selected, eNB4 is selected	42
16	Posterior probability updating at different pulls	44
17	Linear regression model to predict REB.	45
18	Handover of UE with two eNodeB.	46
19	Data flow of UE with two eNodeB during execution of handover.	46
20	Comparison between MAB with entropy approach, MAB approach and classical RSRP method.	47
21	HOF probability reduction under various velocity and TTT	49
22	PP probability reduction under various velocity and TTT	50
23	Throughput at various simulations runs of classical and MAB approach.	51

LIST OF TABLES

Table No.	Name of Table	Page No
1	Types of Nodes in Heterogeneous Cellular Networks	3
2	Specification of different elements in HetNets	20
3	System Parameters for simulations	29
4	Bernoulli distribution of UE latching among the 10 eNodeB	30
5	Entropy measures for load balancing.	45
6	Segmentation of UE in HetNets	48
7	Results of handover campaign	52

LIST OF ABBREVIATIONS

3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
ANN	Artificial Neural Network
BS	Base Station
CRE	Cell Range Expansion
CWND	Congestion Window
dB	Decibel
dBi	decibels relative to isotropic radiator
EUTRA	Evolved Universal Mobile Telecommunications System
HetNet	Heterogeneous Networks
HMM	Hidden Markov Model
HO	Hand Over
HOF	Handover Failure
HOPE	Handover Proximity Entity
Hyst	Hysteresis
IID	Independent and Identically Distributed
IMSI	International Mobile Subscriber Identity
LPN	Low Power Node
LTE	Long Term Evolution
MAB	Multi Armed Bandit
MADM	Multi Attribute Decision Making
MBS	Macro Base Station
MCMC	Markov Chain Monte Carlo
MDP	Markov decision process
MLB	Mobility Load Balancing
MM	Mobility Management

MUE	Mobile User Equipment
OFDM	Orthogonal Frequency Division Multiplexing
PBS	Pico Base Station
PDF	Probability Distribution Function
PP	Ping Pong
QoS	Quality of Service
RB	Resource Block
REB	Range Expansion Bias
RL	Reinforcement Learning
RLF	Radio Link Failure
RNTI	Radio Network Temporary Identifier
RRU	Remote Radio Unit
RSRP	Reference Symbols Received Power
RSS	Received Signal Strength
SINR	Signal to Interference plus Noise Ratio
SMA	Simple Moving Average
SMC	Sequential Monte Carlo
SONs	Self - Organizing Networks
TS	Thompson Sampling
TTT	Time To Trigger
UC^3	Unified Central Congestion Control
UCB	Upper Confidence Bound
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
VA	Valuable Access
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WPM	Weighted Product Method

Chapter 1

INTRODUCTION

1.1 Background and Motivation

Heterogeneity is one of the main properties of emerging next generation wireless networks. Recent trends in wireless communication indicate that wide area cellular networks can offer high speed and reliable multimedia services to mobile and fixed end users. In this regard, various access technologies, such as worldwide interoperability for microwave access, wireless fidelity, universal mobile telecommunications system, and long term evolution can be incorporated with each other in order to achieve a high performance. In heterogeneous networks (HetNet), a variety of low power base stations (BSs) such as micro, pico and femto BSs can work in overlay with a macro cell BS. In fact HetNet provides high quality services. In order to benefit such positive points, some challenges must be alleviated at the beginning. The most challenging issue in HetNet is a seamless handover (HO) between various stations from different network tiers.

Handoff is the process of transferring an ongoing voice or data from one cell to another as user moves through a coverage area of a cellular system. Processing handoff is key for telecommunication providers to meet required quality of service for users and to reduce cost. Handover techniques in HetNets include intra technology handover referred to as horizontal HO, and inter technology handover referred to as vertical HO, which mainly aim to apply an appropriate handover initialization strategy. These techniques can improve resource utilization and user quality of service (QoS) in HetNets. Handoff is especially important in a congested inner city environment with small cell sizes where efficiency can greatly be improved by avoiding many unnecessary handoffs due to frequent back-and-forth switching between base stations (known as ping-pong effect).

The deployment of pico and / or femto BSs within the macro cell, indeed, can provide higher connection speed and better coverage to the mobile users located at the border of the macro cell or in regions with high traffic demand. While increasing the efficiency of the cellular networks, HetNets also raise several technical challenges related to user management. An important aspect is related to the management of user mobility that, differently from the

classical cellular networks, has to deal with cells of widely varying coverage areas. In general, Handover (HO) process is triggered by the User Equipment (UE), which periodically measures the Reference Symbols Received Power (RSRP) from the surrounding cells. When the difference between the neighboring cells RSRP is higher than a fixed HO hysteresis value, then the HO process starts, as exemplified in Figure 1 if this condition holds for a period of time equal to the Time- To-Trigger (TTT) parameter, the HO is finalized and the UE connects to the BS with the strongest RSRP.

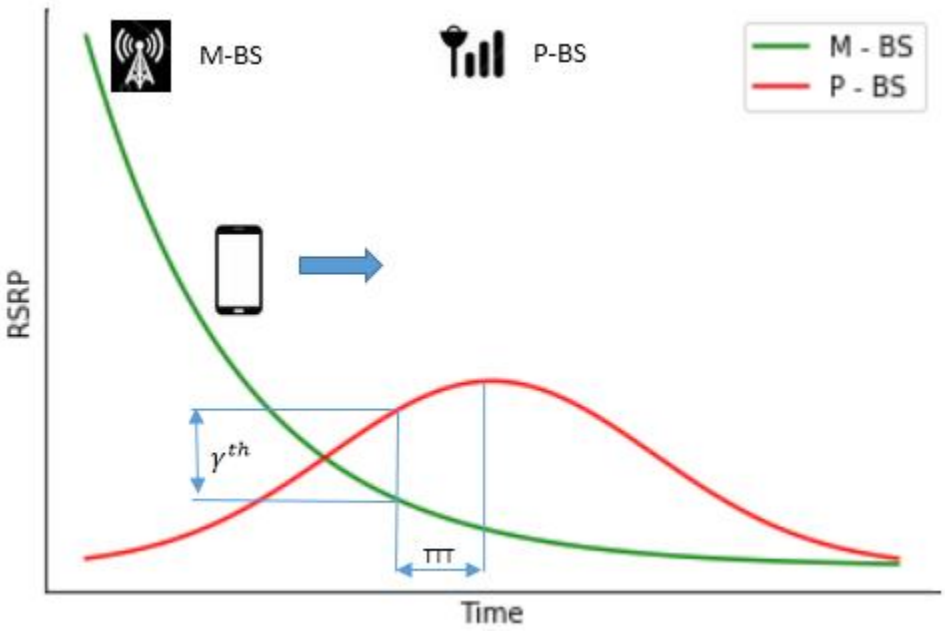


Figure 1: Example of the decay of the power profile from the Macro Base Station and Pico Base Station as the UE moves away from the Macro Base Station towards the Pico Base Station.

The main idea behind a HetNet is to improve spectral efficiency per unit area with the deployment of a diverse set of non-conventional, low-power nodes such as pico BSs, femto BSs and relays within the areas covered by the existing macro cellular infrastructure, over the same frequency spectrum.

A HetNet structure is depicted in Figure 2. Macro BSs, which are deployed in a planned layout to provide basic coverage, transmit at high power levels (5W to 40W). Pico BSs are usually intended for outdoor deployments to alleviate “dead-spots” (no-coverage zones) and “hot-spots” (localities of higher traffic demand).

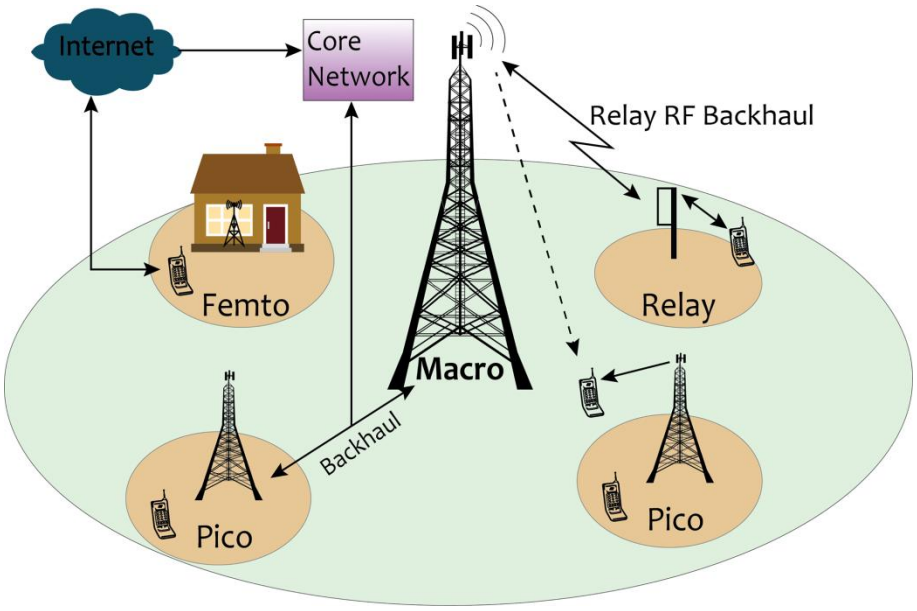


Figure 2: Heterogeneous Cellular Network.

Table 1: Types of Nodes in Heterogeneous Cellular Networks

Node	**Transmit Power**	**Features**
Macro	5W – 40W	Operator deployed, open access, dedicated backhaul
Pico	250mW – 2W	Operator deployed, same access and backhaul features as macro
Femto	< 100mW	Consumer deployed

HetNets entail a number of challenges in terms of capacity, coverage, mobility management (MM), and mobility load balancing (MLB) across multiple network ties. Mobility Management is essential to ensure a continuous connectivity to mobile UE while maintaining quality of service (QoS).

Poor Mobility Management approaches may increase handover failure (HOF), radio link failure (RLF), and ping-pong (PPs), and results in unbalanced load among cells. This entails low resource utilization efficiency and hence deterioration of the user experience. In order to solve this problem, while minimizing PPs, mobility parameters in each cell need to be carefully and dynamically optimized according to cell traffic loads. It is essential to optimize handover parameters such as time to trigger (TTT), range expansion bias (REB), and hysteresis margin in order to answer the question - "when to handover the UE to which cell".

In this thesis work, seamless handover is performed by reinforcement learning approaches like Multi Armed Bandit approach and Bayesian Bandit and features like velocity of UE, entropy, and range expansion bias are considered.

1.2 Problem Statement

In recent years, with the exponential growth of traffic in wireless communication environment, there are much more strict requirements in network capacity and QoS for mobile wireless networks. The wireless system must have the capability to provide high data transfer rates, quality of services and seamless mobility. With the evolution of different technologies i.e. user equipment and applications, the demand of higher data rate and seamless connectivity is increasing exponentially. This higher demand cannot be fulfilled by any single wireless communication technology. Utilization of technologies like wireless local area network, universal mobile telecommunications system, and long term evolution in different scenario will result the end user satisfaction.

Traditional handover mechanism is performed by RSRP i.e. the UE is connected to the node that has the strongest RSRP. When the connection have to switch between heterogeneous networks for performance and high availability reasons, seamless vertical handoff is necessary. The requirements like bandwidth of the network, handoff latency, speed, network cost, network conditions, power consumption, entropy and user's preference as well as trajectory of motion must be taken into consideration during vertical handoff.

1.3 Objectives

- To evaluate the handover performance of HetNets when Multi-Armed Bandit Based Learning Approach with entropy is used as deciding factor for load balancing.
- To compare the handover efficiency of Reinforcement Learning approach which uses entropy for load balancing with Multi-Armed Bandit Based approach and classical RSRP handover approach.

Chapter 2

LITERATURE REVIEW

2.1 Literature Review

The deployment of small cells in HetNets raises new challenges in relation to the Handover process and the mobility management. In fact, the performance of a mobile user within a HetNet scenario highly depends on the setting of the handover parameters in relation to other context parameters such as the channel conditions and the user position and speed. A Markov - based Framework to model the user state during handover was presented by Francesco Guidon et al. [1] derived an optimal context-dependent handover criterion.

Zhixiong Ding et al. [2] introduced motion trajectory models for mobile user equipment (MUE) to provide basis of handover simulation and then they provide an effective handover scheme based on the prediction of the UE motion curve and the pre-handover mechanism.

Multi Attribute Decision Making (MADM) is one of the successful used methods in the literature to solve decision making problems. The problem of access network selection has been addressed by decision making methods based on available network information. However the quality of information is not considered. Weighted Product Method (WPM) is an MADM method that penalizes the unreliable attributes in making a decision. Peyman TalebiFard et al. [3] introduced an algorithm for a context-aware network selection that is based on modified WPM for access network selection. It uses a weight distribution method based on sensitivity analysis of WPM for the most influential criteria based on the state of user at a given time.

A. M. Miyim et al. [4] proposed Fast handover proximity entity (HOPE) scheme which support fast and efficient handover in heterogeneous cellular networks. The proposed method demonstrated half handover time as against what was obtained with legacy handover; furthermore they recorded a speed of 1/3 higher than what was obtained from legacy handover.

A. Habibzadeh et al. [5] introduced a new HO Decision-Making Algorithm which is based on joint traffic and propagation metrics is proposed for HetNets that can reduce the number of unnecessary handover in comparison with the traditional handover algorithms especially when the femto cell assignment probability is relatively high.

Malka N. Halgamuge [6] proposed that a mobile user behavior can be modeled by a Hidden Markov Model (HMM) as well as they formulated the handoff problem as an optimization problem of base station scheduling that minimizes a cost function that involves the HMM state estimation error and base station measurement cost.

Ili Nadia Md. Isa et al. [7] tuned the two handover parameters Hysteresis (Hyst) and Time-To-Trigger (TTT). And the simulation result has shown that the network performance is better after optimizing the Hyst and TTT of the handover parameters.

Dang Feng et al. [8] proposed an adaptive weight vertical handover algorithm, which can select the most suitable network as target handover network, according to the type of working application and the adaptive calculating weight vector according to user's preference.

Hiroyuki Koga et al. [9] proposed an optimized handover mechanism based on the Unified Central Congestion Control (UC^3) cloud controlling network architecture. In this paper the key aspect is to calculate the congestion window (CWND) value at the cloud server using information collected by wireless access networks.

Abhijit Bijwe et al. [10] proposed vertical handoff algorithms using neural networks. To achieve vertical handoff, ANN is one of the tools which are explored. ANN helps in taking the handoff decision based on RSRP, bandwidth, cost, network delay etc. ANN consists of input, hidden, output layer. In proposed method mobile terminal performs periodical measurement of RSRP and bandwidth samples of two different networks (e.g. cellular, WLAN) and vertical handoff is taken. More appropriate vertical handoff decision is taken as more number of parameters is considered (e.g. RSRP, bandwidth).

Kudo et al. [11] got the results of the number of outage UEs and average throughput which shows that after thousands of trials, the Q-learning approach can perform better than no learning schemes. They showed that the proposed method can decrease the number of outage UEs and improve average throughput at almost all ratios of RBs. Moreover, it can largely enhance the cell-range UE throughput compared with the schemes using a common bias value.

Wang et al. [12] proposed a direction prediction mechanism for LTE networks in order to lower the unnecessary handover. They proposed a direction prediction scheme with a simple cosine function to predict the moving direction of UEs. The eNBs, in front of the moving direction of UEs, are designated candidate eNBs for handover. Then, a target eNB is selected from the

candidate eNBs for handover through an angle-based dynamic weight adjustment scheme. Simulation results revealed that under various velocities of UEs, their proposed handover scheme apparently outperformed the standard scheme from the aspect of average handover times, number of packet loss, and average transmission delay.

R. Sasikumar et al. [13] calculated the effective bias value for each mobile station independently, by using a fuzzy logic inference system. Their proposed solution takes multiple attributes like signal strength of pico and macro cells, speed and direction of mobile stations, battery level and traffic requirements as input parameters and an effective bias value was calculated for each mobile station. The simulation experiments demonstrate that the best bias value will be in the range of 0 to 20 and the optimal bias value also depends on the results of our study, it is also observed that it is possible to reach the near optimal bias value even without knowing the distribution of UE. The proposed solution decrease the number of outage UE at almost all ratios of RS compared to fixed bias method.

Weyu Li et al. [14] proposed two SON aspects – load balancing and handover parameter optimization which can achieve a better coordination. This new method tunes the hysteresis according to a key indicator, radio link failure ratio, with realistic consideration, thus avoiding the possibility that load balancing has a bad influence on the network performance. With the proposed method, which is simpler and easy to realize, the network handover performance and load balancing effect are both guaranteed compared with conventional solutions.

Long Li et al. [15] has proposed a hierarchical network selection scheme for user-cell association in HetNets. The proposed scheme utilizes MADM to calculate the network ranking scores, where the computational complexity is reduced by avoiding unnecessary handovers and related calculation. The performance of the proposed scheme in terms of average UE wideband SINR and spectral efficiency is compared with the conventional max-SINR scheme through system-level simulations. The results have shown that the system performance is enhanced by the proposed scheme while achieving load balance for the HetNet. UE wideband SINR is improved up to 80% of UEs in simulated scenarios.

Simone et al. [16] evaluated the mobility performance in heterogeneous networks with macro and pico cells on the same carrier frequency. The study is based on different deployment scenarios and mobility parameters. The first handover parameters that are found promising for

macro-only scenarios are used for the considered case (co-channel deployment of both macro and pico). Considerable improved performance can be achieved by making the handovers fasters for small cells, e.g. by lowering the value of the time to trigger. This lowers the link failure rate, while increasing the rate of handover ping-pong (short stay). The optimal value of TTT achieving the lowest failure rate while minimizing the handover ping-pong proves to be different for macro and pico, hence dependent on the cell size.

Mishra et al. [17] proposed an efficient regression based scheme to predict a near optimal bias value that attempts to reduce blocking probability and improve load fairness index in the system. The simulation results verify that, in comparison to static bias, the proposed scheme also improves the cell edge user throughput, along with the target criteria.

Yuefeng Peng et al. [18] used three metrics – radio link failure (RLF) rate, handover failure (HOF) rate, short time of stay (short ToS) to evaluate and analyze the handover performance in HetNets. Then, they proposed scheme number 1 – optimize pico-macro handover and scheme number 2 – optimize macro-pico handover, to solve the issue of handover performance deterioration when UE moves in medium speed. In other words, two schemes can separately optimize pico cell leaving and attaching. Furthermore, these two schemes can be used jointly to further improve the mobility robustness in HetNets.

Jeffrey G. Andrews et al. [19] has surveyed and compared the primary technical approached to HetNet load balancing: (Centralized) optimization, game theory, Markov decision processes, and the newly popular cell range expansion (a.k.a. biasing), and have drawn design lessons for OFDM-based cellular systems.

Xiaofei Wang et al. [20] has surveyed and analyzed that AI-based techniques are proved to be able to acclimatize and be competent for the smart improvement of HetNets by the self-“X” features. In the same work they have discussed the state-of-the-art AI-based techniques for evolving the smarter HetNets infrastructure and systems, focusing on the research issues of self-configuration, self-healing, and self-optimization, respectively. A detailed taxonomy of the related AI-based techniques of HetNets is also shown by discussing the pros and cons for various AI-based techniques for different problems in HetNets.

Klaus I. Pedersen et al. [21] has proposed scheme that characterizes the UE devices autonomously decide small cell addition, removal, and change without any explicit signaling of

measurement events to the network or any signaling of hand - over commands from the network. Hence, the proposed solution effectively offloads the network from having to perform frequent small cell handoff decisions, and reduces the signaling overhead compared to known network controlled mobility solutions.

Kinan Ghanem et al. [22] studied the HO preparation and execution and the HO completion in E-UTRA. The effects of ping-pong phenomenon in LTE networks were investigated. A novel ping-pong avoidance scheme to detect the ping-pong type of movement and keep the old path for a short time in E-UTRA was also presented. The presented scheme distinguished between the general and the ping-pong type of movement. The performance evaluation of the algorithm showed that keeping the old path in the case of ping-pong movement can reduce the rate of ping-pong HO and its undesirable effects and enhance the HO quality indicator.

Meryem Semsek et al. [23] proposed an effective model for analyzing handover failures in small cell deployments, considering all important mobility management parameters of interest. They have considering a linear mobility model for UEs, HF probabilities for macro cell and pico cell UEs are derived in closed form for various scenarios. The analysis is then extended to fast-fading and shadowing scenarios: relevant statistics in a fading scenario are extracted from a 3GPP compliant system level simulator to facilitate semi-analytic expressions for HF probabilities.

Reinforcement learning approach with entropy measures is being used in this thesis work, to effectively handover the UE from macro cell to pico cell and vice-versa. In this work velocity, range expansion bias, and entropy is taken as a primary features to offload / evict the traffic from macro cell to pico cell and vice-versa. The contributive approaches being made are as follows:

- The work mainly focuses on context-aware scheduling and load balancing solutions. The load balancing methods uses entropy and REB in a HetNet scenario, while in context-aware scheduling the UE association process is solved.
- To implement the load balancing method, two reinforcement learning MM approaches; MAB and Bayesian Bandit have been proposed.

Chapter 3

Related Theory

3.1 Reinforcement Learning

Reinforcement learning, like many topics whose names end with “ing”, such as machine learning and mountaineering, is simultaneously a problem, a class of solution methods that work well on the class of problems, and the field that studies these problems and their solution methods. Reinforcement learning problems [24] involve learning what to do—how to map situations to actions—so as to maximize a numerical reward signal. In an essential way these are closed-loop problems because the learning system’s actions influence its later inputs. Moreover, the learner is not told which actions to take, as in many forms of machine learning, but instead must discover which actions yield the most reward by trying them out. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These three characteristics—being closed-loop in an essential way, not having direct instructions as to what actions to take, and what are the consequences of actions, including reward signals, play out over extended time periods—are the three most important distinguishing features of the reinforcement learning problem.

Online learning [25] is one of the most important characteristics of RL. Some examples of RL are as: fly stunt maneuvers in a helicopter, defeat the world champion at backgammon, control power station, make a humanoid robot walk, and play many different Atari games better than humans.

3.1.1 The Agent-Environment Interface

The reinforcement learning problem is meant to be a straightforward framing of the problem of learning from interaction to achieve a goal. The learner and decision maker is called the agent. The thing it interacts with, comprising everything outside the agent, is called the environment. These interact continually, the agent selecting actions and the environment responding to those actions and presenting new situations to the agent. The environment also gives rise to rewards, special numerical values that the agent tries to maximize over time. A complete specification of

an environment, including how rewards are determined, defines a task, one instance of the reinforcement learning problem.

Generally the agent and environment interact [26] at each of a sequence of discrete time steps, t = 0, 1, 2, 3, … At each time step t, the agent receives some representation of the environment's state, $S_t \in S$, where S is the set of possible states, and on that basis selects an action, $A_t \in A(S_t)$, where $A(S_t)$ is the set of actions available in state S_t. One time step later, in part as a consequence of its action, the agent receives a numerical reward, $R_{t+1} \in R$, and finds itself in a new state, S_{t+1}. Figure 3 shows the agent–environment interaction.

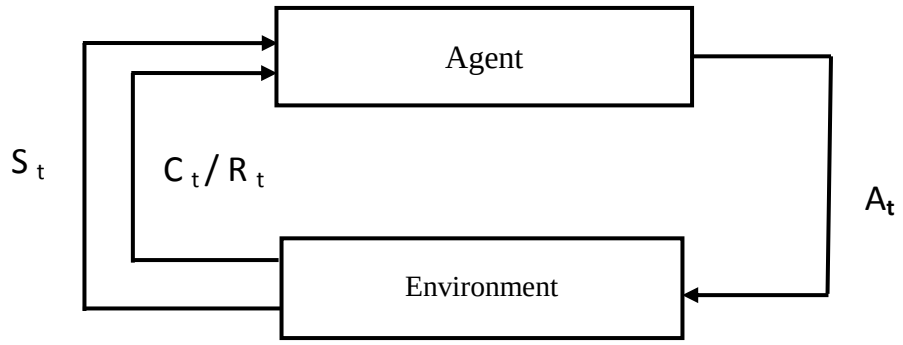


Figure 3: Interaction between agent and environment.

RL has two important components, policy and value function. Policy defines the action of agents at each step, in other words, policy is the mapping from observed state to an action that should be taken. It is expressed as a simple function, a look-up table, or other cases that need more exploration. Policy itself is enough to decide the action of agents. It is represented as a probability π (s, a) of selecting action "a" at state "s". To calculate the policy means to decide π (s, a) of all available actions at every state. The agent's goal is to maximize the total amount of reward it receives over the long run.

Policy π is a mapping from each state s and action "a" to the probability π (s, a) of taking action a when in state "s". Informally, the value of a state "s" under a policy π, denoted by $V^{\pi}(s)$, is the expected return when starting in s and following π. $V^{\pi}(s)$ can be defined formally as

$$V^{\pi}(s) = E_{\pi}\{\sum_{t=0}^{\infty} \gamma^t c_t | s_0 = s\} \quad (1)$$

Where $E_\pi\{.\}$ denotes the expected value given that the agent follows policy π. Note that if the terminal state exists, its value is always zero. The function V^π is referred to as the state-value function for policy π.

3.2 Multi Armed Bandit Algorithm

The stochastic multi - armed bandit problems have been introduced by Robbins and has been used extensively to model the trade-offs faced by an automated agent whose aim is to gain new knowledge by exploring its environment and to exploit its current, reliable knowledge. Such problem arises frequently in practice, for example in the context of clinical trials or on-line advertising. The multi-armed bandit problem offers a very clean, simple theoretical formulation for analyzing trade - offs between exploration and exploitation. Some applications of Multi-Armed Bandit problem are: Internet display advertising: companies have a suite of potential ads they can display to visitors, but the company is not sure which ad strategy to follow to maximize sales. Ecology: animals have a finite amount of energy to expend, and following certain behaviors has certain rewards. Finance: while stock option gives the highest return, under time-varying return profiles. Clinical trials: a researcher would like to find the best treatment, out of many possible treatments, while minimizing losses.

In its simplest formulation (generally referred to as stochastic), a bandit [27] problem consists of a set of K probability distributions ($D_1, \ldots, D_K$) with associated expected values ($\mu_1, \ldots, \mu_K$) and variances ($\sigma^2_1, \ldots, \sigma^2_K$). Initially, the D_i is unknown to the player. In fact, these distributions are generally interpreted as corresponding to arms on a slot machine; the player is viewed as a gambler whose goal is to collect as much money as possible by pulling these arms over many turns. At each turn, $t = 1, 2 \ldots$ the player selects an arm, with index $j(t)$, and receives a reward $r(t) \sim D_{j(t)}$. The player has a two-fold goal: on one hand, finding out which distribution has the highest expected value; on the other hand, gaining as much rewards as possible while playing. Bandit algorithms specify a strategy by which the player should choose an arm $j(t)$ at each turn.

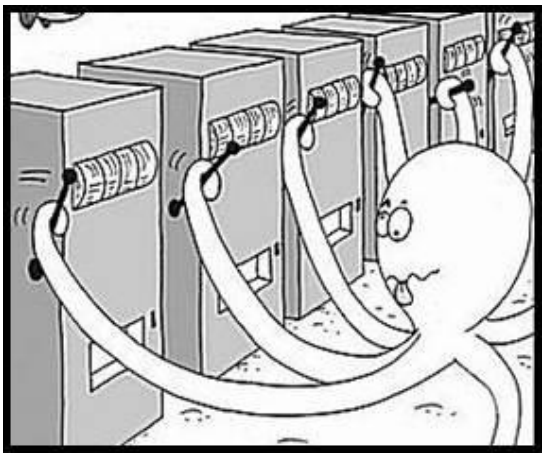


Figure 4: Greedy Octopus pulling N slot machine once in a random order.

Figure 4 illustrates a goofy looking octopus playing multi-armed bandit problem which involves a stylized casino with K one – armed bandit slot machines.

The most popular performance measure for bandit algorithm is the total expected regret, defined for any fixed turn T as:

$$R_T = T\mu^* - \sum_{t=1}^{T} \mu_{j(t)} \quad (2)$$

Where $\mu^* = \max_{i = 1,\dots k} \mu_i$ is the expected reward from the best arm.

Alternatively, the total expected regret can be expressed as

$$R_T = T\mu^* - \mu_{j(t)} \sum_{k=1}^{K} E(T_k(T)) \quad (3)$$

Where $T_k(T)$ is a random variable denoting the number of plays of arm k during the first T turns.

A classical result of Lai and Robbins states that for any suboptimal arm k,

$$E(T_k(T)) \geq \frac{\ln T}{D(p_k \,||\, p^*)} \quad (4)$$

Where $D(p_j \,||\, p^*)$ is the Kullback - Leibler divergence between the reward density p_k of the suboptimal arm and the reward density p* of the optimal arm, defined formally as

$$D(p_k \| p^*) = \int p_j \ln \frac{p_j}{p^*} \quad (5)$$

Regret thus grows at least logarithmically, or more formally, $R_T = \Omega(\log T)$. An algorithm is said to solve the multi – armed bandit problem if it can match this lower bound, that is if $R_T = O(\log T)$.

3.2.1 ∈-greedy

The ∈-greedy algorithm is widely used because it is very simple, and has obvious generalizations for sequential decision problems. At each round t = 1, 2 … the algorithm selects the arm with the highest empirical mean with probability 1- ∈, and selects a random arm with probability ∈. In other words, given initial empirical means $\hat{\mu}_1(0) \ldots \hat{\mu}_K(0)$,

$$P_i(t+1) = \begin{cases} 1 - \in + \in/k & \text{if } i = \arg\max_{j=1,\ldots,K} \hat{\mu}_j(t) \\ \in/k & \text{otherwise.} \end{cases} \quad (6)$$

If ∈ is held constant, only a linear bound on the expected regret can be achieved.

3.2.2 Upper Confidence Bounds (UCB)

The simplest algorithm, UCB, maintains the number of times that each arm has been played, denoted by $n_i(t)$, in addition to the empirical means. Initially, each arm is played once. Afterwards, at round t, the algorithm greedily picks the arm j (t) as follows:

$$j(t) = \arg\max_{(i=1\ldots k)} \left(\hat{\mu}_i + \sqrt{\frac{2 \ln t}{n_i}} \right) \quad (7)$$

At turn t, the expected regret of UCB is bounded by:

$$8 \sum_{i:\, \mu_i < \mu^*} \frac{\ln t}{\Delta_i} + \left(1 + \frac{\pi^2}{3}\right) \sum_{i=1}^{k} \Delta_i \quad (8)$$

$\Delta i = \mu^* - \mu_i$. This O (log n) bound on the regret matches a well-known Ω (log n) bound by Lai and Robbins. Hence UCB achieves the optimal regret up to a multiplicative constant, and is said to solve the multi-armed bandit problem.

The main feature of another algorithm, UCB1-Tuned is that it takes into account the variance of each arm and not only its empirical mean. More specifically, at turns t =1, 2,… the algorithm picks an arm j (t) as

$$j(t) = \arg\max_{i=1\ldots k}\left(\hat{\mu}_i + \sqrt{\frac{\ln t}{n_i}\min\left(\frac{1}{4}, V_i(n_i)\right)} \right) \quad (9)$$

Where,

$$V_i(t) = \hat{\sigma}^2_i(t) + \sqrt{\frac{2\ln t}{n_i(t)}} \quad (10)$$

The estimate of the variance $\hat{\sigma}^2_i(t)$ can be computed as usual by maintaining the empirical sum of squares of the reward, in addition to the empirical mean.

3.2.3 Thompson Sampling

Let K denote the number of arms, and π denote the prior distribution over $\mu = \{\mu_1 \ldots \mu_K\}$, where μ_i is the expectation of arm i's reward. Suppose up to time step t, the agent has chosen action i for $\tau_{i,t}$ times, and received rewards $X_i(t) = \{x_{i,1} \ldots x_{i,t}\}$. Let $X(t) = \{X_1(t) \ldots X_K(t)\}$ be all the observed rewards until time step t.

Thompson sampling [28] selects each arm randomly according to its (posterior) probability to be optimal, which is

$$\forall_i \in [K], P_i(t) := P(\mu_i = \max_j \mu_j \mid X(t)) \quad (11)$$

And $[K] = \{1, \ldots, K\}$ denotes the set of integers from 1 to K. Previous implementation of TS is outlined in Algorithm 1, in which the Lines 6 and 7 essentially draw sample from $P_i(t)$. As seen in Line 6 of Algorithm. 1, this implementation requires inferring the posterior distribution of the mean rewards, which is efficient if the prior is conjugate. However, in practice non-conjugate priors are more flexible in many situations such as where the arms are not independent.

For a non-conjugate prior, one possible solution is to approximate the intractable posterior with a sequential Monte Carlo (SMC) sampler.

Algorithm 1: Thompson sampling

1: Input: Prior distribution π.

2: t = 0.

3: Maintain sets: $X_i = \Phi$, $\forall_i \in [K]$, $X = \{X_1, \ldots, X_K\}$.

4: **while** t < T **do**

5: t = t+1.

6: Draw samples $\mu \sim P(\mu|X)$.

7: $I_t = \arg\max_i \mu_i$.

8: Take action I_t, and receive reward x_t, $X_{it} \cup \{x_t\}$.

9: **end while**

3.2.4 Bayesian Bandit

Assumptions about the reward distribution R has not been made so far, except bounds on rewards. Bayesian Bandit [29] exploits prior knowledge of rewards, p[R]. It computes posterior distribution of rewards $p[R \mid h_t]$, where $h_t = a_1, r_1 \ldots a_{t-1}, r_{t-1}$ is the history. Better performance is achieved if prior knowledge is accurate. It uses posterior to guide exploration – Upper Confidence Bound (Bayesian UCB) and Probability matching (Thompson Sampling).

In the Bayesian UCB, independent Gaussians distributions for reward distribution is considered,

$$R_a(R) = N(r;\mu_a, \sigma_a^2) \tag{12}$$

The algorithm computes Gaussian posterior over μ_a, σ_a^2 (by Bayes law) by

$$p[\mu_a, \sigma_a^2 \mid h_t] \propto p[\mu_a, \sigma_a^2] \prod_{t|a_t=a} N(r_t; \mu_a, \sigma_a^2) \tag{13}$$

Finally, the algorithm picks the action that maximizes standard deviation of all the Gaussian curves.

$$a_t = \arg\max \mu_a + c\sigma_a/\sqrt{N(a)} \tag{14}$$

Mathematically:

Conjugate Prior for Binomial

X | p ~ Binomial (n, p), p ~ Beta (a, b) "Random Variable / Prior"

Find Posterior distribution

$$F(p \mid X=k) = \frac{P(X=k \mid P) f(P)}{P(X=k)} \quad (15)$$

$$= \frac{(n,k) P^k (1-P)^{n-k} \, C \, p^{a-1} (1-p)^{b-1}}{P(X=k)} \quad (16)$$

$$F(p \mid X=k) \propto p^{a+k-1}(1-p)^{b+n-k-1} \quad (17)$$

Therefore $p \mid X \sim \text{Beta}(a + X, b + n - X)$ (18)

Initially Beta distribution [30] was considered, then prior was calculated. Using Beta distribution and priori value, posterior distribution is computed. Still, Beta distribution is achieved. Hence "a Beta prior with Binomial observations creates a Beta posterior".

Beta prior with Binomial data implies a Beta posterior. Suppose X comes from, a well-known distribution, call it f_{α}, where α is possibly unknown parameters of f. f could be a Normal distribution, or Binomial distribution, etc. For particular distribution f_{α}, there may exist a prior distribution P_{β}, such that:

$$P_{\beta}(\text{prior}) * f_{\alpha}(X) \text{ (data)} = P_{\beta'} \text{ (posterior)} \quad (19)$$

Where, β' is a different set of parameters but P is the same distribution as the prior. A prior P that satisfies this relationship is called a conjugate prior. They are useful computationally, and can avoid approximate inference using MCMC and go directly to the posterior. Conjugate priors are only useful for their mathematical convenience: it is simple to go from prior to posterior.

3.3 Exploration – Exploitation Tradeoff

The notions of exploration and exploitation spreads in the research field of optimization and computational intelligence for decades, exploration and exploitation are mathematically defined [31] as follows:

Cited definition 1 – exploration (er): A sampling behavior is exploration iff its sampling point is generated independently of the information acquired by historical sampling points. Mathematically, a sampling X_n at time n is exploration iff

$$P(X_n = x \mid X_{n-1} \dots X_0) = P(X_n = x) \tag{20}$$

Cited definition 2 – exploitation (ei): A sampling behavior is exploitation iff the generation of its sampling point depends upon the information acquired by historical sampling points. Mathematically, a sampling X_n at time n is exploitation iff

$$P(X_n = x \mid X_{n-1} \dots X_0) \neq P(X_n = x) \tag{21}$$

3.4 Entropy

It is a measure of unpredictability of information content. The measure of information entropy associated with each possible data value is the negative logarithm of the probability mass function for the value. Thus, when the data source has a lower-probability value (i.e., when a low-probability event occurs), the event carries more "information" (" surprising ") than when the source data has a higher-probability value. The amount of information conveyed by each event defined in this way becomes a random variable whose expected value is the information entropy. Generally, entropy refers to disorder or uncertainty, and the definition of entropy used in information theory is directly analogous to the definition used in statistical thermodynamics. The entropy can be explicitly written as

$$H(X) = \sum_i P(x_i) I(x_i) = -\sum_i P(x_i) \log_b P(x_i) \tag{22}$$

Where b is the base of the logarithm used. Common values of "b" are 2, Euler's number e, and 10 and the unit of entropy is bit for b = 2, nat for b = e, and Hartley for b = 10. Minimum entropy will occur when certain event will happen with probability 1 as to what the source will do (i.e. no uncertainty). This can only happen if one of the q symbols always occurs with

probability 1 while all the other symbols occur with probability 0 (i.e. they never occur). Since maximum entropy represents the opposite extreme to minimum entropy and can be thought of as the case when the maximum amount of uncertainty as to what the source will do next, it is intuitively obvious that this should occur when all source symbols are equally likely. Furthermore with q equally likely symbols it is expected to require no less than log q bits to represent each symbol. Maximum entropy occurs for the special case of equi probable symbols and $H(S) = \log q$, bits is the maximum entropy.

$$0 \leq H(S) \leq \log q \tag{23}$$

3.5 Heterogeneous Networks Concepts

HetNets, as mentioned formerly, have been introduced in the LTE-Advanced standardization in order to provide a significant network performance leap when other advanced technologies (CA, MIMO, and CoMP) are unable to achieve that, as they are reaching theoretical limits. Such techniques may not always work well either, especially under low SINR conditions, where received powers are low due to attenuation and/or interference might be high, whereas HetNets can do.

Complementing macro cells with LPNs and dedicated indoor solutions based on the 3GPP standard is a good approach to meet the predicted requirements for higher data rates and additional capacity. This approach can include the use of pico cells, femto cells, relays and remote radio units (RRUs), which delivers high per-user capacity and rate coverage in areas covered by LPNs, with the potential to improve performance in the macro network by offloading traffic generated in hotspots. By adding LPNs to the existing macro layer, the operator creates a two-layer cell structure with eNodeBs of different types that is why it is called HetNet [32], heterogeneous in the deployment sense. The degree of integration that can be achieved throughout the HetNet will determine the overall network performance.

HetNets improve the overall capacity as well as provide a cost-effective coverage extension and higher data rates to hot spots such as airports and shopping malls by deploying additional network nodes within the local-area range. In addition, they also increase overall cell-site performance and cell-edge data rates by bringing the network closer to end users. In this way,

radio link quality can be enhanced due to the reduced distance between transmitter and receiver, and the larger number of eNodeBs allows for more efficient spectrum reuse and therefore larger data rates [33].

These LPNs can be either operator deployed or user deployed, share the same spectrum, and may coexist in the same geographical area. Table 2 shows specifications of different elements in HetNets according to typical transmit power, coverage area, typical backhaul features and access [34].

In HetNets, the coordination between macro cells and small cells has a positive impact on the performance of the radio network and consequently on the overall user experience. Coordinated embedded LPNs improve performance, increasing both network data capacity and throughput without the need to split the available spectrum. Coordinating features like joint transmission and reception provides the user with significantly higher speeds than would be possible with separate, uncoordinated, macro and LPNs layers [35].

Table 2: Specification of different elements in HetNets

Type of Node	Typical transmit Power	Coverage	Typical backhaul features	Access
Macro cell	46 dBm	Few Km	S1 interface	Open to all UEs
Pico cell	23-30 dBm	< 100 m	X2 interface with macro	Open to all UEs
Femto cell	< 23 dBm	< 50 m	User local loop	CSG
Relay	30 dBm	300m	Wireless link with donor	Open to all UEs
RRU	46 dBm	Few Km	Fiber link with parent site	Open to all UEs

3.6 Heterogeneous Networks Deployment Challenges

Frequency allocation, backhauling, handover, self-organization and interference reduction are considered the key deployment challenges of HetNets, which are discussed in this section.

3.6.1 Frequency Allocation

Frequency allocation is an essential issue in the HetNets deployment and it should be considered carefully. As the radio spectrum is a scarce resource, it is desirable that macro cells and LPNs will entirely share the same frequency band.

Given that different capacities might be required in the different coverage areas, it is possible to use just a partial spectrum between macro cells and LPNs that are assigned to a part of the whole frequency resource.

3.6.2 Backhauling

Backhauling will be tricky part in the HetNets deployment because of the complex topology of the various types of LPNs deployed along with the macro cells. For instance, the availability of power and network backhauling of pico cells might be difficult and expensive. Conversely, femto cells have lower backhauling costs compared to pico cells, but difficulties in maintaining QoS appear because they are unplanned and so interference might be more difficult to control. Some LPNs may have their own connections to the core network, whereas some other nodes may construct a cluster to concatenate and route the traffic to the core network, and other nodes may relay on relays as an alternative route option.

3.6.3 Handover

Handovers are necessary in order to provide a non-intermittent uniform service when users are moving around different cell coverage. Furthermore, handovers are a means to offload the traffic from highly congested cells by shifting users at the border to the less congested neighbor nodes. However, the situation is different in HetNets due to the large number of small cells and the different types of backhaul links for each type of cell [36]. In addition, the probability of handover failures increases the probability of user outage. For that, the handover parameter configuration for LPNs needs to be carefully planned and probably different from that of the macro cells.

3.6.4 Self-organizing

Self-organizing Networks (SONs) are a step forward towards automated operation in mobile networks, which reduce the Operation and Maintenance (O&M) cost of mobile networks by

using automated and intelligent procedures to replace human intervention without compromising network performance. Some LPNs such as femto cells are user deployable and no cellular operator intervention is needed, this approach is conceptualized by SONs features [36]. SON features of HetNets can be categorized in to three processes:

- **Self-Configuration**, newly deployed cells download required software and self-configure automatically before entering them into the operational mode.
- **Self-healing**, where cells are auto-recovered whenever failures occur.
- **Self-optimization,** where cells monitor the network status and adapt their settings to improve performance and reduce interference.

3.6.5 Interference

The simultaneous use of the same spectrum between different cell layers that run on different values of transmit power creates interference that will become more severe compared to homogeneous networks. For pico cells the interference does not create coverage hole due to open access to all UEs, but that is not true when it is expanded (that will be discussed later in the range expansion and interference coordination sections).

The situation is different for femto cell due to being equipped with the CSG features that result into new and severe interference conditions. Figure 5 illustrates interference scenarios in relation to femto cell deployments. There are two scenarios that create severe interference when macro UE (MUE) does not belong to femto cell CSG and being close to it [37]. In DL, MUE is being jammed by femto cell. Frequency of the occurrence of this issue can be reduced by femto cell power control, with or without macro UE assistance, but it cannot completely solve the problem [38]. In UL, femto cell is being jammed by MUE since MUE is power controlled by the macro cell MUE will cause likely strong bursty interference to femto cell. Noise padding technique which is a method of wireless communication includes detecting UL interference in a received uplink transmission of a UE, where the received UL transmission is padded with noise based on the detected interference and also based on a frequency domain partition [39]. This technique can smooth out interference in this case, but it also decreases capacity at serving femto cells and increases interference to the neighboring cells. In case the MUE is closer to

femto cell that the UE that is served by femto cell, noise padding cannot solve the problem and the UE served by femto cell would experience outage [40].

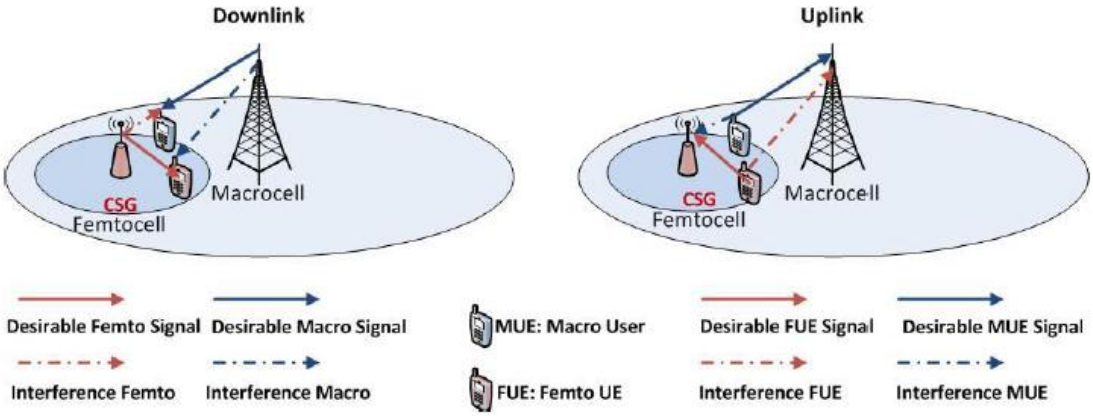


Figure 5: Macro - Femto Interference Scenarios [Source [41]].

3.7 Cell Range Expansion

Cell range expansion (CRE) is a technique to expand a pico cell range virtually by adding a bias value to the pico received power, instead of increasing transmit power of pico base station(PBS), so that coverage, cell-edge throughput, and overall network throughput are improved. Many studies have focused on inter-cell interference coordination (ICIC) in CRE, because macro base station's (MBS's) strong transmit power harms the expanded region (ER) user equipments (UEs) that select PBSs by bias value. Optimal bias value that minimizes the number of outage UEs depends on several factors such as dividing ratio of radio resources between MBSs and PBSs. In addition it varies from UE to another. Most articles use the common bias value among all UEs determined by trial-and-error method. In this thesis work, Multi Armed Bandit / Bayesian Bandit algorithm is proposed to determine the bias value of each UE where each UE predicts its bias value from linear regression model that minimizes the number of outage UEs during handover.

LPNs coverage is quite limited by its transmission power and the strong interference from macro cells, which means that only a small percentage of users can benefit from LPN deployment in cell edges where there are no many UEs. This leads to a state of coverage unbalance and for that, a new technique is required to increase HetNets efficiency, offload the more macro cell traffic, i.e. attract more UEs to LPNs and solve the UL and DL coverage

unbalance. Moreover, the performance of LPNs is significantly improved if UEs are allowed to connect to a weaker SINR LPNs, which refers to extend LPNs boundaries for load balancing purposes. This improves LPNs performance, since more UEs can connect to LPNs and take advantage of the spectrum offered by them, and multiple LPNs can reuse the disused resources on the macro side, allowing for cell-splitting gains.

For these purposes, RE is introduced for LPNs, pico cells, with a positive bias to them in the cell selection. RE is considered a key design feature to enhance HetNets efficiency, which adds an offset to the pico cell received signal strength (RSS) in order to increase its DL coverage footprint.

Figure 6 illustrates the RE concept as follows. The natural LPN boundaries in DL and UL are different in HetNets, as opposed to a homogeneous and correctly planned network. In the DL, the DL SINRs observed from the macro cell and the pico cell are equivalent at a location that is closer to the pico cell, which forms the equal-SINR cell boundary. In the UL, on the other hand, the location of the natural cell boundary is where the path loss to the macro cell and the pico cell are equivalent. This is due to the fact that macro and LPN can reach different maximum power levels, however the UE has the same maximum power for both cases.

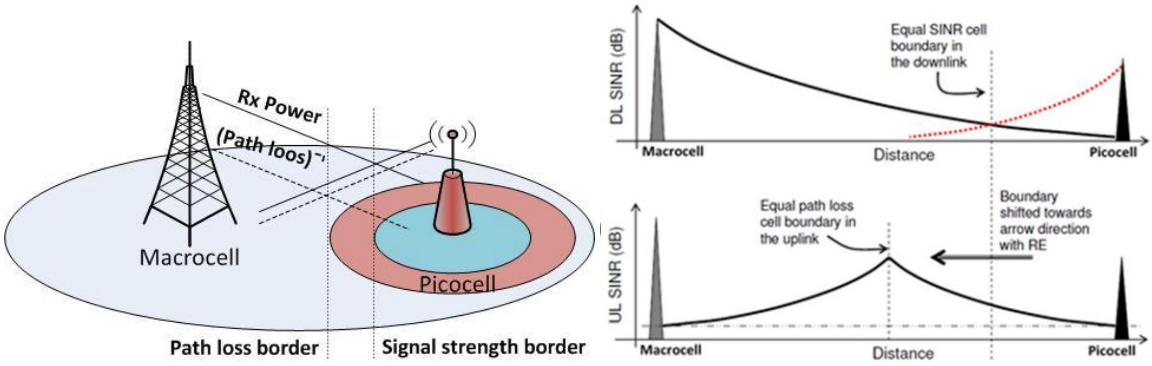


Figure 6: Range Expansion Concept [Source [41]].

In the normal case without RE, the serving cell choice is determined by the highest DL received power, this technique is referred as maximum reference signal received power (MAX RSRP). With RE the serving cell of a UE is selected from the set of neighbor cells Δ according to the rule given as:

$$\text{Serving Cell} = \underset{i \in \Delta}{\text{argmax}} \; (\text{RSRP}_i + \text{Bias}_i) \tag{24}$$

Where RSRP and Bias are expressed in dB, this rule implies that a UE does not necessarily connect to the eNodeB that has the strongest DL received power.

As mentioned, the LPN boundaries in DL and UL are different in HetNets. For that, the best UL cell does not necessarily correspond to the best DL cell. With RE technique, the DL serving cell is determined based on the equation 24, whereas the UL serving cell is defined according to the minimum path loss. The following Figure 7 shows handover in HetNets with RE.

Even though RE significantly mitigates cross-tier interference in the UL, this comes at the expense of reducing the DL signal quality of those users in the expanded region. Such users may suffer from DL SINRs below 0 dB since they are connected to cells that do not provide the best DL RSS, for this reason interference coordination strategies may well help to solve this tradeoff and reduce DL degradation in the cell border. Thus, it is usual to find that RE is jointly designed with ICIC / eICIC schemes.

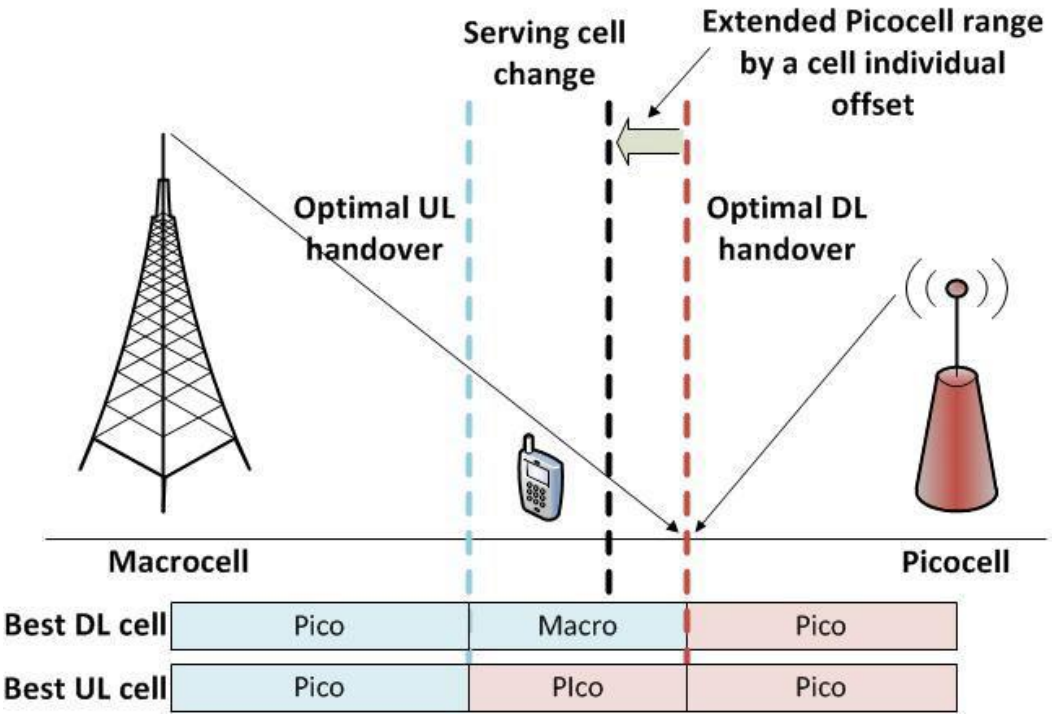


Figure 7: Handover in HetNets with RE [Source [41]].

3.8 Shadowing (Large-scale Fading)

Shadowing is caused by obstacles in the propagation path between the UE and the eNodeB and can be interpreted as the path loss variations induced by irregularities of the geographical characteristics of the terrain with respect to the average path loss obtained from the path loss model.

Shadowing is modeled as a random variable that is added to the propagation path loss. Shadowing value changes as UE changes its position where shadowing is not distance-dependent it is position-dependent. This fact means that, shadowing should not be modeled independently for each UE in the simulation since it is common to have several UEs operating within limited area, which implies proximity between UEs. It is typically approximated by a log-normal distribution of zero mean and standard deviation equal to σ dB, as shadowing effects occur over a large area, in order to be able to capture the dynamics affecting macro-cell diversity in a realistic way a two-dimensional Gaussian process with appropriate spatial correlation is desirable.

3.9 Small-scale Fading (Fast Fading)

Small-scale fading refers to the dramatic changes in signal amplitude and phase that can be experienced as a result of small changes (as small as half a wavelength) in the spatial separation between a receiver and transmitter. It is caused by two main factors. First, the motion between the transmitter and receiver which results in the appearance of Doppler effect and so a parasitic frequency modulation, which is known to be a time-variant mechanism due to motion. Frequency spread due to Doppler translates into time selective fading, which can be categorized as fast fading or slow fading. Second, the multiple paths in the radio signal, which is known to be a time-spreading mechanism, its fading can be categorized to frequency-selective fading and flat fading.

Thus, the final channel response is a combination of several contributions, propagation path loss, shadowing (large-scale fading) and fast fading (small-scale fading).

Chapter 4

Methodology

A qualitative research based on the calculation of different attributes of alternative in "handover decision algorithm" is carried out by using Python and NS3 simulator. Various parameters are fitted into the Multi-Armed Bandit Based Learning algorithm which uses entropy as deciding factor for load balancing and final outputs are analyzed and the UE is handover to the network that have optimum network throughput, less latency and jitter.

4.1 Proposed Framework

The proposed framework is illustrated in Figure 8. User equipment senses the eNodeB channel and it reports the RSRP values to the eNodeB's, it analyzes the velocity of the user equipment. After, analysis of velocity of different UE the eNodeB categorizes the UE based on the analyzed velocity report. The resource block will be allocated to those UE that has least velocity. UE calculates the viable prior probability depending upon direction cosine and distance from neighboring eNodeB's. These prior probabilities are processed by the Bayesian Multi Armed Bandit to produce independent identically distributed (IID) probability function which resembles the reward accumulated during different plays between the agent (UE) and environment (eNodeB). The UE learns and finds the highest accumulated reward distribution and it selects it as the best eNodeB. Then, Range Expansion Bias is calculate through generalized linear regression model which enhances the existing coverage of the conventional pico cell. Informational entropy is calculated from the posterior distribution. The user equipment's are handed over to the eNodeB that has highest accumulated reward / posterior probability distribution and the channel that has least entropy. Finally, the results are analyzed.

4.2 Model network

A model network with four macro cell – three sectors per macro cell and, two pico cells per sector of the macro cell and a UE moving in different trajectory is considered as shown in Figure 9 each macro cell and pico cell has different values for attributes such as cost, packet lost, delay, and received level, velocity of UE and data throughput. The calculated value for the different parameters such as RSRP and velocity of the UE determines whether it is better to connect the UE to MBS or PBS. This decision is made by integrating the simulated parameter

values obtained from the NS3 to the reinforcement learning algorithm which gives two states whether to connect to the neighboring MBS or PBS.

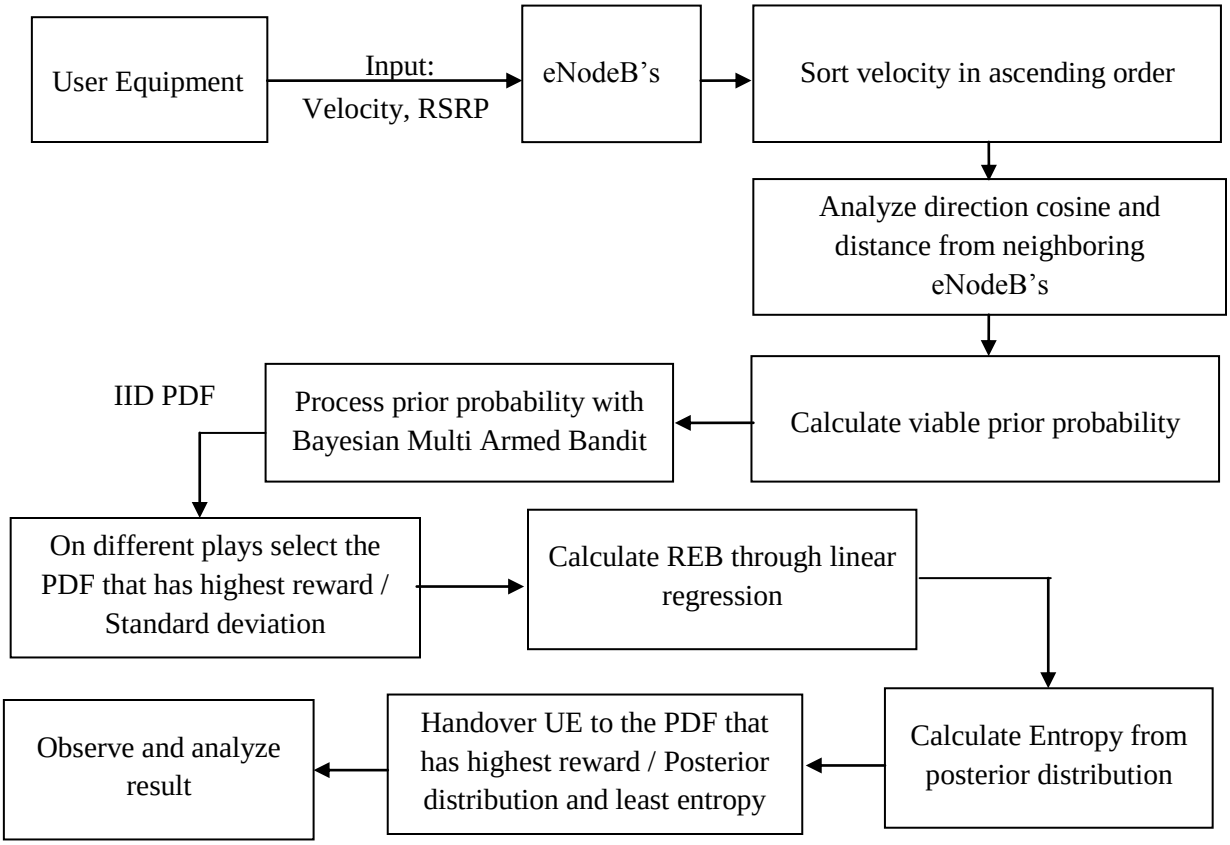


Figure 8: Proposed framework for handover using reinforcement learning approach.

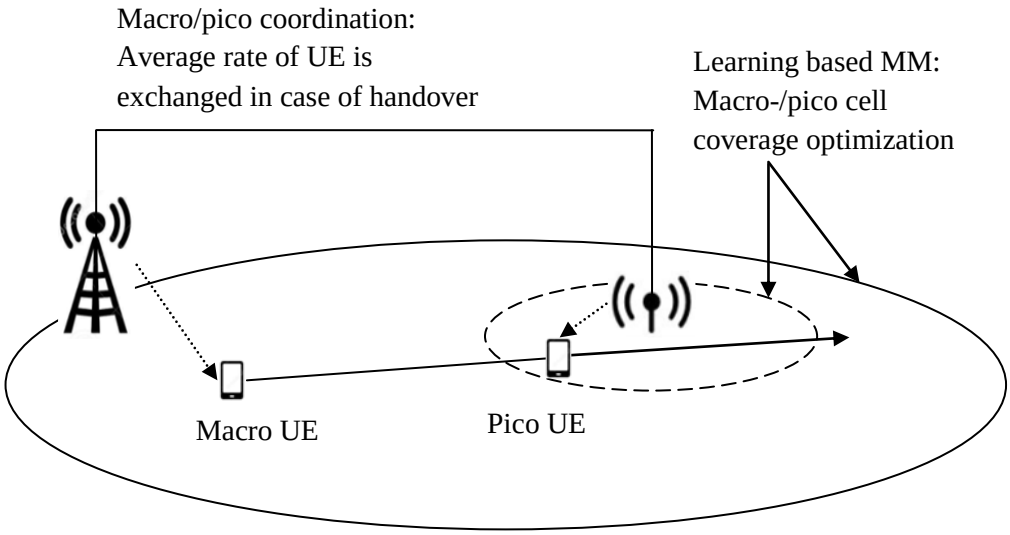


Figure 9: Model Network based on velocity calculation and history based scheduling

4.3 Data Collection:

Macro eNodeB and Pico eNodeB is used for analyzing handover in this thesis work. These two eNodeB have their attribute with own significant values. In this work creation and simulation of the eNodeB is carried out in the NS3 software, and as a result some values are achieved. Consequently these values are used for evaluating the handover process in the heterogeneous networks.

Table 3: System Parameters for simulations

Parameter	**Value**
Center Frequency	2GHz
Bandwidth	10MHz
Number of cells	4 macro cells, 3 sectors per macro-cell, 2 low power node-cells per cell
Inter site (PBS distance)	500m
Macro cell coverage radius	1Km
Pico cell coverage radius	150m
MBS Tx power	46dBm
PBS Tx power	30dBm
Macro Path loss model	$128.1 + 37.6 \log_{10}(R)$ dB (R(km))
Pico Path loss model	$140.7 + 36.7 \log_{10}(R)$ dB (R(km))
Measurement interval	40ms
TTT	160ms
Handover preparation delay	50ms
Handover execution time	40ms
UE speed	3, 30, 60, 120 and 150Km/hr
Number of UEs	50 per speed

Table 4: Bernoulli distribution of UE latching among the 10 eNodeB [Source https://www.superdatascience.com/machine-learning]

eNodeB 1	eNodeB 2	eNodeB 3	eNodeB 4	eNodeB 5	eNodeB 6	eNodeB 7	eNodeB 8	eNodeB 9	eNodeB 10
1	0	0	0	1	0	0	0	1	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
1	1	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0

As, seen from Table 4. Excel sheet at time instant t = 0, UE will try to access resource of eNodeB1, eNodeB5 and eNodeB9 simultaneously. Similarly, at time instant t = 1, UE will try to access resource of eNodeB9 only.

4.4 The Beta Distribution

The Beta distribution is very useful tool in Bayesian statistics. A random variable X has a Beta distribution, with parameters (α, β), its density function is:

$$F_x(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (25)$$

Where, B is the Beta function. The random variable X is only allowed in [0, 1], making the Beta distribution a popular distribution for decimal values, probabilities and proportions. The value of α and β, both positive values, provide great flexibility in the shape of the distribution.

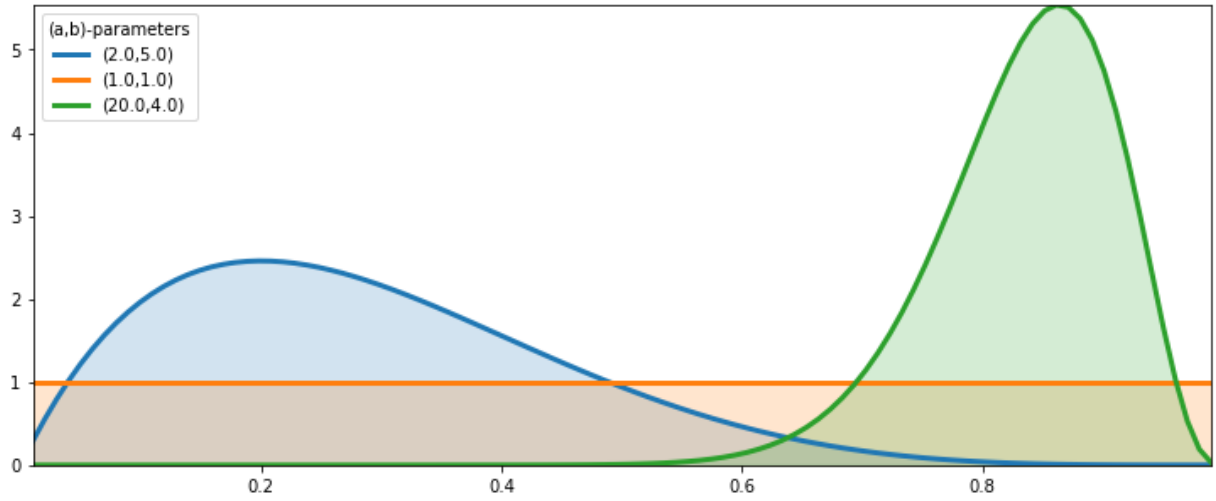


Figure 10: Beta distribution for different values of α and β.

In Figure 10, parameter (1, 1) presents a flat distribution, which is a Uniform distribution. Hence the Beta distribution is a generalized form of the Uniform distribution.

There is an interesting connection between the Beta distribution and the Binomial distribution. Suppose we are interested in some unknown probability p. We assign a Beta (α, β) prior to p. Then we observe some data generated by a Binomial process, let's say X ~ Binomial (N, p), with p still unknown. Then our posterior is again a Beta distribution, i.e., p | X ~ Beta (α + X, β + N - X). Succinctly, one can relate the two by "a Beta prior with Binomial observations creates a Beta posterior". This is very useful property, both computationally and heuristically.

4.5 Linear Regression to Predict REB

The proposed scheme relies on linear regression to extrapolate coverage area such as total number of UEs is close to optimal and in congruence to the target load. The linear regression model predicts the REB value to handover more UEs from MBS to PBS. Every PBS periodically collects the UE measurement from all the associated active and idle UE. The report contains RSRP and other signal quality measurement values. In this work linear model is being used to improve execution speed. However, higher degree polynomial could be used for curve fitting purpose which results in higher accuracy but it may give rise to serious problem like over-fitting. So, simple and generalized linear regression model is used to estimate the bias value. Following normal equation is used to calculate the co-efficient vector

$$\theta = (X^T X)^{-1} X^T Y \quad (26)$$

Now given a new feature element x and co-efficient vector θ the target value y can be computed as follows:

$$y = \theta^T * x \quad (27)$$

Due to asymmetric nature of the equal received signal strength boundary, the effect of REB is not uniform in all directions around the PBS. This can be seen in Figure 6, where at the same distance from the PBS, REB is needed to extend coverage of UEs in the direction of the MBS, whereas on the opposite side the UEs are within coverage of PBS even without REB.

4.6 Signal to Interference plus Noise Ratio (SINR)

Based on long term parameters, the power received by a point P on 2D network layout from a given eNodeB in dB is:

$$P_{Rx}(\text{eNodeB}) = P_{tx}(\text{eNodeB}) + G_{eNodeB}(\theta) + G_{UE} - PL(R) - P_{Shad} \quad (28)$$

Where, P_{tx} (eNodeB) denotes the eNodeB transmit power in dB, G_{eNodeB} (θ) denotes the eNodeB antenna gain in dBi, G_{UE} denotes UE antenna gain in dB which is equal to zero and P_{Shad} denotes the power loss due to obstacles.

Now, SINR can be computed for the whole points in the network layout, as follows:

$$\text{SINR} = \frac{P_{Rx}(\text{eNodeB})}{\sum_{interferers} P_{Rx}(\text{eNodeB}) + P_{therm}} \quad (29)$$

Where, P_{therm} is the thermal noise in dB. SINR values are saved as a SINR 2D map. Figure 11 illustrates the 2D SINR map that is produced after using previous equation. The Figure 11(a),

shows 2D SINR map without shadowing not wrap-around, as seen in the Figure 11 the highest SINR values are located close to macro cells in their antennas directions and reach 20 dB, these values decrease gradually away from eNodeBs and reach the lowest value which is -5 dB at sector's borders. The Figure 11(b) shows 2D SINR map after applying wrap - around and adding shadowing.

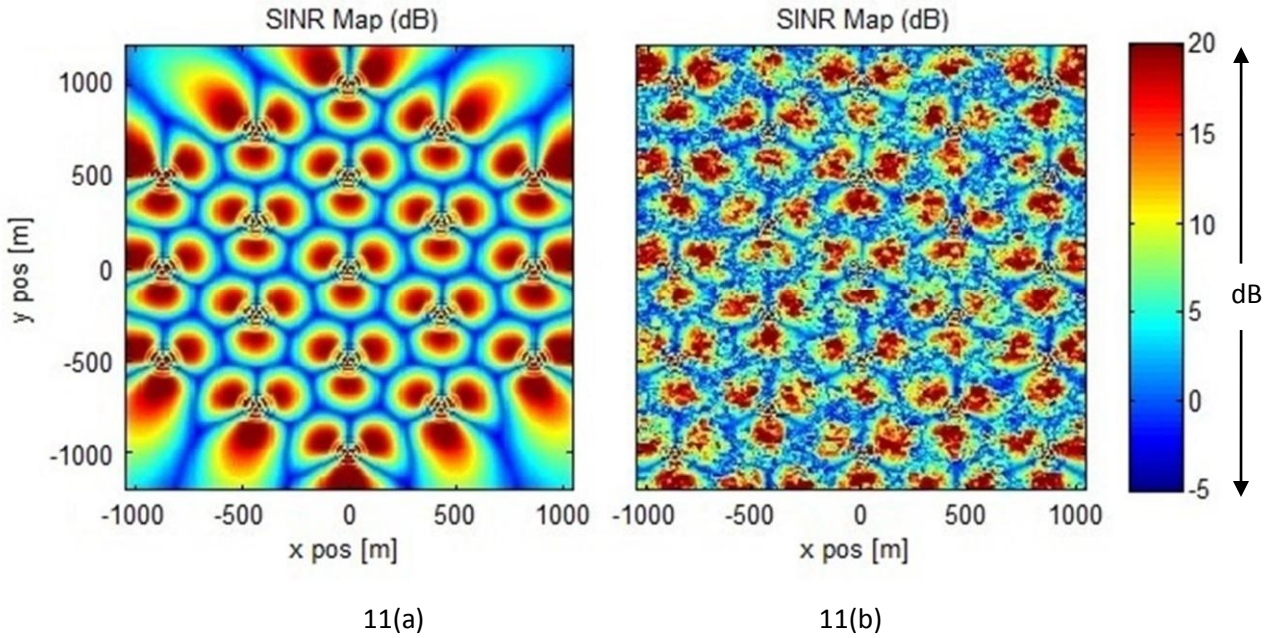


Figure 11: 2D SINR map without shadowing and without wrap-around, with shadowing and wrap around.

4.7 Flowchart

In this section, optimization approach for maximizing the total rate of each cell is described. The load balancing approach is solved by the proposed learning-based MM approaches presented in section 4.8 and 4.9. Both MAB and Bayesian Bandit based learning MM, result in REB β_k value optimization and in load balancing $\emptyset_{k,tot}(t_n)$. Based on the estimated instantaneous load, the context-aware scheduler selects, for each RB a UE considering its history and velocity as described in section 4.7. This results in each UE's instantaneous rate $\emptyset_{k,tot}(t_n)$ and the RB allocation vector $\alpha_{i(k)}(t_n) = [\alpha_{i(k),1}, \dots, \alpha_{i(k),R}]$ containing binary variables $\alpha_{i(k),r}$, and indicating whether UE i (k) of BS k is allocated at RB r or not. The inter-relation between the selected context parameters (UE's history and velocity), the scheduling function, the described optimization formulation, the rationale behind the methodology is as

follows. Within the proposed MM approaches, load balancing is carried out by optimizing the REB and entropy values. And history - based UE scheduling is carried by means of the proposed context-aware scheduler. Combination of both load balancing and history based UE scheduling yields in reduction of the HOF and PP probability via optimal REB value selection and the proposed context-aware scheduler. Here, the load balancing procedure yielding the optimal REB value incurs wideband SINR enhancement and HOF reduction. The context-aware scheduler on the other hand schedules UE based on the highest estimated achievable rate of each UE according to its instantaneous channel condition and its history, which leads to long-term fairness among UE. Both approaches, i.e., load balancing and history-based scheduling, yield throughput enhancement. Additionally, the velocity-based ranking property of the context-aware scheduler reduces the PP probability since low velocity UE are prioritized over high-velocity UE.

Figure 12 illustrates the flowchart of the proposed learning-based MM approaches. Initially UE senses the channels broadcasted by eNodeB based upon velocity of UE, RSRP and Bandwidth of channel, it makes a viable guess of the prior probability. Then resource block (RB) is allocated to the UE based on the velocity segmentation i.e. RB is allocated to the UE that has least velocity. The velocity rate and resource block occupancy is exchanged between macro and pico cell using X2 interface. Bayesian Bandit is used to find the optimized eNodeB channel. Then the algorithm checks if the sensed posterior PDF has highest posterior probability if so it predicts REB value by linear regression else it loops back to check the channel that has highest posterior distribution. After that it checks the channel that has least entropy. Finally, stopping criterion is met and UE is handed to the eNodeB with best posterior PDF.

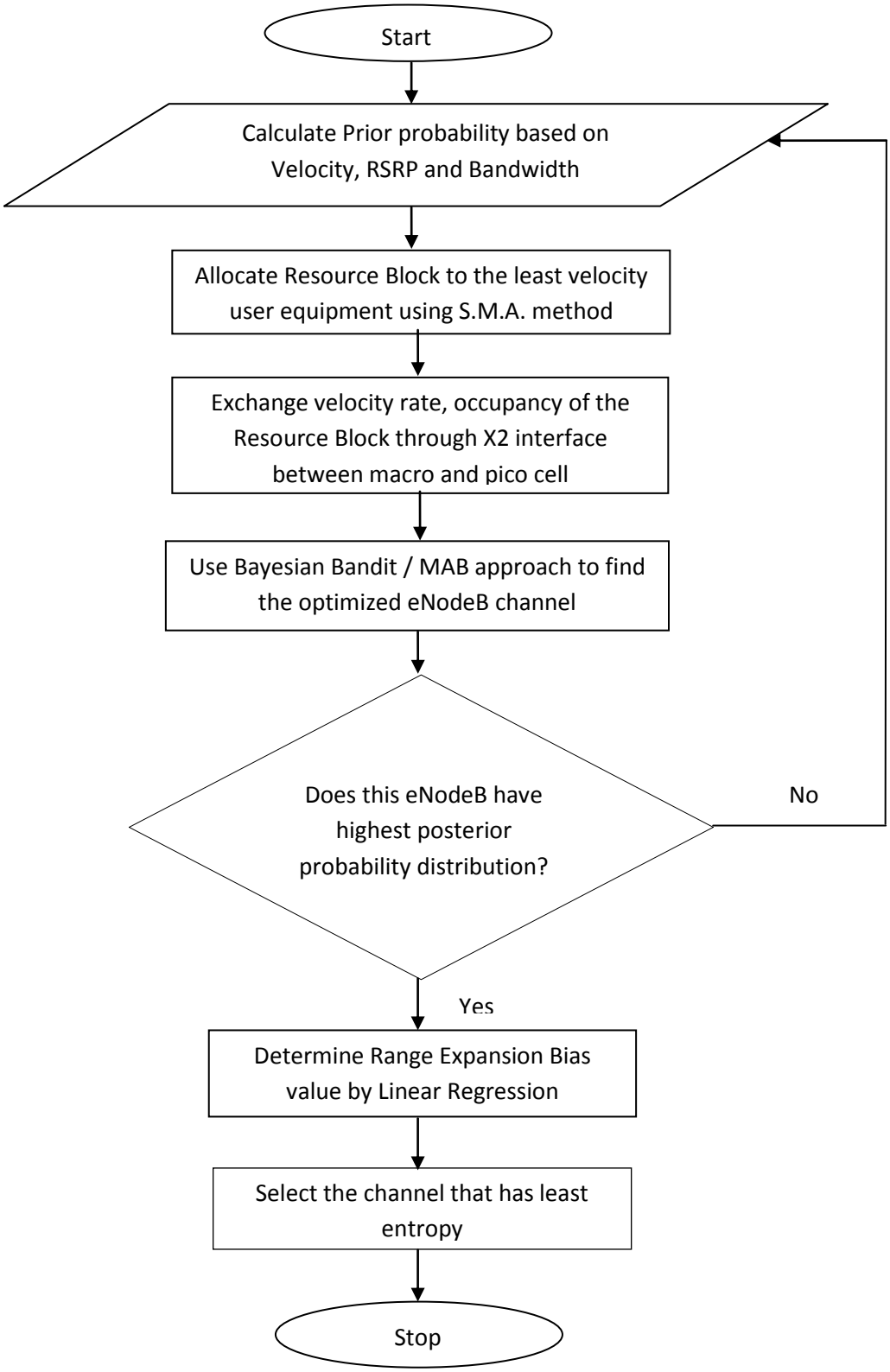


Figure 12: Flow chart of the proposed learning-based MM approaches.

4.8 A context-aware scheduler

The proposed MM approach considers a fairness-based context-aware scheduling mechanism. At each RB r, a UE i (k)* is selected to be served by BS k according to the following scheduling criterion:

$$\text{i (k)}_{\text{r}} = \text{sort } \{ \min (v\ i\ (k)) \} (\arg\max (\text{i (k)} \in \text{U}_{\text{k}} \frac{\emptyset_{i(k),r}(t_n)}{\overline{\emptyset_i}(t_n)}) \quad (30)$$

Where, sort min (v_{i} (k)) sorts the candidate UE according to their velocity starting with the slowest UE. After the sorting operation, if more than one UE can be selected for RB r, the UE with minimum velocity is selected. The rationale behind introducing a sorting / ranking function for candidate UE according to their velocity is that high-velocity UE will not be favored over slow moving UE. This has two advantages: 1) High-velocity UE might pass through the Pico cell quickly and should therefore not be favored to avoid PPs, and 2) the channel conditions of low-velocity UE changes slowly which may result, especially for slow-moving cell-edge UE, in poor rates if they are not allocated to many RBs.

The scheduler defined in (30) will allocate some (or even all) resources to a newly handed over UE since its average rate in the target cell is zero. To avoid this and enable a fair resources allocation among all UE in a cell, a history-based scheduling approach has been proposed as follows. Via the x2-interface, Macro and Pico cells coordinate, so that once a macro UE i (m) is handed over to a pico-cell p, the UE's target cell p and source cell m exchange information. In particular, UE i (m)'s rate history at time instant t_{n} is provided to pico cell p in terms of average rate, $\overline{\emptyset}_{i(m)}(t_n)$, such that the UE's (which is named as i(p) after the handover) average rate at picocell p becomes

$$\overline{\emptyset}_{i(p)}(t_n + T_s) = \frac{T\overline{\emptyset}_{i(m)}(t_n) + \overline{\emptyset}_{i(p)}(t_n + T_s)}{T+1} \quad (31)$$

In the above equation, a moving average is considered from macro-cell to pico cell, whereas in the classical MM approach, a UE's rate history is not considered and is equal to zero. In other words, in the classical proportional fair scheduler, the average rate $\overline{\emptyset}_i(t_n)$ in 30 is $\overline{\emptyset}_i(t_n) = \overline{\emptyset}_{l(k)}(t_n) = 0$ when a UE is handed over to cell k, whereas it can be redefined according to 31, i.e. $\overline{\emptyset}_i(t_n) = \overline{\emptyset}_{l(p)}(t_n + T_s)$. The proposed MM approach, instead considers the previous rate when UE i (m) was associated to the macro cell m in the past. The incorporation of a UE's

history enables the scheduler to perform fair resource allocation even in the presence of a sequence of handovers. Since handovers occur more frequently in HetNets due to small cell sizes, such a history-based scheduler leads to fair frequency resource allocation among the UE of a cell. More specifically, UE recently joining a cell will not be preferred over other UE of the cell since their historical average rate will be taken into account.

4.9 Multi-Armed Bandit Based learning Approach for load balancing

The objective of the MAB approach is to maximize the overall system performance. MAB is a machine learning technique based on an analogy with the traditional slot machine (one armed bandit). When pulled at time t_n, each machine / player provides a reward. The objective is to maximize the collected reward through iterative pulls, i.e. learning iterations. The player selects its actions based on a decision function reflecting the well-known exploration – exploitation trade-off in learning algorithms.

The set of players, actions and the utility function for our MAB based MM approach is defined as follows:

- Players: Macro BSs M = {1, …., M} and pico BSs P={1,….,P}.

- Actions: $A_k = \{\beta_k\}$ with $\beta_m = [0, 3, 6]$ dB and $\beta_p = [0, 3, 6, 9, 12, 15, 18]$ dB being the CRE bias. Higher bias values for pico cells have been considered due to their low transmit power.

- Strategy:
  1) Every BS learns its optimum Cell Range Expansion (CRE) bias value on a long-term basis considering its load:

$$\emptyset_{k,tot}(t_n) = \sum_{i(k) \in Uk} \sum_{r=1}^{R} \alpha_{i(k),r}(t_n).\, \emptyset_{i(k),r}(t_n) \qquad (32)$$

  This is inter-related with the handover triggering by defining the cell border of each cell.

2) A UE is handed over to BS k if it fulfills the condition.

$$P_{l}(i(l)) + \beta_{l} < P_{k}(i(k)) + \beta_{k} + m_{hist} \quad (33)$$

With $\{l, k\} \in K$, m_{hist} is the UE or cell-specific hysteresis margin, β_{k} (β_{m}) is the REB of BS k(l), and $P_{k}(i(k))$ (or $P_{l}(i(l))$) [dBm] is the i (k)-th (or i (l)-th) UE's RSRP from BS k(l) after TTT.

3) Resource Block based scheduling is performed based on

$$i(k)_{r}^{*} = \min \text{sort}(v_{i(k)}) \left(\arg\max i(k) \in U_{k} \frac{\emptyset_{i(k),r}(t_{n})}{\bar{\emptyset}_{i}(t_{n})}\right) \quad (34)$$

- Utility Function: The utility function in MAB learning is a decision function composed by an exploitation term represented by players k's total rate and exploration part considering the number of times an action has been selected so far. Player k selects its action $a_{j(k)}(t_{n}) \in A_{k}$ at time t_{n} through maximizing a decision function $d_{k, a j(k)}(t_{n})$, which is defined as:

$$D_{k, a j(k)}(t_{n}) = u_{k, a j(k)}(t_{n}) + \sqrt{\frac{2 \log(\sum_{i=1}^{|A_{k}|} n_{k,a_{i(k)}}(t_{n}))}{n_{k,aj(k)}(t_{n})}} \quad (35)$$

Whereby $u_{k, a j(k)}(t_{n})$ is the mean reward of player k at time t_{n} for action $a_{j(k)}$, $n_{k, a j(k)}(t_{n})$ is the number of times action $a_{j(k)}$ has been selected by player k until time t_{n}, and $|.|$ - represents the cardinality.

During the first $t_{n} = |A_{k}| . T_{s}$ player k selects each action once in a random order to initialize the learning process by receiving a reward for each action. For the following iterations $t_{a} > |A_{k}| . T_{s}$ action selection is performed according to MAB algorithm. In each learning iteration the action $a^{*}_{j(k)}$ that maximizes the decision function in (35) is selected. Then the parameters are updated, whereby the following notation is used: $s_{k, a j(k)}(t_{n})$ is the cumulative reward of player k after playing action $a_{j(k)}$ and indicator function (i = j) is equal to 1 if i = j and zero otherwise.

4.10 Bayesian Bandit

The Bayesian solution begins by assuming priors on the probability of winning for each bandit. Since, a UE is trying to communicate through three eNodeB that is near to it so it calculates a viable prior which is from 0 to 1. The algorithm is as follows:

For each round,

1. Sample a random variable X_b from the prior of bandit b, for all b.
2. Select the bandit with largest sample, i.e. select bandit B = arg max X_b.
3. Observe the result of pulling bandit B, and update your prior on bandit B.
4. Return to 1.

Computationally, the algorithm involves sampling from N distributions. Since the initial priors are Beta (α=1, β=1) (a uniform distribution), and the observed result X (a win or loss, encoded 1 and 0 respectively) is Binomial, the posterior is a Beta (α=1+X, β=1+N-X).

This algorithm doesn't discard losers, but it picks them at decreasing rate as it gathers confidence that there exists a better bandit. This follows because there is always a non-zero chance that a loser will achieve the status of B, but the probability of this event decreases as it plays more rounds.

4.11 Tools Used

- PYTHON
- NS3 SIMULATOR
- GNU OCTAVE

Chapter 5

Result and Analysis

5.1 Path Loss models

Hata model, Macro path loss model and Pico path loss model are illustrated respectively in Figure 13. It shows path loss increases as UE gradually moves away from the serving eNodeB.

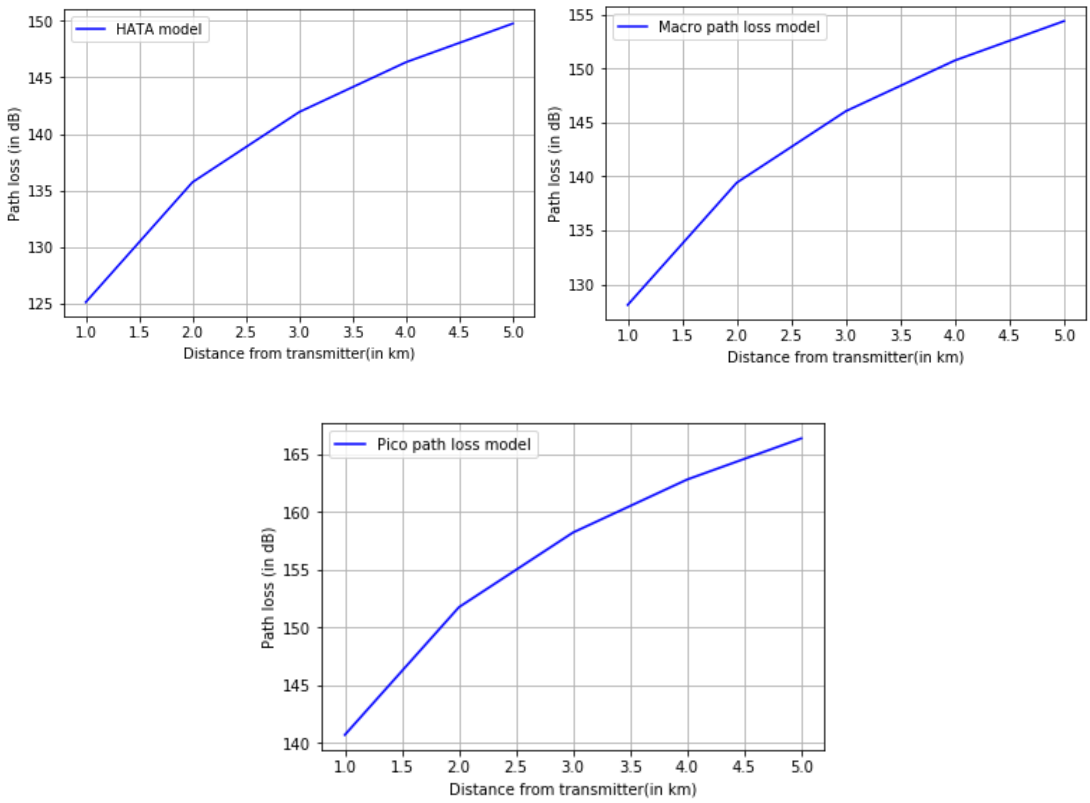


Figure 13: Different path loss model.

5.2 Exploration and Exploitation Dilemma

Online decision-making involves fundamental choice – exploitation: make the best decision given current information and exploration: gather more information. The best long-term strategy may involve short-term sacrifices. Exploitation part of reinforcement learning gathers more information to make the best overall decisions. The ∈-greedy algorithm continues to explore forever if ∈ = 0. The ∈-greedy algorithm selects highest reward value with probability

1-∈ and with probability ∈ it selects a random action. The constant ∈ ensures minimum regret. It has linear total regret. In Figure 14, when ∈ = 0, it's a greedy algorithm and it never exploits so greedy algorithm can lock onto a suboptimal action forever. In the next case when ∈ = 0.1 it explores for 10% of total time / data and exploits for 90% of total time / data.

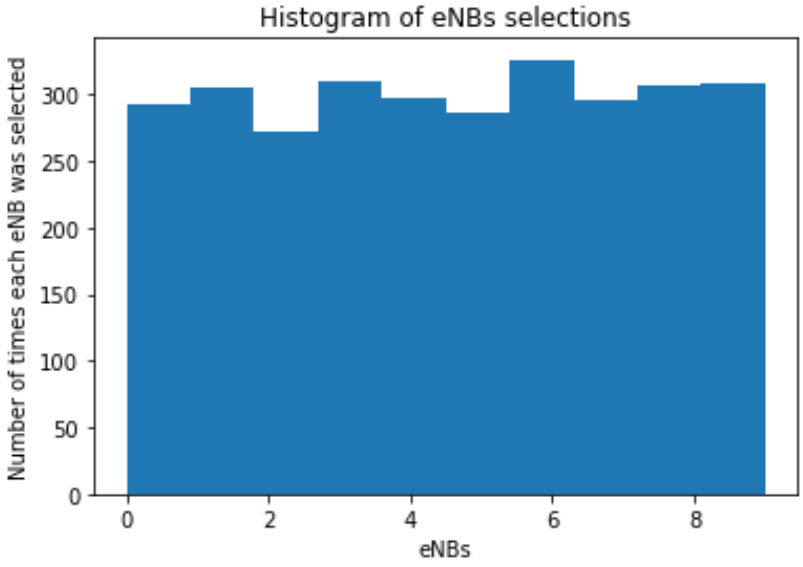
Figure 14: Exploration - Exploitation dilemma.

5.3 Application of dataset to UCB algorithm

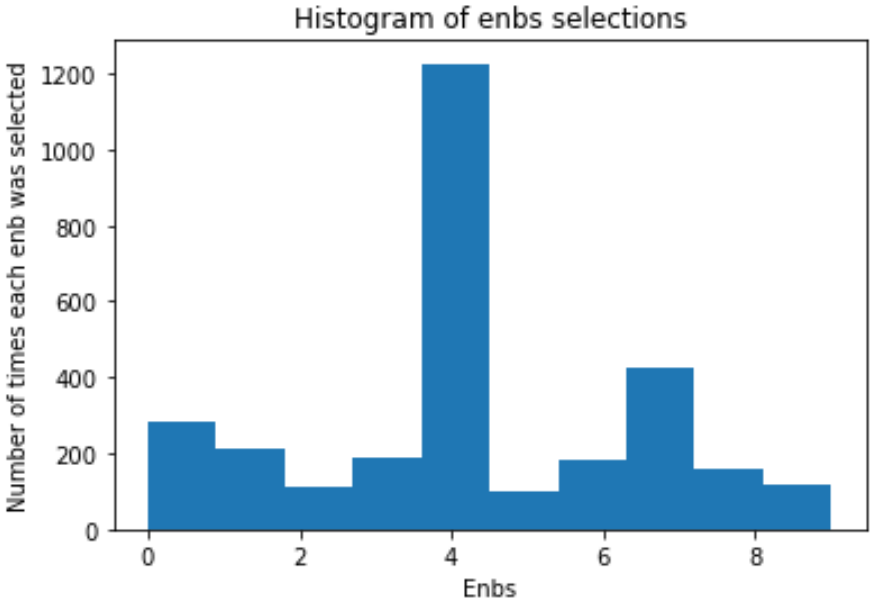
As, seen from Table 4. Excel sheet at time instant t = 0, UE will try to access resource of eNodeB1, eNodeB5 and eNodeB9 simultaneously. Similarly, at time instant t = 1, UE will try to access resource of eNodeB9 only. The UCB tries to maximize the reward value; this reward value is calculated through utility function at each time instant. If the current reward calculated at time instant t = 200 is less than t = 199 then the highest cumulative reward up to time instant t = 199 is used for the further exploitation and exploration else, the current reward value calculated at time instant t = 200 is used for further exploitation and exploration.

The Figure 15(a) illustrates if none of the machine learning algorithm is applied then all the eNodeB are selected in uniform manner. But, if reinforcement learning approaches like UCB - which is a subpart of machine learning technique, is applied then the eNB4 has highest

cumulative reward after several pulls. The UE chooses eNB4 because it has the highest reward value. As seen from Figure 15 (b) the eNodeB4 has the highest reward value / utility function calculated value so the UE chooses eNodeB4 as the optimized Node with highest reward and correspondingly UE will access resource of eNodeB4 because it is the best available option among all eNodeBs.



15(a)



15(b)

Figure 15: (a)Uniform distribution of eNBs selected, (b) eNB4 is selected

5.4 Posterior updating through different pulls

In this thesis work, three eNodeB have been considered for simulating Bayesian Bandit reinforcement learning algorithm. Whenever, a UE is trying to handover with neighboring eNodeB it calculates cost, channel losses, shadowing, velocity, history traces, entropy, RSRP, REB and SINR values. Based on these features UE estimates the prior probability.

Visualization of Multi Armed Bayesian Bandit learning algorithm is presented in Figure 16. The algorithm sequentially learns the best solution after several pulls for the optimization problem. The dashed lines in Figure 16 represent the true hidden probabilities, which are 0.85, 0.60, and 0.75. This Figure could be extended to more dimensions, but the figure significantly suffers, so the probability density function was considered for three dimensions where the visualization of figure is very lucid and practical.

The Bayesian Bandit reinforcement learning algorithm picks those action with following three cases 1 which is more uncertain about an action and value, 2 The actions that are more important to explore, and 3 The actions that could turn out to be the best action. Figure 16 illustrates if the algorithm picks orange bandit action then the action is less uncertain about the value, this bandit is more likely to pick another action until the bandit picks the best action.

Bayesian Bandit algorithm chooses the best bandit (or more accurately, becoming more confident in choosing the best bandit). For this reason, the distribution of the orange bandit is very wide (representing ignorance about what the hidden probability might be). On the flipside the Bayesian Bandit algorithm learns reasonably confident that it is not the best arm, so the algorithm chooses to ignore it.

From Figure 16, it can be seen after 1000 pulls, the majority of the “blue” distribution action has maximum standard deviation among three Gaussian distribution, and the algorithm chooses it as the best arm.

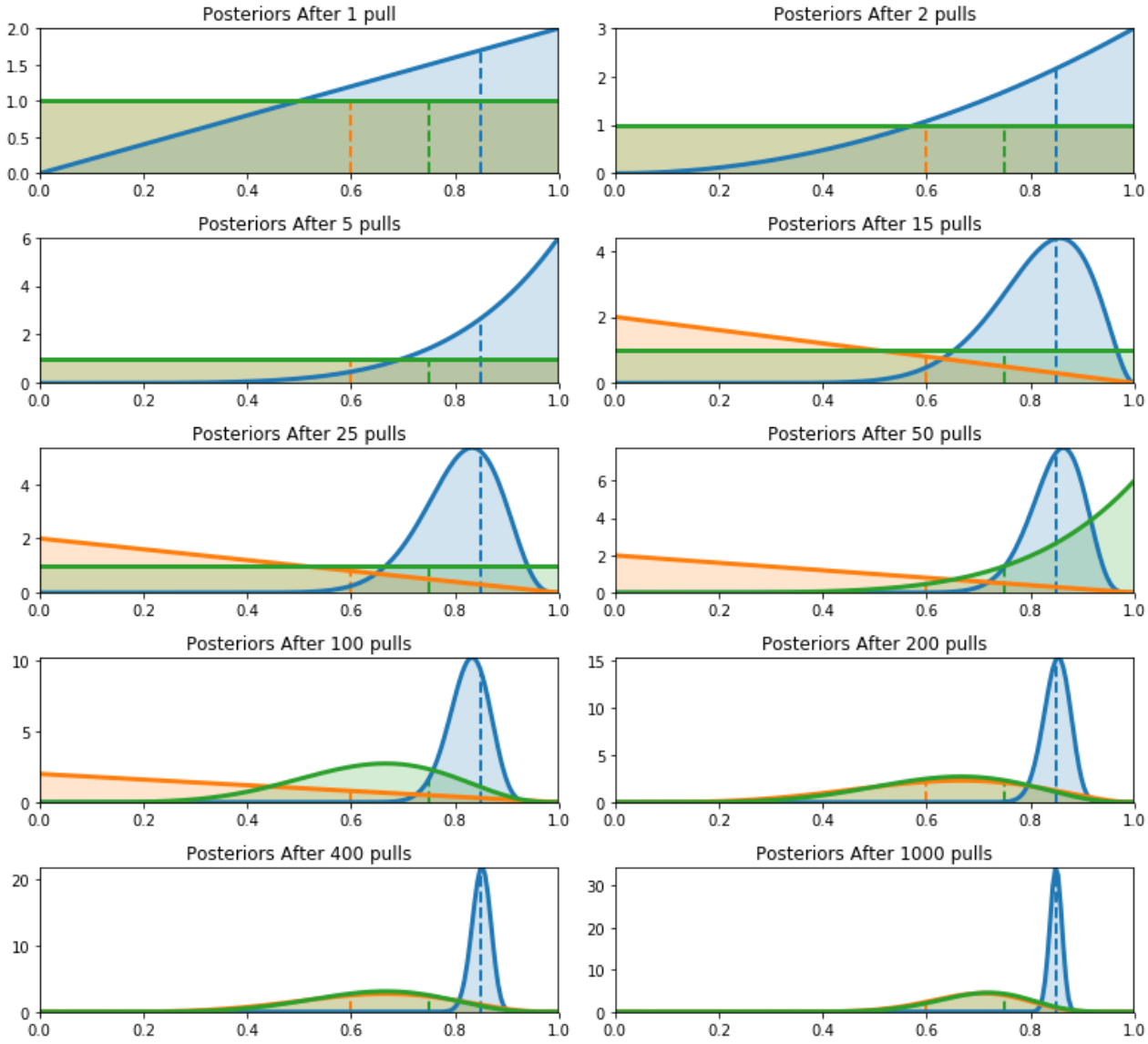


Figure 16: Posterior probability updating at different pulls.

5.5 Linear regression to predict REB and entropy to predict the load balancing

Range Expansion Bias is predicted through linear regression. Figure17 illustrates linear curve fitting model for five posterior probability distribution function. After several pulls of the arms the Gaussian distribution of the action is achieved and UE selects the action with highest standard deviation. UE calculates the posterior probability of the best arm then, it predicts bias value in accordance with linear regression. After all these processes, the algorithm looks for the

information entropy to handover UE to the eNodeB, higher entropy signifies higher randomness and more bandwidth whereas lower entropy signifies lower randomness and less bandwidth. The UE uses entropy for switching purpose from one eNodeB to the neighboring eNodeB. For switching from one eNodeB to the next eNodeB, significantly less bandwidth is required. Table 5 illustrates, the effect of UE offloading in macro and pico eNodeB for varying entropy values.

Table 5: Entropy measures for load balancing.

Posterior Probability	Entropy	Effect (Handover UE to)
0.8	0.257542	Pico eNodeB
0.7	0.360201	Macro eNodeB

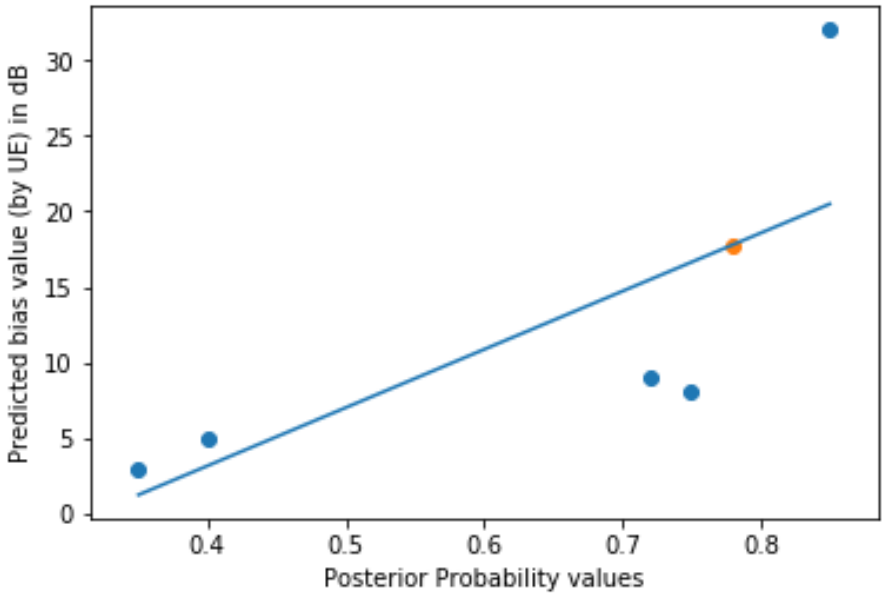


Figure 17: Linear regression model to predict REB.

5.6 Simulation of HetNet in NS3

Single, UE and two eNodeB is simulated in NS3, Figure 18 illustrates initially at 0.0399286 sec UE with IMSI 1 is connected to CellId 1 with RNTI 1 of eNodeB1. At 0.1 sec handover is started with IMSI 1 and RNTI 1 to cellID 2. As can be seen after 0.0104 sec time lapse UE previously connected to cellID 1 with RNTI 1 is doing handover to cellID 2. At, 0.107214 sec handover is successful to cellID 2 with RNTI 2. Finally, after 0.117929 sec handover is

completed, delinating that UE with IMSI 1 and RNTI1 is seamlessly connected to cellID2. Figure 19 illustrates initially UE communicates with both eNodeB since UE is trying to handover from eNodeB1 to eNodeB2 there is less data rate between UE and eNodeB1 and higher data rate between UE and eNodeB2. After some instant of time UE will be handover to eNodeB2 with higher data rates.

```
0.028 /NodeList/4/DeviceList/0/LteUeRrc/ConnectionEstablished UE IMSI 1: connect
ed to CellId 1 with RNTI 1
0.0399286 /NodeList/2/DeviceList/0/LteEnbRrc/ConnectionEstablished eNB CellId 1:
 successful connection of UE with IMSI 1 RNTI 1
0.1 /NodeList/2/DeviceList/0/LteEnbRrc/HandoverStart eNB CellId 1: start handove
r of UE with IMSI 1 RNTI 1 to CellId 2
0.104 /NodeList/4/DeviceList/0/LteUeRrc/HandoverStart UE IMSI 1: previously conn
ected to CellId 1 with RNTI 1, doing handover to CellId 2
0.107214 /NodeList/4/DeviceList/0/LteUeRrc/HandoverEndOk UE IMSI 1: successful h
andover to CellId 2 with RNTI 1
0.117929 /NodeList/3/DeviceList/0/LteEnbRrc/HandoverEndOk eNB CellId 2: complete
d handover of UE with IMSI 1 RNTI 1
```

Figure 18: Handover of UE with two eNodeB.

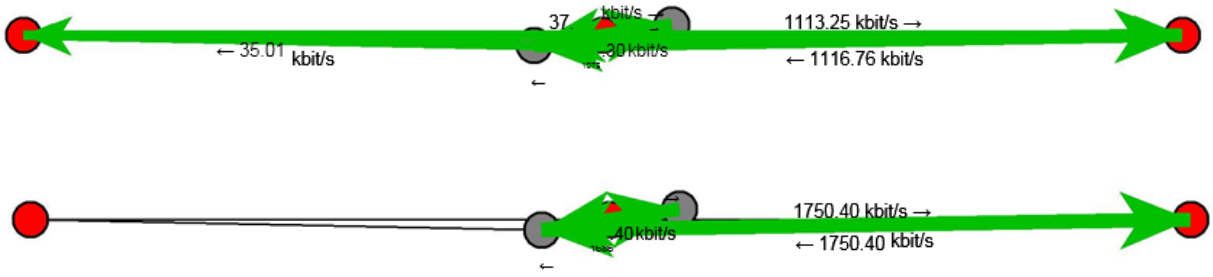


Figure 19: Data flow of UE with two eNodeB before and execution of handover.

5.7 UE Throughput and sum-rate

In this section analysis of increasing and decreasing of REB value is considered. After analysis it has been seen that when REB bias value for pico cell is increased, the cell-edge UE throughput is enhanced. The proposed learning - based MM approach outperforms, the classical approach up to five times for 10 UE per macro cell. Interestingly, the Multi Armed Bandit / Bayesian Multi Armed Bandit - based MM with entropy measures yield higher cell-edge UE throughput for smaller number of UE.

The sum – rate (throughput) versus UE density per macro cell for TTT = 40 ms values is presented in Figure 20 for 40 ms TTT value, the classical approach yield very low sum-rates, while the proposed approach lead to significant improvement of up to 86 % for the same TTT value. The proposed MAB MM with entropy measures approach converge to significantly larger sum-rates than the MAB MM and classical MM approach. Interestingly, the sum-rate performance of the proposed MM approach depends on the TTT values. The reason for this lies in the convergence behavior of the learning algorithms. For smaller TTT values, handover is executed faster and the BS has to adapt its REB strategy to the new cell load before convergence.

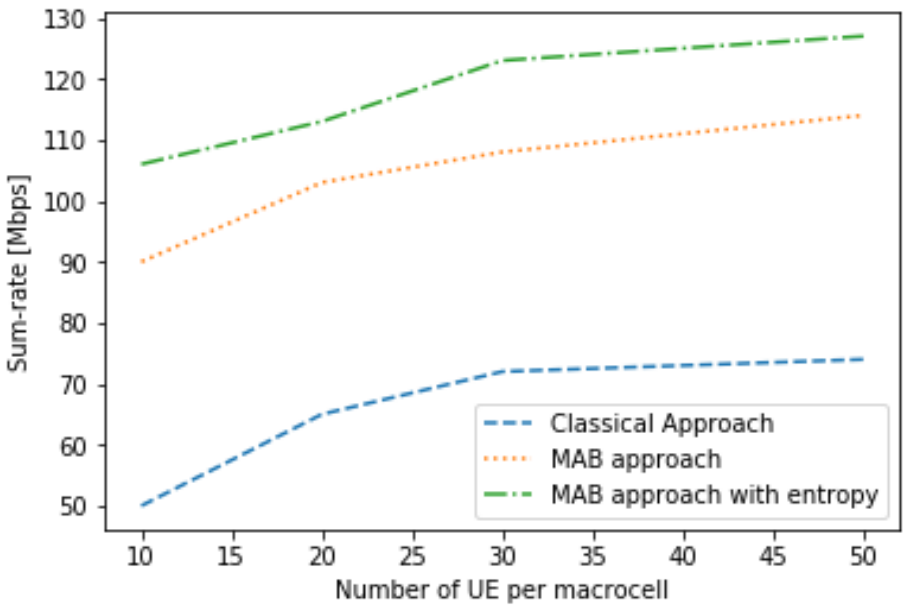


Figure 20: Comparison between MAB with entropy approach, MAB approach and classical RSRP method.

5.8 Segmentation of UE in HetNets

UE traverses randomly in the heterogeneous network. Table 6 illustrates, low velocity UEs are preferred by the pico cell whereas high velocity UEs are preferred by the macro cell. When UE with velocity less than 30 Km/hr traverses the heterogeneous networks it will access the resources of pico cell on the flipside when the UE with velocity more than 30 Km/hr then it will access the resource of macro cell.

Table 6: Segmentation of UE in HetNets

Velocity of UE in Km/hr	Effect
31.6	UEs access the resource of macro cell
26.6	UEs access the resource of pico cell

5.9 HOF and PP probability

Mobility management approaches will enhance the network performance in terms of UE throughput and sum rate, in parallel it will reduce handover failure rates and PP probabilities. After simulation, results of the HOF probability and PP probability are shown in Figure 21 below. In this thesis study same velocity for each UE per simulation is considered, and result of each velocity is presented separately. Figure 21 depicts the HOF probability for different TTT values. As it can be seen, the proposed learning based approaches yield improvements in terms of HOF probability. The HOF probability for UE at 3 km/h speed is similar for the classical MM approach, MAB approach and proposed MAB based MM approach with entropy measures. For higher velocities in which more HOFs are expected, the HOF probability obtained by proposed approach is significantly lower than in the case of MAB MM and classical MM approaches. Interestingly, the proposed methods lead to almost constant HOF probabilities for velocities larger than 30 km/h. For UE at 120 km/h, the HOF probability of the MAB based MM approach with entropy measures is half of the classical approach for TTT = 40 ms, increasing TTT values, the trend between the proposed MM approaches and the classical approach remains similar. This is because the pico cell coverage is small, and thus the macro cell UE quickly run deep inside the pico cell coverage before the TTT expires, significantly degrading the signal quality of the macro cell UE before the handover is completed. In this case, HOFs are alleviated with smaller TTT values. Reducing TTT values may decrease the HOF probability but increase PP probability. Hence, HOFs and PPs must be studied jointly.

Figure 22 shows PP probability for various values of TTT. It can be observed that the number of PPs is reduced with larger TTT values. In addition, for lower velocities, all MM approaches yield similar PP probabilities for all TTT values. For higher velocities, the PP probability is decreased by the proposed MM approaches by up to a factor of two (TTT = 40 ms).

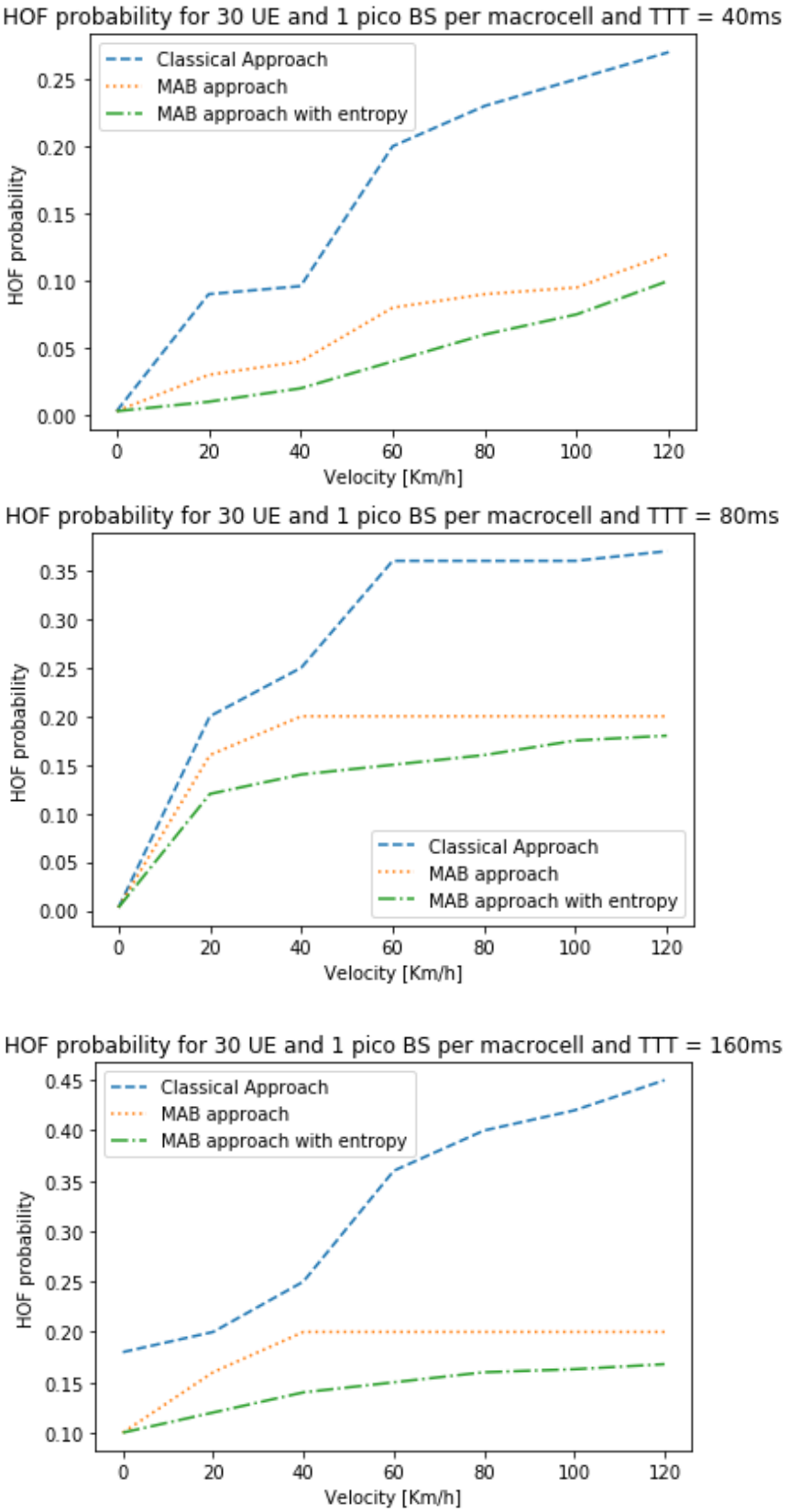


Figure 21: HOF probability reduction under various velocity and TTT.

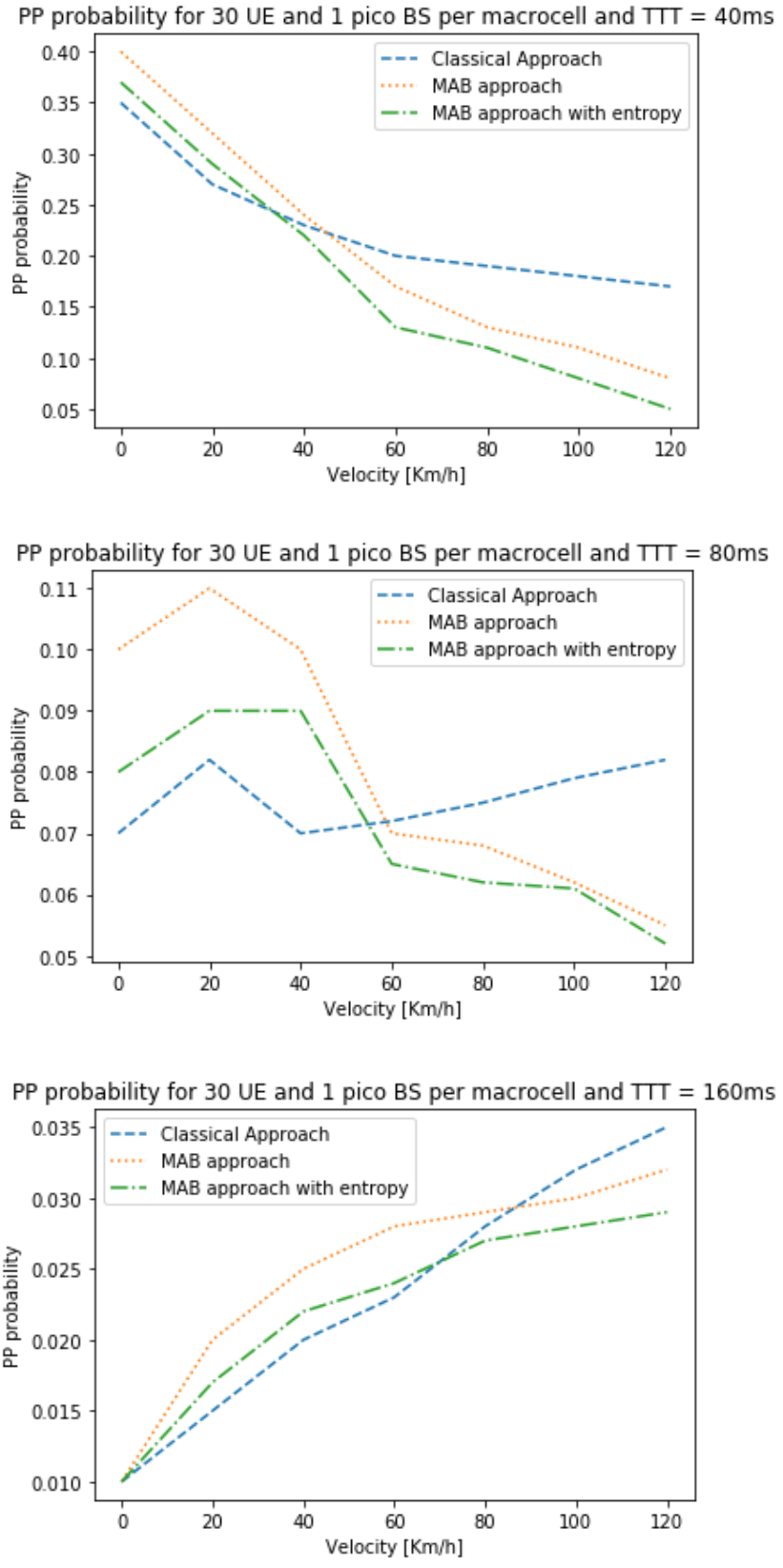


Figure 22: PP probability reduction under various velocity and TTT.

5.10 Validation of Result

Simulation of classical RSRP, learning based handover algorithm and Multi Armed Bandit based Mobility Management with entropy measures algorithm of the LTE module is carried out in ns-3. Then, the effect of each handover algorithm is analyzed and compared. In this simulation campaign the lena-dual-stripe example program is being used. The different handover algorithms are classified based on user average throughput.

After hours of running, the simulation campaign will eventually end. Next some post-processing on the produced simulation output is done to obtain meaningful information about it.

The GNU Octave tool is used to assist the processing of throughput and number of handovers. Sample GNU Octave script is demonstrated below:

```
% RxBytes is the 10th column
UlRxBytes = load ("a2-a4-rsrq-UlRlcStats.txt") (:,10);
UlAverageThroughputKbps = sum (UlRxBytes) * 8 / 1000 / 50
```

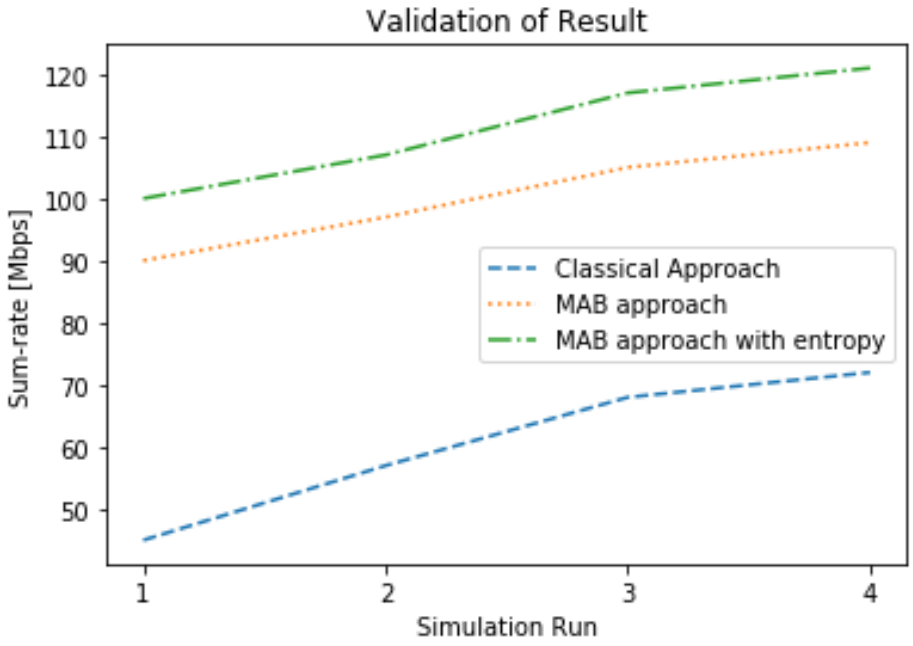


Figure 23: Throughput at various simulations runs of classical, MAB approach and MAB based MM with entropy measures.

Figure 23 validates Multi Armed Bandit based reinforcement learning approach with entropy measures has higher throughput compared to MAB and classical approach at different simulation run.

Table 7 shows the complete statistics after post-processing on every individual simulation run. The values shown are the average of the results obtained from RngRun of 1, 2, 3, and 4.

Table 7: Results of handover campaign

Statistics	Classical Approach	MAB Approach	MAB Approach with Entropy
Average DL system throughput	20, 509 kbps	101, 248 kbps	112, 980 kbps
Average UL system throughput	5, 706 kbps	28, 420 kbps	35, 830 kbps

The results show that handover algorithm that uses reinforcement learning with entropy measures approach improves user throughput significantly.

Chapter 6

Conclusion

There is a risk of the handover failure which may result in a radio link failure in LTE networks, because it adopts a hard handover scheme. System performance is being aggravated by unnecessary handovers. It leads to transmission delay, packet loss, and signal overhead, which may seriously affect the performance of a real-time application. Multi armed bandit and Bayesian multi armed bandit based MM approaches with entropy measures and a history-based context-aware scheduling method for Heterogeneous Networks is being proposed in this thesis in order to lower the unnecessary handover. An efficient linear regression based scheme is used to predict a near optimal bias value that attempts to reduce blocking probability and improve load fairness index in the system. Information entropy is used to evict the user equipment from overcrowded cell to the cell that has relatively less traffic, better throughput and higher signal to interference noise ratio. The multi armed bandit based learning aims at system performance maximization. The proposed learning based approach outperforms the MAB and classical mobility management in terms of user equipment throughput. In average, a gain of up to 86 % is achieved for user equipment throughput, while the handover failure probability is reduced to a factor of two by the proposed reinforcement based mobility management approaches. Simulation value of user equipment's throughput validates the proposed scheme is better over the classical RSRP and Multi Armed Bandit based handover approach.

Chapter 7

Limitations and Future Works

One of the major concerns of learning-based approaches is their convergence behavior in dynamic systems. In terms of cell-center UE throughput the MAB-based MM approach converges slower than other learning based MM approach, but it converges to a larger cell-center UE throughput since it aims at system performance maximization.

For further study the evaluation of handover that uses deep Q-learning can be used. In this approach UE keeps traces of Q-table when they move to another PBS coverage area, and by this method it will help a learning algorithm to converge faster. The required learning time should be studied for realizing this system because if it takes too much time to converge, it cannot be used in the real system.

REFERENCES:

[1] Francesco Guidolin et al., "A Markov-Based Framework for Handover Optimization in HetNets," IEEE, 2014.

[2] Zhixiong Ding et al., "An Effective Handover Scheme in Heterogeneous Networks," IEEE, 2016.

[3] Peyman TalebiFard and Victor C.M Leung., "A Dynamic Context-Aware Access Network Selection for Handover in Heterogeneous Network Environments," IEEE, 2015.

[4] A.M.Miyim et al., "Fast Handover Proximity for Heterogeneous Cellular Networks," IEEE, 2015.

[5] A. Habibzadeh et al., "A Novel Handover Decision-Making Algorithm for HetNets," IEEE, 2015.

[6] Malka N. Halgamuge et al., "Handoff optimization Using Hidden Markov Model," IEEE, 2011.

[7] Ili Nadia Md Isa et al., "Handover Parameter Optimization for Self-organizing LTE Networks," IEEE, 2015.

[8] Dang Feng et al., "A Multi-attribute Vertical Handover Algorithm based on Adaptive Weight in Heterogeneous Wireless Network," IEEE, 2014.

[9] Hiroyuki Koga et al., "Improved Handover Using Cloud Control in Heterogeneous Wireless Networks," IEEE, 2015.

[10] Abhijit Bijwe et al., "Vertical Handoff algorithms using Neural Networks," CEEE, 2013.

[11] Toshihito Kudo et al., "Cell range expansion using distributed Q-learning in heterogeneous networks", EURASIP Journal on Wireless Communications and Networking, 2013.

[12] Hsiu-Lang Wang et al., "A moving direction prediction-assisted handover scheme in LTE networks", EURASIP Journal on Wireless Communications and Networking, 2014.

[13] R. Sasikumar et al., "An Intelligent Pico Cell Range Expansion Technique for Heterogeneous Wireless Networks", Indian Journal of Science and Technology, 2016.

[14] Weyu LI et al., "A Dynamic Hysteresis-adjusting Algorithm in LTE Self-Organization Networks", IEEE, 2012.

[15] Long Li et al., “A Hierarchical MADM-based Network Selection Scheme for System Performance Enhancement”, IEEE, 2014.

[16] Simone Barbara et al., “Mobility Performance of LTE Co-Channel Deployment of Macro and Pico Cells”, IEEE, 2012.

[17] Mishra et al., “Enhancing the Performance of HetNets via Linear Regression Estimation of Range Expansion Bias”, IEEE, 2013.

[18] Yuefeng Peng et al., “Mobility Performance Enhancements for LTE-Advanced Heterogeneous Networks”, IEEE, 2012.

[19] Jeffrey G. Andrews et al., “An Overview of Load Balancing in HetNets: Old Myths and Open Problems”, IEEE, 2014.

[20] Xiaofei Wang et al., “Artificial Intelligence-Based Techniques for Emerging Heterogeneous Network: State of the Arts, Opportunities, and challenges”, IEEE, 2015.

[21] Klaus I. Pedersen et al., “Mobility Enhancements for LTE-Advanced Multilayer Networks with Inter-Site Carrier Aggregation”, IEEE, 2013

[22] Kinan Ghanem et al., “Reducing Ping-Pong Handover Effects In Intra E-UTRA Networks”, IEEE, 2012.

[23] Meryem Semsek et al., “Analysis of Handover Failures in Heterogeneous Networks with Fading”, IEEE, 2016.

[24] Richard S. Sutton and Andrew G. Barto, “Reinforcement Learning: An Introduction”, The MIT Press, 2016.

[25] David Silver, “Introduction to Reinforcement Learning”, University College London, 2015.

[26] Cameron et al., “Bayesian Methods for Hackers probabilistic: Using Python and PyMC”, 2013.

[27] Volodymyr et al., “Algorithms for the multi-armed bandit problem”, Journal of Machine Learning Research 1, 2000.

[28] Yichi Zhou et al., “Racing Thompson: an Efficient Algorithm for Thompson Sampling with Non-conjugate Priors”, arXiv, Cornell University, 2017.

[29] David Silver, “Exploration and Exploitation”, University College London, 2015.

[30] Professor Joe Blitzstein, “Beta Distribution”, Harvard University, 2013.

[31] J. Chen, et al., “Optimal Contraction Theorem for Exploration-Exploitation Tradeoff in Search and Optimization”, IEEE Transactions on Systems Man and Cybernetics Part a-Systems and Humans, vol. 39, pp. 680-691, May 2009.

[32] Stefan Parkvall et al., “LTE-Advanced – evolving LTE towards IMT-Advanced”, Ericsson Research, Sweden 2011.

[33] Ericsson, “HetNets – the solution to managing end-user expectations of capacity and speed”, PRESS information, 2012.

[34] Xiaoli Chu et al., Center of Telecommunications Research, King’s College London, “Inter-Cell interference Coordination for Expanded Region Pico Cells in Heterogeneous Networks”, London, UK, 2011.

[35] 4G Americas, “4G Mobile Broadband Evolution: 3GPP release 10 and beyond, HSPA+, SAE/LTE and LTE-Advanced”, 2011.

[36] H. Claussen et al., “An Overview of the Femto cell Concept”, Bell Labs Technical Journal, vol. 13, no. 1, pp. 221-245, 2008.

[37] M. Lalam, “Interference Management in Co-Channel Femtocell Deployment”, SAGEMCOM, 2012.

[38] R1-094225, “DL Performance with hotzone cells”, Qualcomm Europe, 2010.

[39] H. Xu et al., “Noise Padding Techniques in Heterogeneous Networks”, Patent Application Publication, No, 2011/0250911 A1, 2011.

[40] R1-094226, “UL Performance with hotzone cells”, Qualcomm Europe, 2010.

[41] Mohammad Ahmad Joud, “Pico cell range expansion towards lte-advanced wireless heterogeneous networks”, Universitat Politecnica de Catalunya (UPC), 2013.