“MULTI-AGENT REINFORCEMENT LEARNING BASED ROUTING IN COGNITIVE RADIO NETWORK”

A THESIS SUBMITTED TO

BHARATI VIDYAPEETH UNIVERSITY, PUNE

FOR AWARD OF DEGREE OF

DOCTOR OF PHILOSOPHY IN COMPUTER ENGINEERING

UNDER THE FACULTY OF ENGINEERING & TECHNOLOGY

SUBMITTED BY

MRS. SUNITA SHIVPRAKASH BARVE

UNDER THE GUIDANCE OF

DR. PARAG A. KULKARNI

RESEARCH CENTRE

BHARATI VIDYAPEETH DEEMED UNIVERSITY

COLLEGE OF ENGINEERING, PUNE - 411043

FEBRUARY 2016

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled "**Multi-Agent Reinforcement Learning Based Routing in Cognitive Radio Network**" for the degree of 'Doctor of Philosophy' in the subject of **Computer Engineering** under the faculty of **Engineering and Technology** has been carried out by **Mrs. Sunita Shivprakash Barve** in the Department of **Computer Engineering** at Bharati Vidyapeeth Deemed University, College of Engineering, Pune during the period from August 2010 to February 2016 under the guidance of **Dr. Parag A. Kulkarni**.

Place: Pune
Date :

Prof. Dr. A. R. Bhalerao
Principal, BVDUCOE and
Dean, Faculty of Engineering,
Bharati Vidyapeeth Deemed University,
Pune

CERTIFICATION OF GUIDE

This is to certify that the work incorporated in the thesis entitled "**Multi-Agent Reinforcement Learning Based Routing in Cognitive Radio Network**", submitted by **Mrs. Sunita Shivprakash Barve** for the degree of 'Doctor of Philosophy' in the subject of **Computer Engineering** under the faculty of **Engineering and Technology** has been carried out in the Department of **Computer Engineering**, **Bharati Vidyapeeth Deemed University College of Engineering, Pune,** during the period from August 2010 to February 2016, under my direct supervision/ guidance.

Place: Pune
Date :

Dr. Parag A. Kulkarni
Research Guide

DECLARATION BY THE CANDIDATE

I hereby declare that the thesis entitled **"Multi-Agent Reinforcement Learning Based Routing in Cognitive Radio Network"**, submitted by me to the Bharati Vidyapeeth University, Pune for the degree of **Doctor of Philosophy (Ph.D.)** in **Computer Engineering** under the faculty of **Engineering and Technology** is original piece of work carried out by me under the supervision of **Dr. Parag A. Kulkarni**. I further declare that it has not been submitted to this or any other university or Institution for the award of any degree or diploma.

I also confirm that all the material which I have borrowed from other sources and incorporated in this thesis is duly acknowledged. If any material is not duly acknowledged and found incorporated in this thesis, it is entirely my responsibility. I am fully aware of the implications of any such act which might have been committed by me advertently or inadvertently.

Place: Pune **Mrs. Sunita Shivprakash Barve**

Date : **Research Student**

In memory of my father

To my mother

With love and eternal appreciation

ACKNOWLEDGEMENT

I would like to extend my heartfelt gratitude and appreciation to the people who helped me to bring this study into reality.

I am deeply delighted to show oblige for **Dr. A. R. Bhalerao**, **Dean, Faculty of Engineering and Principal**, **BVDU College of Engineering, Pune** for his timely and invaluable support during my research work.

I wish to express my sincere gratitude to **Dr. Parag Kulkarni**, for his supervision, guidance and encouragement throughout this study. I was privileged to have him as my research guide. Only his exemplary guidance, unceasing encouragement and his indefatigable support to my efforts even in my critical moments secured the final completion of this thesis. His tireless enthusiasm and inspiring guidance during my research work will always remain a source of inspiration to me. I am very grateful for his time, his valuable inputs and appreciation when required.

I am grateful to **Prof. Dr. Sarang Joshi**, Pune Institute of Computer Technology, Pune and **Prof. Dr. J. V. Aghav**, Head, Department of Computer Engineering & IT, College of Engineering, Pune for their constructive critiquing, comments and suggestions for the betterment of my research work.

I am deeply obliged to **Dr. Yogesh J. Bhalerao**, **Principal**, MIT Academy of Engineering, Alandi for his continuous support, visionary guidance and encouragement for my research work. I am thankful to **Prof. Dr. S. M. Khairnar**, **Dean,** Research and Development, MIT Academy of Engineering, Alandi for giving right suggestion at the right time and for being the source of motivation in holding high research standards. I would like to express my gratitude to **Prof. Mrs. Uma Nagaraj**, **Head,** Computer Department, MIT Academy of Engineering, Alandi for her precious advice and being supportive in every way.

I would like to express my gratitude to **Dr. S. B. Deosarkar**, Principal, Vidya Pratishthan’s College of Engineering, Baramati for his invaluable suggestions and introducing me to this new and exciting research area of Cognitive Radio Network.

My deepest thanks to **Dr. Mrs. Sunita Jahirabadkar, Mrs. Bela Jogalekar** and **Mrs. Anagha Kulkarni** for their timely review, brilliant comments and suggestions throughout the writing of my thesis. I would also like to acknowledge all my friends and colleagues without whom this experience would have been incomplete.

I gratefully acknowledge the financial assistance under Research Promotion Scheme from **All India Council for Technical Education (AICTE), New Delhi**.

Finally, I would like to thank my family to whom I owe a great deal. To my late father **Shri. Anil R. Thoke** for showing me that the key to life is hard work. My sister **Mrs. Deepali Shende** and her family for love and support. Very special thank from the bottom of my heart to my mother **Smt. Pramila Thoke** for her boundless love and encouragement throughout my life. She had actively supported me in my determination to find and realize my true potential. My deepest gratitude to my father-in-law **Shri. Bhagwatrao Y. Barve** and mother-in-law **Mrs. Sushila Barve** for their love and continuous support.

The one person who has made this all possible is my husband **Prof. Dr. Shivprakash Barve**. He has been a constant source of support and encouragement. His enthusiasm and integral view on high research standards, greatly inspires me. His unconditional love, patience, and continual support for my academic endeavors enabled me to complete this thesis.

Finally, my children **Aaditya** and **Durva,** you are always been a source of happiness for me. No words can express my love for both of you.
Thank you one and all.

Mrs. Sunita Shivprakash Barve
Research Student

Abstract

Wireless spectrum is amongst the most heavily used and expensive natural resources around the world. Rapid growth in the wireless communication sector offers new wireless applications and services resulting in increased number of users. Most of the spectrum suitable for wireless communication is allocated to license holders or service providers on a long-term basis and for large geographical regions. A Large portion of the assigned spectrum is underutilized for a significant amount of time. A fixed amount of spectrum and growing number of wireless applications or users results in spectrum scarcity. Opportunistic spectrum access using cognitive radio technology enables exploring vacant spectrum bands, thereby reducing spectrum scarcity and improving the spectrum utilization. Spectrum agile cognitive radios can opportunistically identify the vacant portions of the spectrum and transmit on them while ensuring that the services of the license holders are not affected.

Multi-hop cognitive radio network opens up new and unexplored service possibilities enabling the wide range of pervasive communication applications. In the multi-hop cognitive radio network, cognitive nodes sense a wide range of spectrum for getting information about spectrum availability and follow specific policy to opportunistically use vacant portion. However, it will have a significant impact on the routing performance, as the reliable knowledge of topology and channel statistics are not available, especially in mobile and ad hoc cognitive radio network.

This work is proposing context aware, online and opportunistic routing algorithm using Multi-Agent Reinforcement Learning, to deal with uncertainties and dynamic environment of Mobile Cognitive Radio Ad hoc Network. Context awareness in multi-hop cognitive radio network enables sensing of the physical environment and adapt behavior accordingly. The observed context and strategic interaction among the multiple agents help to route the packet successfully from source to destination. The proposed routing scheme jointly addresses, link and relay selection based on transmission success probabilities.

Multi-Agent Reinforcement Learning based online and opportunistic routing meets the application demands of environmental aware computing. It learns from temporal differences without the need for model of the environment and finds strategies to attempt uncertainties. Reinforcement learning based agents are intelligent to learn optimal or near optimal solutions. The main goal is to allow the nodes to observe, learn and respond to the dynamic operating environment in an efficient manner. Nodes are dynamically learning about the spectrum and neighbouring node availability and adapt behavior as per the environment.

The behavior and channel usage pattern of the primary user is analyzed to predict chances of accessing spectrum band opportunistically. Non-deterministic traffic of the primary user produces the stochastic observable outcome in the form of idle and busy states which is characterized using Hidden Markov Model. Predicted channel availability and channel characteristics are used to select the link at every hop.

Every node in Mobile Cognitive Radio Ad hoc Network is an intelligent agent which tries to improve spectrum utilization and network performance. This Multi-Agent Reinforcement Learning based stochastic routing scheme optimally explores and exploits the opportunities in the network using Softmax action selection rule.

This sophisticated learning mechanism successfully explores opportunities in non-stationary environment of Mobile Cognitive Radio Ad hoc Network. Proposed algorithm outperforms the classical routing algorithms by reducing route re-discoveries, packet drops and control overhead. Temporal difference policy evolution successfully deals with the uncertainties and select the reliable path from source to destination. Experimental and simulation results show the effectiveness of this algorithm.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

Symbols	**Description**
APR	Average Per Packet Reward
ARP	Alternative Renewal Process
CA	Channel Availability
CBR	Candidate Best Relay
CC	Control Channel
CN	Cognitive Node
CP	Control Packet
CR	Cognitive Radio
CRN	Cognitive Radio Network
CU	Cognitive User
DC	Data Channel
DP	Data Packet
DSA	Dynamic Spectrum Access
ETT	Expected Time to Transmit
FCC	Federal Communications Commission
GPS	Global Positioning System

Symbols	**Description**
GOS	Generated Observation Sequence
HMM	Hidden Markov Model
ISM	Industrial, Scientific and Medical
LC	Link Cost
MAC	Medium Access Control
MARL	Multi-Agent Reinforcement Learning
MC	Monte Carlo
MCRAN	Multi-hop Cognitive Radio Ad Hoc Network
MDP	Markov Decision Process
OS	Observation Sequence
PD	Propagation Distance
PU	Primary User
RF	Radio Frequency
RL	Reinforcement Learning
RREP	Relay Reply
RREQ	Relay Request
SDR	Software Defined Radio
SU	Secondary User
TD	Temporal Difference
USRP	Universal Software Radio Peripheral
WiFi	Wireless Fidelity
WLAN	Wireless Local Area Network
WN	Willing Neighbour

Chapter 1

Introduction

Chapter 1

Introduction

1.1 Dynamic Decision Making

Dynamically complex and multi-dimensional problems in distributed environment requires dynamic decision making. Dynamic complexity refers to complex time evolutionary behavior and non-linear interaction between different elements of the system (Pruyt E., 2006). The system behavior on multiple dimensions complicates the strategy selection, especially when there is no best strategy on all dimensions over time. Dynamic decision making problems require understanding of control dynamics in complex and real world systems (Cleotilde et al., 2005). The following points differentiate Dynamic Decision Making from classical form of decision making:

- Series of decisions are taken to reach the goal instead of single decision
- Interdependence of previous decisions
- Dynamic or non-static environment
- Real time, on-line decision making

In dynamic decision making, environment changes autonomously or due to previous action selection of the decision maker. Dynamic decision making environment has following unique properties:

Dynamics: Characterized by constant change in an environment and dependency of the system on its previous state. Self correcting loops improve the performance in long run.

Complexity: The number of interacting elements within the system increases complexity which results in difficulty to understand the behavior of the system. Complexity is function of decision maker's ability to observe and analyze relationship and interaction among components of the system.

Obscureness: There are some imperceptible aspects of the system which result in obscureness. The ability to gather knowledge about components of the system helps to take better decisions.

Dynamic Complexity: The environment changes autonomously or due to previous action selection of decision makers which is referred as dynamic complexity. Dynamic complexity in environment makes it hard to understand and control the system.

To solve dynamic decision making problems, it is required to have context awareness of the dynamic environment. Parameters of the system should be adaptively adjusted based on the interactions with the environment and competitive or cooperative users. The existence of other multiple users characterizes these types of problems into multiple agent/decision maker interaction problems. Multiple agent interaction problems predict the system behavior and guide future action selection for all users. The involvement of multiple agents or decision makers having their own interest, goals, views and preferences complicate decision making procedure. High degree of uncertainties and dynamics of the broader context, directly or indirectly impact many current and future decisions. Decisions in distributed and dynamic environment have following characteristics:

- Decisions are taken by multiple agents using current observation sequence and perceived local environment.

- Environmental state transitions and the observed new information is used for further planning and decision making strategy.
- Current decision constrain future decisions.
- Inevitable changes in state of the environment making it difficult to find abstraction of reality.
- It is required to monitor and understand the temporal constraints on decision making in dynamic environment as there are non-linear and stochastic effects.

1.2 Context Awareness in Dynamic Environment

Traditionally each agent is trained to follow predefined set of rules. It is not possible to change these rules dynamically in response to the dynamic environment. Statically defined number of states and actions are not able to identify new states and events in uncertain environment. Statically defined behavior in dynamic environment results in suboptimal performance.

Context awareness refers to the system that senses the dynamic physical environment and adapts behavior accordingly. Any information or knowledge used for characterizing situations is called context (Dey A. K., 2001). Context awareness is concerned with acquisition of context using sensors to perceive situations and providing abstraction. Context awareness helps decision makers to,

- Decide applications behavior or action selection based on recognized context.
- Be able to observe the environment, learn from dynamics and respond to uncertain environment efficiently (Kok-Lim A. Y. et al., 2012).
- Be able to identify new states, situations and analyze them to learn near optimal behavior.
- Achieve adaptive behavior and intelligence by evaluating its actions with the help of interactive feedback from the environment.

1.3 Machine Learning for Context Awareness

Intelligent context aware agent needs to use collected knowledge from experiences, learn from dynamic changes and act optimally in complex and uncertain situations (Kulkarni P., 2012). Context aware systems achieve intelligence using machine learning. Machine learning has many facets like memorization, inference, etc. Learning enables agent to take better decisions in uncertain and dynamic environment.

Development and enhancement of computer algorithms with machine learning, meets the decision-making requirement in practical scenario. Machine learning algorithm collects and analyzes information from multiple sources. It picks up relevant data and uses it to decide behavior in similar situations. Nature of learning is classified into three broad categories depending on learning from, 'Signal' or 'Feedback'. These categories are:

Supervised Learning: Learning on class of examples that is inferring from labeled data. Agent learns the model using training sequence and evaluates the learned model using testing sequence. Every example in training sequence is a pair consisting a vector of input objects and the desired output value consisting of the supervisory signal. The accuracy of learning is defined by prediction of correct output signal for any valid input object.

Unsupervised Learning: Learning from unlabeled data that is finding hidden structure using similarity and differences. Training data provided for learning algorithm is unlabeled and there are no means for evaluating proposed solution. In unsupervised learning, neither explicit target outputs are associated, nor environmental evaluations for each input is given. A large amount of data is studied and analyzed to find statistical patterns to represent structure of the data.

Reinforcement Learning (RL): This class of learning algorithms includes decision making agent which learns by direct interaction with a dynamic environment. The agent learns the best policy that is sequence of actions that maximize total reward by trial and error interactions with the environment. In RL, agent

> needs to discover '*what to do* ' and '*how to map* ' situations corresponding to the action. Reinforcement differs from providing instructions. Instead of telling what to do, it expresses a reasoned opinion on agent's past performance. RL agent uses critic to learn and generate *internal value* of intermediate changes or actions in terms of how changes enable to reach the goal state.

The behavior of the agents can be programmed in advance, but to deal with the dynamic environment, agent should learn new behavior online by observing changing context. Continuous online learning of changing scenarios or context evolve with time to approximate the appropriate policy. Agent's systematic and continuous learning of dynamic environment gradually improves its performance and also the system's performance (Lucian B. et al, 2008). Reinforcement learning is for developing computing agents that can learn optimal behavior by analyzing context to decide policy in dynamic environment.

1.3.1 Reinforcement Learning

Reinforcement learning (RL) is aimed at discovering agent behavior so that a target task is achieved in an unknown environment. Reinforcement learning can be successfully utilized for dynamic environment and multi-agent systems, as it does not require any information about the dynamics of the environment (Abul et al., 2000). RL agent perceives environmental dynamics in the form of state and selects action as per probabilities in that state. As per the action performed, environment changes to a new state and agent receives a scalar reward. The reward is used to evaluate the correctness of action selection and policy under consideration (Hinojosa et al., 2011).

1.3.2 Agent-Environment Interaction

RL is mapping of dynamic scenarios to actions such that numerical reward is maximized (Sutton and Barto, 1998). Agent discovers new behavior by trying new action with more reward. Agent's trial and error interaction experience is used to determine desired behavior representing various dynamic scenarios.

Agent interacts with the external environment. It perceives the state of the environment at time t, where discrete time step $t = 0, 1, 2, 3, ...$. Depending upon state s_t, agent takes an action a_t, at time t. The agent receives a numerical reward $r_{t+1} \in \Re$ and environment changes to s_{t+1} as a consequence of its action at time step $t + 1$. Agent updates action selection probabilities at every time step, representing *agent's policy* at that instance. The agent's policy $\pi_t(s, a)$ is the probability of selecting $a_t = a$ in $s_t = s$ at time t. As per dynamic situation, current context and own experience, the agent changes its policy to increase the reward in long run.

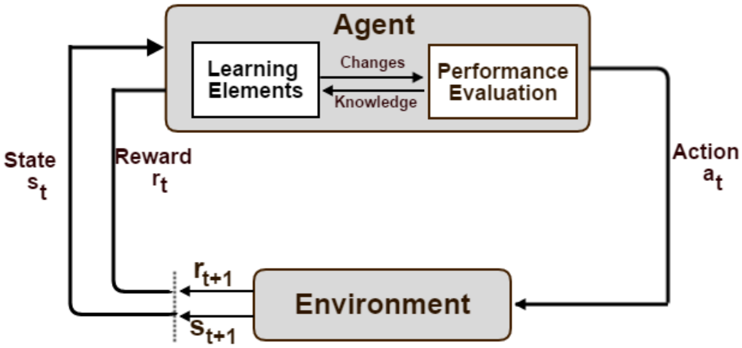


Figure 1.1: Agent-Environment Interaction

The agent and environment interaction is shown in Figure 1.1 with the help of three signals: state, action and reward. Agent evaluates its own actions and provides knowledge to learning elements for improving policy of action selection. This representation of the interaction between agent and environment is widely useful, applicable and sufficient to represent most of the decision-learning problems. Some of the important elements of RL are:

Policy: The policy is definition of action selection probabilities with respect to the particular state. The value function of the state S_t with respect to particular policy π is defined as cumulative numerical reward received by the agent and represented as $V^{\pi}(s_t)$.

Goal and Reward: The goal of the RL agent is formalized in the form of increasing cumulative numerical reward over the long run. Formalizing the idea of goal in terms of reward signal is a distinctive feature of reinforcement learning.

Returns: Is a specific function of the reward sequence up to final time step T, denoted as $R_T = r_{t+1} + r_{t+2} + r_{t+3} + ... + r_T$, where T is final time step. The RL agent tries to maximize the return over *Finite-Horizon* or *Episodic* task.

1.3.3 Markov Decision Process

Markov Decision Process (MDP) is a promising mathematical framework for modeling decision making in the environment where outcome is partially random and partially under control of decision maker/ learner/ agent. MDP is useful for studying optimization problems solved via reinforcement learning. RL task satisfying the Markov property is represented using MDP, where the state and reward in next time step depends only on the current state and action. A finite MDP is defined as a tuple $\prec S, \mathcal{A}, f, \rho \succ$ where:

- Set S represents the state space
- Set $\mathcal{A}$ is action space
- $f: \quad S \times A \Rightarrow S$, *State Transition Probability Function*, where $s_{k+1} = f(s_k, a_k)$
- $\rho: \quad S \times A \rightarrow R$, *Reward Function*

The stochastic process represented using MDP is in one of the state s at time t. Agent chooses action a available in s. The process transits randomly in state s' in next time step $t+1$ and agent gets reward of $R^a_{ss'}$ using reward function ρ. Chosen action in current state s influences the probability of process moving into its new state s'. Specifically, it is denoted by the transition probability $P^a_{ss'}$. Thus the next state s' depends on the current state s and the decision maker's action a. The state transition function f of an MDP process satisfies the Markov property, as given by s and a. The next state s' is conditionally independent of all previous states and actions. These quantities, $R^a_{ss'}$ and $P^a_{ss'}$ specify all the important aspects of the dynamics of MDP (Huang Z. et al., 2011).

1.3.4 Value Function

The value function is the function of the state to represent the effectiveness of the state for the agent and it is defined in terms of expected future reward. As the future benefits/rewards are dependent on choice of action, value function is defined with respect to policy under consideration.

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 the policy π. If agent chooses to follow the policy π and maintain average of actual return for each state encountered, then average will represent the state's value $V^{\pi}(s)$. As this method of estimation involves averaging over random sample of actual returns, it is *Monte Carlo* method of estimating value function.

If there are many states, then maintaining separate average of every state is not practical. Instead, $V^{\pi}(s)$ is maintained as parametrized function which tunes the parameters to better match the observed return. This kind of approach can produce accurate estimate depending on the nature of the parametrized function approximation used.

1.3.5 Action Selection

The important and distinguishing feature of RL is that it evaluates the action taken rather than instructing with correct action, like in other forms of learning. This requires active exploration using trial and error search for finding good behavior. In the simplest form of reinforcement learning task, agent has to choose among n different options or actions. After each action selection, agent receives a numerical reward chosen from a probability distribution as per action selected. Agent's objective is to maximize the expected total reward over multiple runs.

Expected reward with every selected action decides the value associated with that action. Agent maintains the estimate of action values. At any given time, the action with the maximum value is called a *Greedy* action. The selection of greedy action is *exploiting* current knowledge of the values of the actions. Selecting any non-greedy action is perceived as *exploring*, as it enables to improve value estimate of non-greedy action's. Exploitation may maximize the expected reward on an individual instance, but exploration may produce greater total reward in the long run. In any specific case or application, whether it is better to explore or exploit depends on precise values of estimate, degree of uncertainties and the number of remaining plays. Action selection in RL has following properties:

- The reward received by the agent after each action selection gives information about goodness of the action.
- Received reward doesn't say anything about action's correctness or incorrectness, or whether it is the best or worst action selection.
- Correctness is relative property of actions that can be determined only by trying all of them and comparing their reward.
- The agent inherently requires explicit search among the alternative actions using generate-and-test method.
- Agent tries actions, observes their outcomes and selectively retains those that are most effective. This is *learning by selection* instead of *learning by instruction*.

1.4 Cognitive Radio Network

Rapid growth in the wireless communication sector has resulted in increased number of users and number of new wireless applications being offered. These applications demand more spectral bandwidth to support diverse services like Voice Telephony, Web Browsing, Text and Multimedia Messages. A fixed amount of the spectrum versus growing number of wireless applications or users results in spectrum scarcity due to frequency allocation structure of command-and-control. Wireless spectrum is amongst the most expensive and in-demand natural resource around the world. Large portions of the spectrum suitable for wireless communication are allocated to license holders who maintain exclusive rights to their allocated spectrum. According to the statistics of the Federal Communications Commission (FCC), many portions of the allocated radio spectrum are underutilized for a significant amount of time. The temporal and geographical variations in the utilization of the assigned spectrum which range from 15 % to 85 %.

The limited available radio spectrum and the inefficiency in spectrum usage necessitates a new communication paradigm to exploit the existing spectrum dynamically. Dynamic Spectrum Access (DSA) is proposed to be an efficient technique for solving current spectrum scarcity problem. With DSA, unlicensed users may use licensed spectrum bands opportunistically in a dynamic and non-interfering manner. This is achieved by recent advancements in software defined radios, allowing to operate on any frequency band in a wide spectrum range with minimum channel switching delay. This spectrum agile device is known as *Cognitive Radio (CR)*, which is able to opportunistically identify the vacant portions of the spectrum and transmit on them, while ensuring that the Primary User's (PUs) service is not affected (Mitola and Maguire, 1999).

CR nodes networked together which form the Cognitive Radio Network (CRN), provide high bandwidth to mobile users via heterogeneous wireless architectures and dynamic spectrum access techniques. Spectrum scarcity and under utilization problems can be solved by opportunistically accessing the underutilized licensed bands without interfering with existing users.

1.4.1 Cognitive Radio

Cognitive radio is formally defined as (Akyildiz I. et al., 2006),

A cognitive radio is a smart radio, having the ability to sense the external environment, learn from history, and make intelligent decisions to adjust its transmission parameters according to the current state of the environment.

Cognitive Radio is an intelligent wireless communications system using reconfiguration ability and agile functionality of Software Defined Radio (SDR). Cognitive Radio (CR) is built using,

Cognitive Modules: Cognitive(sensor) modules sense the environment, learn from the history of sensed observations (modeler), and adjust the transmission parameters of the SDR. These modules help to find unused portion of the spectrum and select best available channel with its parameter for transmission at particular time and location.

Reconfigurable SDR: SDR provides capability to dynamically change transmission parameters of the radio as well as modulation / demodulation scheme in software. It dynamically programs cognitive radio to communicate on a variety of frequencies according to the radio environment.

CR is having awareness of its environment. It learns from environment about spectrum occupancy, network traffic and adapts to new scenarios based on current situation and previous experiences.

1.4.2 Capabilities of Cognitive Radios

The capabilities of Cognitive Radios are:

- **Frequency Switching:** A CR can dynamically select the operating frequency based on current context of spectrum usage.

- **Adaptive Transmission Characteristics:** Transmission characteristics are updated dynamically to exploit opportunities for spectrum usage.

- **Transmitting Power Control (TPC):** A CR can vary the transmission at allowable limits whenever necessary. It can also reduce it to a lower level for greater sharing among multiple nodes.

- **Geo-Location determination:** A CR can incorporate ability to determine location of own and the location of other transmitters (e.g. by using GPS). The appropriate operating parameters such as the power and frequency are selected as per location.

- **Spectrum negotiation:** A CR may incorporate capabilities to negotiate spectrum access on an ad hoc or real-time basis.

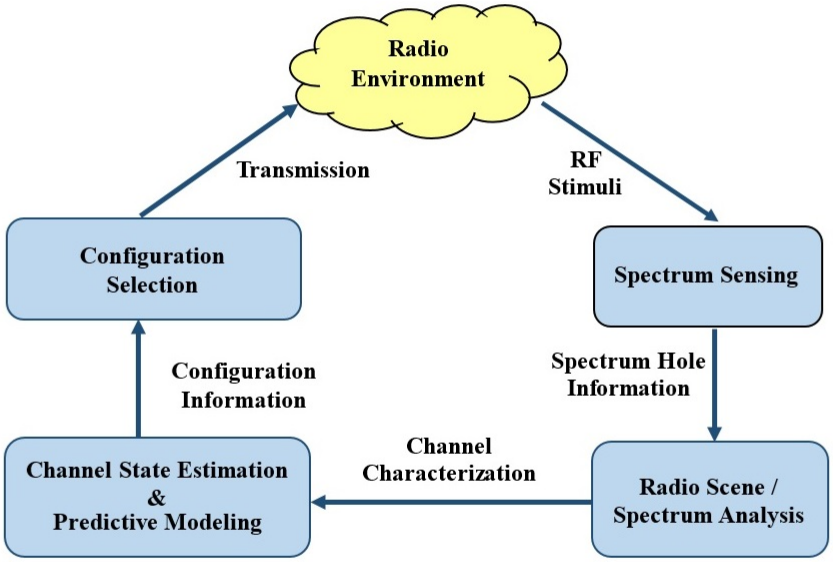

Radio Environment
Transmission
RF Stimuli
Configuration Selection
Spectrum Sensing
Configuration Information
Spectrum Hole Information
Channel Characterization
Channel State Estimation & Predictive Modeling
Radio Scene / Spectrum Analysis

Figure 1.2: Cognitive Radio Cycle

1.4.3 Cognitive Cycle

Cognitive Radios determine their behavior in a reactive or proactive manner based on the external environmental stimuli, as well as their goals, capabilities, experiences and knowledge (Barve S., 2014).

Phases of the cognitive cycle are shown in Figure 1.2. Details of these phases are as follows:

- **Spectrum Sensing**: Here available spectrum bands are monitored and sensed to capture their information e.g. interference, spectrum holes etc.
- **Radio Scene Analysis**: The characteristics of the detected spectrum holes are determined (Cormio and Chowdhury, 2009).
- **Channel State Estimation and Predictive Modeling**: During this phase, capabilities of configuration are discovered based on spectrum sensing results and exploiting past experiences and knowledge (Sharma M. et al., 2007).

- **Configuration Selection**: Best configuration is selected based on the performance requirements of the cognitive user.

1.4.4 Architecture of Cognitive Radio Network

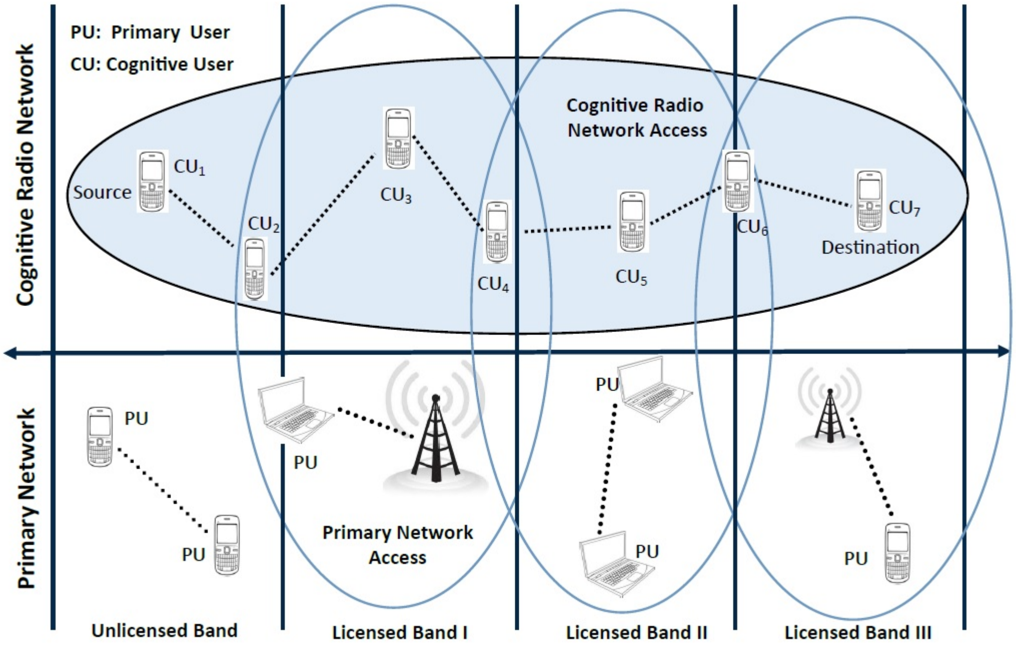


Figure 1.3: Architecture of Cognitive Radio Network

The most general architecture of the CRN is shown in Figure 1.3, which distinguishes two types of users sharing common spectrum:

- **Primary Users (PUs)**: Have exclusive right and priority in spectrum utilization within the band, they are licensed.
- **Cognitive Users (CUs)**: Have access to the spectrum in a non-interfering manner with PU. Secondary user is equipped with CRs having additional functionality of reconfiguration and cognitive capability to explore spectrum access opportunities.

CRN operates in a mixed spectrum environment with both licensed and unlicensed bands. CU should access the spectrum in an opportunistic manner. CRN can be deployed as infrastructure network or ad hoc network,

- Infrastructure based CRNs can access their own base station or base station of the primary network.

- In ad hoc CRN, CUs can communicate with each other on licensed or unlicensed spectrum band.

CUs can operate in both licensed and unlicensed band, with varying functionality. CUs on licensed band can be deployed to exploit the spectrum holes. The main concern for CU on licensed band is detection of the PUs, selection of the best spectrum for communication, interference avoidance with PUs and spectrum handoff in case of PU appearance.

CRN in Unlicensed/ Industrial, Scientific and Medical (ISM band), is deployed for fair spectrum usage policy among different users (Jian Tang et al., 2010). Due to an open spectrum policy of ISM band, innovative applications and technologies have increased causing interference among multiple networks. CUs should cooperate with each other for sophisticated spectrum sharing. Cognitive Radio Network however, imposes various challenges due to the wide range of intermittent spectrum, heterogeneous wireless architecture and diverse service requirements of applications.

1.4.5 Context Awareness in Cognitive Radio Network

This work advocates the use of RL to achieve context awareness. The capability enhancement using RL is of paramount importance for general functionality of CRN. Context awareness using RL has gained tremendous popularity for offering substantial network wide performance enhancement (Kok-Lim A. Y. et al., 2012).

- Every host in CRN should be continuously aware of its surrounding physical environment for taking more accurate decisions.

- Each decision or action selection in CRN require knowledge about PU activity, spectrum occupancy, multiple cognitive users with their strategies and mobility of hosts.

- Action selection based on this gathered context or collected knowledge, helps to improve performance of CRN.

1.5 Routing in Cognitive Radio Network

Multi-hop cognitive radio network opens up new and unexplored service possibilities enabling wide range of pervasive communication applications (Cesana M. et al., 2011). Multi-hop CRN can be constructed by relaying the information through intermediate nodes between source and destination. There are two main reasons behind the need of multi-hop CRNs:

1. Due to opportunistic access and intermittent spectrum availability every source - destination pair cannot share common communication channel. Intermediate nodes between source and destination can be used to handle this heterogeneity. These intermediate nodes can switch between the common channels along the path to reach appropriate destination.

2. Use of shorter distances with less transmission power of channel meets the transparency requirement of primary network, instead of longer distances with more power. Reduced transmission power maintains channel quality and requirement of signal-to-noise-ratio for shorter distances. Therefore, longer distance between source and destination is divided by few shorter distances having same channel quality using intermediate relay nodes.

In Multi-hop CRN, cognitive nodes sense a wide range of spectrum and get information about spectrum occupancy and availability. These cognitive nodes then use specific policies to select one out of the available spectrum bands. The set of available channels from node to node are different. Moreover, the existence of multiple users (Primary or Cognitive) and their varying demands of transmission have great impact on path selection and routing.

Well-designed multi-hop CRN's provide high bandwidth efficiency, extended coverage and ubiquitous connectivity for wireless users. However, special features of

CRN raise several unique challenges due to high fluctuation in available spectrum and incomplete knowledge of the topology, especially for the routing process.

*The **problem of routing** in multi-hop CRNs targets the creation and maintenance of the wireless multi-hop path among cognitive nodes by deciding both relay node and the channel to be used on each link of the path.*

Traditional routing algorithms will be inefficient to handle dynamism of CRN. In CRN, quality of the end-to-end route is not only dependent on throughput, bandwidth, and delay, but also on path stability, interference and behavior of multiple users (PU or CU) competing for same spectrum resources. The dynamic spectrum which is intermittent in terms of both time and space keeps surrounding environment dynamic, resulting in reallocation of spectrum and forwarding nodes. This necessitates the collaboration between spectrum management and routing modules.

1.6 Research Challenges

There exist several research challenges that need to be investigated for providing the efficient routing solution in cognitive radio network:

- **Spectrum Awareness**: Tight coupling between spectrum management module and routing module is necessary for formulating efficient routing solution. The routing module should be context aware and collaborative with the entire cognitive cycle to make more accurate decisions.

- **Exchanging Routing Information**: In highly mobile and dynamic environment, information exchange is the only means of creating view of the overall topology of a network. As the CUs are using spectrum as a visitor, use of common control channel for information exchange can also be affected by PU activities (Jiao and Yuqing, 2010). Modeling of the routing solution by considering the above facts will significantly improve its performance.

- **Dynamic Topology and Intermittent Connectivity**: The connectivity between neighbouring nodes is significantly affected due to the mobility of CU or the appearance of PU. This results in rapid change in reachable neighbouring nodes. For this reason, prediction based topology construction can be used for the creation of more stable paths (Guan Q. et al., 2010).

- **Route Maintenance**: The selected route may become unavailable due to intermittent spectrum or mobility of the node. The routing should be reactive to these spectral and topological changes. Re-routing not only changes the spectrum to be used but also changes the participant nodes in path formation. Therefore, it will be beneficial to design spectrum aware routing solution.

- **CUs Behavior**: Spectrum Utilization among multiple CU at every hop along the path can be improved by jointly studying their behavior. The main challenge is that each cognitive node must explicitly consider other cognitive nodes, and coordinate their behavior with each other, such that it results in a coherent joint behavior.

1.7 Motivation and Problem Definition

Motivations behind the study of routing in Cognitive Radio Network are:

- CRN is a complex and dynamic system with various uncertain factors such as unstable topology, intermittent channels and node availability which affect performance in time varying and complex manner.

- Rule or Policy based actions are not able to deal with continually changing environment.

- Context awareness of environmental dynamics is necessary for every agent or node before selection of any action.

- Every agent should be capable of learning in an online and incremental manner for predicting future behavior on the basis of past experiences and current statistics.

Routing in CRN is different from routing in classical wireless network, due to intermittent spectrum resource availability and unstable topology. Conventional routing metrics such as hop count, congestion etc, are not sufficient for taking routing decisions in CRN. Routing is a challenging task especially in multi-hop CRN due to the diversity in channel availability, neighbouring nodes availability and behavior of licensed user.

1.7.1 Application Demands

Following are some specific application demands for providing efficient routing solution in CRN,

Environment and Context Aware Computing: Agent should be aware of its current state of the environment, that is spectrum and neighbouring node availability in network.

Strategies to Attempt Uncertainties: Agent should be able to address sudden, automatic and autonomous changes of the environment, which are not under control of decision maker.

Need for Adaptive and Opportunistic Routing: Agent should be able to observe, learn and act to optimize own performance as well as system's performance by improving utilization of dynamically available spectrum.

Learn Without Model of the Environment: Considers no or limited knowledge of the topology and simultaneously learns channel statistics and optimal routes.

Learning has been a core idea in cognitive radio since its origin, as one of the steps in cognitive cycle proposed by Mitola (Mitola and Maguire, 1999). It is advantageous to bring the power of machine learning for dynamic channel selection, dynamically learning channel statistics and routing in CRN. There is a need to improve performance by incorporating continuous learning and enabling nodes to adapt their routing strategies accordingly. RL based agents can add intelligence to the system which learns and analyzes for finding optimal or near optimal solutions. As the environment in CRN is dynamic and unstable, there is a need to learn from or evaluate every action selected by the agent. Instead of learning from outcome of series of action selection or from outcome of particular episode, agent should be able to measure difference between temporally successive predictions.

1.7.2 Problem Statement

> ***To design and implement Multi-Agent Reinforcement Learning based spectrum aware opportunistic routing in Cognitive Radio Network such that average per packet reward is maximized.***

1.8 Proposed Solution and Major Contributions

The proposed routing scheme utilizes a reinforcement learning framework to opportunistically route the packets even if reliable knowledge about channel statistics and network model is not available. The characteristics of proposed solution are:

- It jointly addresses the issues of learning and routing in an opportunistic context, where the network structure is characterized by the transmission success probabilities.

- It is a stochastic routing scheme that optimally explores and exploits the opportunities in the network.

- Every node is intelligent agent, able to observe, learn and respond dynamically in an efficient manner.

- The proposed solution opportunistically routes the packets even in the absence of reliable knowledge about channel statistics and network topology.

- Multi-Agent Reinforcement Learning (MARL) and Hidden Markov Model (HMM) based channel and relay selection algorithm enable the network users to identify the target spectrum band while achieving performance requirements and maximizing spectrum usage efficiency.

- Proposed algorithm considers no knowledge of the topology and learns channel statistics and efficient routes simultaneously.

1.8.1 Research Contributions

The main contributions are as follows:

1. Link selection based on probability of the channel being available and forecasting using time series prediction.
2. Markov Decision Process formulation using Temporal Difference (TD) implementation of Reinforcement Learning with incomplete and erroneous information of topology.
3. Strategic interaction among multiple agents for context awareness in dynamic environment.
4. Adaptive and opportunistic routing algorithm based on the transmission success probabilities.
5. Softmax Action Selection for balancing exploration and exploitation in relay selection process.

1.9 Thesis Organization

The thesis is organized as follows:

- **Chapter 2** provides context of the field and intellectual progression of cognitive radio networks. A review of the state of the art research work is given. The chapter demonstrates current issues being debated and how they are addressed by existing literature.

- **Chapter 3** defines the problem under consideration with research objective and approach used to solve the defined problem.

- **Chapter 4** gives overview of research methodology used. The purpose is to present benefits of proposed methodology towards dynamic environment of cognitive radio network.

- **Chapter 5** provides novel multi agent reinforcement learning based online and opportunistic routing algorithm for CRN in detail.

- **Chapter 6** focuses on implementation details and analysis of channel selection generating the context of environmental statistics.

- **Chapter 7** shows implementation details and analysis of MARL based online and opportunistic routing in mobile and ad hoc cognitive radio network.

- **Chapter 8** discusses unique characteristics and performance improvement gained with context awareness in channel and node selection using MARL based routing.

- **Chapter 9** draws conclusions and discusses future research directions.

Chapter 2

Background and Literature Review

Chapter 2

Background and Literature Review

2.1 Background

Cognitive Radio Network (CRN) is a network in which each node is equipped with Cognitive Radio (CR). CR is aware of the dynamic environment and it adaptively adjusts operating parameters based on interaction with the environment and other users of the network. These networks are offering tremendous performance and operational benefits by providing high bandwidth to mobile users via dynamic spectrum access techniques. CRN, however, imposes several research challenges due to broad the range of available spectrum and diverse quality of service requirement of applications (Cesana M. et al., 2011).

Most of the recent research on CRN was concentrating on the physical and Medium Access Control layer, focusing on efficient spectrum sensing, management and sharing techniques. In addition to this, routing is also important challenge for realization of CRN, especially in the networks with multi-hop communication requirements. The design goal of multi-hop CRN requires integration of cognitive principles and the rules of interaction among multiple nodes. The set of wireless nodes should form a social network which must be modeled and analyzed as one entity, in order to optimize the network functions.

Cognitive nodes sense wide range of spectrum to identify spectrum holes of various underutilized spectrum bands. It then uses specific policies to select one of the spectrum band for transmission. In Multi-hop CRN, the set of available channels from node to node are different and not static. Moreover existence of Primary Users (PU) and Cognitive Users (CU) with their varying demands of transmission have great impact on path selection.

Traditional routing algorithms will be inefficient to handle dynamism of CRN:

- In CRN, quality of the end-to-end route is not only dependent on throughput, bandwidth, and delay but also on path stability and presence of multiple users competing for same spectrum resource.
- The dynamic spectrum which is intermittent in terms of both time and space is responsible for changes in surrounding environment.
- More amount of control overhead due to reallocation of the spectrum and rerouting.

Dynamic Spectrum Access and routing are challenging problems due to coexistence of both primary and cognitive users. Every node at each hop should be able to adapt to changing spectrum resource, learning about the spectrum occupancy and its history and making decisions on the suitability of the available spectrum resource. Therefore, it is essential to study the intelligent behavior and interaction of multiple network users competing for spectrum resource.

The aim of this chapter is, to provide detailed insight of routing problem in multihop CRN in the context of dynamism due to intermittent time and space availability of spectrum. The basic objective is to provide overall view of the field with focus on routing under consideration of multiple users sharing same spectrum resource.

2.2 Continuous Evolution of Cognitive Radio Network

'Cognitive radio', was proposed by Joseph Mitola III (Mitola and Maguire, 1999), as novel approach in wireless communication described as computationally intelligent wireless devices understanding users communication requirements and provide wireless services and radio resources as per these requirements. The cognitive radio with its continuous evolution is shown below in Figure 2.1,

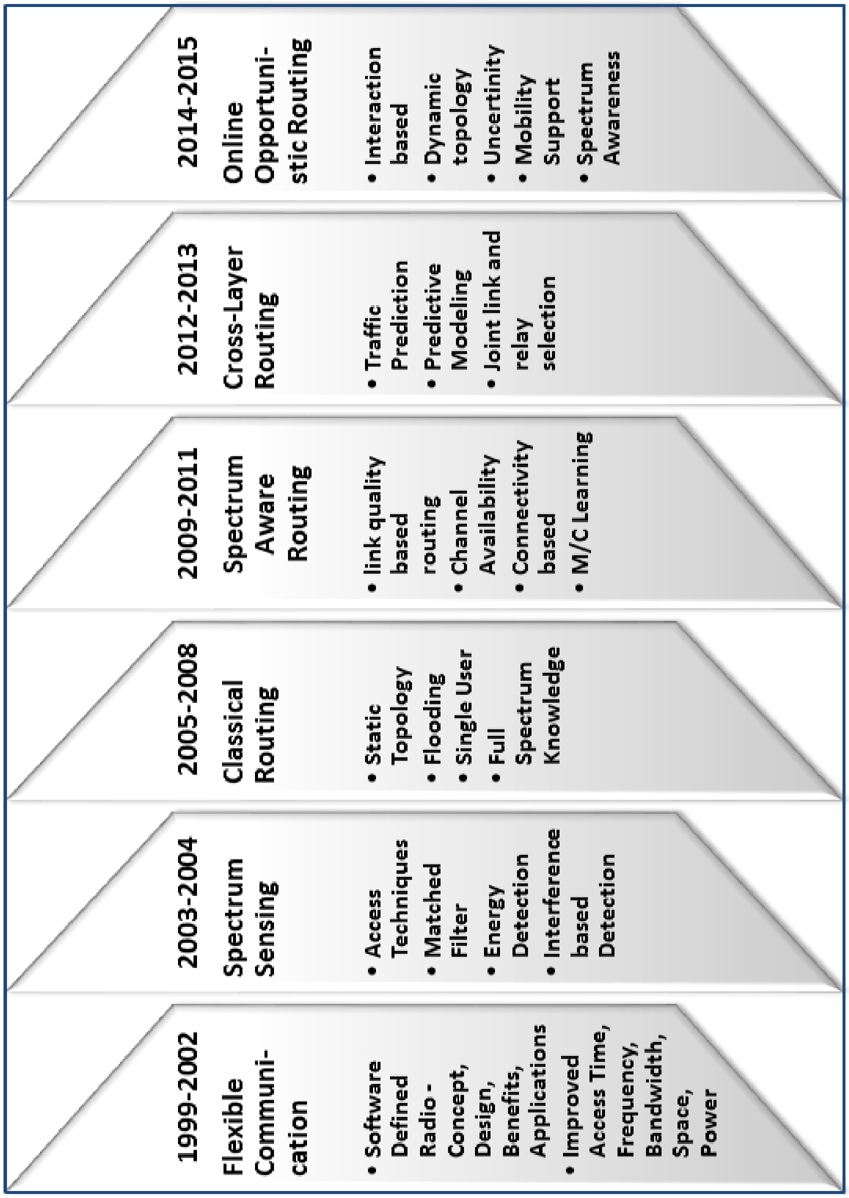


Figure 2.1: Continuous Evolution

Cognitive Radio Network evolved as a solution to spectrum scarcity of a very important and costly natural resource. After it was proposed by Joseph Mitola III, various researchers had worked from practical implementation to various issues in communication protocol at different layers. Table 2.1 to Table 2.9 is showing continuous evolution in the field of Cognitive Radio Network from year 1999 to year 2015.

Table 2.1: Continuous Evolution of CRN: 1999-2005

Title	**Author**	**Publication**	**Findings**
Cognitive Radio: Making Software Radios More Personal	Joseph Mitola and Gerald Q. Maguire	IEEE Personal Communications 6(4), **1999**	Radio-Domain-Aware Intelligent Agent, Radio Knowledge Representation Language, Cognitive Radio as Chess Game
The Software Radio Concept	Enrico Buracchini	IEEE Communications Magazine, **2000**	Benefits of Cognitive Radio to Manufacturer, Operator and Users, Increased H/W Lifetime
Cognitive Radio for Flexible Mobile Multimedia Communications	Joseph Mitola	Mobile Networks and Applications 6, **2001**	Flexible pooling of radio spectrum, model based reasoning, Observe-Think-Act
FCC Report of the Spectrum Efficiency Working Group	Engelman R. et al.	Federal Communications Commission, **2002**	Improved access through Power, Time, Frequency, Bandwidth and Space, Permitting other users uses, Adjusting regulations as technology develops
Cognitive Radios will Adapt to Users	Costlow. T.	IEEE Intelligent Systems 18(3), **2003**	Adaptability of Cognitive Radio and its applications are explored, Various research challenges are discussed
Implementation Issues in Spectrum Sensing for Cognitive Radios	Cabric D. et.al	Signals, Systems and Computers 1, **2004**	Spectrum Sensing through Matched Filter, Energy Detection and Cyclo-stationary Feature Detection
Ad-hoc Cognitive Radio - Development to Frequency Sharing System by using Multi-hop Network	Fujii T and Suzuki Y	IEEE International Symposium DySPAN, **2005**	Frequency sharing with adaptive route selection as per the surrounding radio environment in multi-hop ad-hoc cognitive radio network

Table 2.2: Continuous Evolution of CRN: 2006-2008

Title	Author	Publication	Findings
Outage Performance of Cognitive Wireless Relay Networks	Kyounghwan L. and Yener A	IEEE Global Telecommunications Conference, **2006**	Intra-cluster cooperation scheme along with the system level cooperation via relaying through cognitive nodes to improve the outage performance
Applications of Machine Learning To Cognitive Radio Networks	Clancy C. et.al.	IEEE Wireless Communication, **2007**	First-order logic to represent state of environment and actions, Incorporate learning engine into predicate calculus based reasoning engine
Enabling Open-Source Cognitively-Controlled Collaboration Among Software-Defined Radio Nodes	Troxel G. D. et. al.	Elsevier Journal on Computer Networks 52, **2008**	Real-time software defined data radio using open source GNU radio platform. Fine grained cognitive control radio for future wireless communication
Spectrum Sharing for Multi-hop Networking with Cognitive Radios	Hou Y. T. et. al.	IEEE Journal on Selected Areas in Communication 26(1), **2008**	Modeling problem of spectrum sharing, scheduling, interference control and flow routing as mixed integer non-linear problem solved using sequential fixing of integer variables
SAMER: Spectrum Aware MEsh Routing in Cognitive Radio Networks	Pefkianakis I. et. al.	IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, **2008**	Builds run-time forwarding mesh adapting to the dynamic spectrum conditions and link availability
Channel Modeling Based on Interference Temperature in Underlay Cognitive Wireless Networks	Sharma M. et al.	IEEE International Symposium on Wireless Communication System, **2008**	HMM is trained with observed interference temperature values to predict future channel availability to select preferable channel
Neural Network-Based Learning Schemes for Cognitive Radio Systems	K.Tsagkaris et. al.	Elsevier Journal of Computer Communications 31, **2008**	Assist to predict capabilities for specific radio configuration, Extending learning by feeding time zone information

Table 2.3: Continuous Evolution of CRN: 2009

Title	Author	Publication	Findings
Joint Design of Spectrum Sharing and Routing with Channel Heterogeneity in Cognitive Radio Networks	Ma M. and Tsang D.	Elsevier Journal of Physical communication 2(1-2)	Maximize flow percentage of traffic demand, cross layer optimization framework for joint spectrum sharing and routing as mixed integer linear programming
ASAR: Ant-based Spectrum Aware Routing for Cognitive Radio Networks	Bowen L. et.al.	Proc. of International Conference on Wireless Communications and Signal Processing	Biologically inspired approach, F-ants to discover spectrum feasible paths and B-ants to update routing information
Search: A Routing Protocol for Mobile Cognitive Radio Ad-hoc Networks	Chowdhury and Felice et.al.	Elsevier Journal of Computer Communications 32	Shortest path based on greedy advancement towards destination traversed on the combination of channel to the destination
Reinforcement Learning Based Spectrum-Aware Routing in Multihop Cognitive Radio Networks	Xia B. et.al.	International Conference on Cognitive Radio Oriented Wireless Networks and Communications	Estimation of available channels on the route with forward and backward exploration
ROPCORN: Routing Protocol for Cognitive Radio Ad-hoc Networks	Talay and Altilar	International Conference on Ultra Modern Telecommunications	Routing based on spectrum availability and load estimation, the link state database is maintained and any change in link state triggers fresh next hop route computation
Multihop cognitive radio networks: to route or not to route	Khalife H et al.	IEEE Network Magazine 23(4)	Appropriate routing approach chosen depending on the specific environment, the traffic to be carried out, availability of the primary bands and their history in considered environment

Table 2.4: Continuous Evolution of CRN: 2010

Title	Author	Publication	Findings
A Geometric Approach to Improve Spectrum Efficiency for Cognitive Relay Networks	Xie M. et al.	IEEE Transactions On Wireless Communications, 9(1)	Geometric Conditions of node are used to select potential relay, Shorter distance less power transmission with concurrency
Routing and QoS Provisioning in Cognitive Radio Networks	How K. C. et al.	Elsevier Journal of Computer Network, 55(1)	Multi-metric Route selection considering switching delay and queuing delay, Transmit Power Control for traffic priority
Routing in Cognitive Radio Networks:Challenges and Solutions	Cesana M et al.	Elsevier Journal of Ad Hoc Networks 9(3)	Extensive overview of the routing in CRN, under consideration of two main categories: Full and Local spectrum knowledge
IPSAG: An IP Spectrum Aware Geographic Routing Algorithm Proposal for Multi-hop Cognitive Radio Networks	Badoi C. I. et al.	IEEE 8th International Conference on Communications	Geographic routing based on hop by hop forwarding with local and global information.
Cross-Layer Routing and Dynamic Spectrum Allocation in Cognitive Radio Ad Hoc Networks	Lei D et al.	IEEE Transaction On Vehicular Technology 59(4)	Taking decision based on locally collected spectrum and power allocation information, opportunistically calculates the next hop depending upon queuing and spectrum dynamics
Prediction-Based Topology Control and Routing in Cognitive Radio Mobile Ad Hoc Networks	Guan Q et al.	IEEE Transactions On Vehicular Technology,59(9)	Mobility based link Prediction, Distributed prediction based topology control, Distributed Dijkstra algorithm preserves global connectivity

Table 2.5: Continuous Evolution of CRN: 2011

Title	Author	Publication	Findings
CRP: A Routing Protocol for Cognitive Radio Ad-Hoc Network	Chowdhury and Akyildiz	IEEE Journal on Selected Areas in Communications 29(4)	Primary receiver protection with increased distance and transmission time or Reduced distance with overlap between PU and SU
End-to-end Protocol for Cognitive radio Ad hoc Networks: An Evaluation Study	Marco D. F. et al.	International Journal on Performance Evaluation 68(9)	TCP performance with factors like spectrum sensing cycle, interference from primary user and channel heterogeneity
Joint Routing and Spectrum Allocation for Multi-Hop Cognitive Radio Networks with Route Robustness Consideration	Shih C. et al.	IEEE Transaction on Wireless Communication 10(9)	The aggregate throughput and the robustness of routes determined by proposed scheme guarantees a basic level of robustness for a set of routes
Multi-cast Communications in Multi-Hop Cognitive Radio Networks	Gao C. et. al.	IEEE Journal on Selected Areas in Communications 29(4)	The goal is to support a set of multicast sessions with a given bit rate requirement with minimum network-wide resource. Cross-layer approach with joint consideration of scheduling and routing
Stability-Capacity-Adaptive Routing for High Mobility Multi-hop Cognitive Radio Networks	Huang X. L. et al.	IEEE Transactions On Vehicular Technology 60(6)	Bird-flocking mobility model, Link availability probability integrated with clustering, Node importance degree based on common channel

Table 2.6: Continuous Evolution of CRN: 2012

Title	Author	Publication	Findings
Spectrum-Aware Opportunistic Routing in Multi-Hop Cognitive Radio Networks	Liu Y et al.	IEEE Journal On Selected Areas In Communications 30(10)	Exploit the geographic location information and discover the local spectrum access opportunities
Opportunistic Spectrum Access in Cognitive Radio Networks: Global Optimization Using Local Interaction Games	Xu Y. et.al.	IEEE Journal of Selected Topics In Signal Processing 6(2)	Global optimization for distributed channel selection using local heuristic game and congestion games
Reactive Routing for Mobile Cognitive Radio Ad hoc Networks	Cacciapuoti A S et al.	Elsevier Journal of Ad-Hoc Networking 10	Cognitive ad-hoc on demand distance vector routing with inter and intra route spectrum diversity
On Routing and Channel Selection in Cognitive Radio Mesh Networks	Mumey B. et al.	IEEE Transactions on Vehicular Technology 61(9)	Selecting the channels on given routing path such that the end-to-end throughput is maximized using dynamic programming-based approach
Probabilistic Quality-Aware Routing in Cognitive Radio Networks under Dynamically Varying Spectrum Opportunities	Badarneh O. S. and Salameh H. B.	Elsevier Journal on Computers and Electrical Engineering 38	Joint probabilistic routing and channel assignment protocol, selecting the path with the maximum probability of success among all possible paths
Adaptive Opportunistic Routing for Wireless Ad Hoc Networks	Bhorkar A. et al.	IEEE/ACM Transactions On Networking 20(1)	Jointly learn and route in an opportunistic context characterized by the transmission success probabilities

Table 2.7: Continuous Evolution of CRN: 2013

Title	Author	Publication	Findings
Resource Allocation in Cognitive Radio Relay Networks	Liang and Chen	IEEE Journal on Selected Areas In Communications 31(3)	Centralized proportional fair scheduling with the effect of transmit power control and volatility of usable frequency bands
Self Adaptive Routing for Dynamic Spectrum Access in Cognitive Radio Networks	Talay A. C. and Altilar D. T.	Elsevier Journal of Network and Computer Applications 36	Aims to choose optimal routes at the outset of routing and retain optimal route by the use of route adaptation and route preservation
Interference Aware Routing Using Network Formation Game in Cognitive Radio Mesh Networks	Zhou Y. et al.	IEEE Journal on Selected Areas in Communications 31(11)	Distributed algorithm for the network formation game to minimize aggregate interference from the Secondary users to the Primary users
A Cross-Layer QoS-Aware Communication Framework in Cognitive Radio Sensor Networks for Smart Grid Applications	Shah G.A et al.	IEEE Transactions on Industrial Informatics 9(3)	Traffic flows are differentiated in priority classes according to QoS needs and service queues are maintained attributing delay, bandwidth and reliability of data.
SURF: A Distributed Channel Selection Strategy for Data Dissemination in Multi-hop Cognitive Radio Networks	Rehmani M. H. et al.	Elsevier Journal on Computer Communications 36	Distributed channel selection strategy for efficient data dissemination based on primary user unoccupancy and number of cognitive radio neighbours using the channels.
A Routing Protocol for Cognitive Radio Ad Hoc Networks Giving Consideration to Future Channel Assignment	Wu C. et al.	First International Symposium on Computing and Networking	AODV based protocol chooses a route by considering the effect on the primary users, available channel bandwidth and link reliability

Table 2.8: Continuous Evolution of CRN: 2014

Title	Author	Publication	Findings
A Resource Intensive Traffic-Aware Scheme Using Energy-Aware Routing in Cognitive Radio Networks	Bourdena A. et al.	Elsevier Journal of Future Generation Computer Systems 39	Traffic aware scheme for minimizing energy consumption and maximizing resource exchange between secondary communication nodes
QoS Multi-cast Routing Protocol Oriented to Cognitive Network Using Competitive Co-Evolutionary Algorithm	Wang X. et al.	Elsevier Journal on Expert Systems with Applications 41(10)	A QoS multi-cast routing protocol based on the cognitive behavior where each node maintains local information
Metric-Based Taxonomy of Routing Protocols for Cognitive Radio Ad hoc Networks	Abdelaziz and El-Nainay	Elsevier Journal of Network and Computer Applications 40	Routing protocols are classified according to the routing metric into six main categories: delay based, link stability based, throughput based, location based, energy-aware, and combined or multi-metric
Scalable Dynamic Routing Protocol for Cognitive Radio Sensor Networks	Spachos P and Hantzinakos D.	IEEE Sensors Journal 14(7)	Accurate channel model is built to evaluate the signal strength in different areas, leading to energy-efficient, resource-constrained, and spectrum-efficient protocol
TIGHT: A Geographic Routing Protocol for Cognitive Radio Mobile Ad Hoc Networks	Jin X. et al.	IEEE Transactions on Wireless Communications 13(8)	The optimal and suboptimal modes route a packet along optimal and suboptimal trajectories to the destination
Cognitive routing for multi-hop mobile cognitive radio ad hoc networks	Lee and Lim	IEEE Journal of Communications and Networks 16(2)	Mobility-aware cognitive routing examines the risk level of each node against interference regions and selects the most reliable path for data delivery using a Markov predictor

Table 2.9: Continuous Evolution of CRN: 2015

Title	**Author**	**Publication**	**Findings**
SACRP: A Spectrum Aggregation Based Cooperative Routing Protocol for Cognitive Radio Ad hoc Networks	Ping S. et al.	IEEE Transaction on Communications 63(6)	Spectrum aggregation based cooperating routing protocol for cognitive radio ad-hoc network. Two cooperating protocols, class A for power minimization and Class B for reducing the end-to-end delay.
An Efficient Relay Selection Strategy for Random Cognitive Relay Network	Bang J. et al.	IEEE Transaction on Wireless Communications 14(3)	Reduced selection complexity and limited candidate relays to reduce feedback burden and achieves lower outage probability of complete network
Combined Channel Assignment and Network Coded Opportunistic Routing in Cognitive Radio Networks	Qin Y. et. al.	Elsevier Journal of Computers and Electrical Engineering	Heuristic algorithm to select forwarding candidates and assign channels with nonlinear programming optimization model
SMART: A SpectruM-Aware ClusteR-based rouTing Scheme for Distributed Cognitive Radio Networks	Saleem Y. et al.	Elsevier Journal of Computer Networks 91	Based on the network conditions secondary users form cluster and adjust it through cluster merging and splitting to search route the clustered network
Distributed Resource Allocation in Cognitive and Cooperative Ad hoc Networks Through Joint Routing, Relay Selection and Spectrum Allocation	Ding L. et al.	Elsevier Journal on Computer Networks 83	The algorithm aims at maximizing the network throughput through local control actions and information through joint optimization of routing, relay assignment and spectrum allocation
Neighbour Discovery in Traditional Wireless Networks and Cognitive Radio Networks: Basics, Taxonomy, Challenges and Future Research Directions	Khan A. A. et al.	Elsevier Journal of Network and Computer Applications 52	Comprehensive taxonomy for neighbour discovery protocol in traditional wireless networks and cognitive radio networks

2.3 Routing in Cognitive Radio Network

The main objective of routing in multi-hop CRN is creation and maintenance of the route in open spectrum phenomenon. Secondary users can co-operate with each other to form heterogeneous multi hop network across multiple primary networks. Multi-hop CRN consists of intermediate nodes for relaying the information between sender and the receiver. If physical capabilities of secondary users are efficiently exploited, it can sense, switch and transmit over many different bands. Multi-hop CRN has challenges of reducing the interference to primary user and manage intermittent spectrum resource along the path.

Secondary users can access both unlicensed or licensed bands. Nodes in CRN can communicate with each other in two different ways that are through *infrastructure* or *without infrastructure* (that is with each other directly) (Akyildiz I. et al., 2006).

- Infrastructure based CRN have their own secondary base-station. Base station is fixed infrastructure component with cognitive capabilities and provides single hop access to all secondary users.
- Secondary users can communicate directly with each other in a multi-hop manner without infrastructure through ad-hoc connection on licensed or unlicensed band.

The problem under consideration for this study is routing in ad-hoc multi-hop networks.

Definition: Problem of routing in multi-hop cognitive radio network is creation and maintenance of wireless multi-hop paths among secondary users by deciding both forwarding node and spectrum to be used on each link along the routing path.

Wireless multi-hop network can be modeled as finite set of nodes consisting of P number of Licensed/Primary Users (PU) and S number of Secondary/Cognitive Users (CU) as shown in Figure 2.2. There are N orthogonal channels denoted by set H. Every node periodically scans wide range of spectrum to gather information about dynamic changes in the surrounding. The *spectrum sensing* function enables the cognitive radios to adapt to its environment by detecting spectrum holes and opportunistically transmitting on them.

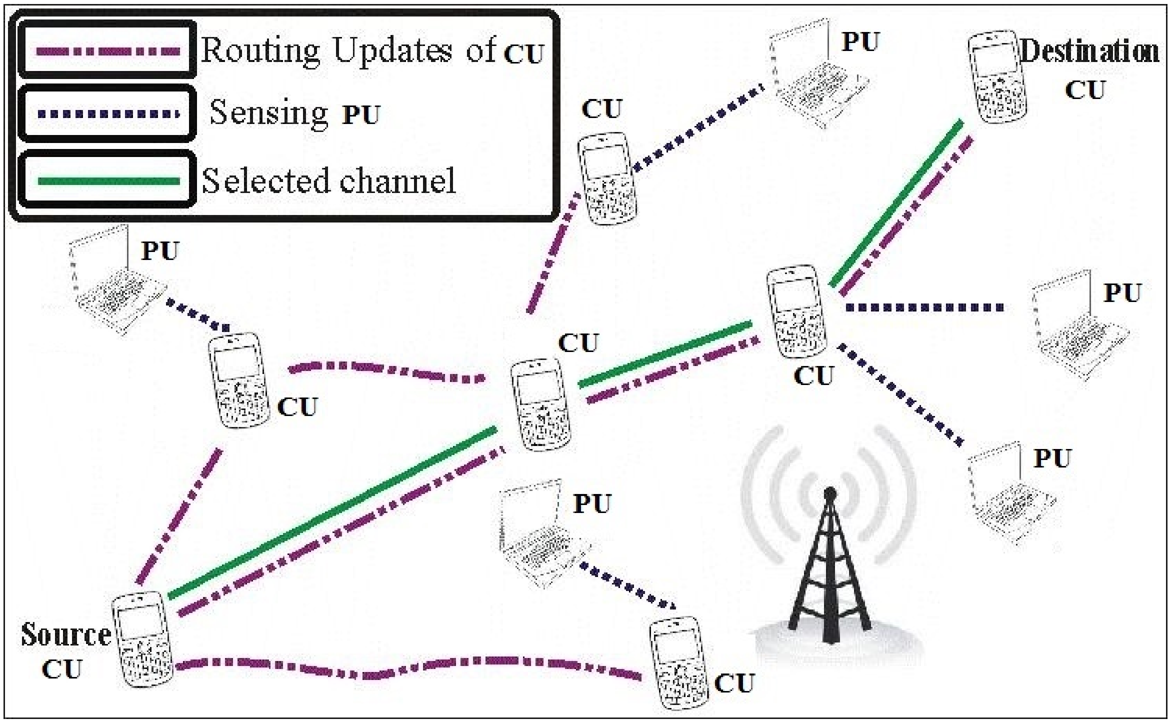


Figure 2.2: Multi-hop Cognitive Radio Network and Routing

These available spectrum holes have different characteristics varying over time. Spectrum analysis helps to learn the characteristics of different spectrum bands. This helps to select appropriate spectrum band as per users requirement. Spectrum band or channel is characterized by parameters such as frequency, bandwidth, propagation delay, estimated transmission time etc. These parameters are used to decide *capacity* of each channel, which is important factor for spectrum characterization.

In multi-hop CRN, set of available channels at any one node may not be same as another node, due to local spectrum availability. The set and the number of available channels at node i are denoted by Hi and ni respectively. Each secondary node i (where $1 \leq i \leq S$) has programmable radio interface. A radio interface is able to tune to a wide range of channels. Spectrum bands are having different capacities and opportunities. For every link l_{ij} between secondary node i and j, the set of available common channel H_I, can be constructed by,

$$H_I = H_i \cap H_j \tag{2.1}$$

Among all available common channels from H_I, one optimal channel is selected as per the quality of service required by the application. Channel selection should not increase interference to the PU. In collaboration with channel selection, routing should select stable and non-interfering path from source to destination with minimum control overhead. Topology, channel availability and network dynamics are shared among all secondary nodes in the form of routing updates. Routing in wireless multi-hop network includes deciding the path and intermediate/relay nodes with spectrum on each link from source to destination.

2.3.1 Design Issues and Existing Solutions

The dynamic spectrum which is intermittent in terms of both time and space necessitates collaboration between routing and spectrum management. Cross-layer approach for routing and spectrum management is proposed to determine the operating spectrum on each hop (Jiao and Yuqing, 2010).

There are many existing solutions for spectrum aware routing using conventional algorithms for multi-hop cognitive radio networks. Dynamic changes in the topology are captured using the link availability predictions based on the interference to the PU (Guan Q. et al., 2010).

Cognitive Routing Protocol (CRP) gives explicit protection to the PU receiver (Chowdhury and Akyildiz , 2011). It allows two classes of routes based on the service differentiation in cognitive radio network. Routing is achieved in two phases that are spectrum selection phase and next hop selection phase.

Online opportunistic routing selects forwarding links based on the locally identified spectrum access opportunities (Bhorkar A. et al., 2012). Multiuser diversity is achieved in the relay selection process which allows the sender to coordinate with multiple neighbouring nodes.

Best relay node with the highest forwarding gain is selected using reinforcement learning to adaptively learn good routes (Xia B. et al., 2009). Multi-agent reinforcement learning algorithm is used to evaluate the desirability of choosing various transmission parameters (Wu C. et al., 2010).

The important ***design issues*** for spectrum aware routing are summarized below,

- ***Common Control Channel:*** Routing algorithms need to broadcast messages for neighbour and route discovery. Due to lack of common control channel, broadcasting is a major problem in CRN which affects the communication and coordination among network nodes.

- ***Intermittent Connectivity:*** Reachable neighbours of the node change rapidly due to, i). Available spectrum may change or become unavailable as primary user starts exploiting the same spectrum. ii). Once a node selects particular configuration, it is no longer accessible on another channel.

- ***Re-routing:*** Re-routing is major design concern due to intermittent connectivity between different nodes. The routing algorithm should minimize the routing overhead resulted from re-routing.

Following sections discuss several spectrum and route selection algorithms with focus on the aforementioned design issues. All considered routing solutions are broadly categorized into following classes:

- Spectrum Aware Routing Protocols

- Full Spectrum Knowledge Based Routing

- Flooding based Classical Routing Protocol

- Opportunistic Routing

2.3.2 Spectrum Aware Routing Protocols

In the view of taking accurate routing decisions, every Cognitive Node (CN) should be aware of its surrounding physical environment. Selection of reliable path dynamically requires collaboration between routing module and spectrum management functionality. The information about spectrum availability is provided to routing modules

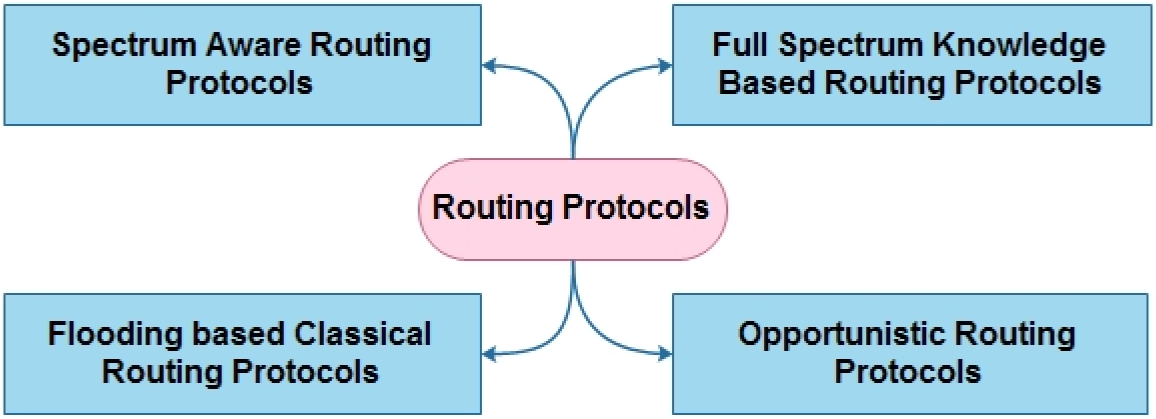

Spectrum Aware Routing Protocols
Full Spectrum Knowledge Based Routing Protocols
Routing Protocols
Flooding based Classical Routing Protocols
Opportunistic Routing Protocols

Figure 2.3: Classification of Routing Protocols

through external central entity or it is collected locally by each cognitive node. Routing protocol design for the multi-hop network should be highly coupled with entire spectrum management cycle (Cesana M. et al., 2011). Spectrum awareness can be achieved by measuring **interference** on each channel.

Interference Temperature Model based Channel Selection for End-to-End Routing

Each mesh node measures the interference temperature for each channel locally and disseminates these measured values with other nodes within its interference range (Sharma M. et al., 2007). The interference temperature model is used for computation of available channels and selecting best one, in static multi-hop multi-channel wireless mesh networks. Based on the collected statistics all nodes compute a set of available channel for transmission.

Once available channels are computed, it is required to select proper channel for transmission. Prerequisite of channel selection for link between two neighbouring nodes m and n is that both the nodes must have non-empty intersection of their available channel sets:

$$AC_m \cap AC_n \neq \phi \tag{2.2}$$

End-to-end routing metric is designed using minimum per-hop link cost for each hop on the end-to-end path. Per-hop link cost is not only dependent on transmission delay and switching cost but also on channel stability factor which is average amount of time for which channel should remain available for transmission. The routing metrics of end-to-end route r consisting of p hopes is formalized as:

$$RM(r) = (1 - \delta) * \sum_{i=1}^{p} LC_{ci} + \delta * (1 \leq j \leq C^{Xj}) \tag{2.3}$$

Route metrics $RM(r)$ consider summation of LC_{ci}, which is cost of using channel c on hop i, Xj is the total number of times channel j is used in route r, and C is a total number of channels in the spectrum pool. Link cost for the channel c is summation of Expected Transmission Time of frame ($ETTc$), Cost of Channel Switching (SCc) and Channel Stability Factor (SFc) given as:

$$LC_c = ETT_c + SC_c + SF_c \tag{2.4}$$

Incorporating **dynamics of the environment** by finding availability of spectrum will make the link selection more suitable for the cognitive radio network.

Routing under Consideration of Environment Dynamics

In multi-hop CRN, the topology and connectivity maps are determined by the available spectrum holes and their instantaneous variations. The objective is to consider activities of primary user for giving efficient routing solution(Khalife H. et al., 2009). According to the primary user activity over the primary channels, cognitive environment is classified into three categories:

- **Static**: Holding time offers relatively static wireless environment, Once frequency band is available it can be exploited for an unlimited period of time.
- **Dynamic**: The intermittent availability of the exploited spectrum band seriously affects the services offered to the CU.

- **Highly Dynamic**: Due to highly active PUs, spectrum bands are not available for whole communication duration. The potentially possible solution is to opportunistically transmit over any available channel.

In static multi-hop CRN, primary frequency band is available for a duration that exceeds the communication time. Routing design in these types of network is accounting for intra cognitive node interference and PU interference.

In order to find available and stable path in dynamic CRN, issues like route stability, exchanging control information and channel synchronization should be handled carefully. Routing should be assisted by a metric from lower layer which should reflect the spectrum availability and its quality.

In highly dynamic environment a complete opportunistic solution is suggested where every packet can be forwarded over opportunistically available channels to constitute potential solution. Centralized synchronization window can be used for sharing control information, which consists of fixed time slot where all nodes are tuned to different frequencies and exchange all possible control information. The control information can also be shared on low frequency common control channel.

Dynamically changing environment requires every CN to be **spectrum aware** for opportunistically using it for transmission.

Spectrum Aware Opportunistic Routing

Cognitive routing is coupled with spectrum sensing and sharing in the multi-hop cognitive radio network is investigated in (Liu Y. et al., 2012). Geographic location information is used to discover spectrum access opportunities to improve transmission performance. In relay selection step, the sender selects the next hop from the candidate neighbouring secondary users.

Geographic information and local spectrum usage statistic can improve the routing performance if it reflects the actual channel occupancy at that instance. Observing local channel statistics, analyzing it and deciding best channel for transmission needs to be handled with intelligent algorithm. Spectrum and node selection intelligently **balance load and traffic** at every hop.

Traffic-Aware Scheme for Energy-Efficient Routing

The objective is to prepare resource intensive traffic aware routing protocol that enables energy conservation and efficient data flow. Secondary users coordinate with each other to deal with heterogeneous spectrum availability in distributed CRN (Bourdena A. et al., 2014).

The message exchange among multiple cognitive nodes is similar to AODV routing protocol like normal wireless network (RFC, 2003). Backward Traffic Difference (BTD) estimation methodology is used for energy efficient routing. BTD is bounded by hop-by-hop and end-to-end delay limitations of transmission.

Each node separately runs the traffic aware mechanism using BTD. The BTD estimation affects the sleep-time duration and enables energy conservation of nodes. The spectrum resource usage statistics in dynamically changing scenarios should be incorporated in routing solution for correctly representing the topology and spectrum availability.

2.3.3 Full Spectrum Knowledge Based Routing

The routing approaches implemented with the assumption of having complete information of spectrum occupancy map. These approaches are leveraging benefits of theoretical tools for searching efficient roots. Spectrum occupancy maps can be maintained to indicate its availability over time and space.

Route Stability and Opportunistic Routing

All nodes in CRN compute the cost to reach all possible destinations. This cost reflects the highest spectrum availability with the minimum hop count. This requires collecting information of local spectrum availability from all nodes for building mesh of shortest hop count path between all pairs of network nodes.

Full spectrum aware routing is necessary to balance between long-term route stability and short-term opportunistic performance (Pefkianakis I. et al., 2008). The objective is to select routes with highest spectrum availability and quality. For achieving optimal routing, it is necessary to consider work of Physical, MAC and Network layer.

The spectrum aware optimal routing is defined as path with minimum hop count, increased end-to-end throughput and exploitation of spectrum opportunities. This routing protocol creates forwarding mesh which is adjusted periodically according to the spectrum dynamics, and opportunistically routing packets across this mesh. It is built by computing a cost $Cost_i$ for each node i which represents spectrum availability of the highest spectrum path whose length is less than H hops. The value of H complements with $Cost_{max}$ which is the maximum cost to the destination.

Collecting information from all nodes about spectrum availability and its duration increases traffic of control messages throughout the network. Moreover, computed paths based on periodically collected information may not reflect **actual scenario of spectrum availability and overall topology.**

Topology Control Based Routing

Prediction Based Cognitive Topology Control (PCTC) (Guan Q. et al., 2010), captures the network topology dynamically based on link prediction for providing efficient and opportunistic link management and routing solution. This prediction of the link availability, duration can be used to construct a more reliable topology which results in reduced rerouting. Prediction of link-availability duration is dependent on the interference to primary users and the mobility of cognitive users. The principle of PCTC is to preserve the reliable path with maximum path weight for any pair of nodes under a connectivity guarantee.

Topology construction process consists of three steps neighbour information collection, path search and neighbour selection. The distributed localized Dijkstra topology control is used which aims at constructing an efficient topology which preserves the global network connectivity. Resultant topology preserves reliable links for deciding global reliable path between any two nodes.

PCTC requires knowledge of connectivity of all nodes. Periodic broadcast messages are used to construct connected graph. This connected graph is used with classical mobile ad-hoc network routing protocol for route formation. It is distributed protocol and every node constructs topology based on collected information. The predicted link availability at every node may differ as per local spectrum usage. Applying classical routing protocol on resulting predicted topology may increase the overhead of topology reconstruction and route reformation, due to dynamic changes and **uncertainty of link availability**.

Route Robustness for Joint Routing and Spectrum Allocation

The objective is to study joint routing and spectrum allocation problem in multi-hop CRN with channel heterogeneity and dynamics (Shih C. et al., 2011). It is important to select robust route experiencing less frequent interruptions due to appearance of the primary user. Algorithm jointly determines which route to be used and spectrum allocation on every hop to minimize the system throughput.

A basic level of robustness for a set of routes called as "Skeleton Set Formation Scheme", is same as Breadth-First-Search for each flow from each source to destination. Every node in the skeleton set satisfies robustness constraints. Channels between two nodes are allocated after selecting end to end route.

This algorithm is using graph based approach for selecting skeleton set of routes therefore, it requires information of complete topology and full spectrum knowledge in advance. Frequent changes in topology and spectrum availability make it difficult to find complete information of overall topology.

Cumulative Delay and Node Capacity Based Routing

Route discovery scheme is integrated with node-importance-based clustering scheme and node contraction scheme (Huang X. et al., 2011). The high mobility nodes have adverse impact on data transmission due to a high link-disrupting rate. Cluster formation and control channel selection schemes are beneficial to find the stable routes to realize the multi-hop connection. Proposed routing scheme integrates the on-demand routing with dynamic channel assignment. The path is determined with largest path availability probability.

Control channel selection is integrated with on-demand route discovery in high mobility, multi-hop multichannel environments. Clustering is used to mitigate effects of spectrum heterogeneity, as it is reducing the network scale and routing protocol overhead. The routing in multi-hop CRN includes intra-cluster and inter-cluster routing processes. The intra-cluster routing occurs in single cluster while inter-cluster routing occurs in multiple clusters. Due to high mobility nature, spectrum route request and route reply are used to discover paths between nodes via the control channel message exchanges.

2.3.4 Flooding based Classical Routing Protocol

In classical wireless routing protocols, flooding algorithm is used to send route request packets to all neighbouring nodes except the one it arrived on until it reaches the destination. Flooding of route request packet results in multiple fixed paths from source to destination. Routing in cognitive radio network can be extended by classical routing protocols using flooding to find path from source to destination. More than two neighbours of any requesting node may create a broadcast storm.

SACRP : Spectrum Aggregation based Cooperative Routing Protocol

Spectrum aggregation based cooperating routing protocol for cognitive radio adhoc network (Ping S. et al., 2015), has two classes of cooperating routing protocols. class A is for power minimization and Class B is for reducing the end-to-end delay.

Class A aggregates multiple channels and selects suitable relay node by flooding route request on the common control channel. The route request is containing information of aggregated channels of each hop, the number of hops and relay to be used on each hop. The destination node selects the path with minimum hop count.

Class B selects relay nodes with better channel conditions reducing the number of re-transmission and end-to-end delay. The method of setting multiple complete paths from source to destination increases the control overhead due to flooding messages. Selection of one of the path may not be able to complete transmission requirement, as the network conditions may change dynamically.

Search: A Routing Protocol with Geographic Forwarding

The optical path is found by geographic forwarding on each channel with greedy forwarding for advancement towards destination and primary user region avoidance (Chowdhury and Felice, 2009). The route is set up by transmitting route request on each available channel. It is forwarded by each intermediate node towards destination by adding hop ID, current location, time stamp and flag showing current propagation mode that is greedy forwarding or primary user avoidance. The basic idea of greedy advancement towards the destination is dependent on the assumption of location of destination.

Search attempts to find the shortest path by using focus region to select forwarding node on a straight line towards the destination. Greater advancement towards destination requires more transmission power. Increased transmission power affect increase interference to PU.

Minimum Delay-Maximum Stability Route through Opportunistic Service Differentiation

The objective is to create minimum delay maximum stability route for end-to-end traffic flow and for providing differentiated services to different traffic priorities defined by flow duration and flow priority (Kiam H et al., 2011). It is achieved with the help of four different modules, namely, Route Discovery, Route Decision, Opportunistic Routing and Route Maintenance.

In route discovery, node discovers all possible paths between a source node and destination node. All the available paths are sorted according to the delay which will be used by route decision module to select the path that can satisfy flow duration of the considered end-to-end traffic flow. After selection of the route, opportunistic routing module selects candidate forwarding node according to packet priority. Broadcasting of route request packets helps to discover route.

Service differentiation is achieved with different traffic priority. Higher priority packets have higher candidate selection range and higher chances of being received at nodes near the destination. Route selection is based on received route reply packets via multiple paths having information of route, delay and minimum expected time for transmission. This increases the route discovery message overhead. As spectrum availability is intermittent in cognitive radio network, using large portion spectrum opportunities for control information transmission instead of data transmission will not increase spectrum utilization in desired efficient manner.

2.3.5 Opportunistic Routing

Cognitive radio network as a dynamic spectrum access network requires routing algorithm which is able to exploit network opportunities. Classical routing algorithm chooses a static route by flooding the route request packets. On the other hand, in opportunistic routing, packets are transferred from next forwarding node until it reaches the destination. The choice of next forwarding node is dependent on service requirement and network dynamics. Opportunistic routing is class of routing protocols designed to route the packet as per network conditions, availability of link and neighbouring nodes.

Reinforcement Learning based Spectrum Aware Routing

Spectrum-aware routing protocol using Q-learning is used to learn good routes (Xia B. et al., 2009). In this scheme, every node x stores table of $Q_x(y, d)$, the number of channels available to the destination through x for neighbour y and destination d. These values can be used to select best next hop and can be updated while routing using forward exploration.

Another algorithm proposed in this paper, Dual Reinforcement Routing is improving the performance and time to learn good route with the help of forward and backward exploration. In a nutshell, both of these algorithms exchange information with neighbours to adaptively learn good routes which have more available channels from just local information, gradually incorporating more global state of the environment. These algorithms outperform the spectrum aware shortest path routing but network sharing among competing operators is not considered which will affect spectrum availability at every hop.

In dynamic and uncertain environment of cognitive radio network, Q-values stored and updated during previous transmission **may not be able to represent current situations of channel availability**.

Cross Layer Routing and Dynamic Spectrum Sharing

Distributed and localized algorithm for joint ROuting and dynamic Spectrum Allocation (ROSA) for multi-hop CRNs is taking decision based on locally collected spectrum and power allocation information (Lei D. et al., 2010). Time slotted common control channel is used to collect network statistics and data channel for data communication. For reducing the probability of selecting congested intermediate relay node, differential backlog is used. Differential backlog is difference of queue backlog in given sessions. Feasible next hop is decided if it is having positive advancement toward destination.

ROSA opportunistically calculates the next hop depending upon queuing and spectrum dynamics, therefore, each packet will select different path. If the destination is in the range of transmitter it will be selected directly. The transmitter may select a node other than destination if there is no available low interference mini-band between the transmitter and the destination. Channel and node selection with locally available information without analyzing channel and node usage statistics may have chances of failure. This requires **learning better opportunities** among all possible options for improving overall performance.

IP Hop by Hop Geographic Routing Protocol

IP spectrum aware routing protocol takes all information of channel statistics from lower two layers i.e results of sensing operation (Badoi C. I. et al., 2010). The data packet's header is containing information needed to select next hop. Each node takes local decisions based on its local neighbourhood information. The intersection of sensed available channels by two nodes (current and next hop node), is used to find the channel for transmission as per quality of service requirement.

IPSAG uses the local information of position and channel opportunities and takes hop-by-hop geographic routing decision. The criterion used for the creation of neighbourhood is the Euclidian distance. It is represented as a circle with the core and the radius given by the maximum Euclidian distance between the core and secondary node. If a packet encounters pre-existing systems like WiMAX or WiFi, it will be routed according to the systems way of routing, based on the tunneling routing approach, otherwise secondary node route according to IPSAG. The proposed algorithm is **not able to deal with channel heterogeneity** along the complete path.

Relay based Routing with Energy Consumption and Interference Control

Relaying is a promising solution to enhance the system capacity at low cost (Xie M. et al., 2010). For deriving maximum reachable distance of secondary user in one-hop, following two conditions are evaluated:

- PU Transparency: The incurred interference does not interrupt the licensed user under QoS requirement.
- SU Reliability: The designated secondary user is able to successfully decode the data to meet its own QoS.

Both of the above conditions are satisfied by reducing the transmission power of secondary users on shorter distance. This is achieved by multi-hop relaying with multiple short hops. The relay route is, $SU_s - SU_1 - SU_2 - .... - SU_d$, where SU_s and SU_d denotes source and destination respectively and SU_i is the i^{th} intermediate relay node. These relay routes are providing an alternative path for direct route with single long hop of SU_s to SU_d.

Multi-hop cooperative relaying is achieved through two routing algorithms, Nearest Neighbour Routing and Farthest Neighbour Routing. Nearest neighbour routing attempts to find nearest neighbour inside the sector which improves channel quality and saves energy consumption at every node.

Farthest neighbour routing is used to find the farthest neighbour with in the sector consisting of few long hops in the complete path which is good for reducing the delay but tend to consume more energy to achieve good channel quality. This also **increases interference to the primary user**.

Relay Selection Strategy with Lower Outage Probability

The objective is to propose an efficient relay selection strategy with reduced selection complexity and limited candidate relay to reduce feedback burden (Bang J. et al., 2015). It deals with lower outage probability of complete network for designing efficient and reliable design of CRN.

Performance of network can be enhanced by allowing nodes to transmit in two hops. Inactive secondary nodes support other secondary source nodes dealing with channel impairments. Channel heterogeneity is also handled successfully using inactive secondary nodes.

Cooperative multi-hop routing instead of only two hop paths can improve reliability of the CRN and overall throughput. For leveraging all benefits of the CRN, routing should be cooperative and context aware of surrounding environment.

2.4 Channel Selection and Traffic Prediction in CRN

Cognitive nodes are designed to utilize spectrum bands opportunistically and in non-interfering manner by sensing unused spectrum portion called as spectrum/white holes. Cognitive nodes should identify transmission opportunities and intelligently determine ongoing primary users activities to avoid interference to the primary user. Hence, it is important to precisely estimate and model primary user activities for efficient spectrum usage by cognitive users and overall spectrum utilization.

Effective modeling of primary user activities leads to precise prediction of their future state. It is achieved by learning from the patterns and history of spectrum utilization. This enables to analyze all available spectrum bands and select best out of them for transmission. Moreover, selecting best and stable channel on every hop leads to reliable routing of packets from source to destination. Thus, reliable channel selection and traffic prediction play vital role in spectrum management, routing and overall performance of CRN.

Neural Network based Learning Scheme

Neural network based scheme is proposed to discover data rate capability of specific radio configuration (Tsagkaris K. et al., 2008). Evaluation of all possible configurations is necessary for cognitive radio, for selecting best configuration as per requirements. Constructing and training a network with additional input reflecting configuration such as location information, user preferences or constraints improves accuracy in prediction.

Machine learning algorithm which continuously learns from dynamic environment improves the performance over the long run. The integration of a learning engine with cognitive radio is very important especially for channel estimation and predictive modeling phase. It helps for improving stability and reliability of the discovery and evaluation of configuration capabilities.

With this view, cognitive radio can use many different learning techniques ranging from pure look-up table to arbitrary combinations of machine learning techniques that includes artificial neural network, evolutionary / genetic algorithms, reinforcement learning and **Hidden Markov Model** etc.

Channel Modeling Based on Interference Temperature

Hidden Markov Model (HMM) is used to predict the interference dynamics on any channel in future time slots (Sharma M. et al., 2008). HMM is trained with interference temperature values of wireless channel. These values are used to predict the interference temperature dynamics on the channel in future. This prediction is used to select preferable channel for communication.

HMM is trained with the help of observed outcome of channel interference temperature obtained using simulation. The interference conditions are varied from one channel simulation to another by changing the probability of threshold value in simulation. Different threshold values for each channel ensures different interference conditions on the channel.

Simulating channel occupancy by just varying probability of the channel being busy and idle is not representing real world channel statistics. Therefore, selecting preferable channel for communication based on generated predicted availability may fail to protect primary user from interference. Moreover, implemented HMM may **trap in local maxima and may not be able to explore all opportunities** in network.

Traffic Prediction using Seasonal Auto-Regressive Integrate Moving Average

The objective is to estimate channel availability in CRN by predicting the traffic pattern of the primary user (Li and Zekavat, 2009). Wireless traffic prediction is implemented using classical model for discrete time series that is SARIMA (Seasonal Auto-regressive Integrate Moving Average).

The set of observations of the number of call arrivals in different time intervals with different periods can be considered as discrete time series. SARIMA model is fit for non stationary call arrival process. Traffic prediction technique reduces the channel switching rate of cognitive user and the interference to the primary user. The traffic pattern of primary user is different and heterogeneous. The proposed method **does not support multiple types of patterns**.

Predictive Methods for Inference of Availability of Spectrum

Spectrum sensing consumes huge energy for scanning all channels. Predictive methods for inference of availability of spectrum hole can reduce energy consumption of secondary user by sensing only that channel which is predicted as idle (Tumuluru V. K. et al., 2010). This also helps to improve spectrum utilization.

In this effect, two adaptive schemes MLP predictor and HMM for channel status prediction are presented and compared qualitatively. Multilayer perception network is fully trained model requiring fewer past observations or occupancy history. MLP is trained only once and used to predict future channel availability. The MLP predictor should be retrained periodically for better performance in **time varying traffic scenario**.

Selective Opportunistic Spectrum Access

Selective Opportunistic Spectrum Access (SOSA) scheme is proposed to estimate the probability of a channel appearing idle with available statistics to choose the best spectrum sensing order (Yuan G. et al., 2010). This enables secondary user to sense and select target spectrum band in an optimum order to maximize spectrum efficiently and fulfill quality of service requirement. The channel ranking method takes statistical traffic characteristics as the critical input parameter. Mean value of the OFF and ON periods of all primary channels is represented by α and β respectively.

Probability P_n of the n^{th} primary channel appearing idle in the next time slot is given by,

$$P_n = \frac{\alpha_n}{\alpha_n + \beta_n} \tag{2.5}$$

The secondary user uses the probability based ordering strategy to arrange the sensing sequence for next time slot. This allows the secondary user to operate in a discontinuous spectrum environment.

First Difference Filtering and Correlation for Primary User Traffic Data

Primary user activity is assumed to follow Poisson traffic model with exponentially distributed inter-arrivals (Canberk B. et al., 2011). Using Poisson modeling, primary user activities are modeled as smooth and burst-free traffic. This results in missing of some available but unidentified opportunities of spectrum.

This method model the primary user traffic in more efficient and accurate way using first difference filtering and correlation. This is achieved by arranging the observed sequence of primary user traffic data into clusters which are enhanced with temporal correlations. Spiky and bursty characteristics of the signal are more accurately distinguished by employing clustering. This leads to defeating fluctuations more precisely.

Time Series Prediction for Opportunistic Channel

Two strategies for opportunistic channel selection are proposed to exploit underutilized licensed spectrum in CRN (Tan X. et al., 2013). Past observations and knowledge of primary spectrum are used to predict near future busy probability of primary spectrum. This is useful for selecting suitable channel for data transmission results in reduced collision and switching probability.

Strategy A, allows occupying only one spectrum unit for transmission, thus the data transmission rate of the secondary user is limited by bandwidth. Spectrum B is flexible and allows many idle spectrum units for data transmission. Using any spectrum band without cooperation among many secondary users may increase interference. It is very important to build coordination mechanism to avoid interference among secondary users.

Level of Primary User Activity and Amount of Structure in Observed Data

Regularities of the channel utilization undoubtedly influence the performance of learning algorithm in dynamic channel selection in cognitive radio network. The objective is to improve learning performance of opportunistic dynamic channel selection (Macaluso I. et al. , 2013). This is achieved by characterizing the primary user activity using Lempel-Ziv complexity. The performance of learning algorithm is influenced by the amount of structure in the observed data of primary user activity. More structured data implies more effective learning.

The rate of production of new pattern in a sequence is measured which is the most important feature for real time data. Quantifying regularities in primary user activity data can improve performance of reinforcement learning to decide which channels should be opportunistically explored. Thus, benefits of adaptive learning algorithm depend strongly on the pattern of utilization of channel by the primary user.

Channel Reservation Policy for Primary and Secondary Users

Opportunistic spectrum access achieves higher spectrum utilization. Channel reservation policy for the primary user is proposed to reduce hand off, blocking probability and interference-free transmission of the secondary user (Chakraborty and Misra , 2015).

This policy suggests reserving the suitable number of idle channels for incoming primary user. The secondary user can access only unreserved idle channels. As long as primary user is having available reserved channel, primary user is not allowed to use unreserved channels. The secondary user continues transmission in unreserved channel for longer duration resulting in reduction of spectrum handoff and dropping probabilities.

For practical implementation of cognitive radio network, this architectural framework highlights the need for coordination among Medium Access Control (MAC) protocols, channel assignment schemes and spectrum hand-off. Spectrum broker or spectrum controller is considered as most significant element of CRN architecture and channel reservation policy.

Dynamic Channel Availability as per User Relative Position

Channel availability estimation strategy, is designed to explicitly consider features of mobile scenario of network nodes (Cacciapuoti A. S. et al., 2015). In mobile scenario, the channel availability varies dynamically over time due to changes in users relative positions. The proposed strategy estimates channel availability based on the relative distances between primary and cognitive user.

Each cognitive user is aware of its location and periodically checks the position of primary users. Cognitive user activities are organized in fixed time slots representing sensing time and transmission time. If the cognitive node is outside the range of primary user for complete time slot, then the channel is considered as available in that time slot.

Accessing Spatial Spectrum Holes for Relay based Cognitive Cellular Network

Users at the cell edge in cellular network experience degraded quality of the service due to poor signal strength and severe interference. This is resolved by using cognitive radio features in relay based cellular network (Mankar P. et al. , 2015). Use of cognitive radio to deal with poor signal strength in cellular network improves performance significantly.

Detected spectrum holes are used by relay nodes to forward the received signal from base station to interference limited user. Application of this kind requires selection of relay cognitive node for forwarding information, localization of primary transmitter for selecting channel for transmission and cognitive node assignments to the end user on edge. Machine learning helps to dynamically assign channels in different cells for improving the performance.

2.4.1 PU Activity Model: A Comparison

Primary user activities are represented using observation sequence. This observation sequence is modeled and analyzed using different time series analysis techniques. Table 2.10, gives details of the different primary user activity models and their initial parameter set. Following points are summarized from the review of the various techniques of primary user activity modeling and channel availability prediction:

- The estimation and modeling of primary user activity is very crucial for the performance of cognitive radio network (Saleem and Rehmani, 2014).

- Primary user activity model should capture bursty and spiky fluctuations by temporal correlations (Canberk B. et al., 2011).

- The cognitive node should learn continuously from variations in network conditions and complexity of network activity (Macaluso I. et al. , 2013).

- The number of parameters required for multilayer perception network are more and it should be trained repeatedly for time-varying traffic scenarios (Tumuluru V. K. et al., 2010).

- Seasonal Auto Regressive Integrated Moving Average (SARIMA) model is not able to learn from uncertain, bursty and spiky discrete time series of primary user activities (Li and Zekavat, 2009).
- Hidden Markov Model is able to learn online and it is repeated continuously in dynamic environment (Rabiner L. R. , 1989).

The different models used for analyzing primary user activities are compared with their properties in following table,

Table 2.10: Primary User Activity Models

Property	**Multilayer Perception Network** Tumuluru V. K. et al., 2010	**Hidden Markov Model** Sharma M. et al., 2008	**SARIMA Model** Li and Zekavat, 2009
Adaptive Parameters	Weights w_{ji}	Probabilities A, B, π	Probability Distribution
Training Data sets	Large	Moderate	Small
Model specifications	Non-Linear Discrimination	Non-Linear Discrimination	Linear Discrimination
Training Criteria	Minimize Mean Square Error	Maximize Likelihood $(P(O\|\lambda))$	Minimize Mean Square Error
Mode of Training	Offline, Only Once	Online, Repeated	Iterative
Limitations	Local Minima Problem	Continuous Resource Requirement	Only Burst-free Time Series

2.5 Limitations of State-of-the-Art Techniques

This chapter presents, an overview of the state of the art spectrum aware routing protocol in CRN. The existing work in multi-hop CRN and routing is discussed in previous sections. The performance and services of various routing protocols required to address unique challenges of CRN are elaborated in detail. For analytical evaluation of the routing protocol, they are categorized into following categories:

- Spectrum Aware Routing Protocols
- Full Spectrum Knowledge Based Routing
- Flooding Based Classical Routing Protocol
- Opportunistic Routing

As per the review of existing literature, following are some identified limitations, deficiencies, or gaps in existing knowledge that need to be addressed:

Assumption of full spectrum knowledge: Several routing protocols are designed with the assumption of availability of full spectrum knowledge to the network nodes. The spectrum occupancy map representing time and space channel availability is maintained by a central entity (Ma M. and Tsang D., 2009). These spectrum maps are used to compute routes centrally or these maps are given to networking nodes for finding routes in ad hoc and distributed manner. These architectural models are suitable for static networks where spectrum availability is known between any pair of networking nodes (Shih C. et al., 2011).

Due to intermittent spectrum availability in space and time domain, creating and maintaining spectrum maps centrally may not represent actual conditions of spectrum availability. Monitoring, collecting and analyzing spectrum availability dynamics at different geographic positions over a time requires ample amount of time. This may lead to misinterpreting some of the spectrum opportunities and missing some actual ones.

Static paths fail to deal with uncertainties of environment: Conventional algorithms attempt to find a fixed and static path from source to destination using intermediate nodes (Bhorkar A. et al., 2012). Participation of any intermediate node in routing process is dependent on local spectrum opportunities available at that node. Deciding fixed path on the basis of available information may fail because of sudden changes in spectrum availability as the primary user starts its activity using same spectrum resource. An outage of intermediate nodes and unavailability of intermediate links may lead to route failure.

Imperfect primary user activity modeling: Cognitive users are utilizing spectrum when primary users are not using it. Cognitive users have to vacate spectrum band as primary user appear and start transmission at any instance of time. Therefore, there is no guarantee about spectrum availability for the entire duration of transmission of the cognitive user.

If cognitive users are able to model primary user activities by monitoring, analyzing and learning their spectrum utilization and its history, they can predict the future state of the channel and suggest preferable channel for communication. The channel usage pattern undoubtedly influences the performance of the learning algorithm in dynamic channel selection of cognitive radio network. Therefore, activity model should represent actual behavior of primary user and should be able to find patterns in channel occupancy to provide more accurate prediction of channel availability (Saleem and Rehmani, 2014).

Extensions of classical routing protocol: Path discovery in classical routing protocol starts by broadcasting route request message to its neighbours. Each neighbouring node forwards this request to its neighbouring node until the destination is reached. Multiple copies of route request are received at the destination via multiple paths. One of the path is then selected for transmitting packets from source to destination.

In cognitive radio network, classical routing is extended by selecting complete paths and channel to be used along the path with some additional routing metric. As the destination is configured to particular spectrum band as per its spectrum availability, a source has to broadcast route request on all possible channels with a hope that it will reach destination. This definitely increases control message overhead in network which results in inefficient use of available spectrum opportunities (How K. C. et al., 2011).

Lack of interaction among network nodes: In cognitive radio network, cognitive nodes are struggling for utilizing the same limited spectrum resource. If multiple cognitive users start utilizing same channel, it will greatly affect QoS requirement and increase interference to the primary user. For improving spectrum utilization, it is essential to have efficient spectrum sharing schemes. Interaction among network users can help to understand willing neighbours to participate in routing as per their spectrum availability. Strategic interaction among network nodes definitely increases spectrum utilization (Cesana M. et al., 2011).

Inefficiency to deal with the dynamic environment: To deal with the dynamic environment in an efficient manner, the user should be enabled to understand context by observing, learning and responding to the complex environment. On the other hand, traditional wireless systems, have predefined set of rules to follow. In CRN, every host should be continuously context aware of its surrounding physical environment for taking more accurate decisions. Each action selection in CRN requires knowledge about licensed user activity, spectrum occupancy, multiple cognitive users with their strategies and mobility of all hosts. Action selection based on this gathered context learning or collected knowledge helps to improve performance of cognitive radio network (Saleem Y. et al., 2015).

Off-Policy generalized policy iteration: Generalized policy iterations for evaluating policy at the end of every episode called as off-policy learning, leading towards local minima problem. State of the environment(spectrum or relay availability) changes frequently due to intermittent spectrum availability. There is a need to evaluate and update policy frequently to deal with the dynamic environment. Agents in cognitive radio network should learn from temporal differences of every state change. This helps to take more real time decisions as per network and channel statistics (Bhorkar A. et al., 2012).

2.5.1 Related Work

The two state-of-the-art techniques having characteristics of environmental aware computing are implemented and analyzed. The first technique offers routing paths from source to destination with consideration of characteristics of available links. On the other hand, second technique explores routing opportunities with local information and dynamic topology construction.

Adaptive Opportunistic Routing

This approach considers distributed, adaptive and opportunistic routing algorithm with zero knowledge of network topology and channel statistics (Bhorkar A. et al., 2012). The performance is measured in terms of per packet reward. The proposed scheme jointly addresses the issues of learning and routing in an opportunistic context. Every action selection for deciding path is based on transmission success probabilities.

It assumes no knowledge of topology and channel statistics. Reinforcement learning is used at every node to adapt to routing strategies and exploiting the statistical opportunities in network. The opportunistic routing decisions are made in an online manner by choosing the next relay based on actual transmission outcomes as well as the rank ordering of neighbouring nodes. It is described in terms of initialization and four stages of transmission, reception, acknowledgment and adaptive computation stage.

Reinforcement learning is used for selecting best candidate forwarding node and minimum cost links resulting in reduced cost of the end-to-end routing path, maximizing average per-packet reward. Routing decisions are made in distributed manner via the following three-way handshake between node i and its neighbours $N(i)$.

1. At time n, node i transmits a packet.

2. The set of nodes S_n^i, successfully receiving the packet from node i, transmit acknowledgment containing *Estimated Score* to node i,

3. Node i compares *Estimated Score* of all S_n^i and announces node $j \in S_n^i$ as next transmitter or Termination decision T in forwarding packet.

This routing approach does not consider closely related issues like congestion control and throughput of the network. The performance of this routing protocol can be improved by using channel statistics information for selecting stable links.

Joint spectrum-route selection with service differentiation

Cognitive Routing Protocol (CRP), is designed to categorize routes according to the service requirement. *Class I* routes provide better cognitive network performance. On the other hand, *class II* routes aim to achieve a higher measure of protection for primary user (Chowdhury and Akyildiz , 2011).

In *class I* route, end-to-end latency is the key consideration issue. Selected link should support highest propagation distance and longest allowed duration for transmission. The algorithm is designed to find the best link k out of N possible spectrum bands to maximize the optimization function O:

$$O_{class-I} = D_k.T_f^x \tag{2.6}$$

Where D_k specifies the distance covered with given transmission power and T_f^x is fractional time for transmission considering different sensing schedule of the neighbouring users. The optimization function aims to maximize the number of packets transmitted over maximum possible distance.

The optimization function for the *Class II* route is to minimize interference to the primary users. This is achieved by minimizing the product of overlapping fractional area between CU-PU coverage ranges represented as A_f^x and the propagation distance D_k.

$$O_{class-II} = D_k.A_f^x \tag{2.7}$$

The path selection is based on arriving route request packets at the destination indicating different routing paths. Destination select desired *Class I* route with $max\{\sum_{\forall j} O_{class-I}(j)\}$ or *Class II* route with $min\{\sum_{\forall j} O_{class-II}(j)\}$.

This routing protocol continuously monitor CU's location with respect to the known primary transmitter location. If the displacement of the node towards one or more primary user is determined, it proactively discovers the new path.

Upon receiving route request, every cognitive node updates two fields in the route request, that is choice of spectrum and value of optimization function of *class-I* and

Table 2.11: Routing Protocol Features: Comparative Analysis

Features	**Adaptive Routing Protocol**	**Cognitive Routing Protocol**	**Proposed Routing Protocol**
Spectrum Awareness	NO	YES	YES
Topology Construction	NO	NO	YES
Mobility Support	YES	YES	YES
Uncertainty	YES	NO	YES
Link Quality	NO	YES	YES
Flooding of Route Request	NO	YES	NO
Full Spectrum Knowledge	NO	NO	NO
CU Interaction	YES	NO	YES
PU Protection	NO	Limited	YES

class-II depending upon the class of routes. This route request is forwarded to destination along multiple paths. The destination chooses the final route depending on the arrival time of the route request. The broadcasting of route request on all channels having limited availability affects efficient spectrum utilization.

Table 2.11 give details of features supported through state-of-the-art techniques and characteristic features supported by proposed method.

2.6 Discussion and Future Directions

The reviewed literature and aforementioned limitations of the state-of-the-art technique gives foundation to strongly believe that, there is significant scope for devising a routing protocol that adopts the instantaneous variations in the network environment. To this extent, the routing in multi-hop CRN is considered as an open research issue. The open research issues which need to be addressed for formulating routing solution in CRN are given in following sections.

2.6.1 Exchanging Routing Information

Many routing solutions currently use common control channel for sharing routing updates which are useful in neighbour discovery, route discovery and route establishment. Due to the fact that secondary user is using spectrum as a visitor, the common control channel can also be affected by primary users activity. Moreover, selection of one particular channel as global control channel seriously affects network performance. Information about channel availability should be shared among cognitive nodes locally for deciding control and data channels. In highly mobile and dynamic environment, information exchange is only means of creating a view of the current topology of network.

Moreover, efficient spectrum sharing can be achieved by allowing strategic interaction among cognitive nodes. Cooperation among multiple cognitive nodes increases overall performance and spectrum utilization. Design and modeling of routing solution considering information exchange as important factor will improve the performance of routing.

2.6.2 Dynamic Topology and Unstable Connectivity

Topology and connectivity of the network are significantly affected, due to the high mobility of the network entities and activities of primary users. The reachable neighbours of any node may change their location due to mobility. They will not be used as relaying node as primary users start exploiting network resources. Solution based on static environment and stable topology will not serve the purpose of route selection in intermittently connected network.

Routing in CRN should be reactive taking into account the rapid changes in the neighbouring nodes availability and accordingly next hop selection must be done. The common channels for transmission between any pair of node changes due to primary users activity. As a result, the same neighbour can be viewed with another common channel. Changes in resource availability may affect the network end to end delay and throughput due to resource reallocation. The design of routing protocol should consider the intermittent availability of the spectrum both in terms of time and space which remains a very important design parameter.

2.6.3 Reconfiguration of Routing Path

Due to dynamic topology and intermittent spectrum availability, selected path becomes unusable in the particular area resulting in route failure. Reconfiguration of the routing path by selecting new nodes or different channel for communication can solve this problem. Naturally reconfiguration of paths increases the protocol overhead, therefore, the routing algorithm should be spectrum aware to reduce rerouting. Spectrum awareness means, routing algorithm should learn the good and bad route from previously taken decision. Spectrum fluctuations, channel stability and history should be considered as important metrics in route selection. Moreover, predictions of channel stability and link quality in future will help in the design of effective routing solutions.

To this end, it is suggested that, apart from convergence and spectrum aware properties of routing algorithm, it should be designed to boost the performance by using machine learning and prediction tools. The potential benefits of learning from the previous experience will help in controlling reconfiguration overhead.

2.6.4 Primary User Activity Modeling

The primary user behavior is a key parameter for taking routing decisions in multi-hop CRNs. The autonomy of the primary user will naturally affect the presence of the secondary user. Practically fully cooperative behavior from the primary user cannot be taken for granted. Instead, primary users will only cooperate with secondary users if the cooperation can bring them some benefits (Beibei Wang et al., 2010), otherwise, they will ensure that no secondary user should be able to use primary spectrum. Opportunistic access of spectrum should not increase interference to the primary network. To tackle this challenge, it is important to study, model and analyze the primary users behavior which results in designing efficient, self enforcing routing scheme. The importance of studying the primary user behavior is to help in modeling spectrum sharing among network users with various optimality criteria.

2.6.5 Cognitive User Behavior

CRN is equipped with intelligent users who are able to observe, learn and act to optimize their performance. Fully cooperative behavior of the cognitive user with each other cannot be taken for granted if they belong to different authorities and pursue different goals. Multiple cognitive users compete non-cooperatively and independently with each other for spectrum sharing. They will only cooperate with each other if the co-operation can bring them more benefits.

The optimization of spectrum usage among multiple users is generally a multi-objective optimization problem, which can be studied with different constituents and behavior of the cognitive users. The interaction among multiple cognitive users can be cooperative, non cooperative, stochastic or economic. The comprehensive study can efficiently improve the routing performance for multiple cognitive users along the routing path.

Spectrum sharing among multiple cognitive users at every hop along the path can be improved by jointly studying their behavior by leveraging game theory and multi-agent learning (Sharma and Gopal, 2010). The main challenge in multi-agent learning is that each learning agent must explicitly consider other learning agents, and coordinate their behavior, such that coherent joint behavior results.

2.7 Concluding Remarks

The chapter is intended to provide the context of the field and intellectual progression of cognitive radio networks. The cognitive radio network with its unique challenges is introduced followed by current state of research in the field of routing and channel selection. The current knowledge in the area under investigation is summarized which helped to identify strengths and weaknesses in previous work.

Good routing protocol design in cognitive radio network is highly dependent on properties and characteristics of spectrum band in use. This direct relationship necessities spectrum aware routing protocol, which selects links satisfying quality of service requirement of the secondary user and low interference to primary users. Moreover, the routing algorithm should be able to exploit network opportunities and be able to deal with uncertainties.

The limitations of state-of-the-art techniques demonstrating current issues being debated and how they are addressed by existing literature is addressed. Future directions and important design considerations for efficient routing protocols are discussed to establish a theoretical framework and methodological focus. Literature review presented in this chapter gives a theoretical basis for online and opportunistic routing algorithm and helps to determine the nature of research.

Chapter 3

Problem Definition

Chapter 3

Problem Definition

3.1 Problem Definition

To design and implement multi-agent reinforcement learning based spectrum aware opportunistic routing in cognitive radio network such that average per packet reward is maximized.

3.2 Objectives

- To review the literature and implement state of the art techniques.
- To explore the use of reinforcement learning in dynamic environment and search the optimal strategies for agents.
- To design link selection metric for selecting stable and non-interfering links with the primary user.
- To design and implement online opportunistic routing algorithm based on the transmission success probabilities using Multi-Agent Reinforcement Learning (MARL) and it's performance evaluation.
- To balance exploration and exploitation using soft-max action selection in relay selection process.

3.3 Approach

To deal with dynamics of Cognitive Radio Network and to achieve objective of online opportunistic routing with stable, non-interfering link with primary user, following approach is considered,

- Link Selection:
  - Time-Series Prediction of PU Activities.
  - Characterize the observable output.
  - Learn hidden models for predicting future behavior.
- MARL based Relay Selection:
  - Online and Opportunistic Path.
  - Explore and learn optimal strategies for incompletely known and dynamic environment.
  - Temporally successive prediction.
- Performance Improvement:
  - Balancing Exploration and Exploitation.
  - Softmax Action Selection.

Chapter 4

Proposed Research Methodology

Chapter 4

Proposed Research Methodology

Uncertain factors like unstable topology, intermittent channels and nodes availability affects the performance of dynamic cognitive radio networks in time varying and complex manner. These uncertainties are handled by adding intelligence to the cognitive node for observing, learning and acting accordingly. Context awareness using this intelligence makes cognitive node learn optimal or near optimal behavior in dynamic situation.

As shown in Figure 4.1, context awareness is achieved by three important design considerations in cognitive radio network,

- Strategic interactions among multiple cognitive nodes are modeled using Multi-Agent Reinforcement Learning (MARL). The cooperation and interaction among multiple cognitive nodes help to find stable and reliable path from source to destination.

- Temporal Difference (TD) learning to learn incrementally from temporally successive predictions. The dynamics due to intermittent channel and node availability is captured and incorporated in every decision of route formation.

- Hidden Markov Model (HMM) is used for traffic prediction on primary channels. This helps to decide channel availability, to access it opportunistically without interfering with primary user activities.

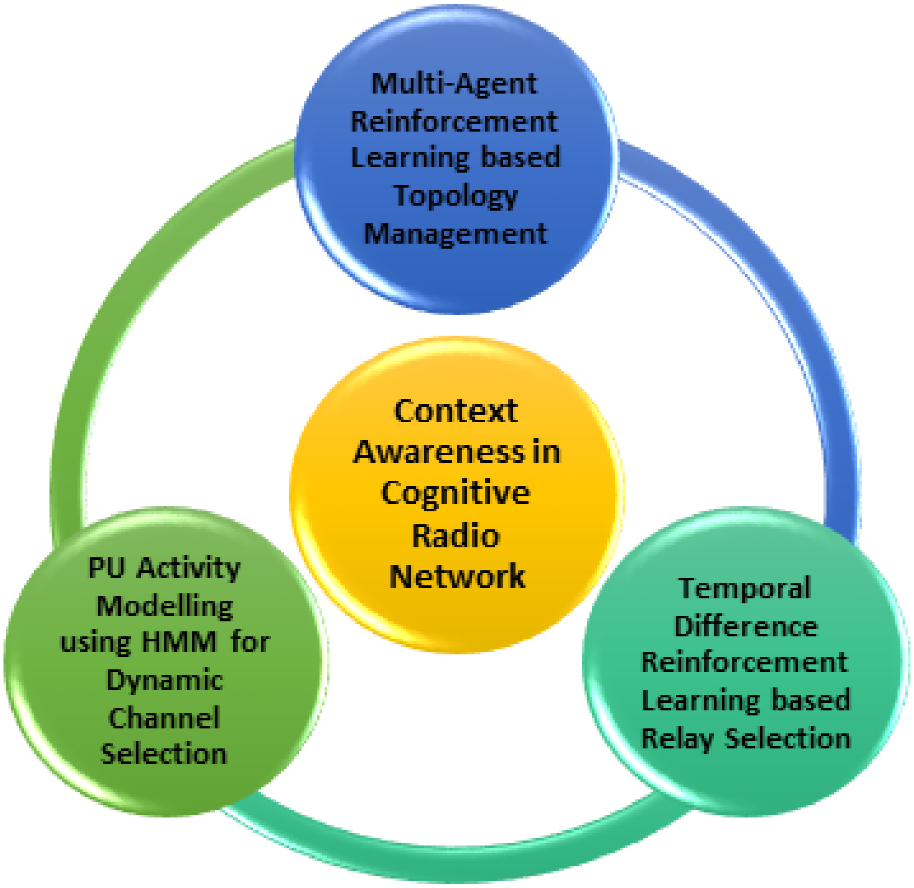

Multi-Agent Reinforcement Learning based Topology Management
Context Awareness in Cognitive Radio Network
PU Activity Modelling using HMM for Dynamic Channel Selection
Temporal Difference Reinforcement Learning based Relay Selection

Figure 4.1: Context Awareness in Cognitive Radio Network

The objective of this chapter is to overview above mentioned important methodological constructs and their need in design of online, opportunistic routing in Mobile Cognitive Radio Adhoc Network (MCRAN). Section 4.1 gives the overview of the MARL. Section 4.2 discusses Temporal Difference Reinforcement Learning with its policy evaluation for dealing with uncertainties of the environment. Section 4.3 gives details of action selection strategies for balancing exploration and exploitation. Section 4.4 provides details of Hidden Markov Model for characterizing primary user's channel usage traffic pattern. Section 4.5 gives concluding remarks on proposed research methodology.

4.1 MARL: Multi-Agent Reinforcement Learning

Context aware, interactive and adaptive dynamic decision problems involving multiple competitive or cooperative users are characterized as multi-agent problem. In multi-agent problems, the state of the environment changes autonomously or action of participating agent. Multi-Agent Reinforcement Learning (MARL) is seen as an approach for modeling complex system interacting with autonomous agents as shown in Figure 4.2.

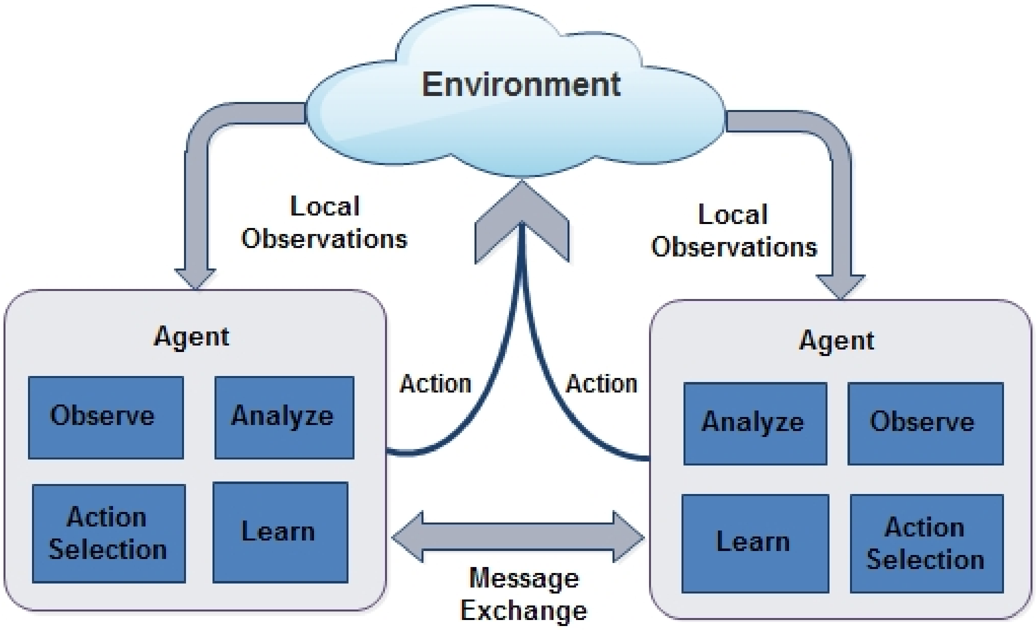


Figure 4.2: Multi-Agent Reinforcement Learning System

Multi-agent dynamic and distributed systems are studied and applied to problem solving in various domains like collaborative decisions, distributed control, resource management among multiple entities and robotic team actions etc (Parunak H.V.D., 1999; Stone and Veloso, 2000). In collaborative multi-agent systems, the best and optimal joint action is decided by interaction with other agents (Parker L. E. , 2002; Pynadath and Tambe, 2002). Agent's optimal strategy is decided by corresponding optimal equilibrium solution. Multi-agent systems are arising as new and natural way of looking at the systems that have temporal and dynamic characteristics (Busoniu et al., 2008). Fast and flexible decentralized multi-project scheduling with dynamic arrival of the project uses auction based system (Adhau et al., 2012). Multi agents policy in adaptive traffic signal controllers for the large-scale urban network converges cooperatively with neighbour's policies (Samah E. et al., 2013).

4.1.1 Need of Multi-Agent Interaction

MCRAN is a complex and dynamic system with various uncertain factors such as unstable topology, intermittent channels and node availability affecting performance in time varying and complex manner. Rule or Policy based actions are not able to deal with the continually changing environment. There is need of multi-agent interaction in MCRAN because,

- There is a lack of infrastructure to centrally manage and coordinate the task of identifying vacant bands and finding available neighbouring nodes.
- Dynamic topological changes due to mobility and intermittent availability of spectrum and nodes, necessitates distributed multiple agents to learn the solution via interactions to a nonlinear and stochastic task (Lunden et al., 2013).
- The strategic interaction among multiple agents helps to achieve performance benefits for cognitive user and avoid interference to primary users.
- Cooperation among multiple nodes brings joint benefits of reliable path from source to destination and improved spectrum utilization.
- Learning new behavior online improves the performance of complete multi-agent system to deal with dynamic behavior of the environment (Busoniu et al., 2008).

4.1.2 Reinforcement Learning in Multi-Agent Systems

The generality and simplicity make Reinforcement Learning (RL) suitable for the multi-agent dynamic environment. Q-learning is used for dynamic team computation method and determining joint action policy for the Markov games (Fang et al., 2013). Hybrid Reinforcement Learning algorithm using Q-learning deals efficiently with over-constrained environments and preventing transitions to undesired terminal states by predicting such state-action pairs (Fernandez et al., 2013). Semi- Markov decision process formulation of the call admission and routing for low orbit satellite using actor-critic with Temporal Difference learning archives higher revenue generation (Usaha and Barria, 2007). RL is applied successfully in multi-agent systems as learning technique which requires nothing about the dynamics of the environment. This trial and error based learning enables agent to transform environmental state to its goal state.

4.1.3 Benefits of Multi-Agent Reinforcement Learning

- Multi-agent system is modeled as Markov process to define learning goal for multiple RL agent.
- Interaction and reasoning among multiple agents help to act rationally, taking best policies for action selection.
- Every learning agent tries to maximize their payoff by minimizing risk in the worst situations (Weiss, 1999).
- Main distinguishing characteristic of MARL is that there is no global control and globally consistent knowledge suitable for distributed systems.
- Distributed data and control bring up inherent advantages of distributed system such as scalability, fault tolerance and parallelism.

The important challenges of defining learning goal for multiple agents, keeping track of other learning agents and coordinating with their behavior is simplified by Markov decision process representation for state transitions of multiple agents.

4.2 Temporal Difference Reinforcement Learning

Fundamental methods for solving the reinforcement learning problem are Monte Carlo and Temporal Difference. These methods learn state-value function for particular policy. The *value function* of the state represents agent's benefits with that state. It is updated by both methods by following particular policy π. The value of the state is the expected return that is expected cumulative discounted reward starting from that state and following the policy π. Both methods require only experience - sample sequences of states, actions and rewards from online or simulated interaction with the environment (Martin M. , 2011).

4.2.1 Monte Carlo Method

Monte Carlo (MC) method is way of solving the reinforcement learning problem based on averaging sample returns. Average of the returns observed after visits to particular state is used to update the state - value function. As number of returns are observed, the average should converge to expected value. In this method experiences are divided into episodes and only on the completion of an episode the value estimates and policies are changed. MC method is thus incremental in an episode-by-episode sense, but not in a step-by-step sense. Monte Carlo learns from the error between predicted and actual outcome. MC methods involve averaging or maximum function over random samples of actual return. It has to wait for entire episode for actual *return*. Then it is used for updating value function, represented by following equation,

$$V(s_t) \Leftarrow V(s_t) + \alpha \left[R_T - V(s_t)\right] \tag{4.1}$$

where R_T represents actual return and α is constant step-size parameter. Every update to $V(s_t)$ is dependent on R_t. Actual reward R_T is known at the end of the episode, therefore, Monte Carlo has to wait for the complete episode for making any update. As shown in Figure 4.3, MC makes every update at the end of the episode after final time step T shown using arrow starting from time-step t to T.

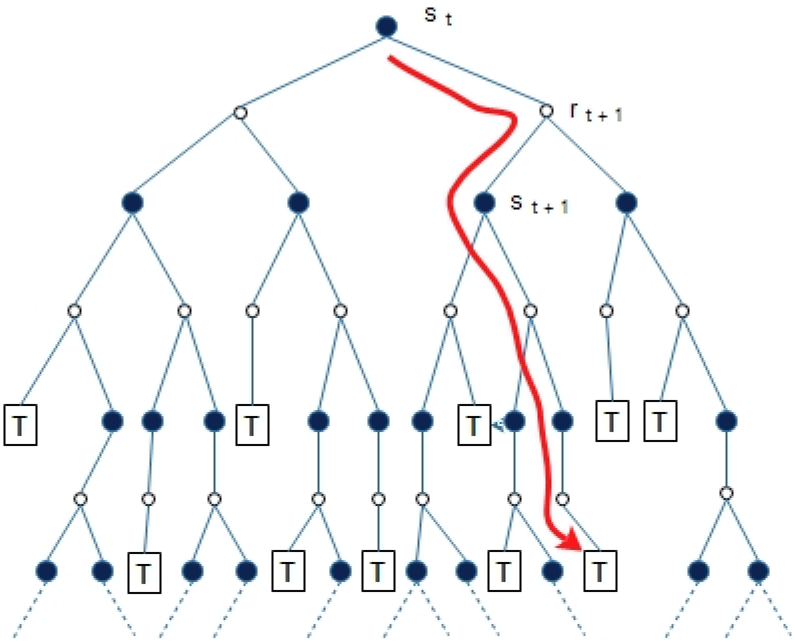


Figure 4.3: Episode - by - Episode Policy Evaluation

4.2.2 Temporal Difference

Temporal Difference (TD) method learns incrementally using past experiences with an incomplete knowledge of the system. The difference between temporally successive predictions is used to update the estimate over a time. TD methods collect experience to solve the prediction problem. TD is called *bootstrapping* method as it updates the value function in part and based on existing estimate. TD method makes useful updates to $V(s_{t+1})$ by using observed reward r_{t+1}. The simplest TD update rule is,

$$V(s_t) \Leftarrow V(s_t) + \alpha \left[r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \right] \tag{4.2}$$

where, α represent step-size. The step size is $\frac{1}{t}$ for processing t^{th} reward for action a, . γ is discount rate, with $0 \leq \gamma \leq 1$. The discount rate represents importance of future reward. TD makes incremental updates after every action selection. Every state change from s_t to s_{t+1} makes update to value function $V(s_t)$, shown using separate arrow in Figure 4.4.

In multi-step prediction problems, experience comes from observation and outcome sequence, with x as the final outcome. The sequence of predictions $P_1, P_2, P_3, ..., P_m$,

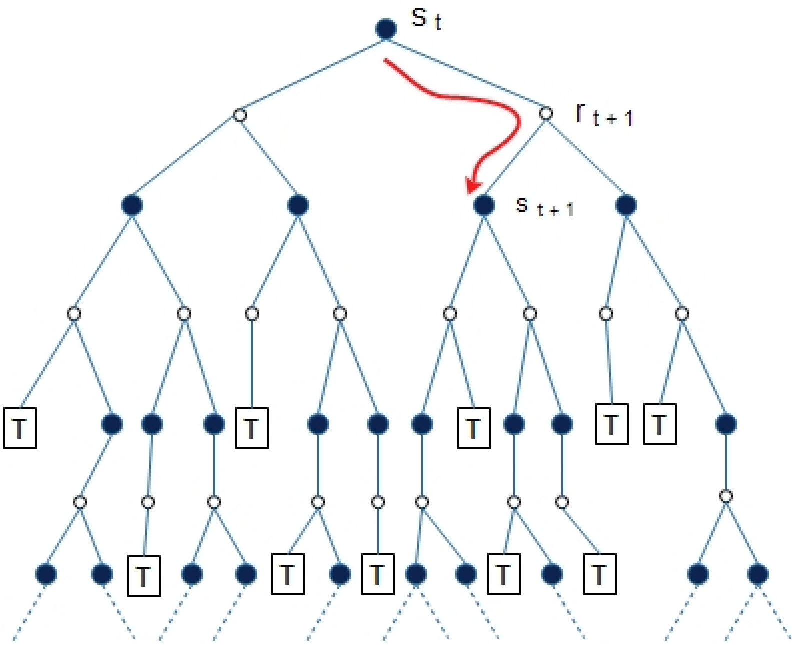


Figure 4.4: Step - by - Step Policy Evaluation

are estimated to predict x. In Monte Carlo method, each increment depends on the error between P_t and x. TD method computes the solution incrementally by computing the error $(x - P_t)$ as sum of changes in the successive prediction represented as:

$$x - P_t = \sum_{k=t}^{m} (P_{k+1} - P_k) \tag{4.3}$$

Temporal Difference Procedural Form: Arbitrarily initializes value function $V(s)$, considering policy π for evaluation

---

```
Step 1: Initialize state $s$
Step 2: Repeat continuously for every step of episode:
    Select action $a$ as per policy $\pi$ for state $s$
    Perform action $a$ in state $s$ and collect reward $r$ with transition to next state, $s'$
    $V(s) \Leftarrow V(s) + \alpha [r + \gamma V(s') - V(s)]$
```

Copy s' to s

Step 3: Stop with s as terminal state

4.2.3 Monte Carlo Versus Temporal Difference

The difference in Monte Carlo and Temporal Difference methods is with approach for solving reinforcement learning problem. The following table summarizes the differences in MC and TD methods,

Table 4.1: Monte Carlo and Temporal Difference: Comparison

Monte Carlo Method	**Temporal DifferenceMethod**
$V(s_t) \Leftarrow V(s_t) + \alpha\left[R_T - V(s_t)\right]$	$V(s_t) \Leftarrow V(s_t) + \alpha\left[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)\right]$
State value update - at the end of the episode	State value update - immediately at the next step
Target - R_T known only at the end of the episode	Target - $r_{t+1} + \gamma V(s_{t+1})$ known immediately after next step
As $R_T = r_{t+1} + r_{t+1} + +r_T$, No learning during episode	Learning occurs during an episode at s_{t+1} and r_{t+1}, it bootstrap
Waits until the end of episode	Online, fully incremental
If bad policy is followed, it continues until the end of the episode	It recognize a bad policy and can learn good policy during an episode
Needs more memory to maintain state values for entire episode	Requires less memory and less peak computation

TD method is suitable for the dynamic environment where actual outcome is difficult to predict. In TD method,

- Prediction about final outcome and actual return is refined gradually with every action selection and state transition.
- This results in decreased *error in estimation* and learns from experience of every temporal change.
- Every update in state value function is temporal difference between next state $V(s_{t+1})$ and $V(s_t)$.

Example - *Error in estimation* of Monte Carlo and Temporal Difference

---

Assumption: Predicted outcome at time step t is $V(s_t) = 30$

Final time step is $T = 6$

Actual Outcome at the end of episode is $V(s_T) = 43$

Table 4.2: Monte Carlo and Temporal Difference: Error in Estimation

Predicted Outcome $V(s_t)$	**Error in Estimation**	
	Monte Carlo	**Temporal Difference**
30	13	10
40	3	-5
35	8	5
40	3	3
43	0	0
43	0	0
Total	27	13

- Error in estimation in Monte Carlo is difference between actual and predicted outcome.
- Error in estimation in Temporal Difference is difference between predicted outcome of next state and previous state.
- The total *error in estimation* of TD is much lesser than MC.
- TD method learns good policy faster compared with MC.

In the view of above facts, state value update in proposed online and opportunistic routing is done using Temporal Difference method. This enables every cognitive node to deal with dynamic environment and learn good policies.

4.2.4 Policy Evaluation

Distinguishing feature of TD is to evaluate the action instead of instructing the agent with correct action. Instructive feedback indicates correct action selection, as in supervised learning. This needs *trial and error* search for evaluating different actions for good behavior. Trying different actions for indicating *how good* they are, need active exploration.

In evaluating feedback, the received reward provides some information about the goodness of the action. Action correctness is relative property with other actions. Therefore, it is determined by trying all actions and comparing their reward. This explicit search among alternative actions is performed by generate-and-test method. In this method, agent selects actions records its outcome and retains only those with most effective outcome. TD is *learning by selection* instead of by *instruction*.

Policy Iteration

The agents goal in TD learning is to improve received total reward over a long run. The state with its probability of selection is called as agents *policy* π. Agent should learn optimal policy to maximize its reward as a result of its experience.

The expected future reward is used by TD to estimate *value function*. A future reward of any agent is dependent on successive actions it selects. Therefore, *value function* is defined in terms of particular *policy* of action selection. Policy π is the probability of action a in state s represented as $\pi(s, a)$. $V^{\pi}(s)$ is the value of a state s following a policy π. It is the represented in terms of expected return,

$$V^{\pi}(s) = E_{\pi}\left\{R_t | s_t = s\right\} \tag{4.4}$$

where E_{π} is expected return following the policy π. TD learns to find the optimal policy for achieving maximum reward over long run. The policy π is better than policy π' if and only if $V^{\pi'}(s) \leq V^{\pi}(s)$ for every state s in S . The *optimal policy* is better policy if compared with all other policies.

The value function V^{π} is used to evaluate π for getting good policy π'. Computation and evaluation of $V^{\pi'}$ results in better policy π''. *Policy iteration* is policy evaluation followed by policy improvement for finding optimal policy. Policy evaluation itself is iterative function. Generalized policy evaluation for TD learning is after every action selection and intermediate reward generation. TD method immediately makes useful update in estimation of V^{π} using the observed reward.

4.2.5 Benefits of Temporal Difference Prediction

Temporal Difference method predominates in multi-step dynamic decision problems over conventional learning methods. Step by step TD prediction is tuned as the final outcome. The benefits of Temporal Difference method are:

- TD method bootstraps from previous guesses or learns estimate in part from other estimates.
- TD methods do not require model of the environment like in dynamic programming.
- TD learns in an on-line and in fully incremental fashion. Monte Carlo is slow as it delays complete learning until the end of the episode for knowing actual

return.

- TD is very fast learning as it learns from every state transition regardless of action selection.
- TD method converges faster than Monte Carlo methods on stochastic task.

4.3 Action Selection

There are three ways to select an action among all alternative actions based on their values,

4.3.1 Greedy Action Selection

Selection of the action with highest estimated action value is greedy action selection. In other words, selection of action a^* at t^{th} play where $Q(a^*) = \max_a Q_t(a)$. This is called as exploiting current knowledge for maximizing immediate reward. Greedy selection method does not try and test any inferior action which may perform better in actual run.

4.3.2 ϵ-Greedy Action Selection

Alternative action selection is to select the best action most of the time. But sometimes, action is selected randomly with uniform probability of ϵ, independent of action-value pair. This is also called as near-greedy action selection. The advantage of this method is in long run for converging to the selection of optimal action. As the number of plays increases, chances of sampling of every action increases, ensuring every action value converges to the best or optimal value.

4.3.3 Softmax Action Selection

ϵ-Greedy selection mitigates the disadvantages of greedy selection, but it chooses equally among all actions. There are chances of selection of worst action instead of best actions. ϵ-Greedy is unsatisfactory for the task with limited number of plays, where worst action selection is very bad. This problem is solved by varying the action selection probabilities using grading function on estimated value. This is softmax action selection rule. The simplest softmax selection rule uses Gibbs or Boltzmann distribution. The probability of action a on t^{th} play is obtained by,

$$P(a) = \frac{e^{Q_t(a)/\tau}}{\sum_{b=1}^{n} e^{Q_t(b)/\tau}} \tag{4.5}$$

Where τ is considered as *temperature* with $0 \leq \tau$. The action selection is equiprobable with high value of temperature i.e selection probability of all actions is nearly same. Low temperature has greater difference in action selection probability. Softmax action selection is the same as greedy action selection as in the limit as $\tau \rightarrow 0$. High temperature causes selecting any action with same probability. As the temperature decreases gradually, softmax starts selecting better action and behaves like greedy approach with low temperature.

4.3.4 Exploration Versus Exploitation

Selection of the greedy action is exploitation of available knowledge of action values. Selection of non-greedy action is called as exploration. It enables to evaluate and update estimate of the non-greedy action values. To maximize the expected reward in one play exploitation can be better choice. Exploration improves performance and produces the greater total reward in long run.

There is chance of having, at least, one action probably better than best action but the agent is not aware of which one. Exploration helps to discover non-greedy actions which are better than greedy action. If reinforcement learning task is having many remaining plays, then exploration can help to find better action by exploring non-greedy actions. Exploration has lower reward in the short run but higher reward in the long run because after discovering better action, agent can exploit them.

With single action selection agent is either exploring or exploiting, therefore, there is conflict between whether to explore or exploit. Exploration and exploitation depend on precise estimate of action value, uncertainties and the number of remaining plays. Softmax action selection is effective way of balancing exploration as the action selection probability is graded function. ϵ-Greedy action selection chooses equally among all actions likely to select worst appearing action. Softmax action selection selects any action as all are equi-probable with high temperature. As temperature decreases gradually, it starts selecting best action. It assures with limit to $\tau \rightarrow 0$, all actions are evaluated for its efficiency and converge to optimal action value of every action.

4.4 Primary User Activity Model

Intelligently determining ongoing primary user activities in licensed spectrum band helps to identify transmission opportunities for cognitive user. This improves spectrum utilization, avoids interference to the primary network and helps to evacuate the band without affecting primary user activities. Precise estimation and modeling of primary network channel dynamics lead to much more effective spectrum usage in CRN (Canberk B. et al., 2011).

4.4.1 Alternative Renewal Process

Spectrum occupancy in primary network is non-deterministic in reality. Therefore, primary channel traffic is modeled using stochastic Alternative Renewal Process (ARP) (Yuan G. et al., 2010). Channel occupancy is perceived as two state repairable system, that is system is either in working (up) state (that is system is available and can be used) or under repair state (down) (that is the system is not available till repaired). A primary channel is treated as a two state repairable system which can either be occupied by a licensed/primary user or not.

As per secondary user perspective, any channel is said to be in,

- *idle state*, when the primary user is not using it.
- *busy state*, when the primary user is using it.

As per two state repairable system, channel's idle state is considered as working state of the system (allowed to be used by cognitive radio network). Channels busy state is considered as under repair state of the system (not allowed to be used by cognitive users). Alternate Renewal Process is the best method to model two state repeatable process. Channel modeled using ARP exist in two states i.e idle and busy. These states occur in an alternate manner. Every cycle consists of two states, one followed by other that is idle state followed by the busy state. This cycle of two states renews, repeats or reoccurs as shown in Figure 4.5,

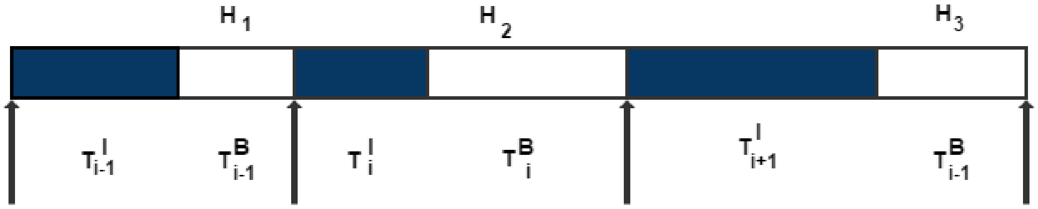


Figure 4.5: Channel Modeling as Alternate Renewal Process

Renewal cycle shows idle duration of the channel using T^I and busy duration using T^B. The Cognitive user collects statistics of primary user channel i.e idle duration of channel and busy duration of the channel. Every cognitive user constructs channel occupancy model as two state ARP. The primary network traffic on every channel alternates between idle state (when not used by primary user) and busy state (when used by primary user). The cycle of idle and busy state renews or repeats itself. The sequence of idle times $\{I_n : 1 \leq n\}$and busy time $\{B_n : 1 \leq n\}$ is i.i.d. (independent identically distributed) sequence. I is generic idle time and B is generic busy time and the distribution of I need not be same as distribution of B.

4.4.2 Discrete Markov Process

The Discrete Markov Process is described as set of states. Agents switch to new or same state according to the set of probabilities associated with the state change. At any time, system is in actual state q_t where, $t = 1, 2, 3, ...$ Probabilistic description needs description of current state and all previous states. As it is Markov chain it is described using only current state and previous state

$$a_{ij} = P[q_t = s_i | q_{t-1} = s_j] \tag{4.6}$$

where, a_{ij} is state transition probability coefficient for all values of i and j. The value of $a_{ij} \geq 0$ represented as,

$$\sum_{j=1}^{N} a_{ij} = 1 \tag{4.7}$$

Hidden model in Discrete Markov Process

There are few applications where agent is observing signal without complete information of source of signal. Agent represents it in the form of observation sequence. Analysis of the observation sequence helps to understand hidden model representing the behavior of source signal.

4.4.3 Hidden Markov Model

Non-deterministic traffic of the primary user produces observable outcome in the form of idle and busy state, which can be characterized as signal. Characterizing real world observations in signal model have following benefits,

1. The theoretical description is used to process signals for generating desired output.

2. Observable outcome represented as signal, is helpful to learn about signal source without source being actually available.

3. Enable to realize practical systems requiring prediction, recognition and identification.

The spectrum occupancy in primary network is non-deterministic and stochastic in reality (Saleem and Rehmani, 2014). The stochastic Hidden Markov Model (HMM) is used as probabilistic function of the Markov chains. The purpose of using HMM is to understand the behavior of the primary user and predict chances of accessing primary band opportunistically. This is achieved by characterizing signal and training HMM with parameters that best account for observed signal. The trained HMM generates observation sequence and evaluates its probability of likelihood with actual sequence.

4.4.4 Elements of Hidden Markov Model

Hidden Markov Model is defined with following terms,

The Number of States, N: The states in HMM are hidden but have significance with respect to observable outcome in that state. The state at time t is q_t, which is one of the state from set of states $S = s_1, s_2, ..., s_N$.

The Number of Distinct Observation Symbols, M: It is the physical output of the system being modeled. The number of possible observation symbols possible per state represented as discrete alphabet size $V = v_1, v_2, ..., v_M$.

State transition Probability, A: For the case where, single step is required to reach any state from any other state is represented by state transition probability metrix $A = a_{ij}$, where

$$A = [a_{ij}] = P\left[q_{t+1} = s_j | q_t = s_i\right], \quad 1 \le i, j \le N \tag{4.8}$$

$a_{ij} \succ 0$ for any state i reached from j in single step. Other types of state transitions would have $a_{ij} = 0$ for one or more pairs of states.

Observation Probabilities: The observation probability distribution for state j for observing v_m, $m = 1, ..., M$ in state s_j

$$B = [b_j(m)] = P\left[O_t = v_m | q_t = s_j\right], \quad 1 \le j \le N \quad and \quad 1 \le m \le M \tag{4.9}$$

Initial state distribution, Π :

$$\Pi = [\Pi_i] \quad where \quad \Pi_i \equiv P\left(q_1 = s_i\right) \tag{4.10}$$

The compact notation for HMM parameter set is given with implicit definition of N and M,

$$\lambda = (A, B, \Pi) \tag{4.11}$$

Given, λ, the model can be used to generate an arbitrary number of observation sequence of arbitrary length. Observation sequence is given as,

$$O = O_1 O_2 ....O_T \tag{4.12}$$

where each observation is symbol from observation symbols set V and length of the sequence is T. The observation sequence generator using HMM is represented in following algorithm,

Given $\lambda = (A, B, \Pi)$ with implicit specification of Number of states N and Number of distinct observation symbols M

1. Use an initial state distribution π, for deciding initial state $q_1 = S_i$
2. Initialize $t = 1$
3. Choose observation symbol $O_t = v_m$ for state S_i, according to the state probability distribution over possible symbols in state S_i i.e $b_i(m)$
4. Transit to new state $q_{t+1} = S_j$ from state S_i, as per the state transition probability distribution from state S_i over all possible states
5. Increment $t = t + 1$ and For $t \prec T$ return to step 3. Else, terminate the procedure.

This procedure is capable of generating observation sequence of arbitrary length. Adjusting HMM parameter set helps to improve HMM performance by maximizing likelihood with original sequence.

4.5 Concluding Remarks

The chapter provides detailed insight of methodological constructs used for designing context aware online, opportunistic routing algorithm for MCRAN. The chapter advocates the use of multi-agent reinforcement learning for taking interactive and adaptive dynamic decisions in MCRAN. The overview of Temporal Difference learning is given to deal with dynamic and uncertain environment of cognitive radio network. TD is used as incremental learning for predicting future behavior on the basis of past experiences and incomplete knowledge of the system. Policy evaluation with action selection strategies to balance exploration and exploitation are discussed.

The Alternate Renewal Process for modeling non deterministic primary user traffic as Discrete Markov Process is discussed. Hidden Markov Model is illustrated in detail as it is used to predict the channel behavior for deciding better routes for transmission. The overall aim of this chapter is to provide conceptual details of methodology and its appropriateness for context and spectrum aware routing in cognitive radio network.

Chapter 5

Research Strategy and Algorithmic Design

Chapter 5

Research Strategy and Algorithmic Design

5.1 Introduction

This work investigates the dynamic decision making for spectrum aware routing in Mobile Cognitive Radio Ad Hoc Network (MCRAN) using multi-agent reinforcement learning and Hidden Markov Madel. The Multi-Agent Reinforcement Learning based opportunistic routing scheme for MCRAN is proposed for jointly addressing the link and relay selection problem. MCRAN is complex and dynamic system with various uncertain factors such as unstable topology, intermittent channels and node availability that affects performance in time varying and complex manner. Rule or policy based actions are not able to deal with the continuously changing environment. Reinforcement learning based agents provide intelligence to the system for finding near optimal solutions.

The Multi-Agent Reinforcement Learning (MARL) based channel and relay selection algorithm achieve good quality of service requirements by selecting reliable and stable path online. The MARL based opportunistic routing scheme for MCRAN, addresses the issue of routing packet in non-stationary and dynamic environment. Every routing decision is made with respect to the current state of the environment and actual transmission success probabilities.

Conventional routing mechanisms attempt to find a fixed path for forwarding packets from source to destination. There are a large number of routing protocols proposed for cognitive radio ad hoc network, which finds a fixed path by using route discovery procedures (Saleem Y. et al.,2015). Such fixed path schemes fail to address the uncertainties and dynamic changes in topology due to intermittent spectrum and route availability.

Routing in MCRAN is defined as the selection of sequence of relay nodes in the absence or erroneous knowledge of topology. Routing should be spectrum aware with following unique characteristics:

Mobility prediction: It is necessary to accurately represent movements of the cognitive nodes. The preferred relay for forwarding packet should not be moving towards the primary user region.

Primary user interference awareness: Interference to primary user should not be increased because of particular link selection.

Opportunistic routing: Partially observable and non-stationary environment necessitates routing in opportunistic and online manner.

Proposed algorithm simultaneously learns channel statistics and good routes with incomplete knowledge of the topology. MARL using Temporal Difference learning is applied to MCRAN for modeling and designing of efficient spectrum and relay selection scheme. Every Cognitive Node (CN) in MCRAN is intelligent agent which is observing, learning and acting to improve its own performance. CN also improves system performance by improving utilization of dynamically available spectrum. The distinguishing characteristics of the MARL based online and opportunistic routing are, Spectrum Awareness, Intelligent Link Selection, Strategic Interaction among Network Nodes, Temporal Difference Policy Evaluation and Exploration using Soft-max Action Selection, as shown in Figure 5.1.

MARL is used to model strategic interaction among multiple CN's for making intelligent decisions on spectrum and route selection. The importance of using MARL in MCRAN is multi-fold,

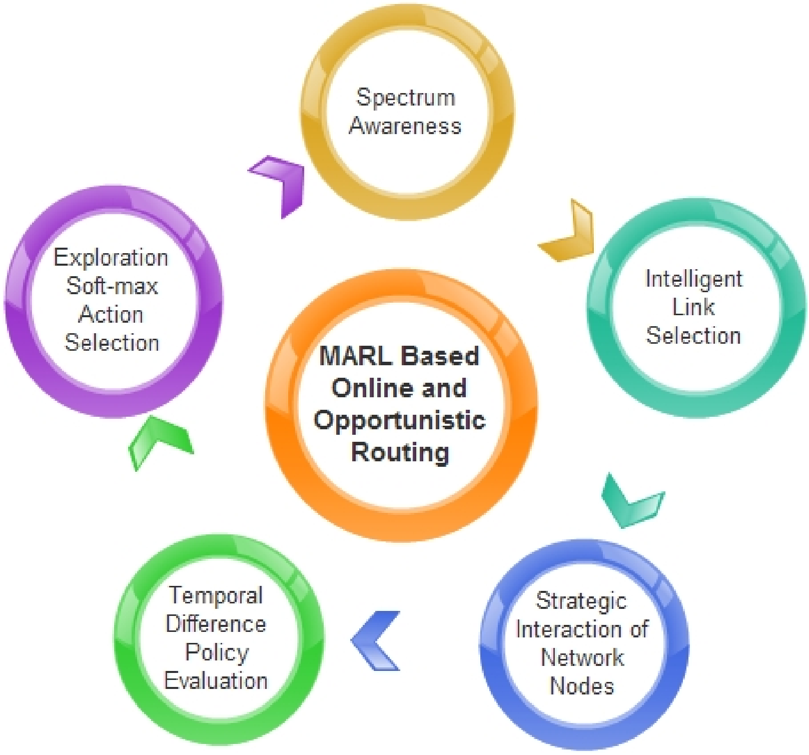

Spectrum Awareness
Intelligent Link Selection
Strategic Interaction of Network Nodes
Temporal Difference Policy Evaluation
Exploration Soft-max Action Selection
MARL Based Online and Opportunistic Routing

Figure 5.1: MARL Based Online and Opportunistic Routing

- Modeling strategic interaction among CNs as multi-agent system to analyze agent's behavior and policies of neighbours to improve reliability and performance of MCRAN.
- MARL enables spectrum aware routing as multi-objective optimization problem under various network settings.
- Flexible, self-organized and distributed approach using MARL enables to derive efficient spectrum sharing only with local information.

5.2 System Architecture and Proposed Model

5.2.1 System Description

MCRAN consists of primary and cognitive nodes. Primary Users (PU) own the M different channels. In each time-slot, these channels are observed by CNs and opportunistically utilize in an intelligent manner. Selection of the best available channel and prominent node towards destination improves the performance by selecting optimum stable links on complete path. By leveraging benefits of reinforcement learning, performance is improved with observed reward for successful completion of the task. Figure 5.2 shows, the entire opportunistic spectrum access and the node selection process is divided into following four sub-processes:

- Observation: Each CN observes channel activities concerning the surrounding network and samples monitor PN activities on every channel in vector Q.
- Traffic Pattern Prediction: Observed vector Q is used for traffic pattern analysis of M channels. The traffic pattern analysis is used to understand traffic trend and predict Channel Availability (CA).
- Interaction: Multiple CNs share their environmental statistics and their own preferences for participating in route formations.
- Adaptation: This final stage involves selection of idle channel and prominent forwarding agent for data transmission and updating score vectors as per reward received.

5.2.2 System Model

The problem of routing packet from source to destination is defined as selection of intermediate relay nodes and links to be used on every hop (Barve and Kulkarni, 2014). The intermittent spectrum availability and absence of topological information differentiate MCRAN from normal mobile network. Therefore, the problem of finding reliable path from source agent s to destination agent d is designed with no or limited knowledge of environmental dynamics. The total number of channels are represented using M. The set of Cognitive Nodes, $CN = \{CN_0, ..., CN_n\}$ interacts for exchanging local observations. Every agent CN_i computes the Channel Availability CA, of all detected channels C, where $C \subseteq M$. CA is calculated using predicted sequence of channel. Predicted sequence is generated by modeling PU's traffic using Hidden

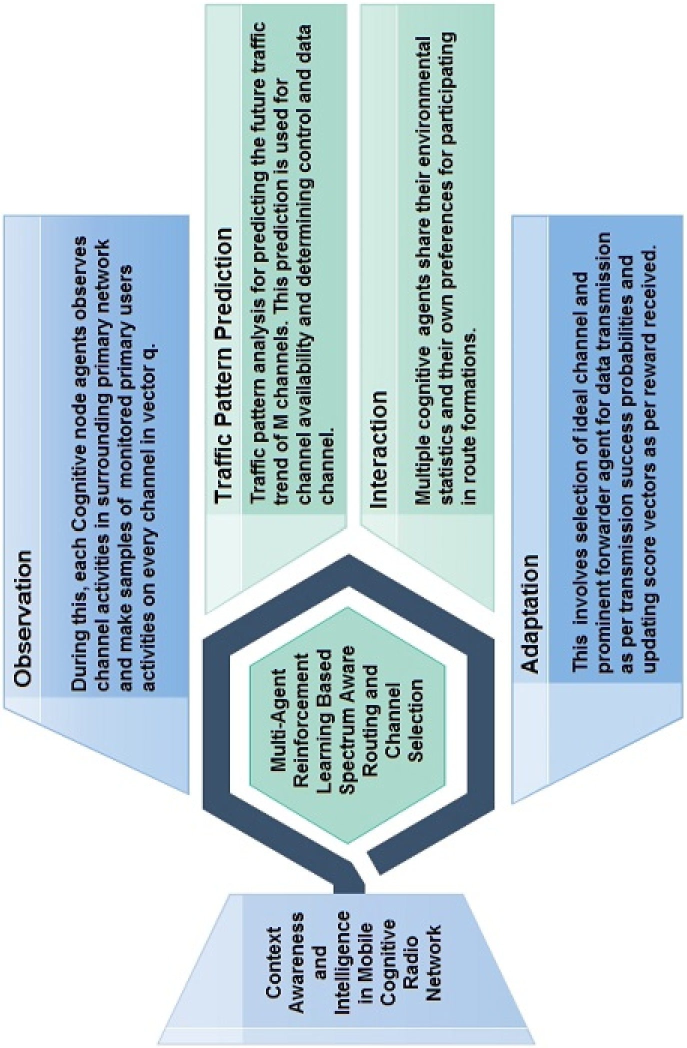

Observation
During this, each Cognitive node agents observes channel activities in surrounding primary network and make samples of monitored primary users activities on every channel in vector q.
Traffic Pattern Prediction
Traffic pattern analysis for predicting the future traffic trend of M channels. This prediction is used for channel availability and determining control and data channel.
Interaction
Multiple cognitive agents share their environmental statistics and their own preferences for participating in route formations.
Adaptation
This involves selection of ideal channel and prominent forwarder agent for data transmission as per transmission success probabilities and updating score vectors as per reward received.
Multi-Agent Reinforcement Learning Based Spectrum Aware Routing and Channel Selection
Context Awareness and Intelligence in Mobile Cognitive Radio Network

Figure 5.2: Spectrum Access and Routing Stages

Markov Model(HMM). The objective is to select link with lowest transmission cost represented by LC_c on channel c where $c \in C$. The lowest link cost represents a link having more CA, best propagation distance, lowest end-to-end delay and switching cost.

In the path formation process, one of the CN called source node s, initiates interaction with neighbouring CNs, $N(s)$ within its vicinity. Nodes in the range of propagation distance of the source receive relay request. Interesting neighbouring nodes N_s, $N_s \subseteq N(s)$, respond with their estimated best score to work as relay agent. One of the willing neighbour is selected as the next candidate forwarding node

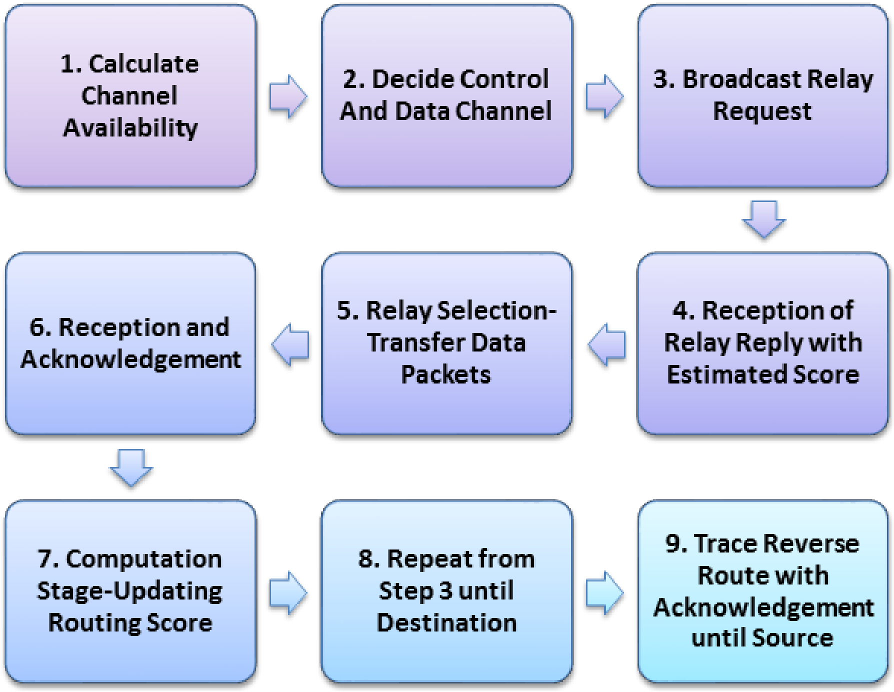

1. Calculate Channel Availability
2. Decide Control And Data Channel
3. Broadcast Relay Request
4. Reception of Relay Reply with Estimated Score
5. Relay Selection-Transfer Data Packets
6. Reception and Acknowledgement
7. Computation Stage-Updating Routing Score
8. Repeat from Step 3 until Destination
9. Trace Reverse Route with Acknowledgement until Source

Figure 5.3: Flow diagram of Link and Relay Selection

using the estimated best score and agents own experience of neighbour as forwarder or relay. After selection of the next forwarder, s sends data packet on data channel having better CA compared to all detected channels. The process of relay agent selection and channel selection is repeated by every relay agent until packet is not received at the destination. After successful reception at the destination, the destination sends an acknowledgment to the source agent s by tracing the path backward until the source is reached.

Transmission from agent j incurs the transmission cost LC_c^j on channel c. Transmission cost consists of the expected time to transfer link layer frame, channel switching cost and channel availability. After successful reception of the packet at the destination, it gets reward of $\Re$. The detailed flow of the multi-agent reinforcement learning based routing algorithm is shown using Figure 5.3.

The problem of selecting random sequence of h hops/agents for the packet generated at the source s with index i for destination d, where Average Per Packet Reward

(APR) is maximized. APR is represented as,

$$APR = \frac{1}{P}\left(\sum_{i=0}^{P}\left(\Re - \sum_{j=0}^{h} LC_c^j\right)\right) \tag{5.1}$$

Where, P is the number of packets successfully transmitted, $\Re$ is the reward received and h is the number of hops to reach the destination. LC_c^j is the cost of the link used at every hop. Selection of the lowest cost link at every hop increases the reward achieved at the destination. Multi-agent reinforcement learning based opportunistic and spectrum aware routing is designed to maximize Average Per Packet Reward (APR).

5.3 Observation: Primary User Activity Monitoring

One of the most important requirement in MCRAN is to sense spectrum holes. CN agent periodically scans and senses multiple channels for detecting spectrum holes. The CN agent monitors and samples primary user activity. Local observations of CN agent detect signal from primary transmitter. Following hypothesis model is used for collecting samples of primary network activities on various channels,

$$Q = \begin{cases} n(t) & H_0 \\ s(t) + n(t) & H_1 \end{cases} \tag{5.2}$$

where, $q(t)$ in vector Q is signal received at time t at sample frequency f_s. Vector Q is represented as,

$$Q = \{q(1), q(2), q(3), ...q(t)....q(p)\}$$

H_0 is null hypothesis stating the absence of primary user signal in a certain spectrum band. Existence of primary user signal is shown using H_1 as an alternative hypothesis (Akyildiz I. et al., 2006). $s(t)$ is unknown signal of primary user which is independent and identically distributed and $n(t)$ is the Additive White Gaussian

Noise (AWGN).

Cognitive node compares the integrator output $q(t)$ with threshold λ for every channel c. At every instance t, the node records an observation symbols OS_t as per following condition,

$$OS_t = 0 \quad \text{if} \ q(t) \leq \lambda; \qquad OS_t = 1 \quad \text{if} \ q(t) \geq \lambda \tag{5.3}$$

Every CN agent continuously senses and records such observation and creates observation sequence for T time slots,

$$OS = OS_1, OS_2, ..., OS_T, \quad where \ \ OS_t \in [0,1] \quad \forall t = 1....T \tag{5.4}$$

More number of 1's in this sequence indicate more traffic on the primary channel. On the other hand more number of 0's indicate spectrum holes or white spaces available for dynamic and opportunistic spectrum access.

Spectrum Sensing: The Energy Detection technique of spectrum sensing are implemented using Software Defined Radio (SDR). In SDR, all radio communication related task like signal manipulation and processing is implemented in software instead of hardware. It is implemented using flexible low-cost platform of Universal Software Radio Peripherals (USRP) and GNU Radio. The wide band spectrum sensing/analyzer gives N time samples of $q(t)$ sampled at center frequency f_s.

The channels of Wireless Local Area Network (WLAN) using IEEE 802.11 protocol are sensed using USRP and GNU radio. The WLAN channels with trademark Wi-Fi uses frequency range of 2.4 GHz. It is divided into a multitude of channels. Three channels in 2.4 GHz frequency range are sensed for generating observation sequence OS of primary user activity for T time slots.

5.4 Traffic Pattern Prediction using Hidden Markov Model

CN agents are having cognitive capability to sense and use the idle resources of the primary network. Intelligent search and exploration of spectrum opportunities for dynamic spectrum access can improve the performance of cognitive node. Each node is responsible for sensing each channel locally and using this information to predict the future behavior of the target channel. Temporarily unoccupied channel is used by cognitive user without creating interference for transmission of primary user. The performance of the cognitive node is optimized by exploring the vacant channel opportunities and avoiding the interference to the primary network.

5.4.1 Primary User Traffic Modeling

Primary network traffic is considered as binary sequence and modeled using stochastic Alternative Renewal Process (ARP) (Yuan G. et al., 2010). Every CN agent constructs channel occupancy model as two state system. The traffic on every channel alternates between idle state (when not used by Primary Network) and busy state (when used by Primary Network). The cycle of idle and busy state renews or repeats itself. The sequence of idle times $\{I_n : n \geq 1\}$ and busy time $\{B_n : n \geq 1\}$ are independent identically distributed sequence i.i.d. I is generic idle time and B is generic busy time, the distribution of I need not be same as the distribution of B. Binary sequence represented using ARP is used to analyze and predict the ongoing activities of PU channels.

Prediction Model

Binary sequence represented using ARP is used as training and testing data for prediction engine. This observed data is used for designing the prediction model which is used to predict the channel occupancy in the form of binary sequence. This predicted binary sequence is used to predict the channel availability. Monitoring and analyzing the ongoing activities of primary user band can avoid interference to primary users. Effective Modeling of these primary user activities can help to opportunistically use and evacuate the channel without affecting the primary users transmission.

5.4.2 Hidden Markov Model(HMM)

Channel dynamics of primary network is modeled using Hidden Markov Model(HMM). It is capable of characterizing real-world observations in terms of observation models. Characterization of such real world observations in terms of observation model is potentially capable of learning about signal source without a presence of the source. The HMM has been used in areas of speech recognition (Rabiner L. R. , 1989). HMM has found application in radio error characterization in digital channel and analysis of internet communication channel. HMM based channel traffic prediction was proposed in the literature with brief analysis of HMM predictor (Sharma M. et al., 2008). Channel status prediction using HMM is used to save the sensing energy and help to exploit the spectrum holes (Tumuluru V. K. et al., 2010).

Spectrum occupancy in primary network is non-deterministic in reality. Stochastic HMM explores spectrum holes more efficiently with lower probability of error in prediction. Every agent in MCRAN periodically scans and senses multiple licensed channels to compute interference temperature and compare it with predefined threshold value. Every node makes such observation and records channel Observation Sequence $OS = OS_1, ..., OS_n$, for n channels, over the time period T. OS is series of 0's or 1's where, $(OS)_t \in [1, 0], \forall t = 1, ..., T$, showing whether the channel is available or not at time t.

HMM characterizes OS into parametric random process. Spectrum occupancy model of primary network is learned using observation sequence samples for predicting the spectrum occupancy in next time slot.

Problem Description:

Given observation sequence OS of channel c for T time slots which is monitored and recorded from actual channel sensing of wireless channel c,

$$OS = \{OS_t\}_{t=1}^{T} \tag{5.5}$$

is used to predict observation sequence for the next time slots on the same channel. Generated Observation Sequence(GOS) using HMM prediction is shown as,

$$GOS = \{OS_t\}_{t=T+1}^{2T} \tag{5.6}$$

The GOS is used for selecting preferred channel for communication.

Initial Parameter Set of HMM:

The HMM has N states, called $s_1, s_2, ..., s_N$ and discrete time-steps $t = 0, 1, ...$ At every time step the system is in one of the available N states. State at time t is denoted by q_t, with $q_t \in \{s_1, s_2, ..., s_N\}$. At every time-step, state q_t is producing, one of the available output symbol as per Emission/Observation Probabilities distribution as shown in following Figure 5.4.

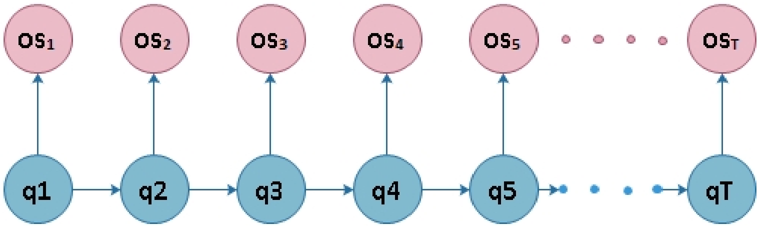


Figure 5.4: State Representation in HMM

The initial parameter set is defined using $\lambda = (A, B, \Pi)$ has implicit specification of N and M as follows,

- ***The Number of States,*** N: The set of states is represented as $S = \{s_1, s_2, ..., s_N\}$ and with $N = 2$ set consist of $\{s_1, s_2\}$.

- ***The Number of Distinct Observation Symbols,*** M: The observation symbols are represented as $V = \{v_1, v_2, ..., v_M\} = \{0, 1\}$ where $M = 2$.

- ***State Transition Probability,*** A: Represents probability of reaching a particular state from any other state with single step. State transition probability matrix,

$$A = [a_{ij}] = P\left[q_{t+1} = s_j | q_t = s_i\right], \quad 1 \leq i, j \leq N \tag{5.7}$$

As the number of states are two that is $N = 2$

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & . & . & a_{1N} \\ a_{21} & a_{22} & . & . & a_{2N} \\ . & . & . & . & . \\ a_{N1} & a_{N2} & . & . & a_{NN} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$a_{ij} \succ 0$ for any state i reached from j in single step. Other types of transitions have a_{ij} is zero for all remaining transitions. Figure 5.5 shows state transitions with state transition probabilities,

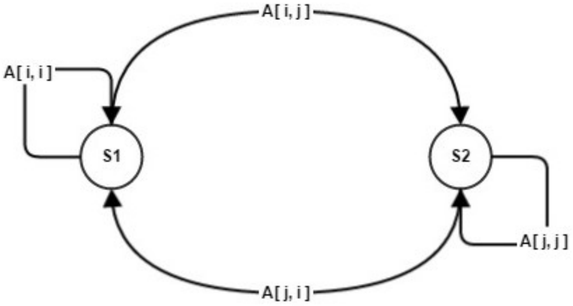


Figure 5.5: State Transitions in HMM

- ***Observation Probabilities***: The observation probability distribution for state j for observing v_m, $m = 1, ..., M$ in state s_j

$$B = [b_j(m)] = P\left[O_t = v_m | q_t = s_j\right], \quad 1 \leq j \leq N \quad and \quad 1 \leq m \leq M \tag{5.8}$$

As the number of symbols are two that is $M = 2$

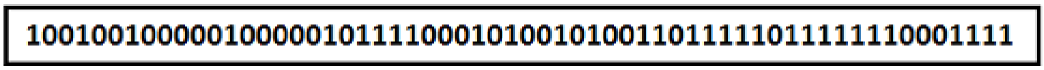
$$B = [b_{ij}] = \begin{bmatrix} b_1(1) & b_1(2) & . & b_1(M) \\ b_2(1) & b_2(2) & . & b_2(M) \\ . & . & . & . \\ b_N(N) & b_N(2) & . & b_N(M) \end{bmatrix} = \begin{bmatrix} b_1(1) & b_1(2) \\ b_2(1) & b_2(2) \end{bmatrix}$$

- ***Initial state distribution,*** Π:

$$\Pi = [\Pi_i] \quad where \quad \Pi_i \equiv P(q_1 = S_i) \equiv [P(q_1 = S_1), P(q_1 = S_2)] \tag{5.9}$$

Initial Parameter Set of HMM (*Example*):

Observation Sequence of 60 symbols, $\{OS_t\}_{t=1}^{T}$, where $T = 60$, is taken for defining initial parameter set $\lambda = (A, B, \Pi)$.

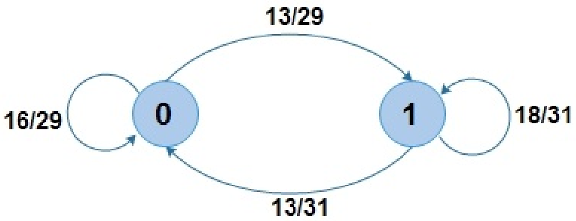


100100100000100000101111000101001010011011111011111110001111

No. of symbol '0': 29 **No. of symbol '1': 31**

Transitions in above Observation Sequence with respect to two symbols 1 and 0 are:

1. $0 - 0 = 16/29$
2. $0 - 1 = 13/29$
3. $1 - 0 = 13/31$
4. $1 - 1 = 18/31$



$$A = \begin{bmatrix} 16/29 & 13/29 \\ 13/31 & 18/31 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}, \qquad \Pi = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$$

Maximum Likelihood Estimation-Learning Model:

The objective of channel availability estimation is learning an HMM from observed data. Maximum likelihood estimation approach is used to calculate λ^* that maximizes the likelihood of the sample training sequences, $OS = \{OS_t\}_{t=1}^{T}$ with improvement in $P\left(OS|\lambda\right)$.

Maximum Likelihood Estimation: *Independent and identically distributed samples of, $OS = \{OS_t\}_{t=1}^{T}$, are drawn from probability distribution $P\left(OS|\lambda\right)$. The objective is to find λ that makes OS_t to $P\left(OS|\lambda\right)$ as likely as possible.*

$$OS_t \approx P\left(OS|\lambda\right)$$

The expectation modification procedure with given $\lambda = (A, B, \Pi)$, recalculate λ at each step using likelihood probability of occurrence of each state s_j after state s_i. This alteration step is repeated until convergence during which $P\left(OS|\lambda\right)$ never decreases. Flow diagram of HMM is shown in Figure 5.6.

Training of HMM and data processing by various components for channel selection are shown in Figure 5.7, 5.8 and 5.9. Algorithm for traffic prediction using HMM is as follows,

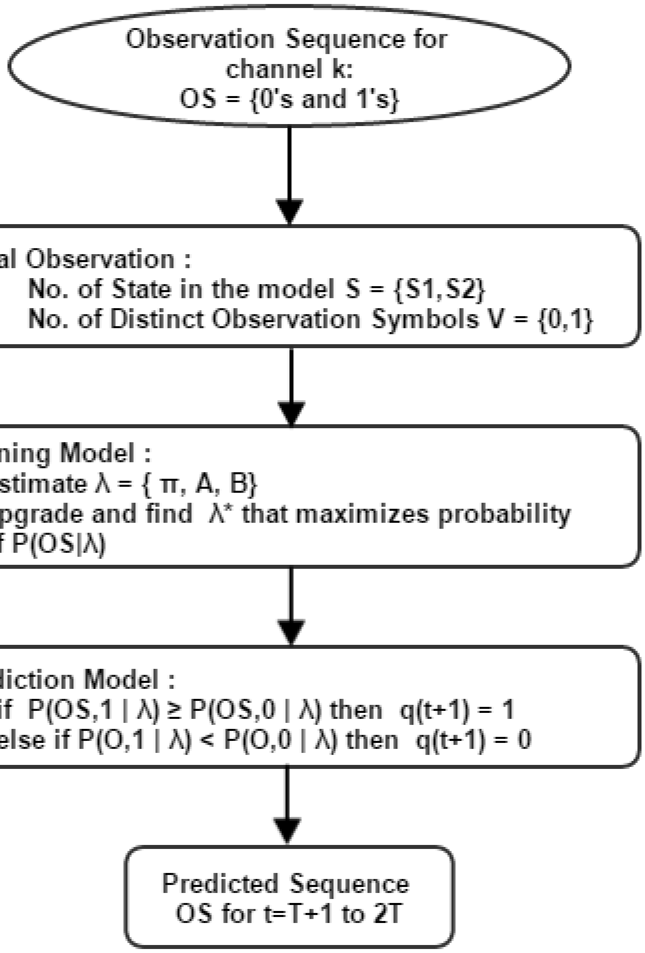

Observation Sequence for
channel k:
OS = {0's and 1's}
Initial Observation :
No. of State in the model S = {S1,S2}
No. of Distinct Observation Symbols V = {0,1}
Training Model :
Estimate λ = { π, A, B}
Upgrade and find λ* that maximizes probability
of P(OS|λ)
Prediction Model :
if P(OS,1 | λ) ≥ P(OS,0 | λ) then q(t+1) = 1
else if P(O,1 | λ) < P(O,0 | λ) then q(t+1) = 0
Predicted Sequence
OS for t=T+1 to 2T

Figure 5.6: Flow Diagram of HMM

Traffic Prediction Algorithm

```
Given a training set of observation sequence for $k$ channels,
Estimate parameter set $\lambda = (A, B, \Pi)$ and learn the model that maximizes the
probability of generating $OS$ for every channel, that is to find $\lambda^*$ that maximizes
$P(OS|\lambda)$.
1: Input
      -The Observation Sequence $X = \{OS_1, OS_2, ....OS_k\}$ for $k$ channels
      -$OS$ is series of 0's or 1's where, $OS_t \in [1,0], \forall t = 1, ..., T$
2: Initialize
      -Decide the distinct observation symbols $v = \{0,1\}$
      -Decide the number of states in the model $S = \{s_1, s_2\}$ where $s_1$ is idle
      -and $s_2$ is busy
3: Estimate $\lambda = (A, B, \Pi)$
      -A: Probability of State Transition between idle and busy
      -where $a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$
      -B: Observation Probability between 0's and 1's at time t
      -$\Pi$: Probability of $S_i$ to be the first state in the observation sequence
4: Training Model
      -Predict $q_{T+1}$ based on past $OS_k = \{q_1, q_2, ....q_T\}$
      -Update $\lambda$ to maximize the likelihood probability of generating $OS$ that is $P(OS|\lambda)$
      -Probability of $P(OS,1|\lambda)$ and $P(OS,0|\lambda)$ is calculated to decide state at $q_{T+1}$
      -if $P(OS,1|\lambda) \geq P(OS,0|\lambda)$ then $q_{T+1} = 1$ else $q_{T+1} = 0$
      -Repeat step 3 and 4 for maximizing the probability of generating $X$ for every
       channel that is to find $\lambda^*$ that maximizes $P(OS|\lambda)$
5: Prediction
      -Generate the observation sequence for the same channel
      $GOS = \{OS\}_{t=T+1}^{2T}$ using $\{OS\}_{t=1}^{T}$
```

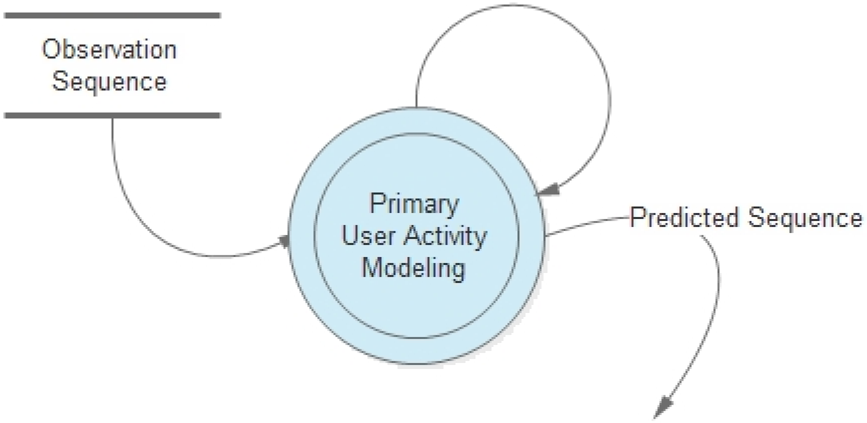


Figure 5.7: Data Flow Diagram of HMM Level 0

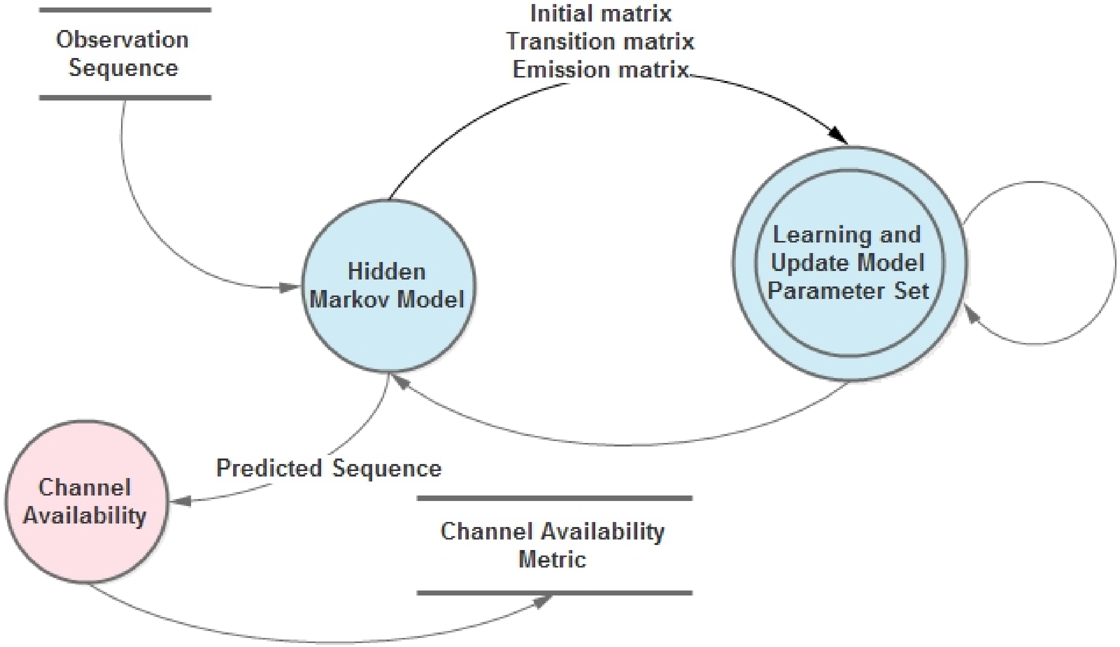


Figure 5.8: Data Flow Diagram of HMM Level 1

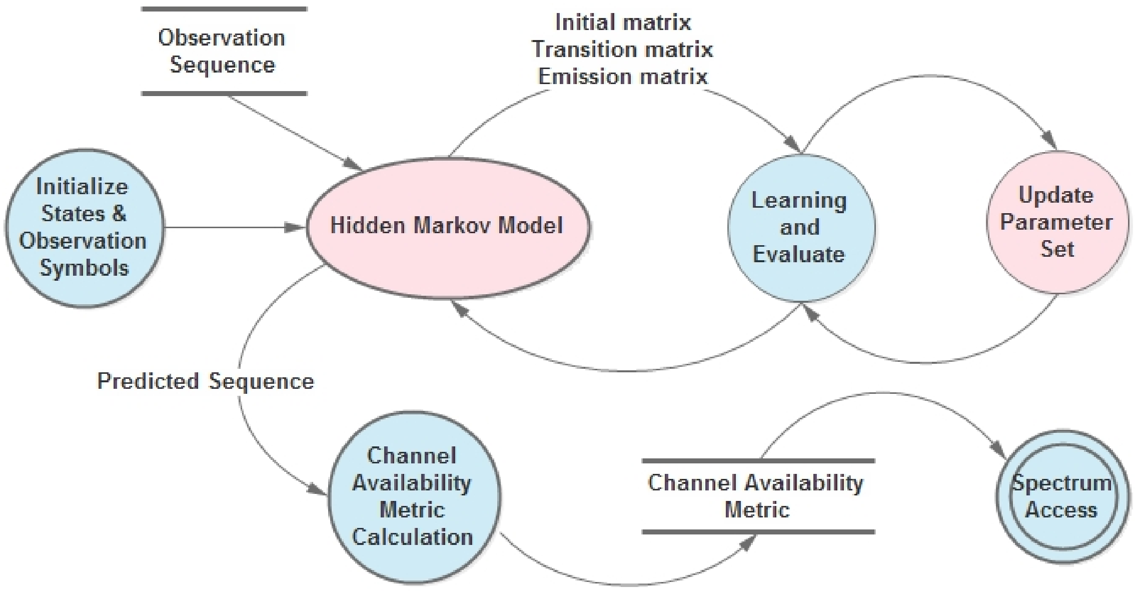


Figure 5.9: Data Flow Diagram of HMM Level 2

5.5 Link Selection in MCRAN

The route setup in MARL based routing requires selecting reliable link based on local environmental observation and other link characteristics. Prerequisite of link selection between two neighbouring nodes is that both the nodes must have a non-empty intersection of their available channel sets.

Specific metrics under consideration for MCRAN during link selection stage are: Channel Availability, Spectrum Propagation, PU Protection, Expected Time to Transmit and Channel Switching Cost. The link selection stage selects links with the lowest transmission cost denoted as LC_c. The lowest link cost represents a link with the highest availability for transmission, best propagation distance, lowest end-to-end delay and switching cost.

5.5.1 Channel Availability

Every node should compute the set of available channels using their Observation Sequence (OS) (Sharma M. et al., 2007). Every node in MCRAN periodically scans and senses multiple primary channels to detect primary signal and compare it with predefined threshold value. Every node makes such observation and records channel Observation Sequence $OS = \{OS_1, OS_2, ..., OS_T\}$ for every channel c, over the time period T. OS is series of 0's or 1's with, $OS_t \in [1, 0] \quad \forall t = 1, ..., T$.

The value of OS_t shows, the channel is available or not at time t. The more the number of 1's in the OS, the higher the interference indicated for the channel at the node. Channel Availability (CA) is obtained by computing the amount of time, the channel is available for a particular time period. CA is calculated using following equation,

$$CA_c = A_c^0 + \frac{|GOS_{H_c}|}{GOS_{H_c}^1} \tag{5.10}$$

where,

H_c : Trained HMM for channel c

GOS_{H_c} : Binary sequence generated by H_c

$|GOS_{H_c}|$: Length of sequence GOS_{H_c}

$GOS_{H_c}^1$: Generated observation sequence containing number of 1's, $GOS_{H_c}^1$

A_c^0 : Average number of 0's between two 1's in the Generated Observation Sequence A_c^0

Availability of channel c is higher with more number of 0's and lower for less number of 0's. Higher value of CA requires,

- Few slots with primary signal strength exceeding the threshold.
- A larger gap between two symbols of 1's in the generated observation sequence.

Channel Availability: An Example

An HMM is trained to characterize each channel and H_c is Trained HMM for c^{th} channel.

$GOS_{H_c} = 0100101100011 =$ Observation Sequence generated by H_c for channel c

Number of 0's between two symbol of 1's, at position 2 and 5 = 2

5 and 7 = 1

7 and 8 = 0

8 and 12 = 3

12 and 13 = 0

$$A_c^0 = (2+1+0+3+0)/5 = 1.2, \qquad GOS_{H_c}^1 = 6, \qquad |GOS_{H_c}| = 13$$

$$CA_c = 1.2 + (13/6) = 3.37$$

The node selects the channel with highest value of channel availability (CA), as the most preferable channel for communication. Every cognitive node calculates CA for every channel c and ranks channels using its value. The information about the set of available channels is shared with neighbouring nodes to agree upon common channels for communication.

The most available channel between two nodes is selected as the Control Channel (CC), for sending Control Packet (CP) and the remaining channels are used as Data Channel (DC), for actual data transmission. $C_{ij} = C_{ij}^{CC} + C_{ij}^{DC}$ Where, C_{ij} is representing a set of available channels between two nodes i and j that is an intersection of available channels between node i and j. C_{ij}^{CC} is Common Control Channel and C_{ij}^{DC} are Data Channels between any two nodes i and j.

5.5.2 Mobility Model

In order to design effective routing protocol, it is necessary to develop and use mobility model that accurately represents the movement of CN (Abedi and Berangi, 2013). If CN is moving towards the primary user region, it affects the spectrum and relay availability. Mobility model can be used to predict the movement of the CN towards primary region. The random movement of the CN can be modeled using random mobility models such as Random Waypoint, Constant Velocity Random Direction and Random Gauss-Markov Mobility Models (Ariyakhajorn et al., 2006).

Proposed algorithm requires appropriate decent mobility model which will attempt to mimic the movements of the real mobile node by allowing past velocities and directions to influence future velocities and directions. Random Gauss-Markov Mobility Model is used to describe the object trajectory as it can capture the correlation of object velocity in time (Liu B. et al., 2011).

In Gauss-Markov mobility model the nodes move randomly and their speed and direction changes are tracked using their previous values. For using Gauss-Markov mobility model, each cognitive node is assigned with both initial speed and direction as well as average speed and direction. After the fixed interval of time, a new speed and direction will be calculated for each node which follows the next course until the next time step.

Mobility prediction

A link between any two nodes is available if nodes are within the transmission range of each other. If the cognitive node is moving towards the primary user region, the probability of link availability decreases significantly due to increased interference in the primary network, shown in Figure 5.10 (Guan Q. et al., 2010). Suppose D_{la} is predicted time for which a link will be available and $P(D_{la})$ represents the probability that this link's availability may last till the end of duration D_{la}. The value, D_{la} is influenced by possible changes in velocity and direction of every CN.

Only predicting link duration $[D_{la}, P(D_{la})]$ is not sufficient for MCRAN, since the links between two CN is affected not only by its mobility but also by activities of PU. In this situation, the changing distance between PU and CN should be monitored for the link quality prediction. Therefore another pair of $[\widehat{D_{la}}, P(\widehat{D_{la}})]$ is proposed to predict link availability before node moves into interference region of PU. Similar to D_{la}, $\widehat{D_{la}}$ is the predicted time until a CN is out of the interference region of a PU, with its probability $P(\widehat{D_{la}})$. The final link selection is revealed by the combination

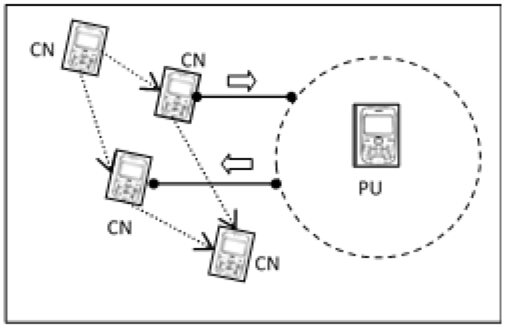


Figure 5.10: Distance and Mobility Prediction

of $[D_{la}, P(D_{la})]$ and $[\widehat{D_{la}},\, P(\widehat{D_{la}})]$ to enable the cognitive link prediction.

The node moving away from PU's region should have more preference as the probability of link availability is more. Thus, link selection for relaying information is influenced by mobility pattern, link characteristics and its availability.

Remark 1 It is assumed that each cognitive user knows its location and PU region using GPS (Global Position Systems). In most geographical routing protocols, a source node knows the location of its corresponding destination node.

5.5.3 PU Protection

A link is considered available if the pair of nodes associated with it are within the transmission range of each other and out of the interference region of any primary user in MCRAN. The former is common condition in wireless ad hoc network while the latter is the specific requirement in MCRAN. Consequently, link availability should be determined by $[D_{la}, P(D_{la})]$ and $[\widehat{D_{la}},\, P(\widehat{D_{la}})]$ together. The link availability duration D_c is calculated as,

$$D_c = min\left\{[D_{la} \times P(D_{la})], [\widehat{D_{la}} \times P(\widehat{D_{la}})]\right\}$$

Minimum of, products of spectrum availability for duration D_{la} , its associated probability, $[D_{la} \times P(D_{la})]$ and duration of link availability before node move into the PU region, its associated probability, $[\widehat{D_{la}} \times P(\widehat{D_{la}})]$. Thus, the PU protection is achieved by selecting only those links which are away from the interference range of

PU.

5.5.4 Spectrum Propagation

End-to-end latency of routing can be improved by selecting a channel with good propagation characteristics. The frequencies in the lower range have better propagation characteristics as they travel farther with lower attenuation. This can be used to decrease the number of hops to reach the destination. To improve latency in MCRAN, links with better propagation characteristics are selected. Propagation Distance (PD)is calculated by (Chowdhury and Akyildiz , 2011)

$$PD_c = \left[\left(\frac{1}{4\Pi f_c} \right)^2 . \frac{P_t}{P_r} \right]^{1/a} \tag{5.11}$$

Where, l is the speed of light, f_c is the representative frequency, a is the attenuation constant, P_t and P_r are transmitting power and required receiver power. Selection of proper frequency and transmission power for link establishment decreases hop count to reach the destination, which results in improving the end-to-end latency.

5.5.5 Expected Time to Transmit

Expected time to transmit (ETT) is the amount of time required to transmit message/packet from the first bit until to the last bit of message. For any wireless channel packet transmission time is obtained from packet size and bit rate in bit per second (Kyasanur and Vaidya , 2006).

$$ETT = PacketSize/BitRate \tag{5.12}$$

Different links may have different data rate when 'autorate' feature is enabled in the wireless interface device. Wireless communication with IEEE 802.11 standard offers different data rates for different links. The link data rate can be measured using probe packets, or can be read by querying the driver available with all new hardware.

5.5.6 Link Selection Cost

Characteristics of the selected link are,

1. Selected channel should take minimum possible amount of time to transmit a link layer frame, represented by Expected Time to Transmit (ETT).

2. Link availability duration is the probability that it will remain available for the particular transmission time requirement.

3. Channel availability is the average amount of time the channel is available for the transmission.

4. Propagation distance is the measure of travel capacity of electromagnetic radiation with lower attenuation.

With all the above parameters, formulation of cost metrics for channel selection is as follows,

$$LC_c = ETT_c + D_c + 1/CA_c + 1/PD_c \tag{5.13}$$

Where, LC_c is the link cost of using channel c
ETT_c is Expected Transmission Time for c
D_c is link availability duration for c
CA_c is the channel availability for c
PD_c is the propagation distance if using channel c.

Selection of the lowest link cost at every hop on entire routing path from source to destination increases the per packet reward. This causes increase in average per packet reward for P number of successfully transmitted packets.

5.6 Online and Opportunistic Routing

The performance of online and opportunistic routing in uncertain and dynamic wireless environment is improved using RL. RL is used to learn optimal policy on-line from direct interactions with the environment for solving non-linear control problems (Wei-bing and Xian-jia, 2009). Using RL, the agent learns how to achieve the given goal by trial and error interaction with the environment. It improves performance of the agent from the received reward or punishment as a result of the performed action(Nie and Haykin, 1999; EI-Sayed EI-Alfy et al., 2006; Duana Yet al., 2007; Usaha and Barria, 2007).

In MCRAN, every CN is considered as an intelligent agent and mobile wireless network as an environment. The agent tries to improve spectrum utilization and network performance. Temporal difference (TD) implementation of RL, learns directly from raw experiences without model of the environment dynamics (Kaebling L. P. et al., 1996). It uses generalized policy iteration to update estimates, based on the parts of other learned estimates, without waiting for the final outcome. TD method waits until the next time step, at $t+1$ it immediately forms a target and makes useful update using observed reward r_{t+1} and estimates $V(s_{t+1})$,

$$V(s_t) = \lfloor V(s_t) + \alpha \{r_{t+1} + \gamma V(s_{t+1}) - V(s_t)\} \rfloor$$

Where α , is a step-size parameter and the setting for α may be fixed or diminishing over time.

5.6.1 Markov Decision Process for Routing in MCRAN

The task that satisfies the Markov Property, that is all decisions and values are functions of the current state only, is called Markov Decision Process (MDP) (Sharma and Gopal, 2010). MDP is defined by its state and action sets and by the one-step dynamics of the environment. Given the current state and action, one-step dynamics enable us to predict the next state, expected next reward and iteratively all future states and rewards.

MDP is represented using the tuple $\langle S, A, f, \rho \rangle$, where S is the set of possible states of environment; A is the set of available actions; f is the state transition probability function and ρ is the reinforcement function (Sutton and Barto, 1998). In the context of MCRAN, MDP tuple can be represented as follows:

- State Space/Set S: S is a possible set of states of the environment. In MCRAN, states are set of neighbours of every cognitive node s represented as N_s. One of the neighbouring node from set N_s is selected as the next candidate relay node.

- Action Space/Set A: A is a set of agent actions or behavior of the agent at a specific time, allowing it to change from one state to another state. In MCRAN, A is the set of all allowable actions available to every node s that is $A(N_s) = N_s \cup T$. The possible action for any node includes,

  1. Select one of the neighbouring nodes as a candidate relay node for transmission.
  2. Terminate the packet.

- State transition probability $f : S \times A \times S \rightarrow [0, 1]$ is the state transition probability function. As a result, of the action $a_t \in A$ the environment changes its state from S_t to $S_{t+1} \in N_s$ according to state transition probability given by f. The probability of selecting particular node as a forwarding node based on the routing score of the neighbouring nodes.

- Reinforcement or reward function $\rho : S \times A \times S \rightarrow R$ is the reinforcement function. Reward R can be used to evaluate immediate effect of action at the transition from S_t, to S_{t+1}. For CRN, the reward function is defined on the state space N_s and potential decision $a \in A(N_s)$ as,

$$\rho\left(N_s, a\right) = \begin{cases} -LC_c & if\ a \in N_s \\ R & if\ a \equiv T\ at\ destination\ d \\ 0 & if\ a \equiv T\ at\ intermediate\ node \end{cases} \tag{5.14}$$

After receiving the packet at destination the cumulative link cost, LC_c is subtracted from the final reward received at the destination. Selection of the lowest link cost at every hop ultimately increases reward at the destination. Discrimination among the termination events T as follows,

- Positive reward R is obtained if the packet is terminated at the destination after successful delivery.
- The termination of the packet at intermediate node causes zero rewards.

5.6.2 Overview of MARL Based Routing

Routing is a challenging problem in MCRAN due to nodes mobility, intermittent spectrum availability and incomplete knowledge of the environment. Routing is done online and in an opportunistic manner by selecting the most prominent relay node as forwarding node according to its routing score, location and spectrum availability. Proposed scheme makes the decision in distributed manner by following three-way handshake between node s and its neighbours $N(s)$.

- Node s, sends control packet of relay request on common CC to all the neighbouring nodes.
- The subset of neighbouring nodes interested in becoming a relay for the source s send relay reply.
- Node s announces the candidate forwarding node based on routing score and then send data packets.

The Link selection process selects the link with the lowest transmission cost represented by LC_c. The lowest link cost represents a link with more availability, best propagation distance, lowest end-to-end delay and switching cost. In the relay selection stage, source node s broadcasts the relay request over the control channel for destination d, and this request is received by the set of neighbouring relay node $N(s)$ within source s vicinity. neighbouring nodes willing to participate in the routing process respond to s represented by the set $N_s \subseteq N(s)$.

As shown in Figure 5.11, one of the willing neighbours is selected as the next candidate forwarding node using the routing score and node's own experience of neighbour as relay. After selection of the relay, s sends data packet on the most available data channel to the next relay. This process is iteratively done until packet is not received at the destination. After successful reception of packet at the destination, the destination is responsible for sending packet acknowledgment to the source by tracing the path backward to the source. It is assumed that fixed positive reward R is obtained after successful delivery of packet at the destination.

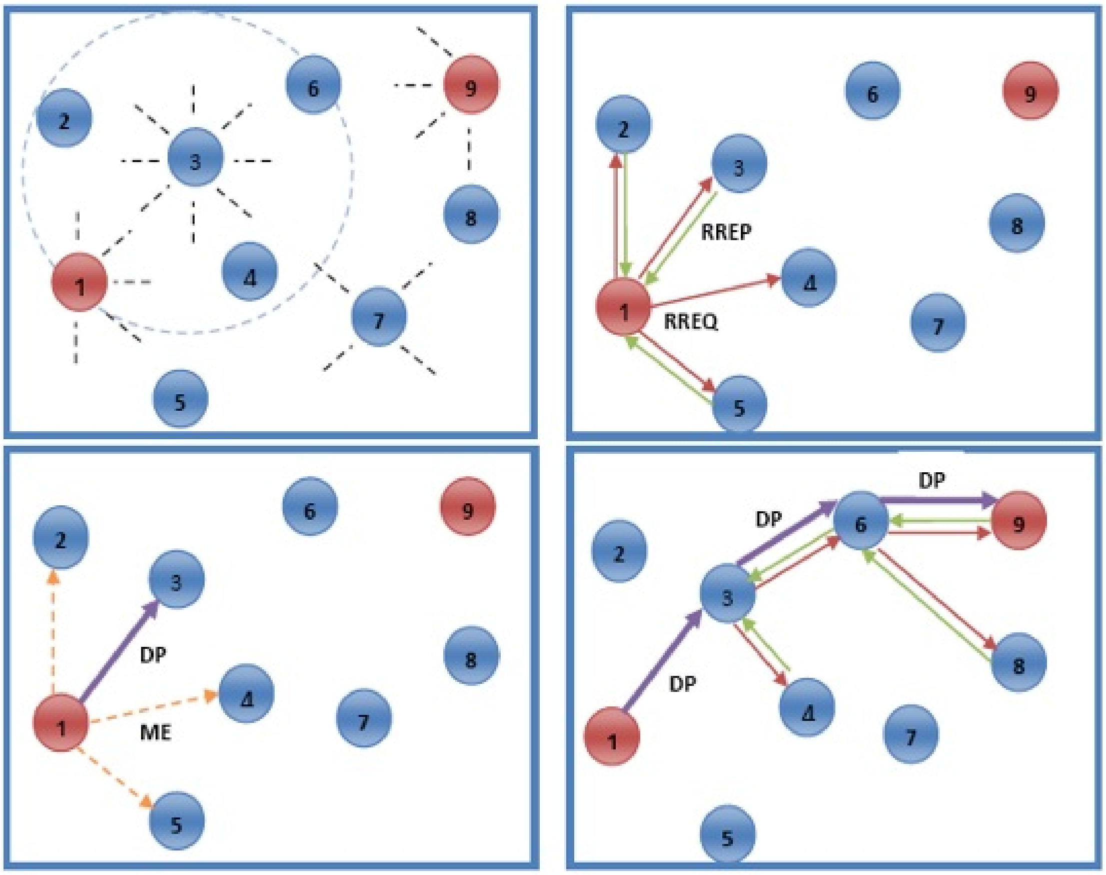


Figure 5.11: Route Formation from Source '1' to Destination '9'

The routing score of any agent is the representation of how many times that agent was interested in route formation and how many times it has been selected and has successfully completed transmission. The routing score is calculated with the help of following three variables,

Willing neighbour $WN\left(s, Ns\right)$ **:** Relay request is replied by the set of neighbour $N_s \subseteq N\left(s\right)$. Every time an agent is willing to participate in the routing process, WN is incremented for every agent in N_s.

Candidate Best Relay (CBR_t^n) : The number of times the node n is selected as a candidate relay node. (CBR_t^n) is incremented after selecting one of the neighbouring nodes as relay for forwarding packets towards destination.

Routing Score Vector $V_t^n\,(s,a)$: The number of times routing decision $a \in A(s)$, for neighbour n has been made till time t for the source s.

After completing selection of the relay and transmission, every node updates V with a reward received computed as follows,

$$V_t^n\,(s,a) \leftarrow V_t^n\,(s,a) + \alpha_{CBR_t^n}\left[r_{t+1} + \gamma V_{t+1}^n\,(s,a) - V_t^n\,(s,a)\right] \qquad (5.15)$$

Where, V represent the routing score for the agent s for selecting the node n as a relay using action a. α, is step-size parameter and the setting for α is influenced by CBR_t^n. r_{t+1} is a reward received by the node s for selecting the node n as relay. γ is representing discount rate $0 \leq \gamma \leq 1$, determines the present value of the future reward. The routing score is updated on every successful relaying.

The problem of routing is a continuous task for P number of packets. Learning from observed reward immediately improves the routing performance with every action selection. Agent conveniently forms a target with $V_{t+1}^n\,(s,a)$ and make a useful update using the observed reward computed using Eq. 5.15.

5.6.3 MARL Based Routing Algorithm

All node share CA information with the help of beacon messages. The control channel is used to send relay request and relay reply and sending the final acknowledgment to the source from the destination. The remaining steps involve the selection of a relay node among the willing nodes. Proposed multi-agent reinforcement learning based routing algorithm is given as follows,

MARL Based Routing Algorithm

Given the set of CNs and channels, find subset of intermediate nodes from source to destination forming path for forwarding packets.
MDP tuple: $\langle S, \mathcal{A}, f, \rho \rangle \Longrightarrow \langle N_s, \mathcal{A}(N_s), f(N_s, \mathcal{A}), \rho(N_s, \mathcal{A}) \rangle$

1. ***Initialization***

   -*CN agent shares available channels using beacon message within their vicinity*

   -*Agree upon Control Channel (CC) and Data Channel (DC)*

2. ***Relay Request***

   -*Source s sends a control packet (CP) containing relay request RREQ on CC for destination d*

   -*Contains destination ID*

3. ***Relay Reply***

   -*Let* $N(s)$ *denote random set of nodes that has received CP*

   -*RREP is sent from node* $N_s \subseteq N(s)$ *includes Routing Score (RS)*

4. ***Selection and Transmission***

   -*Upon reception of RREP from* N_s*, variable Willing neighbour* WN_n *is incremented for node* $n \in N_s$

   -*Node s select routing action* $a \in A(s)$ *using Softmax Action Selection Rule*

   -*Sends data packet on most available common data channel between s and n at the time t*

5. ***Reception and Acknowledgment***

   -*Upon successful reception of DP,an acknowledgment is sent back to the sending node*

   -*Candidate Best Relay,* CBR_t^n *is updated which is number of time node n is selected till time t*

6. ***Computation Stage***

   -*Upon completion of transmission, node s updates score vector V*

   -*The routing decision of node s at time t is based on Routing Score*

   -*Message received from neighbours is used to update the Routing Score Vector*

Initialization

The MARL based routing starts with deciding common channel or links between every pair of node. Suppose C_i indicates available channels for node i and C_j indicate channels available for j. Common channels among the nodes i and j are represented using $C_{ij} = C_i \cap C_j$ with $|C_{ij}| = k$ channels. These common channels are ranked according to their link cost computed using following equation,

$$LC_c = ETT_c + D_c + 1/CA_c + 1/PD_c$$

Where, LC_c is the link cost of using channel c, ETT_c is Expected Transmission Time, D_c is link availability duration, CA_c is the channel availability, PD_c is the propagation distance, if using channel c.

Relay Request

Different pair of nodes has different control channel decided as per their channel availability. Necessary actions performed by various nodes involved in the relay request process are,

- Using Control Packet (CP) Relay Request is sent to neighbouring nodes when any node needs to discover route to destination
- Source agent s send CP on CC for destination d, received by all nodes within its vicinity
- CP contains ⟨*Destination-ID, Source-ID, CP-ID, Hop-Count*⟩

- neighbour node $n \in N_s$ sets up a reverse route entry in its one hope neighbour table for source s

Relay Reply

Subset of nodes N_s from set $N(s)$ send route reply to source s as per their possibility to participate in the routing process.

- Let $N(s)$ denotes random set of node that has received CP
- REP from node $N_s \subseteq N(s)$ contains node's identifier and willingness to participate in route formation

Selection and Transmission

Variable Willing neighbour WN_n is incremented for every relay replay from neighbour. Node s selects routing action of selecting next forwarding node using Soft-max Action Selection Rule using adaptive Routing Score V described in equation no 5.15. Source s sends data packet DP on the most available data channel DC between s and n at time t.

- Variable **Willing neighbour** WN_n is incremented for node $n \in N_s$ if n want to work as relay
- Variable $WN_t(s, N_s)$:Number of times up to time t, N_s has received and replied CP from s
- Node s selects routing action $a(s, n) \in A(N_s^t)$ using **Softmax Action Selection Rule**
- Send data packet on most available common data channel between s and n at time t

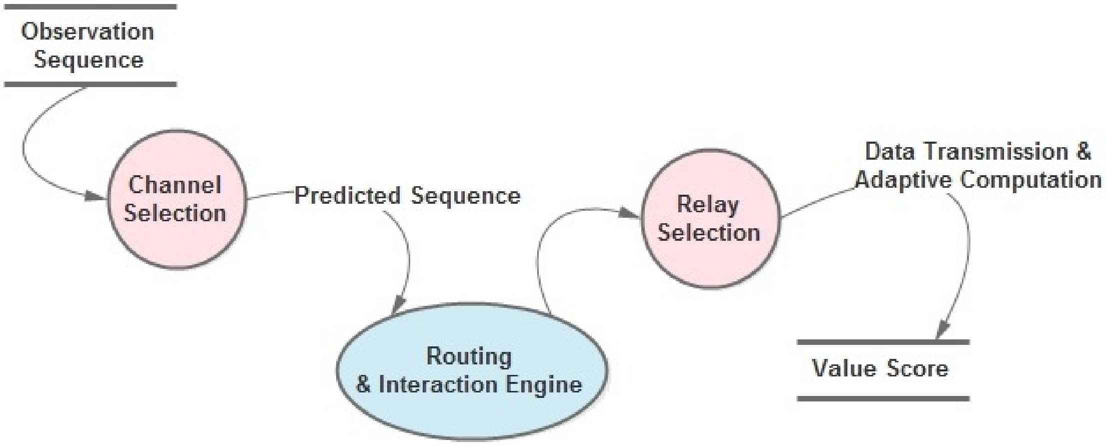


Figure 5.12: Data Flow Diagram for Routing Level 0

Reception and Acknowledgment

After selecting best candidate forwarding node, variable Candidate Best Relay (CBR_t^n), is updated which is number of time up to time t particular node or agent is selected as relay.

- Variable **Candidate Best Relay** is incremented

$$CBR_t(s,n,a) = \begin{cases} CBR_{t-1}^n(s,a)+1 & \text{if } (n,a) = (N_s^t, a_s^t) \\ CBR_{t-1}^n(s,a) & \text{if } (n,a) \neq (N_s^t, a_s^t) \end{cases}$$

- Upon successful reception of DP, an acknowledgment is sent back to the sending node
- A fixed and positive reward R is obtained upon successful delivery of a packet to the destination.
- Packet get $-LC_{c_i}$ as reward upon its Successful reception at intermediate node

End-to-End Routing Metric for p hopes,

$$RM(r) = R - \sum_{i=1}^{p} LC_{c_i} \tag{5.16}$$

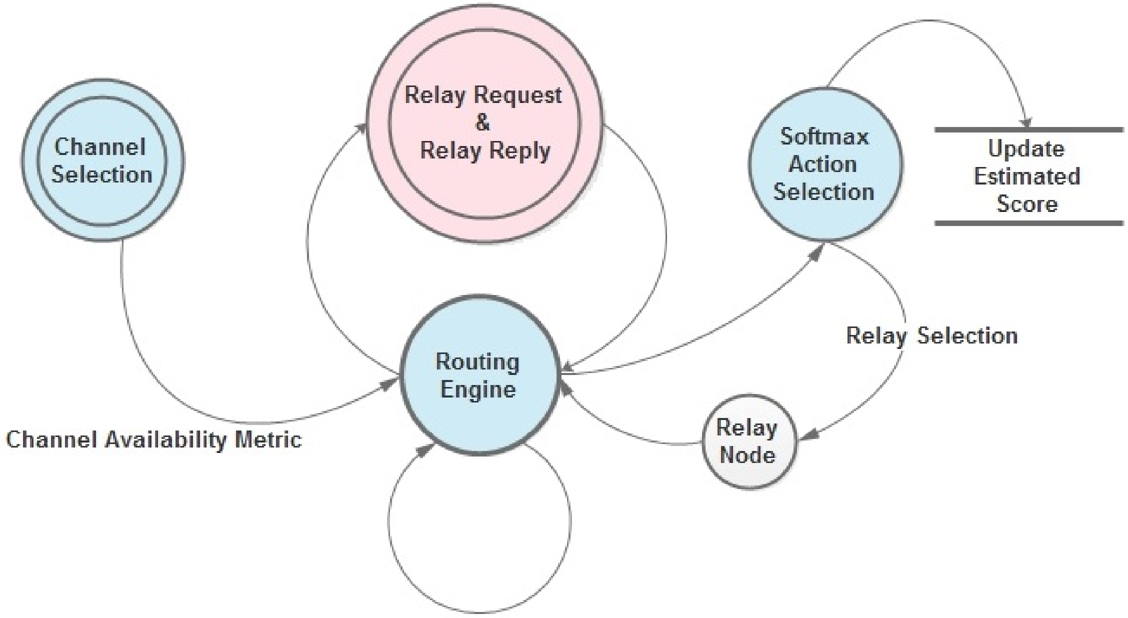


Figure 5.13: Data Flow Diagram for Routing Level 1

Adaptive Computation Stage

Upon completion of transmission, node s updates adaptive Routing Score. The routing decision of node s at time t is based on V Acknowledgment received from neighbours is used to update V **Score Vector**

$$V_t^n\left(s,a\right) \leftarrow V_t^n\left(s,a\right) + \alpha_{CBR_t^n}\left[r_{t+1} + \gamma V_{t+1}^n\left(s,a\right) - V_t^n\left(s,a\right)\right]$$

Routing Score is defined for every packet,

$$RS \leftarrow V_{t+1}(s,N,a) \tag{5.17}$$

Routing and interaction engine selects stable, reliable channels and best relay node for finding efficient routes, as shown in data flow diagram, Figure 5.12. The data flow diagram shown in, Figure 5.13, provides details of information processed by different modules in routing process. The detailed interdependence, data processing and exchanging among various components are shown in data flow diagram, Figure 5.14.

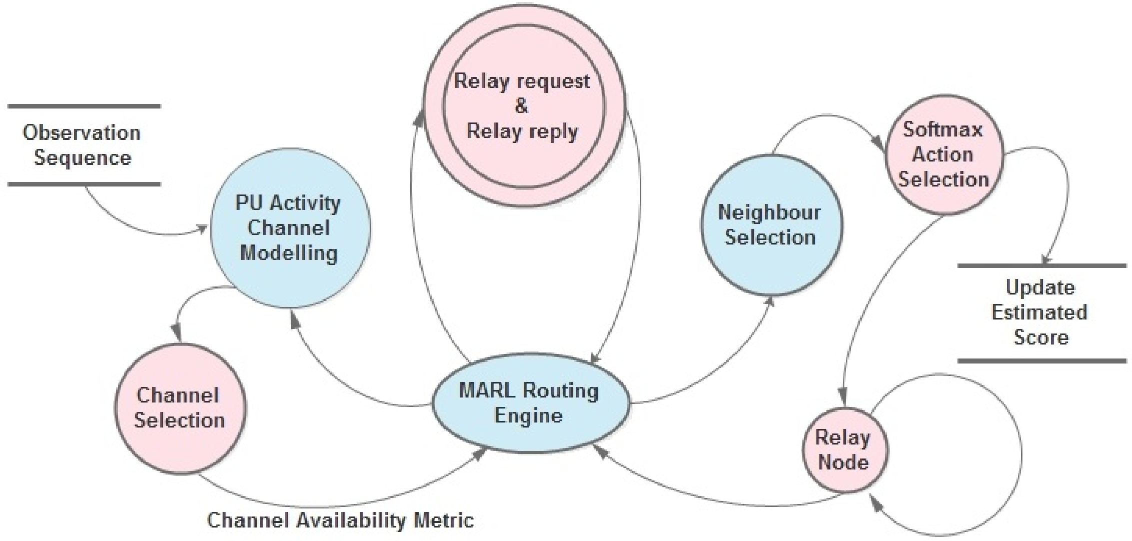


Figure 5.14: Data Flow Diagram for Routing Level 2

5.6.4 Action Selection

Selection of relay using routing score can increase the load on one of the nodes with better score (greedy selection) as compared to other neighbours. If the best node is selected, it is the exploitation of current values of actions. Instead of selecting the top ranked neighbouring node, the softmax action selection rule is used to select one of the non-greedy actions. This enables to explore, as it allows to improve the values of non-greedy actions. Softmax action selection rule is efficient in exploring and exploiting the network opportunities by taking a sequence of decisions which will increase the routing score of every willing neighbour. Suppose estimated value of action a at t^{th} play,

$$Q_t(a) = V_t^n(s, a) \tag{5.18}$$

The softmax action selection method is instinct in balancing exploration and exploitation in online dynamic routing. It can be represented as,

$$P = \frac{e^{Q_t(a)/T}}{\sum_{a=1}^{n} e^{Q_t(a)/T}} \tag{5.19}$$

Where, T is a positive parameter called temperature. High temperature causes equi-probable action selection. This enables exploration by improving the estimate of non-greedy actions. Low temperature causes the greater difference in selection probability for actions that differs in their value estimate. When $T \to 0$, softmax action selection will behave as greedy action selection.

- Softmax Action Selection Rule-Balanced routing action selection
- High Temperature: Action selection is equi-probable **(Exploration)**
- Low Temperature: Greedy action selection **(Exploitation)**

Table 5.1. shows example about how action selection probabilities vary with temperature T.

Table 5.1: Softmax Action Selection Probabilities

STATE	VALUE	$T \to 20$	$T \to 1$
S1	2	0.239	0.041
S2	3	0.252	0.112
S3	1	0.228	0.015
S4	5	0.278	0.831
≈ TOTAL		0.997	0.999

5.7 Concluding Remarks

In this chapter, algorithmic design of spectrum aware hop by hop routing from source to destination using multi-agent reinforcement learning is represented. The detailed algorithm for generating observation sequence of primary user activity to predict channel availability is represented using Hidden Markov Model. Different important characteristics of link with respect to cognitive radio network are elaborated with mobility model. The complete route formation process is explained using MARL based routing algorithm with detail description of every step.

Chapter 6

Implementation and Analysis of Channel Selection

Chapter 6

Implementation and Analysis of Channel Selection

The overall aim of the reinforcement learning based routing algorithm is to select the route from source to destination with maximum available links on each intermediate node. This multi-agent reinforcement learning based routing algorithm is designed and implemented successfully to explore the spectrum and routing opportunities in non-stationary, dynamic and mobile environment.

The implementation of multi-agent reinforcement learning based routing involves two stages i.e link selection stage and online opportunistic route formation stage. Selecting link on every hop with characteristics of good propagation distance, the minimum expected transmission time, more channel availability and duration, results in stable links with more forwarding distance towards destination. Primary user activities should be continuously sensed, monitored and recorded for analyzing channel usage characteristics. This helps to understand primary user's behavior, channel usage statistics and predicting their future channel availability. The following section describes the experimental details of sensing primary user signal for generating Observation Sequence of primary user activity.

6.1 Network Model for Channel Sensing

Five sensing nodes are kept in an area of 100 × 100 meters having minimum distance between any two nodes is 30 meters. Wi-Fi channels belonging to Wireless Local Area Network (WLAN) IEEE 802.11n are sensed with frequency range of 2.4 GHz.

IEEE 802.11n based wireless communication standard devices communicate using wireless distribution of access points in limited area. IEEE 802.11n is a set of specification of communication using Medium Access Control (MAC) and Physical Layer in 2.4 GHz frequency bands.

The frequency range of 2.4 GHz band is divided into 14 different channels separated by 5 MHz band. The overlapping bands cause interference to each other. Interference between any two channels is minimized by avoiding overlapping and making 3 to 4 channel under clear condition. Therefore, 2.4 GHz band with 802.11n offer only three communication channel in non-interfering manner with 22 MHz of theoretical bandwidth. These channels are channel no. 1, channel no. 6 and channel no. 11, shown in Figure 6.1.

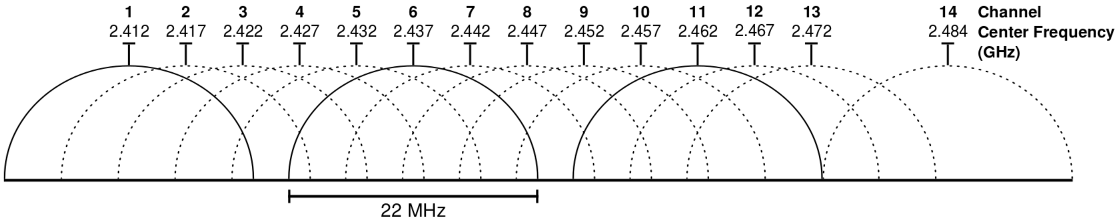


Figure 6.1: Wi-Fi Wireless Channels in 2.4 GHz Range

Channel sensing activity is performed on these three different channels, channel no. 1, 6 and 11 using special programmable radio hardware, Universal Software Radio Peripheral (USRP) and open source signal processing toolkit, called GNU Radio. Real-time spectrum sensing with current environmental conditions increases accuracy in channel availability prediction. Spectrum sensing and hardware implementation are discussed in the following section.

6.2 Implementation Details of Spectrum Sensing

The proper exploitation of the spectrum hole or white spaces in radio spectrum requires robust, fast and accurate methods for their detections. This work uses energy detection technique for adaptively detecting white spaces in radio spectrum (Sobron I. et al., 2015). Energy detection technique detects white spaces efficiently, without the information about primary user. The cognitive radio transceiver is having two important characteristics, 'Cognitive Capability' and 'Re-configurability'.

Cognitive Capability: Enables radio to capture information from surrounding radio environment. Identify the portion of the spectrum that is not used at specific time or location. This helps for selecting best spectrum and appropriate operating parameter for transmission.

Re-configurability: Enables radio to be dynamically programmed with new transmission parameter according to the sensed radio environment. It allows transmit and receive on a variety of frequencies and to use best among various transmission access technologies.

Software Defined Radio (SDR) provides capability to sense dynamic spectrum activities and change transmission parameters of the radio as well as modulation / demodulation scheme in software. SDR is the main building block of cognitive radio for sensing and reconfiguring to the best spectrum band. SDR is defined as a reconfigurable communication platform to perform different functions. The Energy Detection method is implemented using special type of Radio Frequencies (RF) hardware called 'Universal Software Radio Peripheral N210 (USRP N210)' and a software toolkit 'GNU radio'.

6.2.1 Universal Software Radio Peripheral

Universal Software Radio Peripheral, USRP N210 is the best testbed or platform for cognitive radio and software defined radio related experimentation developed by Matt Ettus (Matt Ettus , 2015). The Ettus Research USRP N210 enable to rapidly design and implement powerful, flexible software radio systems, shown in Figure 6.2. Applications requiring high RF performance and great bandwidth can be efficiently developed using USRP N210.

The USRP is reconfigurable hardware designed for RF applications from DC to 6 GHz. USRP interfaces with the computer through gigabit ethernet connection. The connection is used for data transfer between USRP and host processor. This enables personal computer to serve as high bandwidth software radio's. Collectively, GNU radio with its signal processing libraries, a personal computer and Universal Software Radio Peripheral provides an ideal solution for developing software defined radio platform with minimal and generic hardware. USRP consist of two important components motherboard and daughter-board.

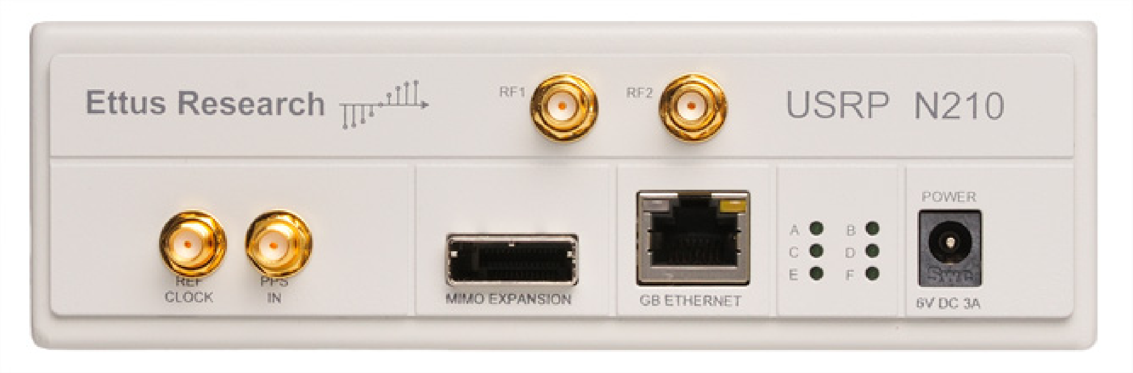


Figure 6.2: Universal Software Radio Peripherals N210

USRP Motherboard

The main components of the USRP motherboard include a 3A-Digital Signal Processing 3400 Field Programmable Gate Array, 100 MS/s (Mega Samples per Second), dual Analog to Digital Converter (ADC), 400 MS/s dual Digital to Analog Converter (DAC). Figure 6.3 shows the block diagram of USRP. USRP N210 reference clock can be disciplined with the worldwide GPS standard using an optional GPSDO module. Using Gigabit Ethernet connectivity, it can stream up to 50 MS/s to and from host applications. The FPGA firmware is reloaded through the Gigabit Ethernet interface.

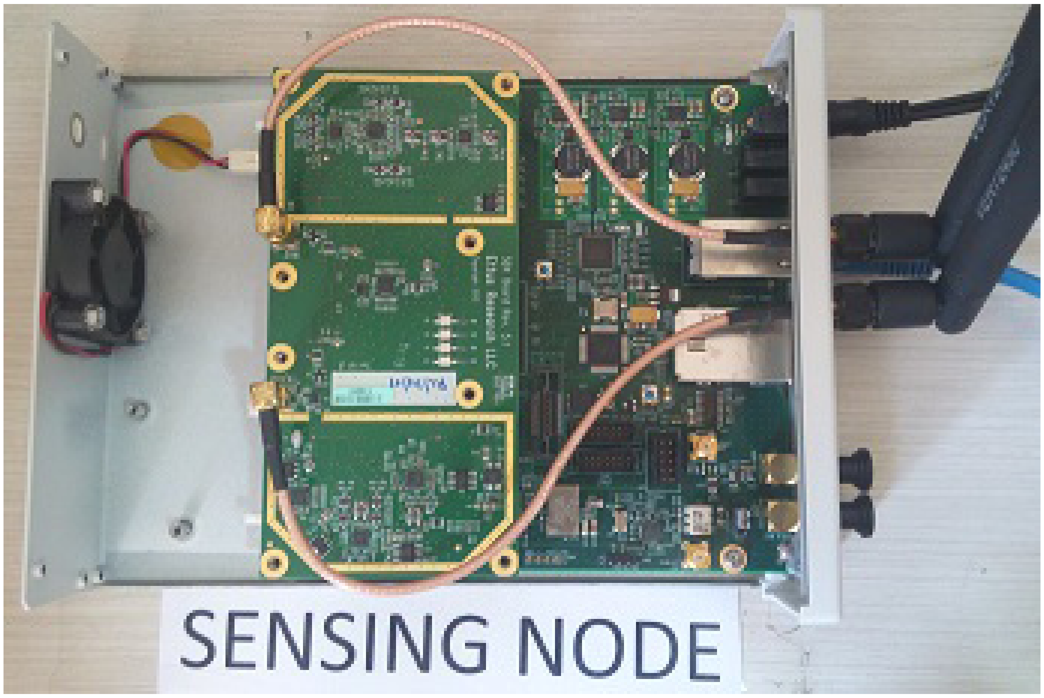


Figure 6.3: USRP Motherboard

USRP Daughter-board

The USRP is paired with daughter-board which is modular component of USRP serves as RF front end. As this focuses on sensing of unlicensed frequencies in 2.4 GHz range (Industrial, Scientific and Medical-ISM radio bands), SBX USRP Daughter-board is used with a wide range for 400 MHz to 4.4 GHz, shown in Figure 6.4. SBX USRP Daughter-board is,

- Wide band transceiver providing 100mW of output power
- The Noise figure of USRP N210 is 5 dB
- Dual band operation is supported by allowing the local oscillators to operate independently for the receive and transmit operation.
- Total bandwidth is 40 MHz.
- Able to access different bands in the 400 MHz-4400 MHz range.

- WiFi, WiMax transceivers and 2.4 GHz ISM band transceivers are few example applications.

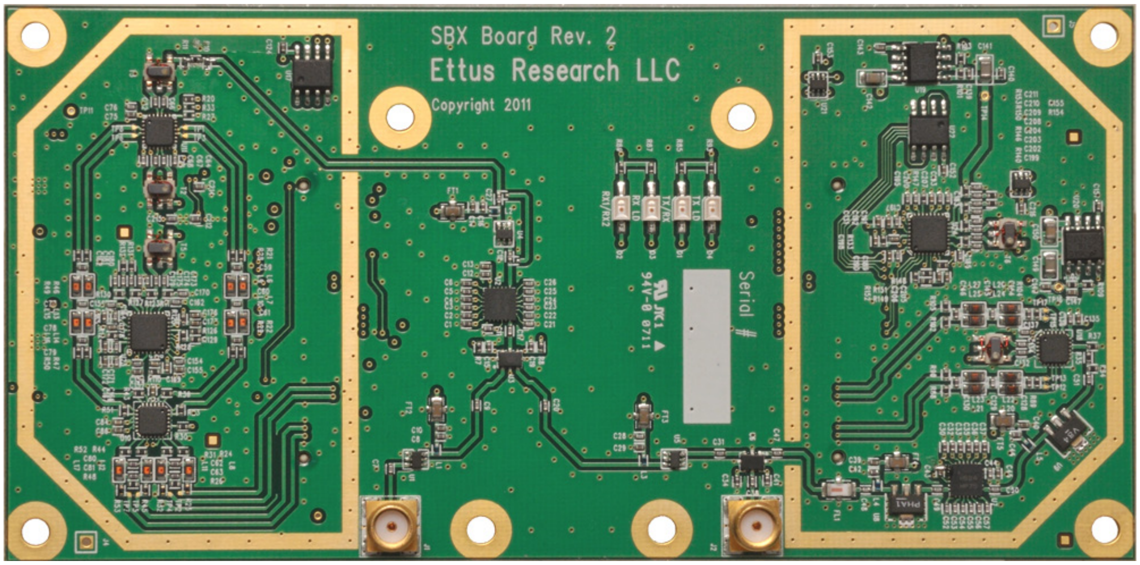


Figure 6.4: SBX- Daughter - board

6.2.2 GNU Radio

Transmission of data on wireless channels requires signal processing for converting it into the digital stream or receive data from the digital stream. Radio system uses dedicated integrated circuits in hardware for performing signal processing/conversion. Software defined radio performs this signal processing using signal processing software instead of hardware. Free and open-source software development toolkit, 'Gnu Radio ' is used to implement this signal processing blocks. GNU Radio can be interfaced with low-cost external RF hardware like USRP (GNU Radio), shown in Figure 6.5.

Many basic elements of radio systems are supported by GNU radio for any wireless application. Some of the elements are modulator, demodulator, encoders, decoders etc. It supports to connect and manage these blocks for transferring data from one block to another.

Python Programming language is used to write all applications of wireless communication and performance critical paths are implemented using C++ with processor floating point extension. This is simple to use and rapid application development environment and offers development of real-time, high throughput wireless application.

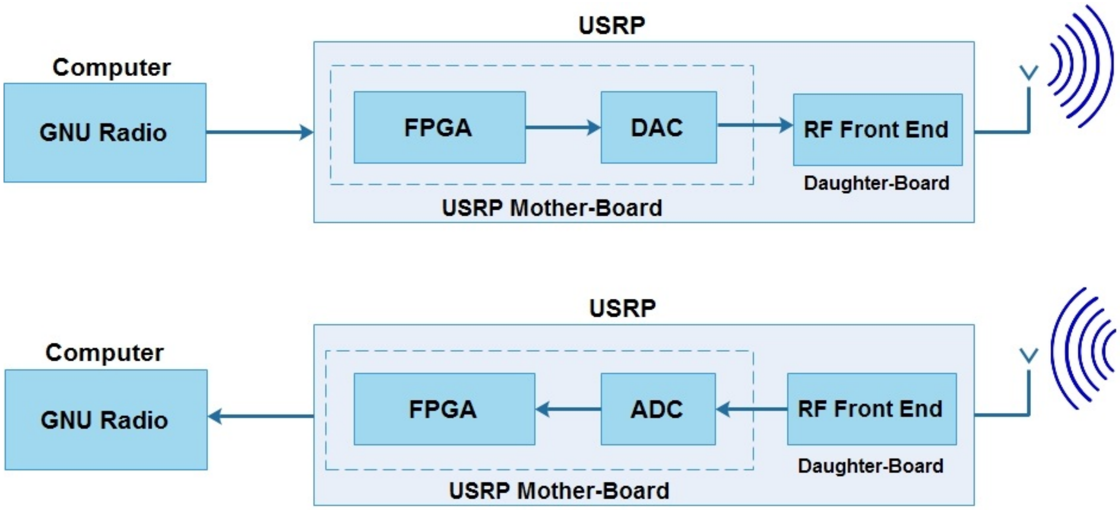


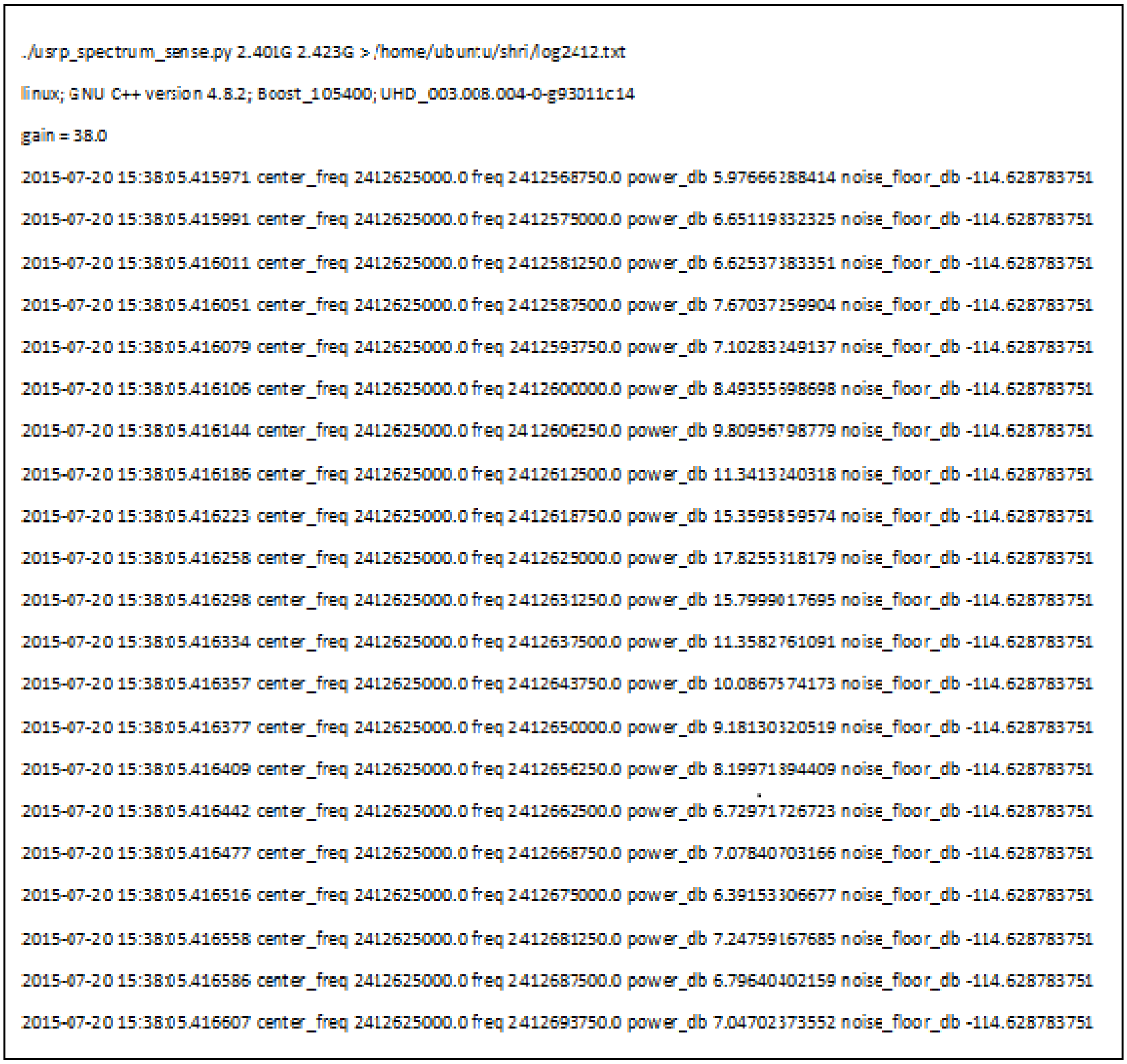
Figure 6.5: GNU Radio and USRP

6.2.3 Spectrum Sensing

The spectrum sensing is implemented as wideband spectrum analyzer. To scan three channels of 2.4 GHz RF spectrum, a spectral window that is bandwidth of 22 MHz is used. The USRP RF front end is tuned to particular channel one by one. USRP scans across the spectrum and makes RF measurement.

Large frequency is analyzed by sweeping over the required frequency range. USRP collects T time samples on center frequency f_s, represented by vector Q. USRP is tuned to different RF center frequencies for collecting samples of three wireless channels. The output of sensing of three Wi-Fi channels is given in Figure 6.6, 6.7 and 6.8.

Every sample in screen shot represents date, time in microseconds, center frequency, signal power recorded at sampling frequency and noise floor. The range of frequencies separated by 22 MHz is given as input to sensing procedure which provides p signal samples for particular center frequency. These samples are used to create Observation Sequence for that channel.



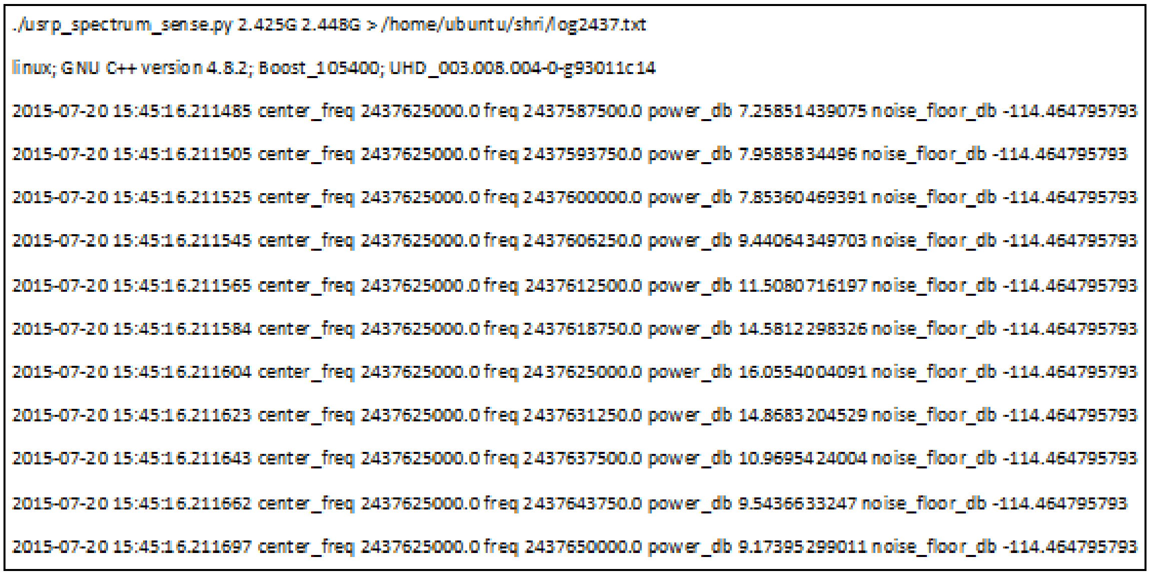
```
./usrp_spectrum_sense.py 2.401G 2.423G > /home/ubuntu/shri/log2412.txt

linux; GNU C++ version 4.8.2; Boost_105400; UHD_003.008.004-0-g93011c14

gain = 38.0

2015-07-20 15:38:05.415971 center_freq 2412625000.0 freq 2412568750.0 power_db 5.97666288414 noise_floor_db -114.628783751

2015-07-20 15:38:05.415991 center_freq 2412625000.0 freq 2412575000.0 power_db 6.65119332325 noise_floor_db -114.628783751

2015-07-20 15:38:05.416011 center_freq 2412625000.0 freq 2412581250.0 power_db 6.62537383351 noise_floor_db -114.628783751

2015-07-20 15:38:05.416051 center_freq 2412625000.0 freq 2412587500.0 power_db 7.67037259904 noise_floor_db -114.628783751

2015-07-20 15:38:05.416079 center_freq 2412625000.0 freq 2412593750.0 power_db 7.10283249137 noise_floor_db -114.628783751

2015-07-20 15:38:05.416106 center_freq 2412625000.0 freq 2412600000.0 power_db 8.49355698698 noise_floor_db -114.628783751

2015-07-20 15:38:05.416144 center_freq 2412625000.0 freq 2412606250.0 power_db 9.80956798779 noise_floor_db -114.628783751

2015-07-20 15:38:05.416186 center_freq 2412625000.0 freq 2412612500.0 power_db 11.3413240318 noise_floor_db -114.628783751

2015-07-20 15:38:05.416223 center_freq 2412625000.0 freq 2412618750.0 power_db 15.3595859574 noise_floor_db -114.628783751

2015-07-20 15:38:05.416258 center_freq 2412625000.0 freq 2412625000.0 power_db 17.8255318179 noise_floor_db -114.628783751

2015-07-20 15:38:05.416298 center_freq 2412625000.0 freq 2412631250.0 power_db 15.7999017695 noise_floor_db -114.628783751

2015-07-20 15:38:05.416334 center_freq 2412625000.0 freq 2412637500.0 power_db 11.3582761091 noise_floor_db -114.628783751

2015-07-20 15:38:05.416357 center_freq 2412625000.0 freq 2412643750.0 power_db 10.0867574173 noise_floor_db -114.628783751

2015-07-20 15:38:05.416377 center_freq 2412625000.0 freq 2412650000.0 power_db 9.18130320519 noise_floor_db -114.628783751

2015-07-20 15:38:05.416409 center_freq 2412625000.0 freq 2412656250.0 power_db 8.19971394409 noise_floor_db -114.628783751

2015-07-20 15:38:05.416442 center_freq 2412625000.0 freq 2412662500.0 power_db 6.72971726723 noise_floor_db -114.628783751

2015-07-20 15:38:05.416477 center_freq 2412625000.0 freq 2412668750.0 power_db 7.07840703166 noise_floor_db -114.628783751

2015-07-20 15:38:05.416516 center_freq 2412625000.0 freq 2412675000.0 power_db 6.39153306677 noise_floor_db -114.628783751

2015-07-20 15:38:05.416558 center_freq 2412625000.0 freq 2412681250.0 power_db 7.24759167685 noise_floor_db -114.628783751

2015-07-20 15:38:05.416586 center_freq 2412625000.0 freq 2412687500.0 power_db 6.79640402159 noise_floor_db -114.628783751

2015-07-20 15:38:05.416607 center_freq 2412625000.0 freq 2412693750.0 power_db 7.04702373552 noise_floor_db -114.628783751
```

Figure 6.6: Sensing Result of Channel No. 1 : 2.412 GHz

6.2.4 Observation Sequence

CN monitors and samples primary user activities. Samples collected for every channel are used for representing primary user activity in particular channel. The sample $q(t)$ in vector Q represents signal power received at time t at sample frequency f_s. Vector Q is represented as,

$$Q = \{q(1), q(2), q(3), ...q(t)....q(T)\}$$



```
./usrp_spectrum_sense.py 2.425G 2.448G > /home/ubuntu/shri/log2437.txt

linux; GNU C++ version 4.8.2; Boost_105400; UHD_003.008.004-0-g93011c14

2015-07-20 15:45:16.211485 center_freq 2437625000.0 freq 2437587500.0 power_db 7.25851439075 noise_floor_db -114.464795793

2015-07-20 15:45:16.211505 center_freq 2437625000.0 freq 2437593750.0 power_db 7.9585834496 noise_floor_db -114.464795793

2015-07-20 15:45:16.211525 center_freq 2437625000.0 freq 2437600000.0 power_db 7.85360469391 noise_floor_db -114.464795793

2015-07-20 15:45:16.211545 center_freq 2437625000.0 freq 2437606250.0 power_db 9.44064349703 noise_floor_db -114.464795793

2015-07-20 15:45:16.211565 center_freq 2437625000.0 freq 2437612500.0 power_db 11.5080716197 noise_floor_db -114.464795793

2015-07-20 15:45:16.211584 center_freq 2437625000.0 freq 2437618750.0 power_db 14.5812298326 noise_floor_db -114.464795793

2015-07-20 15:45:16.211604 center_freq 2437625000.0 freq 2437625000.0 power_db 16.0554004091 noise_floor_db -114.464795793

2015-07-20 15:45:16.211623 center_freq 2437625000.0 freq 2437631250.0 power_db 14.8683204529 noise_floor_db -114.464795793

2015-07-20 15:45:16.211643 center_freq 2437625000.0 freq 2437637500.0 power_db 10.9695424004 noise_floor_db -114.464795793

2015-07-20 15:45:16.211662 center_freq 2437625000.0 freq 2437643750.0 power_db 9.5436633247 noise_floor_db -114.464795793

2015-07-20 15:45:16.211697 center_freq 2437625000.0 freq 2437650000.0 power_db 9.17395299011 noise_floor_db -114.464795793
```

Figure 6.7: Sensing Result of Channel No. 6 : 2.437 GHz

From Figure 6.7, Vector q for channel no. 6 represents signal power received t^{th} sample,

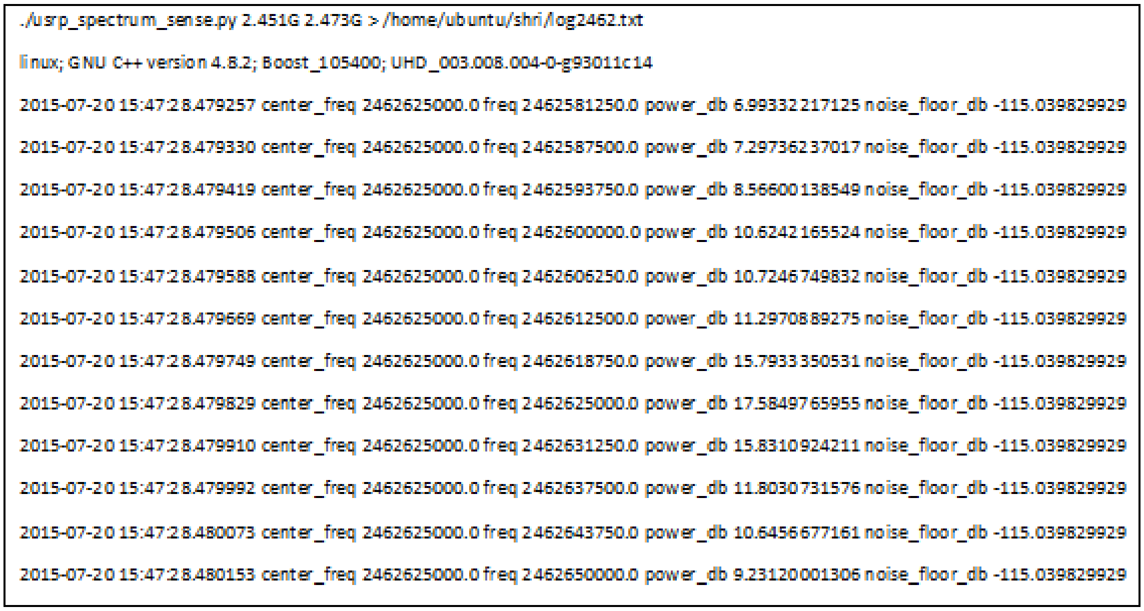
$$q = \{7.258, 7.958, 7.853, 9.440, 11.508, 14.581, 16.055, 14.868, 10.969, 9.543, 9.173\}$$

CN compares the integrator output $q(t)$ with the threshold λ for every channel c. At every instance t, the node records an observation symbol OS_t as per the following condition,

$$OS_t = 0, \quad \text{if } \; q(t) \leq \lambda; \qquad OS_t = 1 \;\; \text{if } \; q(t) \geq \lambda$$

Every cognitive node periodically sense and records such observation and creates observation sequence for T time slots by deciding λ as minimum power required to represent signal generated from primary transmitter. The value of λ is dependent on the total gain. λ is always below the square root of total gain. Monitored Observation Sequence for channel no. 6 is,

$$OS = \{OS_1, OS_2, ..., OS_T\}, \quad where \;\; OS_t \in [0, 1] \quad \forall t = 1....T$$



```
./usrp_spectrum_sense.py 2.451G 2.473G > /home/ubuntu/shri/log2462.txt

linux; GNU C++ version 4.8.2; Boost_105400; UHD_003.008.004-0-g93011c14

2015-07-20 15:47:28.479257 center_freq 2462625000.0 freq 2462581250.0 power_db 6.99332217125 noise_floor_db -115.039829929

2015-07-20 15:47:28.479330 center_freq 2462625000.0 freq 2462587500.0 power_db 7.29736237017 noise_floor_db -115.039829929

2015-07-20 15:47:28.479419 center_freq 2462625000.0 freq 2462593750.0 power_db 8.56600138549 noise_floor_db -115.039829929

2015-07-20 15:47:28.479506 center_freq 2462625000.0 freq 2462600000.0 power_db 10.6242165524 noise_floor_db -115.039829929

2015-07-20 15:47:28.479588 center_freq 2462625000.0 freq 2462606250.0 power_db 10.7246749832 noise_floor_db -115.039829929

2015-07-20 15:47:28.479669 center_freq 2462625000.0 freq 2462612500.0 power_db 11.2970889275 noise_floor_db -115.039829929

2015-07-20 15:47:28.479749 center_freq 2462625000.0 freq 2462618750.0 power_db 15.7933350531 noise_floor_db -115.039829929

2015-07-20 15:47:28.479829 center_freq 2462625000.0 freq 2462625000.0 power_db 17.5849765955 noise_floor_db -115.039829929

2015-07-20 15:47:28.479910 center_freq 2462625000.0 freq 2462631250.0 power_db 15.8310924211 noise_floor_db -115.039829929

2015-07-20 15:47:28.479992 center_freq 2462625000.0 freq 2462637500.0 power_db 11.8030731576 noise_floor_db -115.039829929

2015-07-20 15:47:28.480073 center_freq 2462625000.0 freq 2462643750.0 power_db 10.6456677161 noise_floor_db -115.039829929

2015-07-20 15:47:28.480153 center_freq 2462625000.0 freq 2462650000.0 power_db 9.23120001306 noise_floor_db -115.039829929
```

Figure 6.8: Sensing Result of Channel No. 11 : 2.462 GHz

$$OS = \{0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0\}$$

Every channel i.e channel no. 1, 6 and 11 is scanned, sensed and analyzed to record the strings of 0's and 1's. These strings that is observation sequence is used to find statistics of activities of primary user. Monitoring and recording of wireless channel in the form of observation sequence helps to represent dynamic changes in primary user activity and traffic pattern.

6.3 Training Hidden Markov Model

Observation sequences of channel no. 1, 6 and 11 are obtained having $|Q|$ number of symbols. These $|Q|$ symbols are divided into $|Q|/T$ sub sequences, having length T. For example, from the continuous observation samples of primary user activity $|Q| = 3300$ symbols are taken. These $|Q|$ symbols are divided into 11 sub-sequences of $T = 300$ symbols.

The parameter set for two state HMM is given in Table 6.1. HMM is trained with Idle and Busy state. The first 300 symbols i.e first sub-sequence is used to

Table 6.1: Parameters of Hidden Markov Model

Parameter	**Details**
States of HMM (N)	2
Observation Symbols (M)	2
Observation Symbol Set V	$\{0, 1\}$
Training Sequence	1
Testing Sequences	11
Training Observation Sequence	300 Symbols
Testing Observation Sequence	300 Symbols

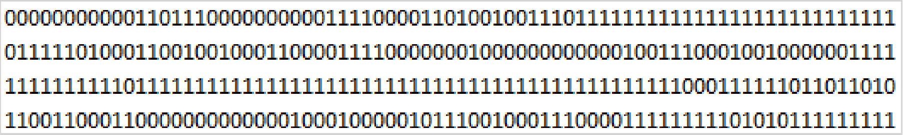
train the Hidden Markov Model (HMM) and remaining sub-sequences are used as test sequences. It is then tested with remaining observation sequence.

HMM training starts with estimating $\lambda = (A, B, \Pi)$, with uniform distributed values. The initial parameter set and training sequence decides time required to train HMM. Initial parameters $\lambda = (A, B, \Pi)$ based on $\{OS_t\}_{t=1}^{T}$ are used to train HMM for predicting $\{OS_t\}_{t=T+1}^{2T}$. λ is updated to $\lambda*$ for increasing log-likelihood of actual sequence and predicted sequence.

Example

This example represents calculations of initial parameter set of HMM for training sub-sequence.

Observation Sequence of 300 symbols, $\{OS_t\}_{t=1}^{T}$, where $T = 300$, is taken from channel observation sequences for defining initial parameter set $\lambda = (A, B, \Pi)$.



```
000000000001101110000000000111100001101001001110111111111111111111111111111
011111010001100100100011000011110000000100000000000010011100010010000001111
111111111101111111111111111111111111111111111111111111111000111111011011010
110011000110000000000000100010000010111001000111000011111111101010111111111
```

Number of '0 ':127 **Number of '1 ':**173

Initial Matrix Π is,

1. Probability of starting with '0 '$127/300 \approx 0.42333$
2. Probability of starting with '1 '$173/300 \approx 0.57666$

$$\Pi = \begin{bmatrix} 0.42333 & 0.57666 \end{bmatrix}$$

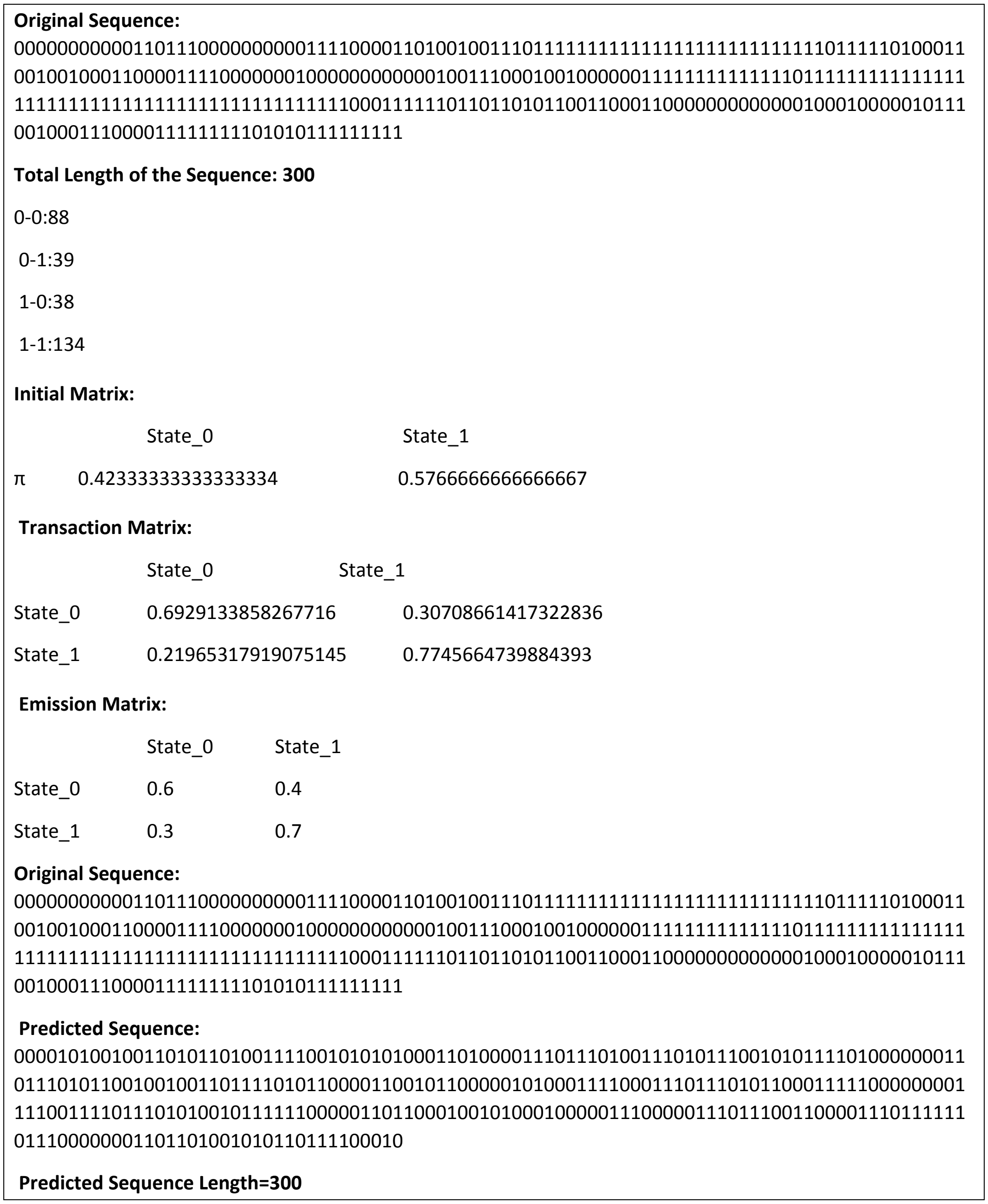
Transitions in above Observation Sequence with respect to number of 1's and 0's are,

1. $0 - 0 = 88/127 \approx 0.69291$
2. $0 - 1 = 39/127 \approx 0.30708$
3. $1 - 0 = 38/173 \approx 0.21965$
4. $1 - 1 = 134/173 \approx 0.77456$

Following matrix represents initial probability of **State Transitions** and **Observations**,

$$A = \begin{bmatrix} 0.69291 & 0.30708 \\ 0.21965 & 0.77456 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}$$

HMM output for actual sequence of 300 symbols with initial parameter set and predicted sequence is shown using screen shot in Figure 6.9.



```
Original Sequence:
00000000000110111000000000001111000011010010011101111111111111111111111111111110111110100011
001001000110000111100000000100000000000010011100010010000001111111111111101111111111111
111111111111111111111111111111100011111011011010110011000110000000000000100010000010111
0010001110000111111111010101111111111

Total Length of the Sequence: 300

0-0:88

 0-1:39

 1-0:38

 1-1:134

Initial Matrix:

                  State_0                      State_1

π        0.42333333333333334                 0.5766666666666667

 Transaction Matrix:

                  State_0                State_1

State_0           0.6929133858267716          0.30708661417322836

State_1           0.21965317919075145         0.7745664739884393

 Emission Matrix:

                  State_0          State_1

State_0           0.6              0.4

State_1           0.3              0.7

Original Sequence:
00000000000110111000000000001111000011010010011101111111111111111111111111111110111110100011
001001000110000111100000000100000000000010011100010010000001111111111111101111111111111
111111111111111111111111111111100011111011011010110011000110000000000000100010000010111
0010001110000111111111010101111111111

 Predicted Sequence:
000010100100110101101001111001010101000110100001110111010011101011100101011110100000011
011101011001001001101111010110000110010110000010100011110001110111010110001111000000001
11100111101110101001011111100000110110001001010001000001110000011101110011000011101111111
0111000000011011010010101101111100010

 Predicted Sequence Length=300
```

Figure 6.9: Initial Parameter Set with Actual and Predicted Sequence

6.4 Prediction using Hidden Markov Model

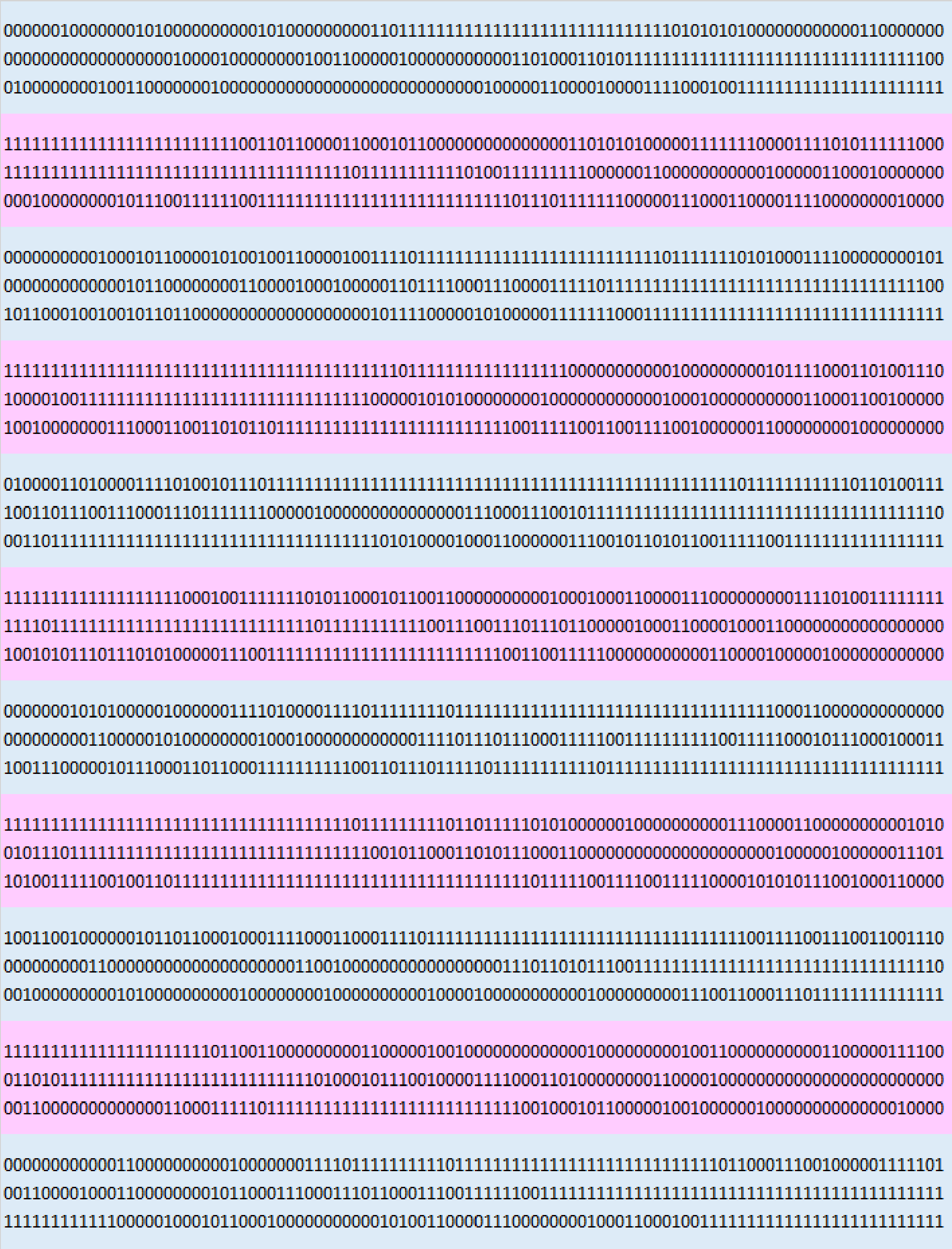
HMM is trained initially with 300 symbols, as first time-slot. Next time slot of 300 symbols is predicted by HMM with maximum likelihood with the actual time slot. Channel Availability (CA) is used for checking accuracy of the HMM for generating maximum likelihood observation sequence. CA is calculated using,

$$CA_c = A_c^0 + \frac{|GOS_{H_c}|}{GOS_{H_c}^1} \tag{6.1}$$

where, GOS_{H_c} is Generated Observation Sequence by HMM trained for channel c, $|GOS_{H_c}|$ is number of symbols in GOS, $GOS_{H_c}^1$, number of 1's in GOS and A_c^0 is an average number of 0's between two 1's in the GOS.

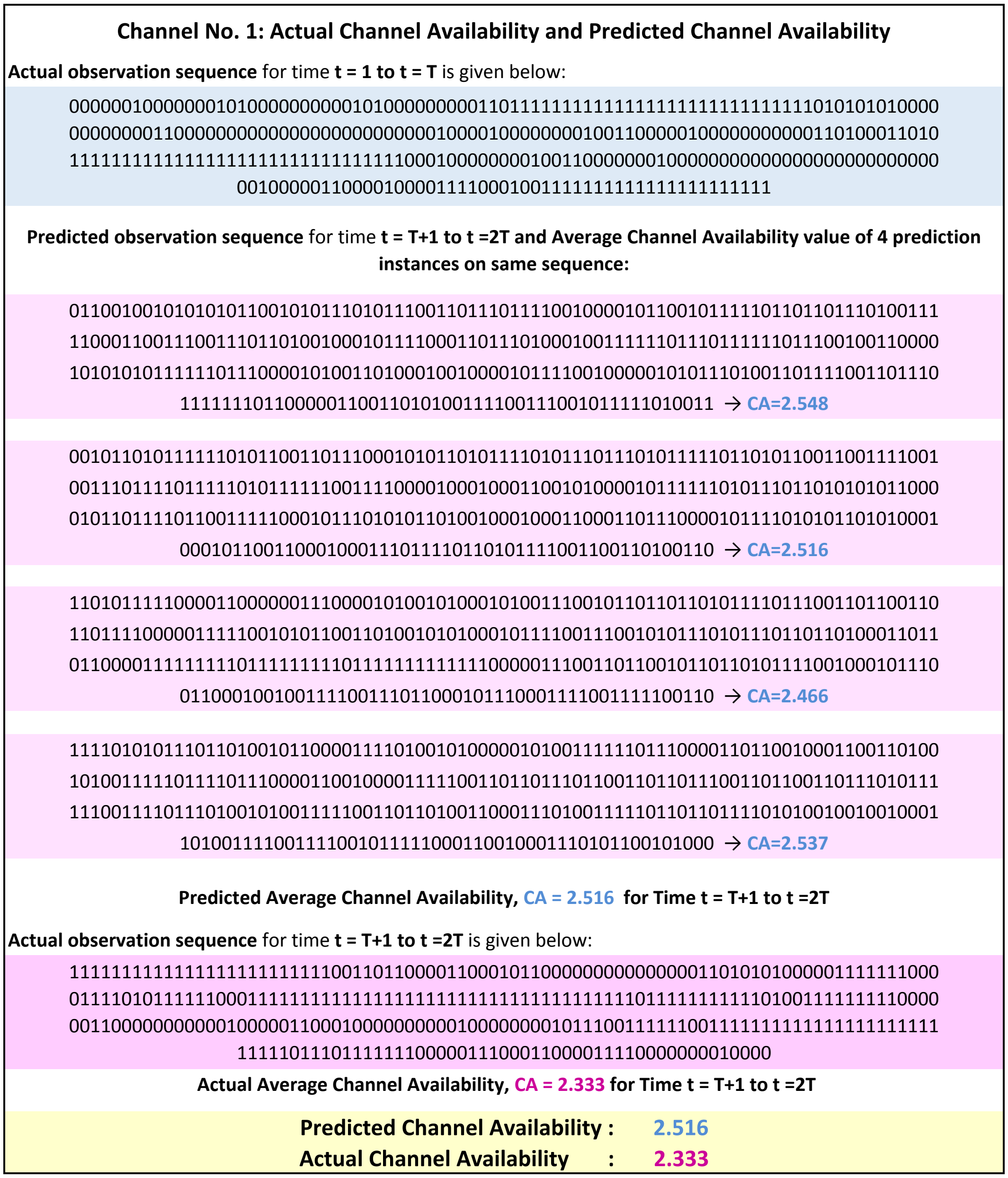
Figure 6.10 shows, 11 sub-sequences of actual OS of channel no. 1 in different colours. Every sub-sequence is of 300 symbols.

Actual sub-sequence of OS for time $t = 1$ to $t = T$ is used to train HMM. Trained HMM is used to predict the GOS of 300 symbols for the next time slot that is $t = T+1$ to $t = 2T$. For improving accuracy of testing, GOS for next time slot is predicted for four times. The average CA of these four predicted observation sequences is compared with CA of actual OS, as shown in Figure 6.11. The overall objective is to increase the maximum likelihood of actual CA with predicted CA.



```
00000010000000101000000000001010000000001101111111111111111111111111111110101010100000000000000110000000
00000000000000000010000100000000100110000001000000000000110100011010111111111111111111111111111111111100
0100000000100110000000100000000000000000000000000000100000110000100001111000100111111111111111111111
1111111111111111111111111001101100001100010110000000000000000110101010000011111110000111101011111111000
1111111111111111111111111111111111111110111111111110100111111111000000110000000000010000011000100000000
000100000000101110011111100111111111111111111111111111101110111111100000111000110000111100000000010000
00000000001000101100001010010011000010011110111111111111111111111111111110111111010100011110000000000101
00000000000001011000000000110000100010000011011110001110000111110111111111111111111111111111111111111100
1011000100100101101100000000000000000000101111000001010000011111110001111111111111111111111111111111
11111111111111111111111111111111111111111101111111111111111110000000000010000000001011110001101001110
10000100111111111111111111111111111111111100000101010000000001000000000000100010000000000110001100100000
1001000000011100011001101010110111111111111111111111111110011111001100111100100000011000000001000000000
010000110100001111010010111011111111111111111111111111111111111111111111111111111101111111111110110100111
10011011100111000111011111110000010000000000000000011100011100101111111111111111111111111111111111111110
00110111111111111111111111111111111111111110101000010001100000011100101101011001111100111111111111111111
1111111111111111111100010011111110101100010110011000000000001000100011000011100000000001111010011111111
111101111111111111111111111111111111101111111111100111001110111011000001000110000100011000000000000000000
1001010111011101010000011100111111111111111111111111111001100111110000000000001100001000001000000000000
0000000101010000001000000111101000011110111111111011111111111111111111111111111111111111100011000000000000000
0000000001100000101000000000100010000000000000111101110111000111110011111111111001111100010111000100011
1001110000010111000110110001111111111001101110111110111111111111011111111111111111111111111111111111111
1111111111111111111111111111111111111111011111111101101111101010000000100000000000111000011000000000001010
0101110111111111111111111111111111111111001011000110101110001100000000000000000000001000001000000011101
1010011111001001101111111111111111111111111111111111111111111011111001111001111100001010101110010001100000
1001100100000001011011000100011110001100011110111111111111111111111111111111111111111001111001110011001110
0000000001100000000000000000000011001000000000000000000011101101011100111111111111111111111111111111110
00100000000010100000000000100000000100000000000100001000000000000100000000011100110001110111111111111111
11111111111111111111111110110011000000000011000001001000000000000001000000000100110000000000110000011111100
01101011111111111111111111111111111101000101110010000111100011010000000001100001000000000000000000000000000
001100000000000000110001111101111111111111111111111111111111001000101100000100100000010000000000000000010000
0000000000001100000000000100000001111011111111111011111111111111111111111111111111011000111001000001111101
0011000010001100000000010110001110001110110001110011111100111111111111111111111111111111111111111111111111
111111111111000000100010110001000000000000101001100001110000000001000110001001111111111111111111111111111
```

Figure 6.10: Observation Sequence for Channel No. 1



Channel No. 1: Actual Channel Availability and Predicted Channel Availability

Actual observation sequence for time **t = 1 to t = T** is given below:

000000100000000101000000000001010000000000110111111111111111111111111111111111010101010000
000000001100000000000000000000000000010000100000000100110000010000000000011010001101 0
111111111111111111111111111111111110001000000000100110000000010000000000000000000000000000
001000000110000100001111000100111111111111111111111

Predicted observation sequence for time **t = T+1 to t =2T and Average Channel Availability value of 4 prediction instances on same sequence:**

011001001010101011001010111010111001101110111100100001011001011110110110111010011 1
110001100111001110110100100010111100011011101000100111111011101111111011100100110000
101010101111110111000010100110100010010000101111001000001010111010011011110011 01110
1111111011000001100110101001111001110010111110100 11 → CA=2.548

0010110101111110101100110111000101011010111101011101110101111011010110011001111001
0011101111011111010111111001111000010001000110010100001011111101011101101010101 1000
0101101111011001111100010111010101101001000100011000110111000010111101010110101 0001
000101100110001000111011110110101111001100110100110 → CA=2.516

110101111110000110000000111000010100101000101001110010110110110101111011100110110 0110
1101111100000111110010101100110100101010001011110011100101011101011101101101000110 11
0110000111111111011111111101111111111111100000111001101100101101101011110010001011 10
0110001001001111001110110001011100011110011111001 10 → CA=2.466

1111010101110110100101100001111010010100000101001111110111000011011001000110011010 0
1010011111011110111000011001000011111001101101110110011011011100110110011011101011 1
1110011110111010010100111110011011010011000111010011111011011011110101001001001000 1
101001111001111001011111000110010001110101100101000 → CA=2.537

Predicted Average Channel Availability, CA = 2.516 **for Time t = T+1 to t =2T**

Actual observation sequence for time **t = T+1 to t =2T** is given below:

111111111111111111111111111110011011000011000101100000000000000001101010100000111111110 00
0111101011111100011111111111111111111111111111111111111101111111111101001111111110000
0011000000000000100000110001000000000001000000000101110011111100111111111111111111111 1
11111011101111111000001110001100001111000000000010000

Actual Average Channel Availability, CA = 2.333 **for Time t = T+1 to t =2T**

Predicted Channel Availability : 2.516
Actual Channel Availability : 2.333

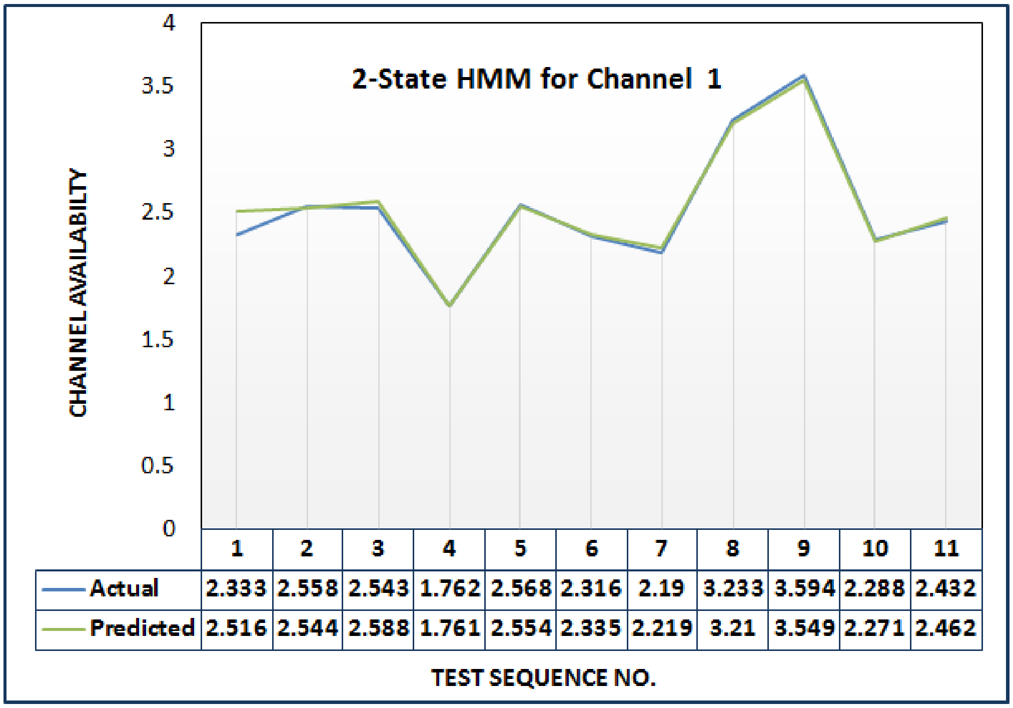
Figure 6.11: Actual Channel Availability and Predicted Channel Availability

Trained HMM is tested separately with 3300 symbols for three Wi-Fi channel that is channel 1, 6 and 11. Observation sequence of channel no. 1 is shown in Figure 6.10. Similar observation sequence for channel no. 6 and channel no. 11 are obtained.

The first actual sub-sequence of every channel with $t = 1$ to $t = T$ is given to HMM. HMM produces (GOS) from time-step $t = T + 1$ to $t = 2T$. Actual sub-sequence from $t = T + 1$ to $t = 2T$ is compared with predicted sub-sequence, GOS from $t = T + 1$ to $t = 2T$ using channel availability CA.

CA of actual observation sequence is compared with channel availability of the predicted observation sequence. The comparative details of WiFi channel no. 1 (center frequency 2.412 GHz) are given in Table 6.2. Tabular representation of actual and predicted channel availability of channel no. 6 (center frequency 2.437 GHz) and 11 (center frequency 2.462 GHz) are given in Table 6.3 and 6.4 respectively.

The objective of prediction using HMM is increasing the maximum likelihood of predicted channel conditions with actual conditions. This helps to predict primary user behavior on particular channel. The graph in Figure 6.12, 6.13 and 6.14 for channel no.1, 6 and 11 respectively, shows that the predicted CA is close to the actual CA that is primary user behavior on all channels.

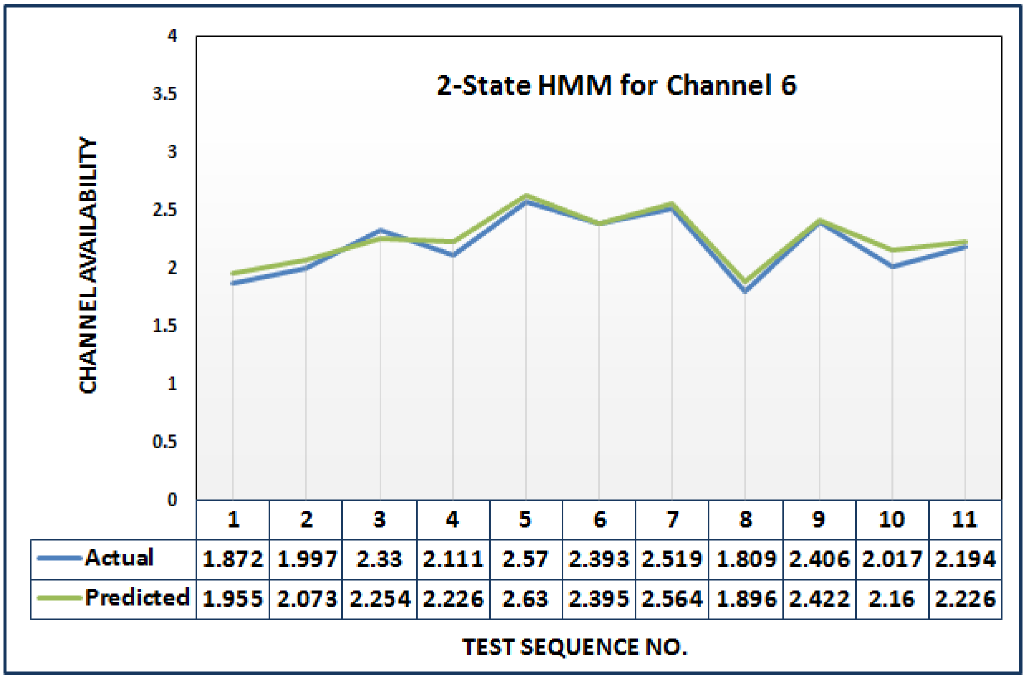


	1	2	3	4	5	6	7	8	9	10	11
Actual	2.333	2.558	2.543	1.762	2.568	2.316	2.19	3.233	3.594	2.288	2.432
Predicted	2.516	2.544	2.588	1.761	2.554	2.335	2.219	3.21	3.549	2.271	2.462

Figure 6.12: Prediction Analysis of 2.412 GHz - Availability Channel No. 1

Table 6.2: Prediction Analysis of Channel 1 : 2.412 GHz

Observation Sequence	Actual CA_1	Predicted CA_1
$t = 1$ to $t = T$	3.880	....
$t = T + 1$ to $t = 2T$	2.333	2.516
$t = 2T + 1$ to $t = 3T$	2.558	2.544
$t = 3T + 1$ to $t = 4T$	2.543	2.588
$t = 4T + 1$ to $t = 5T$	1.762	1.761
$t = 5T + 1$ to $t = 6T$	2.568	2.554
$t = 6T + 1$ to $t = 7T$	2.316	2.335
$t = 7T + 1$ to $t = 8T$	2.19	2.219
$t = 8T + 1$ to $t = 9T$	3.233	3.21
$t = 9T + 1$ to $t = 10T$	3.594	3.549
$t = 10T + 1$ to $t = 11T$	2.288	2.271
$t = 11T + 1$ to $t = 12T$	2.432	2.462

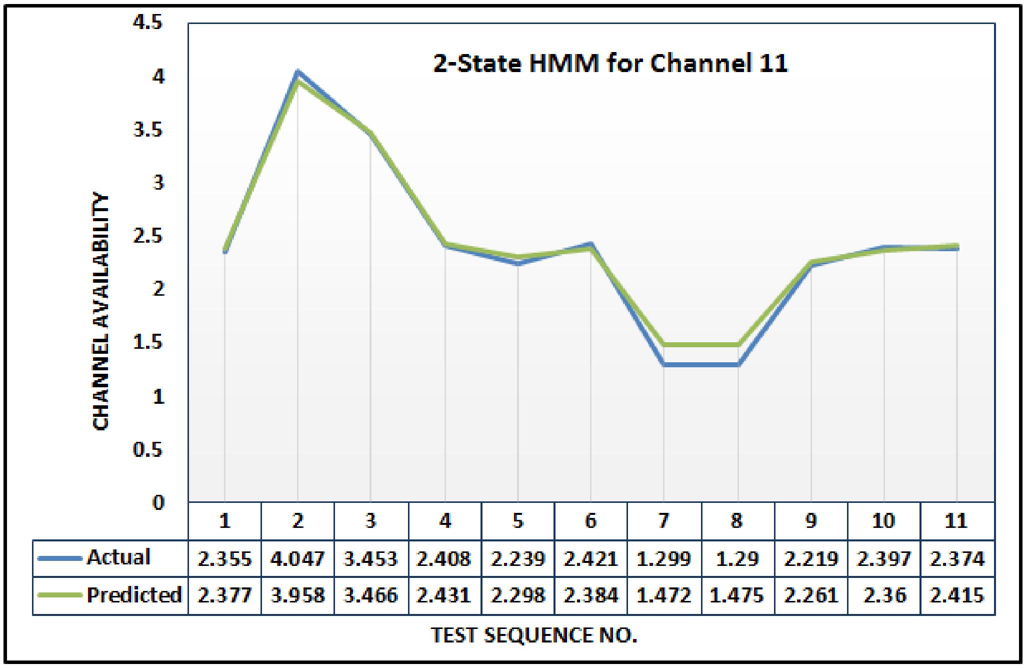


	1	2	3	4	5	6	7	8	9	10	11
Actual	1.872	1.997	2.33	2.111	2.57	2.393	2.519	1.809	2.406	2.017	2.194
Predicted	1.955	2.073	2.254	2.226	2.63	2.395	2.564	1.896	2.422	2.16	2.226

Figure 6.13: Prediction Analysis of 2.437 GHz - Availability Channel No. 6

Table 6.3: Prediction Analysis of Channel 6 : 2.437 GHz

Observation Sequence	Actual CA_6	Predicted CA_6
$t = 1$ to $t = T$	3.880	....
$t = T + 1$ to $t = 2T$	1.872	1.955
$t = 2T + 1$ to $t = 3T$	1.997	2.073
$t = 3T + 1$ to $t = 4T$	2.33	2.254
$t = 4T + 1$ to $t = 5T$	2.111	2.226
$t = 5T + 1$ to $t = 6T$	2.57	2.63
$t = 6T + 1$ to $t = 7T$	2.393	2.395
$t = 7T + 1$ to $t = 8T$	2.519	2.564
$t = 8T + 1$ to $t = 9T$	1.809	1.896
$t = 9T + 1$ to $t = 10T$	2.406	2.422
$t = 10T + 1$ to $t = 11T$	2.017	2.16
$t = 11T + 1$ to $t = 12T$	2.194	2.226



	1	2	3	4	5	6	7	8	9	10	11
Actual	2.355	4.047	3.453	2.408	2.239	2.421	1.299	1.29	2.219	2.397	2.374
Predicted	2.377	3.958	3.466	2.431	2.298	2.384	1.472	1.475	2.261	2.36	2.415

Figure 6.14: Prediction Analysis of 2.462 GHz - Availability Channel No. 11

Table 6.4: Prediction Analysis of Channel 11 : 2.462 GHz

Observation Sequence	Actual CA_{11}	Predicted CA_{11}
$t = 1$ to $t = T$	3.880	....
$t = T + 1$ to $t = 2T$	2.355	2.377
$t = 2T + 1$ to $t = 3T$	4.047	3.958
$t = 3T + 1$ to $t = 4T$	3.453	3.466
$t = 4T + 1$ to $t = 5T$	2.408	2.431
$t = 5T + 1$ to $t = 6T$	2.239	2.298
$t = 6T + 1$ to $t = 7T$	2.421	2.384
$t = 7T + 1$ to $t = 8T$	1.299	1.472
$t = 8T + 1$ to $t = 9T$	1.29	1.475
$t = 9T + 1$ to $t = 10T$	2.219	2.261
$t = 10T + 1$ to $t = 11T$	2.397	2.36
$t = 11T + 1$ to $t = 12T$	2.374	2.415

As per the results, HMM successfully predicts maximum likelihood estimation of actual observation sequence. Predicted observation sequence is used to understand the pattern of spectrum usage in near future. This helps to identify the preferable channel for communication between any two nodes. Results in Figure 6.12, 6.13 and 6.14 for channel no.1, 6 and 11 respectively shows that decision of selecting one of the channels for communication can be done on predicted observation sequence. The computed Channel Availability of predicted observation sequence is almost same as Channel Availability of actual sequence.

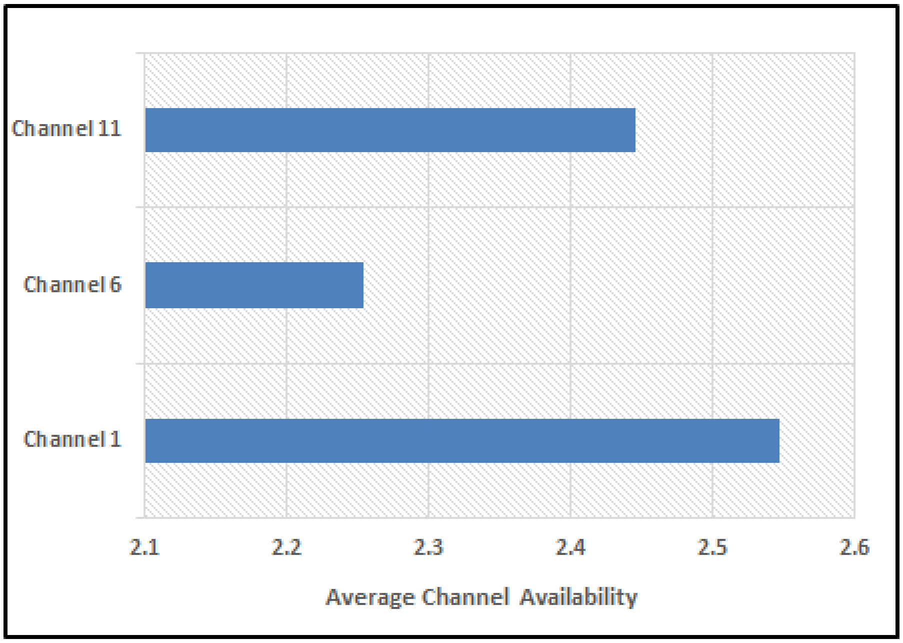


Figure 6.15: Channel Availability

6.5 Performance Evaluation and Channel Selection

Cognitive nodes calculates CA for every channel c and ranks channels using its value. The link selection method uses link cost to select the preferable channel. The objective is to select lowest cost link which is inversely proportional to the value of CA. Availability of the channel is more with generated observation sequence having more number of 0's and larger separation between 1's. CA is dependent on amount of primary user activity on that channel. Comparative analysis in Figure 6.15 shows that channel no. 1 is having good channel availability compared to channel no. 6 and channel no. 11, for the particular instance of observation sequence.

6.6 Concluding Remark

The chapter provides details of implementation and experimental setup for channel sensing and analysis. The network model for sensing three channel in 2.4 GHz range is shown with details of energy detection sensing technique. Special type of programmable hardware is used to sense the wide range of RF frequencies called as USRP N210. Free and open-source GNU radio software development toolkit is used to implement the signal processing blocks for SDR. Sensed signal strength samples are used to create observation sequence to train and test the HMM. The trained HMM generates observation sequence of channel representing future predicted traffic on every channel. As per the result shown, HMM perfectly analyzes the behavior of primary user activities and computes maximum likelihood estimation of channel traffic. The channel traffic prediction is used to decide the availability of every channel to be used by cognitive nodes for communication.

Chapter 7

Implementation and Analysis of MARL Routing

Chapter 7

Implementation and Analysis of MARL Routing

The implementation of multi-agent reinforcement learning based, online and opportunistic routing of the cognitive radio network is cognate to the route formation involving the selection of relay among the neighbouring nodes. The selection of next node is based on successful transmission probabilities and environmental conditions of channel availability. The proposed methodology is compared with two state-of-the-art-techniques, adaptive opportunistic routing and joint spectrum-route selection with service differentiation, described in section 7.1. Section 7.2 gives details of simulation objectives and parameters. The performance evaluation is represented in section 7.3 followed by computational complexity and control overhead of MARL based opportunistic routing for MCRAN in section 7.4. Section 7.5 contains concluding remark on implementation and analysis.

7.1 State-of-the-Art Techniques Implementation

The detailed description of State-of-the-Art technique is given in Related Work (2.5.1.) section. Following are the implementation and performance details:

Adaptive Opportunistic Routing

- Adaptive opportunistic routing exploits the broadcasting nature of wireless transmission by selecting the diverse path to mitigate the impact of poor wireless links.

- Implemented with 16 indoor nodes and equipped with 802.11 radios transmitting at 11 Mbps.

- The route is formed by broadcasting the data packet followed by selecting one of the node as relay.

- Routing decisions are using estimated best score computed using Monte Carlo policy evaluation.

- Estimated best score is updated with the received reward at a particular time.

- Exponential space complexity with respect to the number of neighbours.

Cognitive Routing Protocol, CRP: Spectrum aware route selection with primary user protection and good CRN performance

- Addressing issues like channel characteristics in route selection with primary user protection and cognitive network performance.

- Implemented with 50 cognitive and 9 primary users in the square region of 1000m, equipped with 802.11 radios transmitting at 11 Mbps.

- Class-I route provides better cognitive radio performance by selecting larger separated hopes increasing packet arrival time.

- Class-II route provides better protection to primary user by avoiding interference region.

- Fixed routing paths are formed using ad hoc routing mechanism of wireless network with route preference information on an optimization function.

7.2 Simulation Objective and Parameters

MARL based routing is implemented using java based JiST/SWAN simulation libraries. It efficiently and transparently executes discrete event simulation by embedding semantic with Java execution model. Advantages of JiST/SWAN as a potential alternative over MATLAB, Qualnet, OMNET and NS-2 are:

- JiST entities are regular java class having advantages of platform independence and large library support.
- Produces Equivalent simulation result in less time using less memory (Schoch E. et al., 2008).
- Model for nodes and environment is supported by complete library for the simulation of MANET called SWAN running on JiST engine.

Simulation Objectives

The simulation scenario considers random topology with three channels having goals as follows:

- To reduce collision and interference to primary users and to reduce route rediscovery request frequency by following optimal strategy.
- To select stable and non-interfering links for maximizing throughput.

Simulation Parameters

The realistic random topology of 16 nodes moving in random directions is used for demonstrating the robustness of the algorithm. Three different wireless bands with channel frequency of 2.4 GHz are used. The number of active primary users considered are in the range of 0 to 4. Protocol 802.11n is used to transmit packet of size 1000 bytes at 11 Mbps. The acknowledgment packets are short packets of length 24 Bytes transmitted at 11 Mbps, transmitted at lower rate of 1 Mbps to ensure reliability.

Table 7.1: Network Parameters

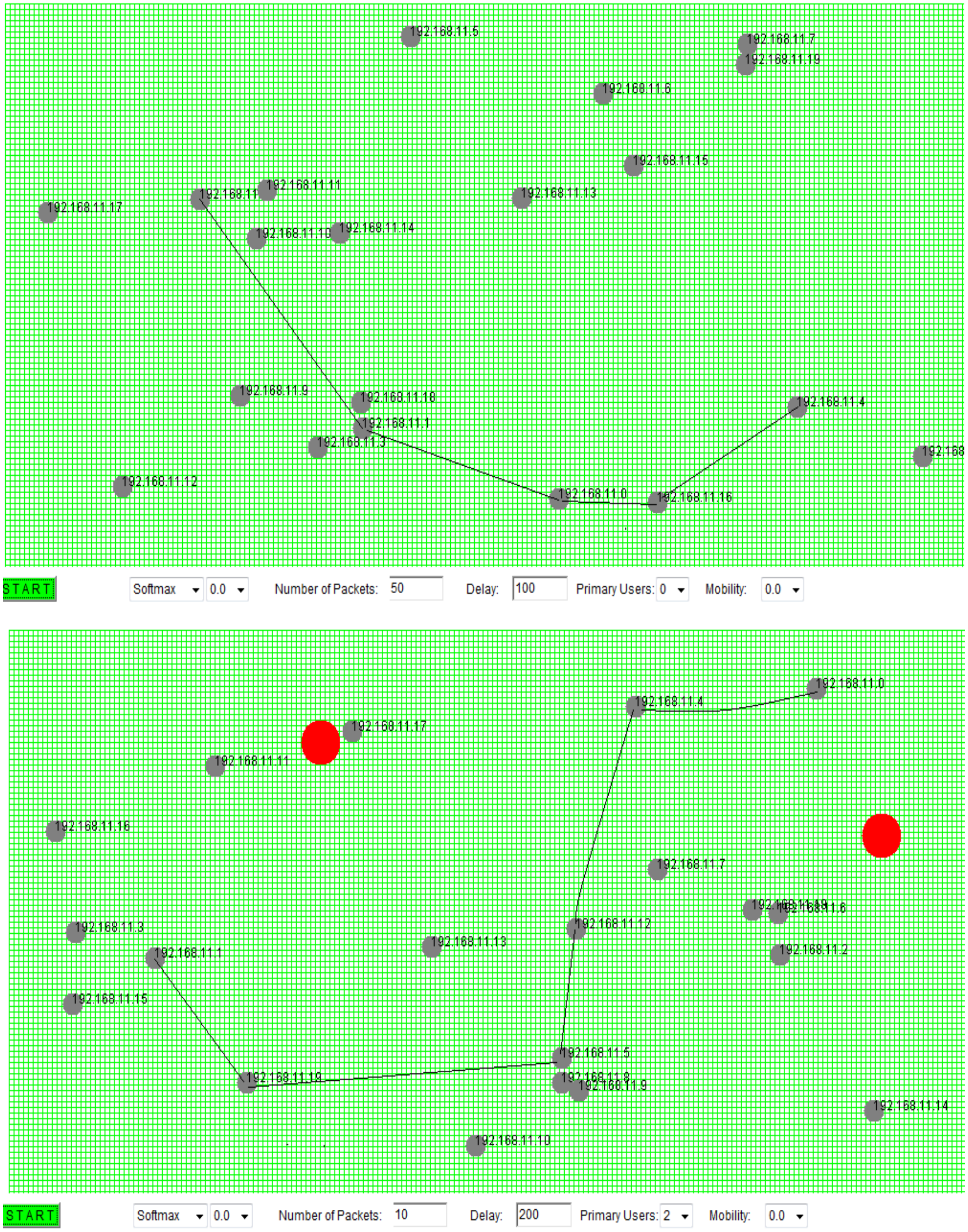
Sr.No.	**Parameter**	**Value**
1	MAC Type	802.11n
2	Routing Protocol	Online Opportunistic Routing Protocol
3	Number of Packets	10 to 1200
4	Number of Cognitive Nodes	20 Nodes
5	Number of PU Nodes	0 to 4 Nodes
6	Max Node Speed	1.6 m/s
7	Channel Type	Wi-Fi Channel (1,6,11)
8	Channel Frequency	2.412 GHz, 2.437 GHz and 2.462 GHz
9	Channel Capacity	11 Mbps
10	Channel Switching Time	$80\mu s$
11	Bandwidth of a Channel	22 MHz
12	Data Payload	1000 Bytes/Packet
13	Network Topologies Used	Dynamic and Mobile

The cost of transmission at every hop is represented as LC_c computed using equation no. 5.13. Simulation and network parameters are given in Table 7.1.

7.3 Performance Evaluation

The performance of MARL based opportunistic routing is investigated using simulation under realistic wireless setting. This implementation demonstrates a robust performance gain using proposed routing algorithm in MCRAN. The evaluation provides details of design parameter and appropriate choice of their values. The proposed algorithm is also investigated with respect to the network parameters and its performance.

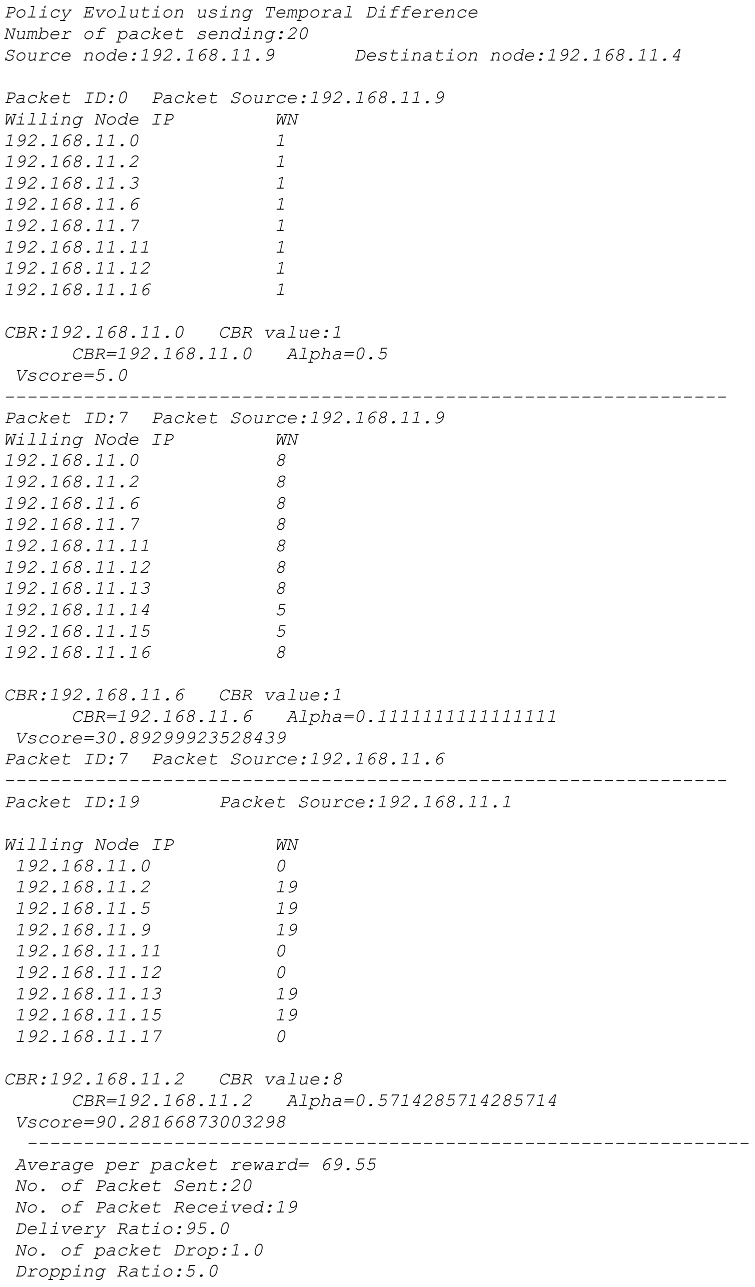
The screen shot of route formation process with and without primary users is



START Softmax 0.0 Number of Packets: 50 Delay: 100 Primary Users: 0 Mobility: 0.0

START Softmax 0.0 Number of Packets: 10 Delay: 200 Primary Users: 2 Mobility: 0.0

Figure 7.1: Route Formation With and Without Primary User



```
Policy Evolution using Temporal Difference
Number of packet sending:20
Source node:192.168.11.9       Destination node:192.168.11.4

Packet ID:0  Packet Source:192.168.11.9
Willing Node IP         WN
192.168.11.0            1
192.168.11.2            1
192.168.11.3            1
192.168.11.6            1
192.168.11.7            1
192.168.11.11           1
192.168.11.12           1
192.168.11.16           1

CBR:192.168.11.0   CBR value:1
      CBR=192.168.11.0   Alpha=0.5
 Vscore=5.0
----------------------------------------------------------------
Packet ID:7  Packet Source:192.168.11.9
Willing Node IP         WN
192.168.11.0            8
192.168.11.2            8
192.168.11.6            8
192.168.11.7            8
192.168.11.11           8
192.168.11.12           8
192.168.11.13           8
192.168.11.14           5
192.168.11.15           5
192.168.11.16           8

CBR:192.168.11.6   CBR value:1
      CBR=192.168.11.6   Alpha=0.1111111111111111
 Vscore=30.89299923528439
Packet ID:7  Packet Source:192.168.11.6
----------------------------------------------------------------
Packet ID:19       Packet Source:192.168.11.1

Willing Node IP         WN
 192.168.11.0           0
 192.168.11.2           19
 192.168.11.5           19
 192.168.11.9           19
 192.168.11.11          0
 192.168.11.12          0
 192.168.11.13          19
 192.168.11.15          19
 192.168.11.17          0

CBR:192.168.11.2   CBR value:8
      CBR=192.168.11.2   Alpha=0.5714285714285714
 Vscore=90.28166873003298
  ----------------------------------------------------------------
 Average per packet reward= 69.55
 No. of Packet Sent:20
 No. of Packet Received:19
 Delivery Ratio:95.0
 No. of packet Drop:1.0
 Dropping Ratio:5.0
```

Figure 7.2: Policy Evolution using Temporal Difference

shown in Figure 7.1 and the output of Temporal Difference policy evolution procedure is shown in Figure 7.2.

Every time before selecting next forwarding node, the source checks link availability and willingness of neighbouring nodes to participate in route formation process. The proposed system provides cognitive capability to routing protocol taking care of learning and routing simultaneously in the dynamic spectrum environment of CRN. Vector of Willing neighbour $WN\left(s, Ns\right)$, Candidate Best Relay (CBR_t^n) and Routing Score Vector $V_t^n\left(s, a\right)$, helps to achieve this objective. The proposed MARL based online opportunistic routing is compared with the CRP and adaptive routing as they are distributed protocols. The performance of the MARL based routing protocol from the viewpoint of distinguishing characteristics of MCRAN and effect of RL learning parameters is investigated in the following subsections.

7.3.1 Effect of Network Parameters

The proposed routing protocol is investigated from the viewpoint of MCRAN and compared with CRP. In the first set of experiments, it is compared by measuring throughput, collision with primary user and route re-discovery. The end-to-end performance is evaluated by varying, (i) Channel availability that is amount of time link is used by primary user, and (ii) Location of source and destination. CRP jointly selects spectrum and routes with service differentiation. Service differentiation is achieved by allowing two classes of routes. class-I provides better CR performance and class-II is designed to provide higher importance to the protection for the primary user.

The proposed MARL based online opportunistic routing lowers the collision to primary user and interference generated through CN operation. In Figure 7.3, it is observed that collision of MARL opportunistic routing with primary node is less as compared to both types of class I and II routes of CRP. The class-I route formation designed to transmit maximum possible number of packets over maximum possible transmission distance. Covering more propagation distance also requires higher transmission power causing collision with identified and unidentified primary users.

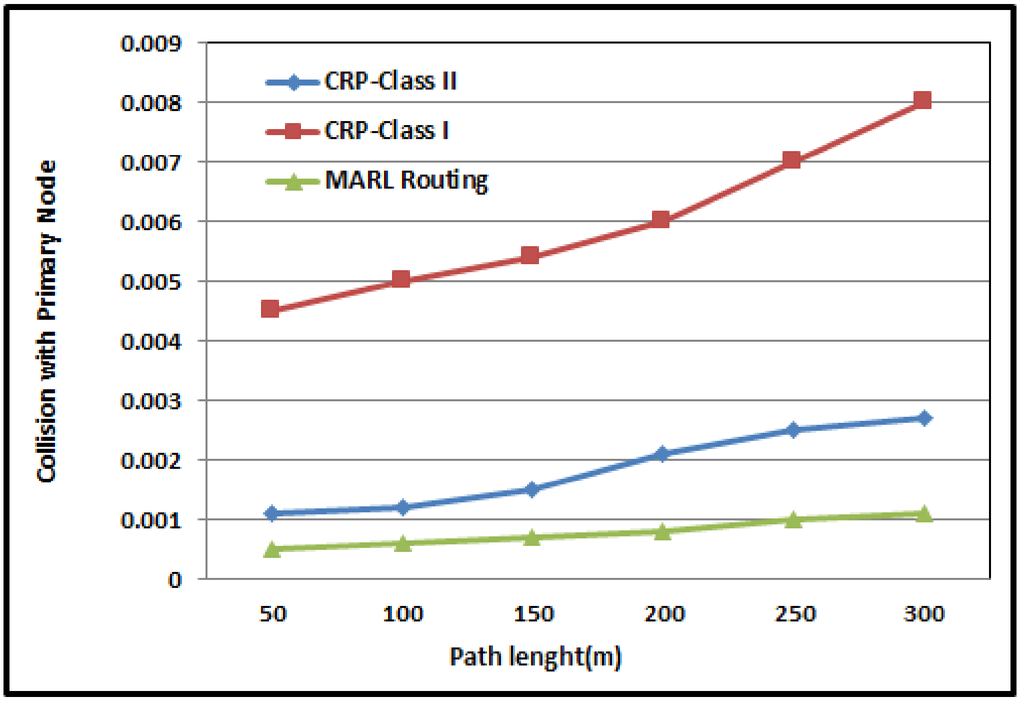


Figure 7.3: Performance Evaluation - Collision with Primary User

It increases interference to the primary user. Class-II selects routes by avoiding fractional area of overlap between coverage ranges of cognitive nodes and primary nodes. Thus, class-II routes have less collision with primary user activities compared to class-I. In MARL based opportunistic routing every agent selects next forwarding agent for the packet as per the current environmental conditions and interaction among multiple neighbouring agents. Therefore, the links which are not creating interference to primary user are selected. Cognitive agents not moving towards primary region are selected.

As given Figure 7.4, route re-discovery as a effect of the failure to transmit packets from one agent to another forwarding agent due to link failure or node unavailability. CRP routing protocol finds the fixed and complete route from source to destination. Failure in one of the link or unavailability of node affects the entire communication from source to destination. This results in finding entire path from source to destination which increases re-routing overhead for all nodes on the entire path. On the other hand, MARL routing finds route in a hop-by-hop manner with consideration of

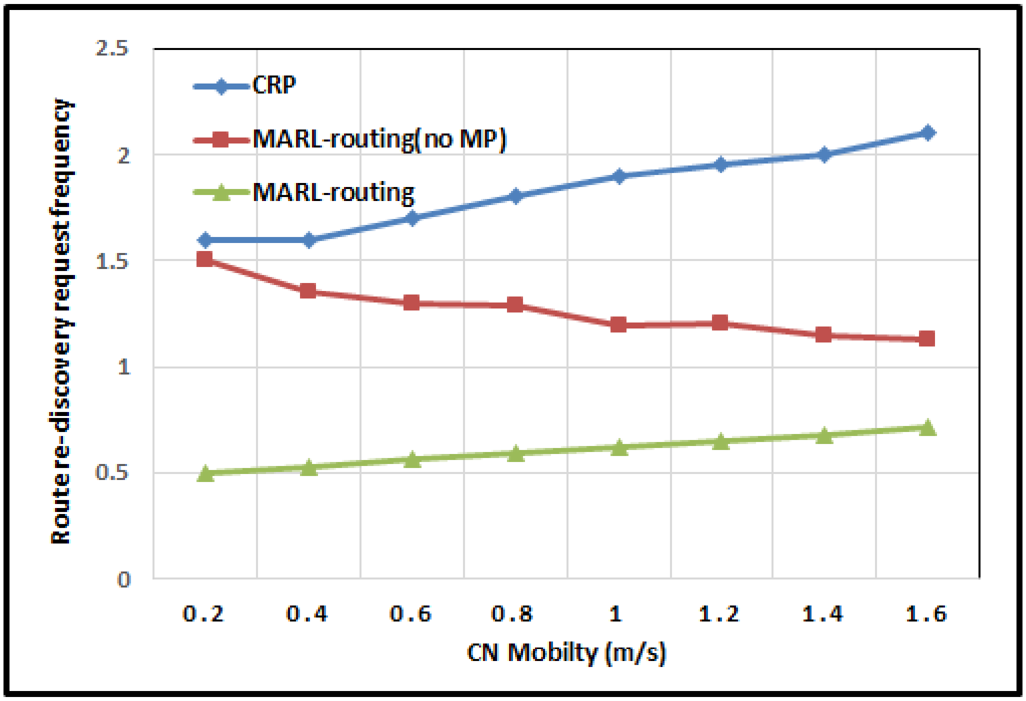


Figure 7.4: Performance Evaluation - Route Re-discovery Request Frequency

environmental statistics. The major advantage of MARL routing lies in the simplicity of route re-discovery. It has proactive and reactive components. During the proactive phase, before selecting any forwarding agent, all agents communicate with each other for understanding neighbouring node's interest to participate in the routing process. Mobility pattern of the cognitive user is used to predict the nodes movement towards the primary user region. Any failure due to local PU activity affects the link selection between any two intermediate nodes. This results in very small amount of route rediscovery overhead. MARL routing with mobility model outperform CRP and in the long run, its routing performance gradually increases as the neighbour behavior is learned by cognitive agent.

The normalized throughput of MARL routing with mobility prediction and without mobility prediction is shown in Figure 7.5. It has been observed that throughput of MARL routing with mobility prediction is better compared to without mobility prediction. If the cognitive nodes are moving towards the PU region, the probability of link availability decreases significantly as the interference is increased to the

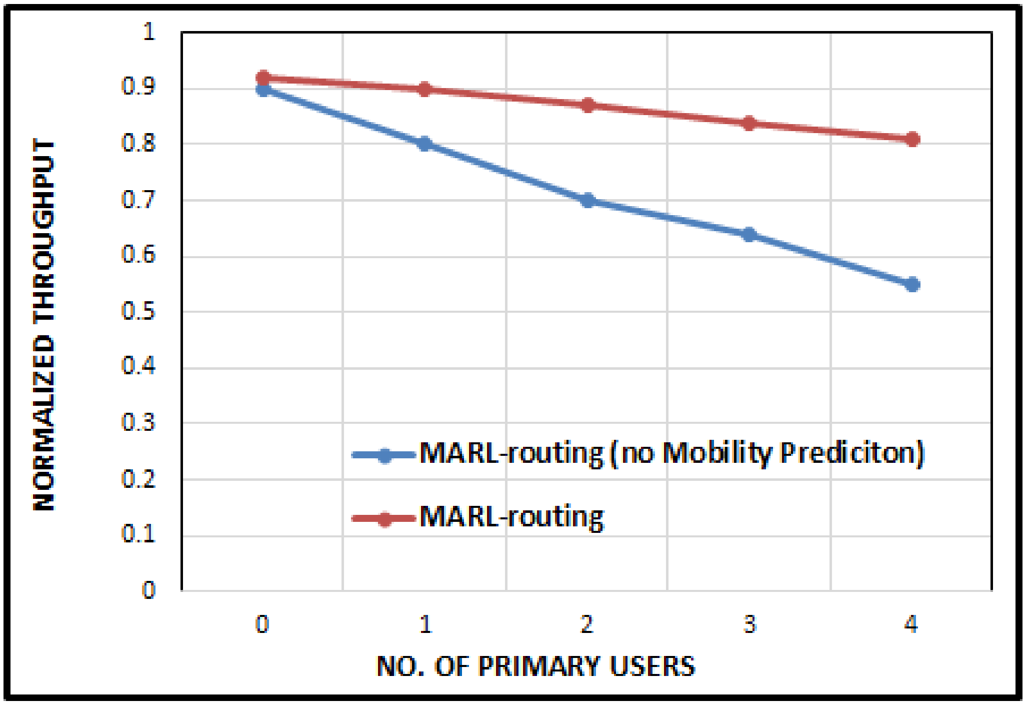


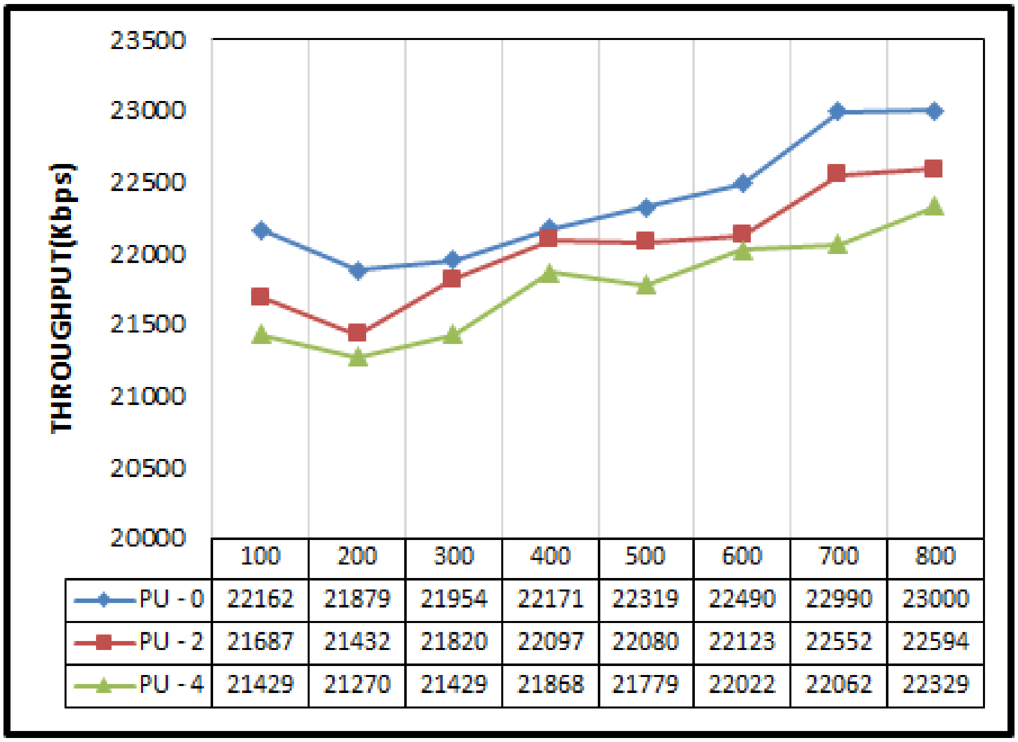
Figure 7.5: Effect of Primary User - Mobility Prediction

primary network.

Selecting links and relay nodes without increasing interference to the primary users is the most important objective of the MARL based routing in MCRAN. An increased number of primary users in the particular region require routing with utmost care by selecting link and relay opportunistically with minimum power and shorter distance affecting end-to-end performance of MCRAN. As the number of primary user increases, routing paths traverse through multiple smaller distance hops and with minimum power, affecting throughput and time required to transmit information. As shown in Figure 7.6, performance gain with no active primary user is more as compared with two or more active primary users.

7.3.2 Effect of Value Function Prediction

The value function is used for deciding relay as a forwarding agent among neighbouring agents. The MARL based routing is considered as the problem of ranking neighbouring agents based on estimating value function.



	100	200	300	400	500	600	700	800
◆ PU - 0	22162	21879	21954	22171	22319	22490	22990	23000
■ PU - 2	21687	21432	21820	22097	22080	22123	22552	22594
▲ PU - 4	21429	21270	21429	21868	21779	22022	22062	22329

Figure 7.6: Effect of Primary User - No. of PU and Throughput

The value of every state representing neighbouring agent is defined as expected return if the particular state is selected. Estimation of state value function is computed using an average or maximum of multiple independent realizations of state selection. This is Monte-Carlo method of estimation resulting in poor quality of the estimate when the variance of the return is high. Moreover, estimating when interacting with system is impossible without introducing some additional bias.

Temporal Difference is the most significant idea of reinforcement learning that addresses this issue and deals with the dynamic environment. Using bootstrapping every prediction is used as target during the course of learning.

Value of the state $V_t(s)$ is the estimate of state s at time t. At every t^{th} step TD(0) update value function as per following computation,

$$\delta_{t+1} = R_{t+1} + \gamma V_t(S_{t+1}) - V_t(S_t),$$

$$V_{t+1}(s) = V_t(s) + \alpha_t \delta_{t+1},$$

$$s \in S$$

α is a step-size parameter, $\alpha_t : t \geq 0$, is non-negative no. chosen to decide step-size. Each update in value function is proportional to temporal differences in predictions.

The estimation using Monte-carlo method is,

$$target[S_t] \leftarrow max$$

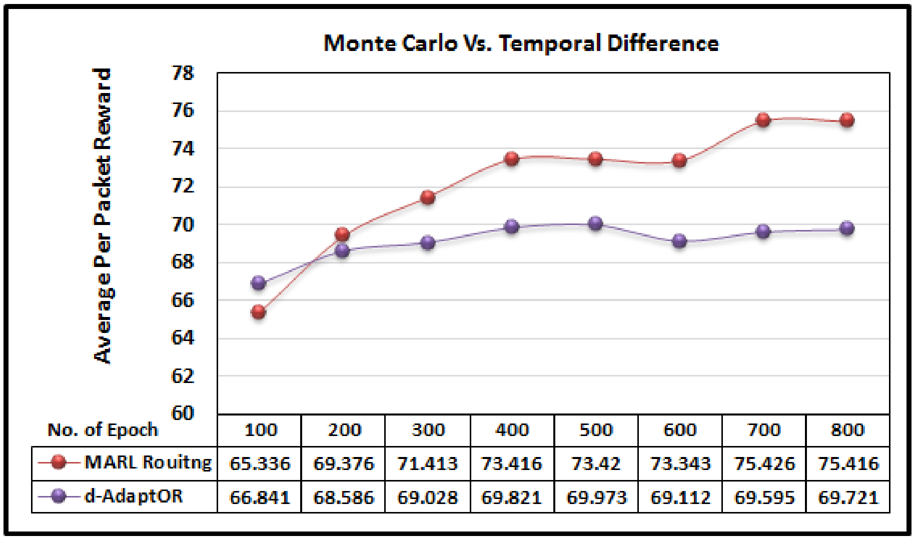
$$V[S_t] \leftarrow V[S_t] + \alpha \cdot (target[S_t] - V[S_t])$$

Thus, updating estimate at every t^{th} step using Monte-Carlo method is,

$$V_{t+1}(s) = V_t(s) + \alpha_t(R_t - V_t(s))$$

where, expected return represented as R_t is computed as,

$$R_t = r_{t+1} + r_{t+2} + r_{t+3} + \cdots + r_T$$



No. of Epoch	100	200	300	400	500	600	700	800
● MARL Rouitng	65.336	69.376	71.413	73.416	73.42	73.343	75.426	75.416
● d-AdaptOR	66.841	68.586	69.028	69.821	69.973	69.112	69.595	69.721

Figure 7.7: Monte Carlo and Temporal Difference Policy Evolution

As shown in Figure 7.7, the Average Per Packet Reward (APR) for P number of packets is calculated using Monte Carlo and Temporal Difference learning. MARL routing is implemented using Temporal difference and d-AdaptOR is implemented using Monte Carlo. Average per packet reward is representing the cost required to travel the whole path from source to destination.

$$APR = \frac{1}{P}\left(\sum_{i=0}^{P}\left(\Re - \sum_{j=0}^{h} lc_m^j\right)\right)$$

APR is computed after the packet is successfully delivered at the destination. Therefore, it learns from transmission success probabilities to find the optimal path from source to destination.

The value function of Monte-Carlo prediction does not represent the actual environmental statistics and successive updates in agents behavior. This affects the selection of neighbouring node as forwarding agent which results in increased cost for transferring packet from source to destination keeping APR low. Temporal difference learns optimal behavior from every interaction and update with current statistics. This results in increased reward over the multiple runs compared to Monte-Carlo. Temporal Difference bootstraps and understand the context to learn behavior for dealing with the dynamic environment.

7.3.3 Effect of Reinforcement Learning Parameters

Temporal Difference methods is implemented in online and fully incremented fashion and thus holds an obvious advantage over Monte Carlo method. TD method makes its update at every time step. It learns from each action selection and transition.

Learning Rate

As TD learns with every state transitions, therefore it is necessary to define the learning rate. Step-size parameter α is used to decide learning rate. As the reward function is random and dependent on environmental statistics, α should change over time. MARL routing based on Temporal Difference learning uses vector *Candidate Best Relay, CBR*, to decide learning rate. *CBR* is updated whenever any agent is selected as a best relay node to forward the packet. The learning rate α is proportionally updated with the value of *CBR* of every agent.

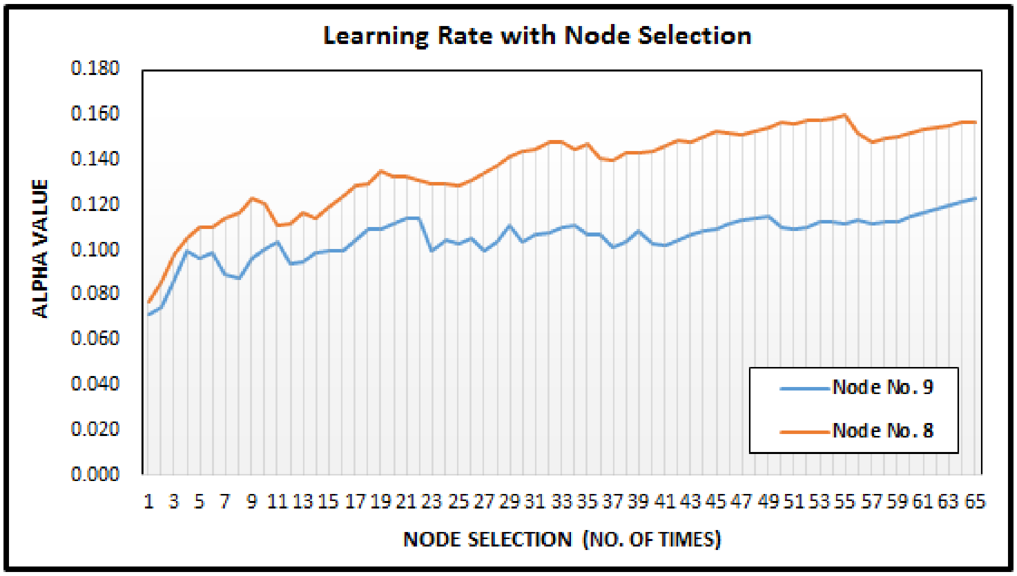


Figure 7.8: Learning Rate with Agent Selection

Suppose, for example, state value prediction $v(s)$ of state s is 8 and with α having value $\frac{1}{2}$, state value decreases to only 4. Reward at every step in dynamic and uncertain environment of MCRAN is different as per the resource availability. Therefore, the learning rate is not fixed and changes over time with every action/agent selection as a relay node. The learning rate α is proportionally increased for node which had contributed successfully to work as the relay in forming the complete path.

Figure 7.8 shows, selection of node no. 8 and 9 as candidate best relay agent among all neighbouring agents. As the score of Candidate Best Relay CBR for both the nodes is increased, the learning rate is also increased proportionally. The value of α is more for the agent which has participated successfully as the relay node more number of times. Therefore, the value of α is slowly improved as the experience about particular node is increased with its selection over multiple runs.

7.3.4 Effect of Action Selection Strategies

To maximize the performance and to produce greater total reward over long run, selection of the action i.e selection of proper forwarding node as the relay is very important.

Selecting neighbouring node with the greatest state action value score is exploiting current knowledge of values of action. This is beneficial in single play or run. On the other hand, selecting non greedy action is an exploration of action values enabling to improve estimate of non-greedy actions. This produces greater reward over the long run.

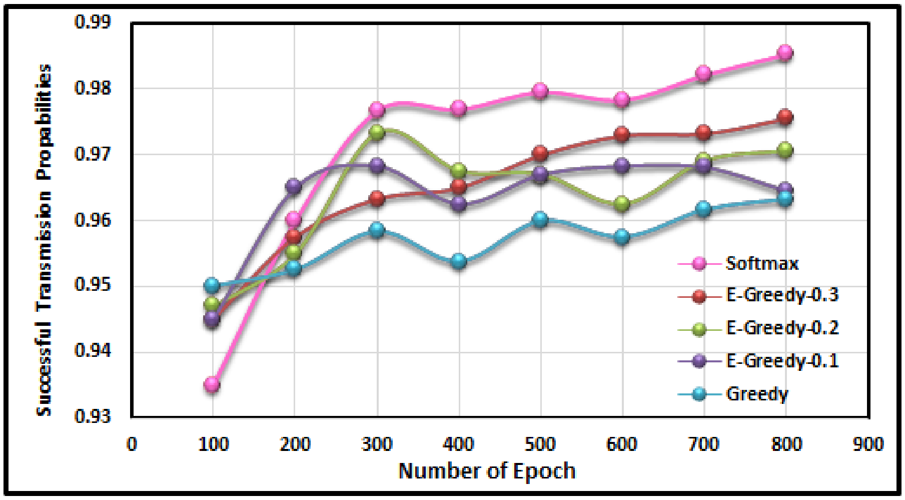
For relay node selection case, whether to explore or exploit is depends upon uncertainties and the number of remaining runs. MARL based routing achieves balancing of exploration and exploitation by using softmax action selection rule. Softmax action selection have various advantages over greedy and ϵ-greedy action value selection:

- Enable to improve state value estimate of non-greedy action i.e state value of all neighbouring nodes.
- Minimize the probability of selecting only one particular relay node every time resulting in balanced action selection and load.
- Vary the action selection probability with uncertainties and number of episodes.

Softmax selection is dependent on the positive parameter τ called *temperature* to decide exploration or exploitation. High value of τ makes all actions euqi-probable and gradual decrease in value of τ, has the difference in selection probability. With τ reaching 0 that is $\tau \rightarrow 0$, action selection using softmax is same as greedy.

Successful Transmission Probabilities

Successful transmission probabilities with two active primary users and 0.2m/s mobility speed of cognitive nodes is shown in Figure 7.9. The performance of softmax selection increases gradually as the episodes are increased. Softmax action selection outperforms greedy and ϵ-greedy action value selection. In the initial run, softmax



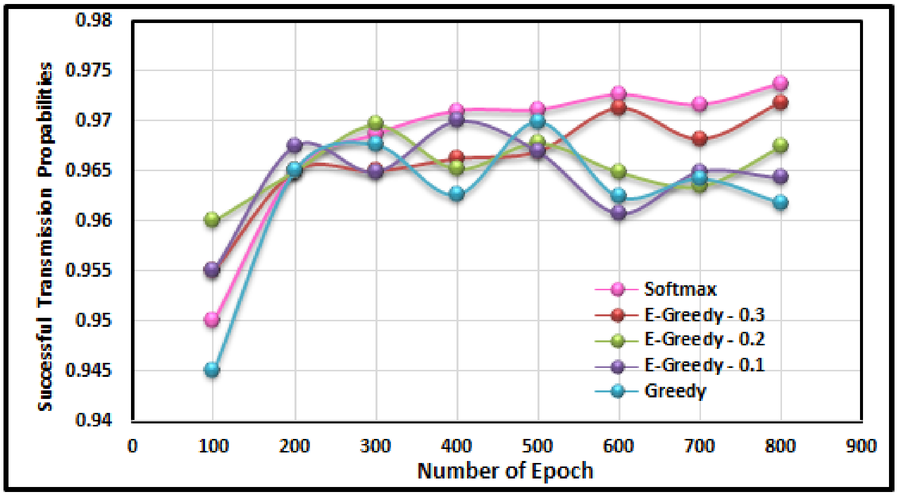
Action Selection	100	200	300	400	500	600	700	800
Softmax	0.935	0.96	0.97665	0.97685	0.9795	0.9783	0.9821	0.98525
E-Greedy-0.3	0.945	0.9575	0.9633	0.965	0.97	0.97295	0.97325	0.97556
E-Greedy-0.2	0.947	0.955	0.9733	0.9675	0.967	0.9625	0.9691	0.9706
E-Greedy-0.1	0.945	0.965	0.9683	0.9625	0.967	0.9683	0.9681	0.9645
Greedy	0.95	0.9525	0.9583	0.95375	0.95995	0.95745	0.96164	0.96325

Figure 7.9: Successful Transmission Probabilities with 2 PU and Mobility 0.2 m/s

action selection selects non-greedy actions resulting in lower successful transmission probabilities. As the number of the epoch are increased and the value of τ is gradually decreased, softmax action selection learns optimal actions which result in increased transmission probabilities.

Improvement with the greedy method compared to other methods is slightly faster at the very beginning, but levels off with minimum value, as greedy does not explore various opportunities. As shown in graph, the greedy method finds optimal actions in only one-third of the task and disappoints for remaining two-third of the play. The greedy method performs significantly poor in the long run as it does not explore and stuck to perform suboptimal actions. The ϵ-greedy method continuously explore to recognize optimal actions and perform better. The $\epsilon = 0.3$ method explores more and finds the optimal action earlier compared to $\epsilon = 0.2$ and $\epsilon = 0.1$.

Effect of action selection methods on transmission probabilities with increased mobility of CN is investigated and represented in Figure 7.10. As the mobility increases, the state value estimate is affected due to uncertainty about the actual location of



Action Selection	100	200	300	400	500	600	700	800
Softmax	0.95	0.965	0.96865	0.971	0.9711	0.9726	0.9716	0.9737
E-Greedy - 0.3	0.955	0.965	0.965	0.96625	0.96695	0.9713	0.9682	0.97185
E-Greedy - 0.2	0.96	0.965	0.96965	0.96525	0.96785	0.96495	0.9635	0.96745
E-Greedy - 0.1	0.955	0.9675	0.96495	0.97	0.967	0.9608	0.96495	0.96435
Greedy	0.945	0.965	0.96765	0.96265	0.96995	0.96245	0.96425	0.96185

Figure 7.10: Successful Transmission Probabilities with 2 PU and Mobility 1.6 m/s

CN. This affects the action selection, results in greater variance in transmission probabilities of the greedy method. Softmax Action selection slowly learns the behavior of mobile multiple CN agents to select action resulting in greater successful transmission probability.

Average Per Packet Reward

The overall performance of the MARL based routing protocol is measured with the help of Average Per Packet Reward, APR. APR represents the quality of links selected between every hop and number of intermediate nodes used to reach the destination. Exploration and exploitation using different action selection methods have a significant impact on the APR. As shown in Figure 7.11, APR received over P number of packets, which are greater for softmax action selection compared to greedy and ϵ-greedy methods.

With the reception of every packet at the destination agent, it receives the fixed reward. The transmission cost from source to destination is subtracted from APR for

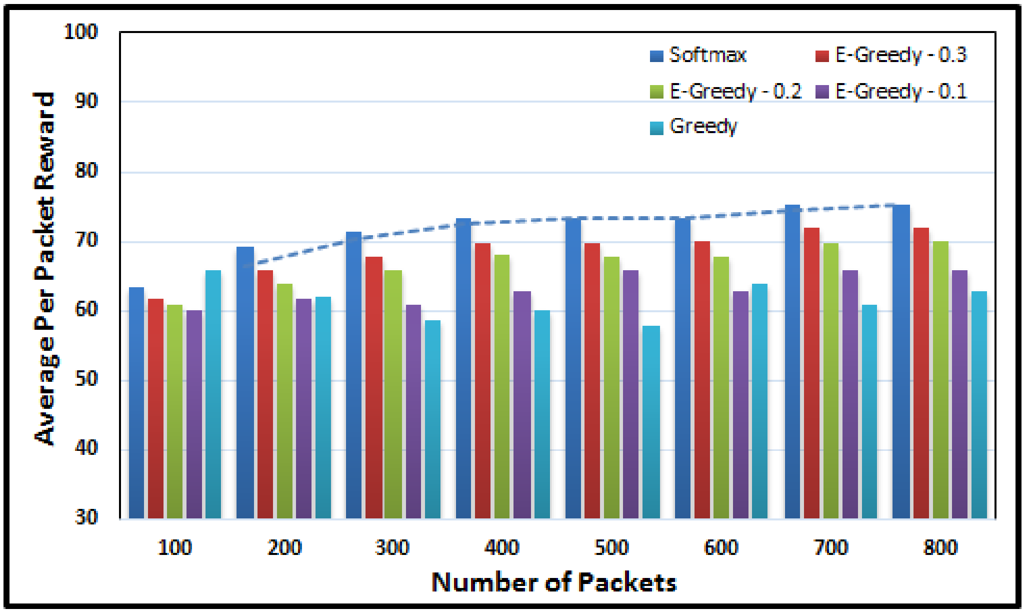


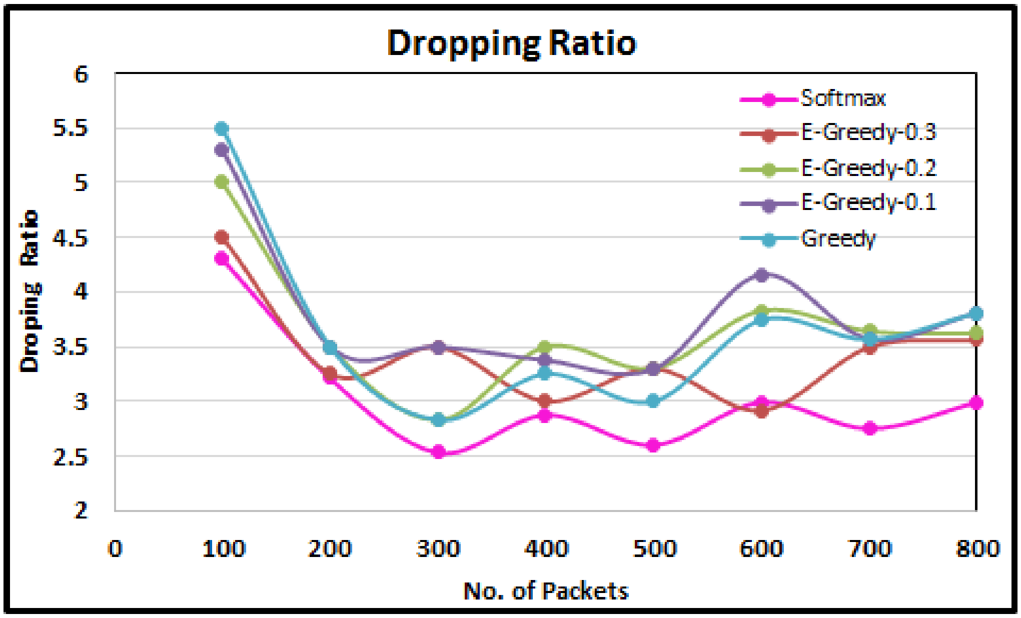
Figure 7.11: Performance Comparison of Action Selection

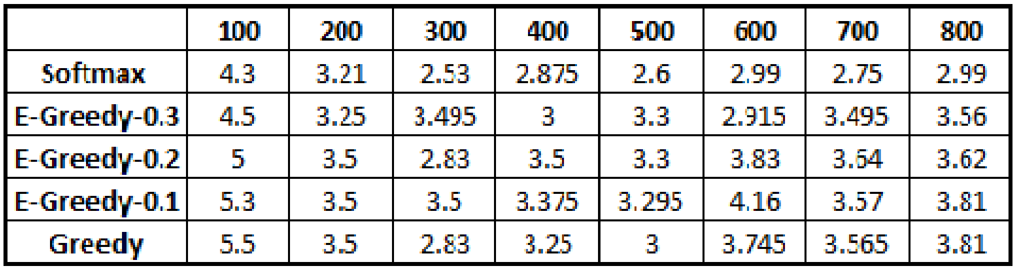
every hop by hop transmission. The maximizing reward is selection of neighbouring agent as advancement towards destination requiring the minimum number of hops and stable, available and reliable links at every hop. As the softmax action selection method slowly learns the behavior of neighbouring agent, it optimally selects the actions.

Packet Drop

Packet drop occurs mainly due to network congestion. If packets arrive at any network element at a rate greater than the rate to send through, then there are huge chances of packet dropping. Packet dropping in MCRAN is mainly due to unavailability of the channel and node moving towards the primary user region. If the particular agent is selected to forward the packet and it is unable to do so due to channel unavailability then there is packet drop.

Figure 7.12 shows reduced packet drop with softmax action selection compared with other action selection strategies. Packet drop with softmax action selection is reduced, as it learns the optimal route over the multiple runs. By exploring action values of multiple agents and learning using vectors candidate best relay, the selection





	100	200	300	400	500	600	700	800
Softmax	4.3	3.21	2.53	2.875	2.6	2.99	2.75	2.99
E-Greedy-0.3	4.5	3.25	3.495	3	3.3	2.915	3.495	3.56
E-Greedy-0.2	5	3.5	2.83	3.5	3.3	3.83	3.64	3.62
E-Greedy-0.1	5.3	3.5	3.5	3.375	3.295	4.16	3.57	3.81
Greedy	5.5	3.5	2.83	3.25	3	3.745	3.565	3.81

Figure 7.12: Action Selection Effect on Packet Dropping

of optimal path with best candidate forwarding node reduces the packet dropping ratio.

Initial runs have a greater error in estimation which is considered as the cost of exploration. The softmax action selection improves gradually over gathered experience. Figure 7.13 shows the error in estimation for three action selection strategies. Greedy action selection has low error in estimation for initial epoch but levels off with the highest probability of error in estimation due to lack of exploration. ϵ-greedy is showing better results compared to greedy action selection. ϵ-greedy learns slowly and outperforms in the long run. Softmax action selection is fast, achieves better results and decreases errors in estimation to best possible level.

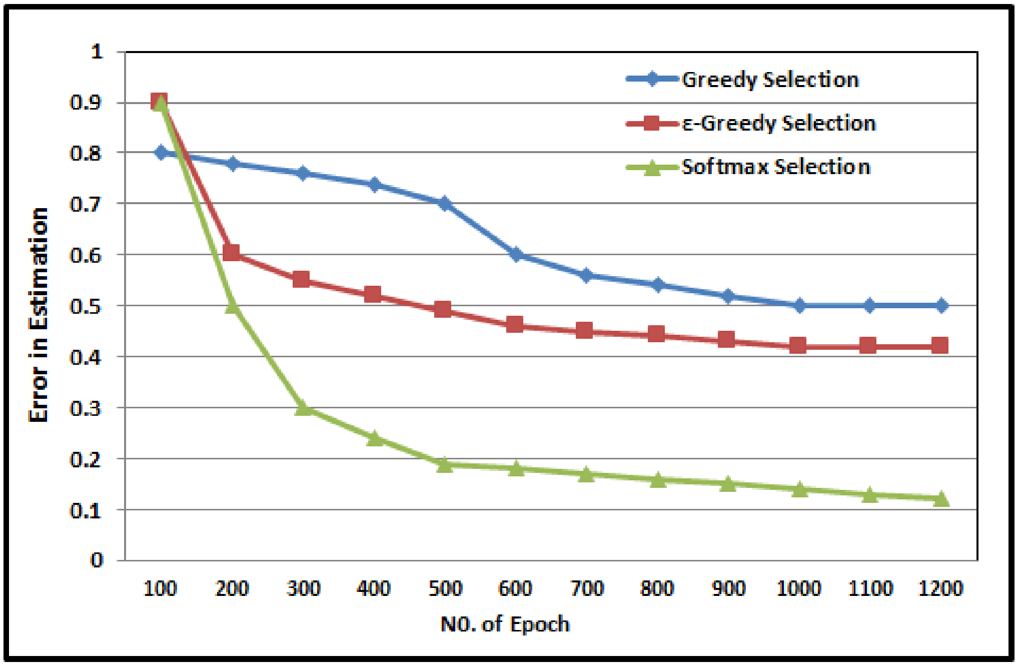


Figure 7.13: Error in Estimation with Action Selection

7.4 Computational Issues

The computational complexity and control overhead of MARL based opportunistic routing for MCRAN is low compared to other heuristic algorithm. The objective of Temporal Difference learning is that to consider correlation between successive predictions for every action selection. TD adjust the prediction so as to match future prediction in more accurate manner. It is applied for relay decision making of predicting the expected long-term reward in dynamic stochastic control environment. Reward function is the function of state (all neighbours), enabling ranking of alternative states for guiding decision making.

Temporal-difference learning is simple and elegant but significant sophistication is required for rigorous analysis of its behavior.

1. ***Complexity:*** The selection of the available neighbouring node based on routing score vector maintained for every neighbour is given in equation no. 5.15,

$$V_t^n\left(s,a\right) \leftarrow V_t^n\left(s,a\right) + \alpha_{CBR_t^n}\left[r_{t+1} + \gamma V_{t+1}^n\left(s,a\right) - V_t^n\left(s,a\right)\right]$$

Where, V represent the routing score for the agent s for selecting the node n as relay using action a. Routing score is updated on every successful completion of relaying. The value estimation of V is influenced by past experiences. Instead of keeping separate average value V, like Monte Carlo, Agent in TD methods maintains V as parametrized equation and adjust its parameter to better match the observed return.

To execute stochastic estimation $V_t^n(s,a)$, the no. of computation required per packet is of the order,

$$O(\max_{s \in S} |N_s)$$

Where, N_s is set of neighbours of source s. The computed value V is dependent on α, step-size parameters and the setting for α is influenced by CBR_t^n. Candidate Best Relay CBR_t^n., is counting variable maintained for every neighbour, representing how many time node n is selected as candidate relay and it had successfully completed transmission of packet.

Every cognitive agent maintains the three vectors for understanding the behavior of the neighbouring node. These vectors are:

- *Willing neighbour $WN\,(s, Ns)$:* Incremented for every agent in N_s who is willing to participate.
- *Candidate Best Relay* (CBR_t^n) : Incremented after selecting one of the neighbouring node n as relay for forwarding packets towards destination.
- *Routing Score Vector $V_t^n\,(s,a)$* : Updated every time when routing decision $a \in A(s)$, for neighbour n has been made up to time t for the source s.

The value Space complexity is proportional to no. of neighbours for each node for storing status information,

$$O(\max_{s \in S} |3 * N_s),$$

2. ***Control Overhead:*** The number of available neighbouring nodes decides num-

ber of relay requests and reply packets. Therefore, control overhead is order of $O(\max_{s \in S} |N_s)$, independent of network size.

3. ***Exploration Overhead:*** Optimal performance of the routing in MCRAN is supported by online opportunistic random routing strategy. The exploration overhead is measured by a number of packets deviated from optimal path. As the number of packets increases, the agent learns optimal policy for selecting best and reliable relay node as next forwarding node. The increased number of packets in long run results in diminished exploration cost.

7.5 Concluding Remarks

The overall aim of the chapter is to analyze the performance of the MARL based routing with Temporal Difference policy evolution in MCRAN. Implementation details and performance of state-of-of-the-art techniques are discussed. The implementation details and network parameters for realistic wireless setting are given with simulation parameters. The performance of the MARL based routing algorithm is evaluated with respect to network parameters, value function prediction, reinforcement learning parameters and action selection strategies. As per the performance evaluation:

- Proposed MARL based online opportunistic routing outperforms the CRP routing to avoid collision with primary users and minimizes the route failures.

- Temporal difference policy evolution successfully deals with the uncertainties due to the mobility of cognitive nodes.

- Proposed algorithm also outperforms d-Adapt routing to deal with the dynamic environment, as it learns directly from every interaction.

- Average per packet reward with proposed algorithm is better compared with Monte Carlo, as it selects every relay with respect to current state of the environment.

- Candidate best relay selection is driven by the agents experience to work as the relay. The learning rate α is proportionally updated with the experience of every agent to explore opportunities and exploit current knowledge of the environment.

- Softmax Action Selection balances the exploration and exploitation by gradually learning the behavior of multiple agents.

- Successful transmission probabilities with softmax action selection are better compared with greedy and ϵ-greedy action selection.

- Gathered experience over a time also reduces the probability of error in estimation resulting in reduced packet drop.

- The computational complexity and control overhead of MARL based opportunistic routing for MCRAN is low compared to other heuristic algorithms.

Chapter 8

Discussion and Summary

Chapter 8

Discussion and Summary

Multi-agent reinforcement learning based online and opportunistic routing for mobile cognitive radio ad hoc network was explored in this thesis. The novel context aware routing algorithm is proposed to explore intermittent spectrum and routing opportunities in the dynamic mobile cognitive radio ad-hoc network. The complete thesis is presented to meet the research challenges given in section 1.6. The proposed algorithm addresses the following research challenges:

- **Spectrum Awareness**: Tight coupling between the spectrum management module and routing module is necessary for designing an efficient routing solution. Therefore, the implementation of multi-agent reinforcement learning based routing involves two stages, link selection stage and online opportunistic route formation stage. Selecting a link on every hop with the characteristics of good propagation distance, minimum expected transmission time, more channel availability and duration results in stable links with more forwarding distance towards the destination.

  The proposed approach makes the cognitive node be continuously aware of the surrounding physical environment for spectrum and relay node availability. It is highly coupled with the entire cognitive cycle to take more accurate decisions. This is achieved by observing real time channel behavior and its usage. Wi-Fi channels having frequency range of 2.4 GHz belonging to IEEE 802.11n are observed and analyzed. Three channels, channel no. 1 (2.412 GHz), channel no. 6 (2.437 GHz) and channel no. 11 (2.462 GHz) are continuously scanned using N210 universal software radio peripheral and GNU radio to make RF measurement and to generate time samples. These time samples of RF measurements are used to observe primary user activities and find the hidden model in channel usage statistics using the Hidden Markov Model.

- **Exchanging Routing Information**: Achieving optimum performance with intermittent spectrum and neighbouring relay availability is impossible without cooperation among multiple cognitive nodes/agents. The proposed algorithm is designed as a multi agent cooperative task to improve the utilization of spectrum and reduces spectrum scarcity. In the mobile and dynamic environment of MCRAN, the overall view of the current topology of a network is created by means of information exchange among multiple cognitive agents.

  Co-operation among multiple agents is achieved through perceptual mechanism by including neighbouring agents as multiple possible states for transferring packets. The problem of routing is considered as selecting optimal path with a minimum number of state transitions from a source to destination with stable and reliable links. Every state transitions or relay node selection is initiated by exchanging the link availability, node's ability and willingness to participate in the route formation process. The Multi-Agent Reinforcement Learning allows nodes/agents to make a route formation decision in distributed manner and adaptive to the context of environmental statistics. This information exchange among all agents makes them adaptive to the dynamic environment of MCRAN.

- **Dynamic Topology and Intermittent Connectivity**: Prediction based node and relay selection is implemented for creation of more stable paths. The appearance of the primary user and mobility of the cognitive nodes significantly affect connectivity. There are rapid changes in the topology of reachable neighbouring cognitive nodes.

  The movement of cognitive node towards the primary user region, significantly affects the spectrum and neighbouring relay node availability. The Random Gauss Markov Model is used to predict the movement of the cognitive node towards the primary region. The node moving away from PU's region has more preference as the probability of link availability is more compared with the node moving toward the primary user region. Thus, link selection for relaying information is influenced by the mobility pattern, link characteristics and its availability.

Dynamically changing topology necessitates that the routing paths should be decided in an online and opportunistic manner. The proposed algorithm considers no knowledge of topology and constructs the path dynamically using interaction with environment. Current observations of the environment, historical knowledge of node's behavior and exploring spectrum opportunities make MARL agent meet the challenges of dynamic topology and intermittent spectrum availability.

- **Route Maintenance**: As the route formation in MCRAN using MARL based routing is adaptive, online and opportunistic, there is less requirement for route maintenance. Before selection of the link between any two nodes, every node predicts its channel availability. It is evaluated using similarity of the actual observation sequence and predicted observation sequence.

  The proposed algorithm gives more importance to the cognitive agent moving away from the primary user region rather than towards the primary user region. This is achieved using the Random Gauss Markov Mobility Model. It was used to describe object trajectory as it could capture the correlation of object velocity in time. This mobility model attempts to mimic the movements of the real mobile node by allowing past velocities and directions to influence the future velocities and directions.

  Instead of finding a fixed path from source to destination as in normal wireless network, the proposed algorithm finds the routing path in hop by hop manner using multi-agent reinforcement learning. This is implemented with the objective of maximizing the average per packet reward having minimum cost link and minimum number of hops on the entire path.

Predicting channel availability, modeling nodes mobility and hop by hop route formation reduces route maintenance. Any route failure in MARL based routing affects the connection between only two nodes. This can be immediately addressed by re-selecting the relay node locally.

- **Cognitive Users Behavior**: Routing in MCRAN is designed to continuously observe and analyze the behavior of neighbouring agents using multi-agent reinforcement learning. All neighbouring agents are considered as available different states for relay requesting agent. Every time selects one of the state using softmax action selection. Each state transition is done with current context and stored information about the neighbouring agent.

  The information maintained by every node contains previous cooperation of multiple agents in the route formation process. Every node maintains list of the neighbouring agents, who had worked as a relay for forwarding packets. Every node also maintains list of willing neighbours, who had shown interest to work as a relay. Value function was used to capture dynamics of the environment and to understand the neighbouring agent's ability to participate in the routing process. Every update in the *Value Function* of state represents Temporal Differences in state transitions.

  Strategic interaction among multiple agents and collected information of neighbours help to understand the agent's ability to participate in the route formation. It results in higher spectrum utilization and reliable connectivity at every hop along the path. Each cognitive node explicitly considers other cognitive nodes and coordinates their behavior with each other to increase the reward of coherent actions.

8.1 Performance of Channel/Link Selection

Selecting stable, reliable and non-interfering channel with primary user is important criteria for designing spectrum aware routing solution in MCRAN. MARL online and opportunistic routing is divided into two important phases, Link Selection and Relay Selection. Link selection phase in the proposed algorithm considers different characteristics of the link to decide suitable links during transmission. The characteristics of the link selection process in MCRAN are:

1. Link selection is based on the local environmental observation and other important link characteristics like Channel Availability, Spectrum Propagation, PU Protection, Expected Time to Transmit and Channel Switching Cost.
2. Random Gauss Markov Mobility Model is used to describe the object trajectory to capture the correlation of object velocity in time.
3. Use of the mobility model accurately represents movement of the CN towards or away from the primary user region affecting the spectrum and relay availability.
4. End-to-end latency of routing is improved by selecting a channel with good propagation characteristics.
5. The selected channel takes minimum possible amount of time to transmit a link layer frame, represented by the Expected Time to Transmit (ETT).
6. Predicted link availability is the probability that the link remains available for a particular transmission time requirement.
7. Link selection decision is taken on very important characteristics that is channel availability.
8. Channel availability is the average amount of time when a channel is available for transmission.
9. Channel availability is computed on the actual and generated observation sequence using Hidden Markov Model.

10. HMM successfully characterizes the observed channel occupancy of the primary user to better understand the channel usage and primary user behavior.

11. RF signal modeled using HMM is potentially capable of learning about signal source i.e primary user without its information and availability.

12. Model of training HMM is completely online and repeated, suitable for dynamic spectrum availability in the cognitive radio network.

13. HMM is trained successfully with a limited number of parameters and moderate data set.

14. HMM continuously improves itself to maximize the likelihood estimation of the actual and generated observation sequence.

15. Actual observation sequence used for prediction is RF measurement of real world signal sensed using RF hardware, Universal Software Peripheral N210.

16. Free and open source GNU Radio is used for digital signal processing for implementing Software Defined Radio.

17. RF measurement samples collected using USRP and GNU radio are used to train HMM for predicting future channel availability.

18. HMM perfectly predicts the future channel availability using the generated observation sequence with maximum likelihood estimation.

8.2 Performance of MARL Routing

MARL based adaptive, online and opportunistic routing was implemented successfully using exploration of successful transmission probabilities in MCRAN. It is compared with the state-of-the-art routing protocols like Cognitive Routing Protocol (CRP) and distributed adaptive routing protocol. MARL based routing achieves good performance benefits without the knowledge of topology and channel availability. The

MARL based adaptive, online and opportunistic routing has following advantages over the state-of-the-art techniques:

1. The performance of online opportunistic routing in uncertain and dynamic wireless environment is improved using Reinforcement Learning.

2. RL learns optimal policy on-line from direct interaction with the environment for solving the multi-step dynamic decision making problem.

3. The agent improves spectrum utilization and network performance using Temporal Difference policy evolution.

4. Temporal difference reinforcement learning deals successfully with dynamic environment as it learns from every interaction and decision making.

5. Policy evolution using TD is completely online and incremental to be adaptive to the uncertainties in MCRAN.

6. TD is a bootstrapping method which learns from previous actions and converges faster than Monte Carlo methods on stochastic task.

7. Average per-packet reward of TD is more over the multiple runs compared to Monte Carlo method.

8. Every decision of routing is done online and in an opportunistic manner by selecting the most prominent relay node as a forwarding node according to its routing score, location and spectrum availability.

9. MARL based online opportunistic routing lowers the collision to primary user and interference generated through CN operation as every agent selects the next forwarding agent for the packet as per the current environment conditions and interaction among the multiple neighbouring agents.

10. MARL routing finds route in hop-by-hop manner with consideration of environmental statistics reducing route re-discovery.

11. Packet drop in MARL based routing is reduced, as it learns the optimal route over multiple runs.

12. Error in estimation improves over gathered experience using TD policy evolution and softmax action selection.

13. Mobility prediction using Random Guass Markov model improves the throughput as it predicts the node's mobility moving towards or away from the primary user region.

14. The learning rate α is proportionally updated with the value of *Candidate Best Relay* of every agent results in selecting more stable and reliable path.

15. The softmax action selection rule is used for balancing the exploration and exploitation.

16. The softmax action selection rule explores non-greedy actions to improve their value function compared to greedy and ϵ-greedy action selection.

17. This results in balanced load among neighbouring agents for forwarding packets.

18. The softmax action selection varies the action selection probabilities with uncertainties and number of episodes.

19. It learns optimal actions as the number of epoch increases, it results in the increased successful transmission probabilities.

20. Action selection with the help of collected experience and cooperation among multiple agents successfully deal with the uncertainties of MCRAN.

21. The strategic interaction among multiple agents and learning over a time improves the spectrum utilization and end to end routing performance.

Chapter 9

Conclusion and Future Research Directions

Chapter 9

Conclusion and Future Research Directions

9.1 Conclusion

The multi-agent reinforcement learning based context aware routing algorithm is designed and implemented successfully to explore the spectrum and routing opportunities in the dynamic environment of mobile and ad hoc Cognitive Radio Network. In contrast to the conventional routing of finding a fixed path from source to destination, the proposed algorithm routes the packets online using strategic interaction among the cognitive nodes. It is designed to select stable, reliable and non-interfering channels with the primary user for providing the spectrum aware routing solution.

Temporal Difference implementation of MARL based routing achieves an optimal performance measured using average per packet reward. It achieves good quality of service requirement of cognitive users by decreasing the route re-discoveries. As per the results, relay selection using softmax action selection rule outperforms when compared with greedy and ϵ-greedy. It varies the action selection probabilities as a graded function of estimated values for balancing exploration and exploitation of network opportunities.

The proposed spectrum aware routing algorithm successfully deals with the dynamic environment by understanding context using Hidden Markov Model. Every cognitive node monitors the primary user activities by sensing signal strength of spectrum bands. The sensed signal observations are used to predict the chances of accessing spectrum band opportunistically. Cooperative and strategic interaction among multiple cognitive agents improve the end-to-end performance of route formation in MCRAN.

This thesis has achieved its overall goal, with respect to the **research objective** stated in chapter no. 3. The major **research contributions** of this work are:

1. **Literature review and the state-of-the-art techniques**: Existing literature is reviewed to provide context of the field and intellectual progression of cognitive radio network especially in spectrum aware routing. This is with the objective of finding current issues being debated, how they are addressed by the existing literature and limitations of the state-of-the-art techniques. The reviewed literature helped to decide research direction, important design considerations and methodological focus for efficient routing protocol.

2. **Explored use of reinforcement learning in dynamic environment and optimal strategies for agents**: Temporal difference implementation of reinforcement learning successfully deals with uncertainties of the dynamic environment. Temporal difference policy evolution enables every agent to find optimal policies in fully incremental and online manner. Every agent finds optimal strategies without model of the environment.

3. **Link selection metric for selecting stable and non-interfering links with primary user**: Stable, reliable and available link selection metrics are used to support quality of service requirement of the cognitive users while reducing interference with the primary user. This is achieved by predicting traffic on every channel using Hidden Markov Model. Predicted traffic using Hidden Markov Model is a maximum likelihood estimation of actual traffic on the same channel. The channel with better predicted availability is selected for transmission.

4. **Online opportunistic routing algorithm based on the transmission success probabilities**: Multi-Agent Reinforcement Learning based routing algorithm is successfully implemented to model strategic interaction among cognitive agents to cooperatively find hop-by-hop routing path from source to destination. The behavior of every neighbouring agent is observed to find candidate forwarding relays with the increased transmission success probabilities.

5. **Balanced exploration and exploitation using soft-max action selection in relay selection process**: The softmax action selection rule efficiently balances exploration and exploitation of network opportunities. The aim is to find the set of all neighbouring agents willing to participate in route formation process. By varying action selection probabilities as graded function it update the routing score of multiple neighbouring nodes. This helps to balance the load of forwarding packet among multiple agents and gradually forms optimal policy for selecting a stable path from source to destination.

9.2 Future Research Directions

The proliferating increase in the wireless application needs breakthrough radio technologies to fulfill the future demands, providing improved spectrum utilization and application performance. Cognitive radio enables pervasive wireless communication application for future wireless world.

The work described in this thesis will pave the way for future explorations in context aware cognitive radio network. Some of the future research directions are suggested as follows:

- Cognitive radio network will soon emerge as the general purpose programmable wireless network. The complex task of building and deploying cognitive radio nodes should be studied as the cross layer optimization problem. All layers in protocol stack adhere to the policy and objectives of cognitive radio network. Interaction among multiple layers and multiple nodes should be studied for successful implementation of dynamic spectrum access.

- The practical implementation of the cognitive radio network requires game theoretic approaches for dynamic spectrum sharing. Therefore, there is a need to study and design proper pay-off functions in non cooperative, economic and stochastic games.

- A combined study of the multi-agent reinforcement learning and game theory will be able to represent the practical possibilities of cognitive radio network. Strategic interaction among primary users and cognitive users for using temporarily unused spectrum holes helps for effective improvement in spectrum utilization.

- Policies can be represented independently from value function using actor-critic implementation of TD learning. Actor-critic method improves the performance by evaluating policy online to decide the correctness of previous actions.

- A natural extension of the multi-agent reinforcement learning can be achieved by including the a reward of another agent in state description. This helps to better understand the state of neighbouring agents and environment for upgrading action selection criteria.

Annexure

ANNEXURE

Laboratory Equipments

USRP N210

The Ettus Research USRP N210 are the highest performing class of hardware of the USRP (Universal Software Radio Peripheral) family of products, which enables engineers to rapidly design and implement powerful, flexible software radio systems. The N210 hardware is ideally suited for applications requiring high RF performance and great bandwidth. Such applications include physical layer prototyping, dynamic spectrum access and cognitive radio, spectrum monitoring, record and playback, and even networked sensor deployment. Features of USRP N210 are:

- Use with GNU Radio, LabVIEW and Simulink
- Modular Architecture: DC-6 GHz
- Dual 100 MS/s, 14-bit ADC
- Dual 400 MS/s, 16-bit DAC
- DDC/DUC with 25 MHz Resolution
- Up to 50 MS/s Gigabit Ethernet Streaming
- Fully-Coherent MIMO Capability
- Gigabit Ethernet Interface to Host

- 2 Gbps Expansion Interface
- Spartan 3A-DSP 3400 FPGA (N210)
- 1 MB High-Speed SRAM
- Auxiliary Analog and Digital I/O
- 2.5 ppm TCXO Frequency Reference
- 0.01 ppm w/ GPSDO Option

SBX USRP Daughter-board:

The SBX is a wide bandwidth transceiver that provides up to 100 mW of output power, and a typical noise figure of 5 dB. The local oscillators for the receive and transmit chains operate independently, which allows dual-band operation. Application of SBX USRP include WiFi, WiMax, S-band transceivers and 2.4 GHz ISM band transceivers. Features of SBX USRP daughter-board are:

- 40 MHz of bandwidth
- Access to a variety of bands in the 400 MHz-4400 MHz range. Example
- 2 quadrature front-ends (1 transmit, 1 receive)
- Transmit Gains: Range: 0-31.5dB
- Receive Gains: Range: 0-31.5dB

Ettus Research™
A National Instruments Company

USRP™ N200/N210 NETWORKED SERIES

FEATURES:

- Use with GNU Radio, LabVIEW™ and Simulink™
- Modular Architecture: DC-6 GHz
- Dual 100 MS/s, 14-bit ADC
- Dual 400 MS/s, 16-bit DAC
- DDC/DUC with 25 mHz Resolution
- Up to 50 MS/s Gigabit Ethernet Streaming
- Fully-Coherent MIMO Capability
- Gigabit Ethernet Interface to Host
- 2 Gbps Expansion Interface
- Spartan 3A-DSP 1800 FPGA (N200)
- Spartan 3A-DSP 3400 FPGA (N210)
- 1 MB High-Speed SRAM
- Auxiliary Analog and Digital I/O
- 2.5 ppm TCXO Frequency Reference
- 0.01 ppm w/ GPSDO Option

N200/N210 PRODUCT OVERVIEW:

The Ettus Research™ USRP™ N200 and N210 are the highest performing class of hardware of the USRP™ (Universal Software Radio Peripheral) family of products, which enables engineers to rapidly design and implement powerful, flexible software radio systems. The N200 and N210 hardware is ideally suited for applications requiring high RF performance and great bandwidth. Such applications include physical layer prototyping, dynamic spectrum access and cognitive radio, spectrum monitoring, record and playback, and even networked sensor deployment.

The Networked Series products offers MIMO capability with high bandwidth and dynamic range. The Gigabit Ethernet interface serves as the connection between the N200/N210 and the host computer. This enables the user to realize 50 MS/s of real-time bandwidth in the receive and transmit directions, simultaneously (full duplex).

The Networked Series MIMO connection is located on the front panel of each unit. Two Networked Series units may be connected to realize a complete 2x2 MIMO configuration using the optional MIMO cable. External PPS and reference inputs can also be used to create larger multi-channel systems. The N200 and N210 are largely the same, except that the N210 features a larger FPGA for customers that intend to integrate custom FPGA functionality.

The USRP Hardware Driver™ is the official driver for all Ettus Research products. The USRP Hardware Driver supports Linux, Mac OSX, Windows.

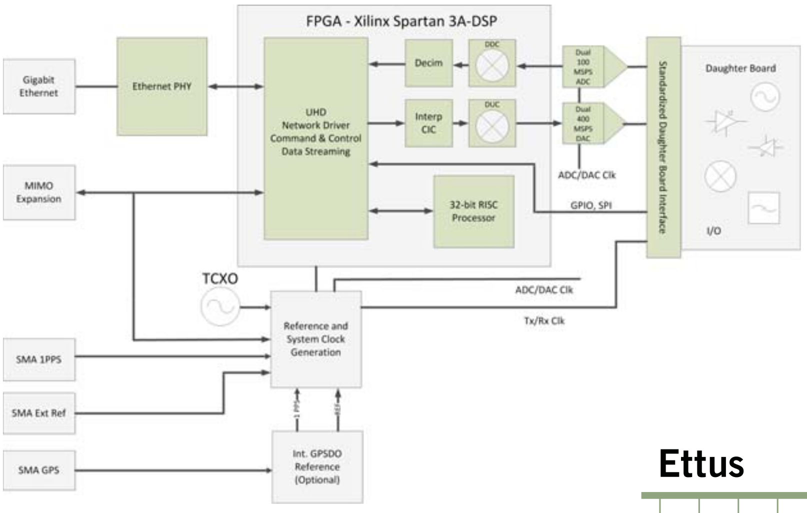
USRP™ N200/N210 NETWORKED SERIES

SPECIFICATIONS

Spec	Typ.	Unit
POWER		
DC Input	6	V
Current Consumption	1.3	A
w/ WBX Daughterboard	2.3	A
CONVERSION PERFORMANCE AND CLOCKS		
ADC Sample Rate	100	MS/s
ADC Resolution	14	bits
ADC Wideband SFDR	88	dBc
DAC Sample Rate	400	MS/s
DAC Resolution	16	bits
DAC Wideband SFDR	80	dBc
Host Sample Rate (8b/16b)	50/25	MS/s
Frequency Accuracy	2.5	ppm
w/ GPSDO Reference	0.01	ppm

Spec	Typ.	Unit
RF PERFORMANCE (W/ WBX)		
SSB/LO Suppression	35/50	dBc
Phase Noise (1.8 Ghz)		
10 kHz	-80	dBc/Hz
100 kHz	-100	dBc/Hz
1 MHz	-137	dBc/Hz
Power Output	15	dBm
IIP3	0	dBm
Receive Noise Figure	5	dB
PHYSICAL		
Operating Temperature	0 to 55°	C
Dimensions (l x w x h)	22 x 16 x 5	cm
Weight	1.2	kg

* All specifications are subject to change without notice.



Ettus Research™
A National Instruments Company

ABOUT ETTUS RESEARCH:

Ettus Research is an innovative provider of software defined radio hardware, including the original Universal Software Radio Peripheral (USRP) family of products. Ettus Research products maintain support from a variety of software frameworks, including GNU Radio. Ettus Research is a leader in the GNU Radio open-source community, and enables users worldwide to address a wide range of research, industry and defense applications. The company was founded in 2004 and is based in Mountain View, California. As of 2010, Ettus Research is a wholly owned subsidiary of National Instruments.

1043 North Shoreline Blvd
Suite 100
Mountain View, CA 94043

P 650.967.2870 www.ettus.com
F 866.807.9801

Bibliography

Bibliography

Abdelaziz Samar and ElNainay Mustafa (2014), “Metric-based taxonomy of routing protocols for cognitive radio ad hoc networks”, Elsevier Journal on Network and Computer Applications 40: 151-163.

Abedi O. and Berangi R.(2013), “Mobility assisted spectrum aware routing protocol for cognitive radio ad hoc networks”, Journal of Zhejiang University-SCIENCE C (Computers and Electronics) ISSN 1869-1951, 14(11): 873-886.

Abul O., Polat F. and Alhaj. R. (2000), “Multi-agent reinforcement learning using function approximation”, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 30(4): 485-497.

Adhau S., Mittal M. L. and Mittal A.(2012), “A multi-agent system for distributed multi-project scheduling: an auction-based negotiation approach”, Elsevier Journal on Engineering Applications of Artificial Intelligence 25(8): 1738-1751.

Akyildiz Ian F., Won-Yeon Lee, Mehmet C. Vuran and Shantidev Mohanty (2006), “Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey”, Elsevier Journal on Computer Network 50: 2127-2159.

Ariyakhajorn J., Wannawilai P. and Sathitwiriyawong C. (2006), “A comparative study of random waypoint and gauss-markov mobility models in the performance evaluation of MANET”, International Symposium on Communications and Information Technologies 18 to 20 Oct 2006: 894-899.

Badarneh Osamah S. and Salameh Haythem Bany (2012), “Probabilistic quality-aware routing in cognitive radio networks under dynamically varying spectrum opportunities", Elsevier Journal on Computers and Electrical Engineering 38(6): 1731-1744.

Badoi C. I., Croitoru V. and Prasad R. (2010), "IPSAG: an IP spectrum aware geographic routing algorithm proposal for multi-hop cognitive radio networks", IEEE 8th International Conference on Communications (COMM): 491-496.

Bang Jonghyun, Lee Jemin, Kim Seokjung and Hong Daesik (2015), "An efficient relay selection strategy for random cognitive relay network", IEEE Transaction on Wireless Communications 14(3): 1555-1566.

Barve Sunita S. and Kulkarni Parag (2014), "Multi-agent reinforcement learning based opportunistic routing and channel assignment for mobile cognitive radio ad hoc network", ACM Springer International Journal on Mobile Networks and Applications 19(6): 720-730.

Barve Sunita S. (2014), "Cognitive Radio Networks: A Tutorial", International Journal on Advanced Computing and Knowledge Discovery, 3(1): 32-37.

Barve Sunita S. and Kulkarni Parag (2012), "Dynamic Channel Selection and Routing Through Reinforcement Learning in Cognitive Radio Networks", In proceeding of IEEE International Conference on Computational Intelligence and Computing Research : 6-12.

Barve Sunita S. and Kulkarni Parag (2012), "A Performance based Routing Classification in Cognitive Radio Networks", International Journal of Computer Applications, 44(19): 11-21.

Beibei Wang, Yongle Wu and K.J Ray Liu (2010), "Game theory for Cognitive Radio networks: an Overview", Elsevier Journal on Computer Networks 54(14): 2537-2561.

Bhorkar Abhijeet, Naghshvar Mohammad, Javidi Tara and Rao Bhaskar (2012), "Adaptive opportunistic routing for wireless ad hoc networks", IEEE/ACM Transaction on Networking 20(1): 243-256.

Bourdena Athina, Mavromoustakisb Constandinos, Kormentzasa George, Pallis Evangelos, Mastorakis Georgec, Yassein Muneer Bani (2014), "A resource intensive traffic-aware scheme using energy-aware routing in cognitive radio networks", Elsevier Journal on Future Generation Computer Systems 39: 16-28.

Bowen Li, Dabai Li, Qi-hui Wu and Haiyuan Li (2009), "ASAR: Ant-based Spectrum Aware Routing for Cognitive Radio Network", International Conference on Wireless Communications and Signal Processing: 1-5.

Buracchini Enrico (2000), "The Software Radio Concept", IEEE Communications Magazine : 138-143.

Busoniu L., Babuska and R., Schutter B.D. (2008), "A comprehensive survey of multi-agent reinforcement learning", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 38(2): 156-172.

Cabric D., Mishra S.M. and Brodersen R. W. (2004), "Implementation issues in spectrum sensing for cognitive radios", IEEE conference on Signals, Systems and Computers 1: 772-776.

Cacciapuoti Angela Sara, Marcello Caleffi and Luigi Paura (2012), "Reactive routing for mobile cognitive radio ad hoc networks", Elsevier Journal on Ad-Hoc Networking 10: 803-815.

Cacciapuoti Angela Sara, Marcello Caleffi, Luigi Paura and Rahman A. (2015), "Channel availability for mobile cognitive radio networks", Elsevier Journal on Network and Computer Applications 47: 131-136.

Canberk Berk, Akyildiz Ian F., and Oktug Sema (2011), "Primary User Activity Modeling Using First-Difference Filter Clustering and Correlation in Cognitive Radio Networks", IEEE/ACM Transaction on Networking 19(1): 170-183.

Cesana, Matteo, Francesca Cuomo and Eylem Ekici (2011), "Routing in cognitive radio networks: challenges and solutions", Elsevier Journal on Ad Hoc Networking 9(3): 228-248.

Chaharsooghi S. K., Heydari J. and Hessameddin Zegordi S. H (2008), "A reinforcement learning model for supply chain ordering management: An application to the beer game", Elsevier Journal on Decision Support Systems 45: 949-959.

Chakraborty Tamal and Misra Iti Saha (2015), "Design and analysis of channel reservation scheme in Cognitive Radio Networks", Elsevier Journal on Computers and Electrical Engineering 42: 148-167.

Chowdhury K. R. and Akyildiz I.F. (2011), "CRP: A routing protocol for cognitive radio ad hoc networks", IEEE Journal on Selected Areas Communication 29(4): 794-804.

Chowdhury K. R. and Felice M. D. (2009), "Search: A routing protocol for mobile cognitve radio ad-hoc networks", Elsevier Journal on Computer Communications 32: 1983-1997.

Chunsheng Xin, Bo Xie and Chien-Chung Shen (2005), "A novel layered graph model for topology formation and routing in dynamic spectrum access networks", First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks DySPAN: 308-317.

Clancy C., Hecker J., Stuntebeck E. and Tim O. (2007), "Applications of Machine Learning To Cognitive Radio Networks", IEEE Wireless Communications: 47-52.

Cleotilde Gonzalez, Polina Vanyukov and Michael K. Martin (2005), "The use of microworlds to study dynamic decision making", Elsevier Journal on Computers in Human Behavior 21(2): 273-286.

Cormio, C. and Chowdhury K. R. (2009), "A survey on MAC protocols for cognitive radio networks", Elsevier International Journal on Ad Hoc Networks 7(7): 1315-1329.

Costlow Terry (2003), “Cognitive Radios will Adapt to Users”, IEEE Intelligent Systems 18(3): 7.

Dey A. K.(2001), “Understanding and using context, personal and ubiquitous computing”, 5(1): 4-7.

Ding Lei, Tommaso Melodia, Stella N. Batalama, John D. Matyjas (2015), “Distributed resource allocation in cognitive and cooperative ad hoc networks through joint routing, relay selection and spectrum allocation”, Elsevier Journal on Computer Networks 83(4): 315-331.

Duan Yong, Qiang Liu and XinHe Xu (2007), “Application of reinforcement learning in robot soccer”, Elsevier Journal Engineering Applications of Artificial Intelligence 20: 936-950.

EI-Sayed, EI-Alfy, Yu-dong Yao and Harry Heffes(2006), “A learning approach for prioritized handoff channel allocation in mobile multimedia network”, IEEE Transaction on Wireless Communication 5(7): 1651-1660.

Engelman Richard et al. (2002), “Report of the spectrum efficiency working group”, Federal Communications Commission-Spectrum Policy Task Force : 1-37.

Fang M., Groen F. C. A., Li H. and Zhang J. (2013), “Collaborative multi-agent reinforcement learning based on a novel coordination tree frame with dynamic partition”, Elsevier Journal on Engineering Applications of Artificial Intelligence 27: 191-198.

Fernandez-Gauna B., Marques. I., Grana. M. (2013), “Undesired state-action prediction in multi-agent reinforcement learning for linked multi-component robotic system control”, Information Sciences 232: 309-324

Fujii T. and Suzuki Y. (2005), “Ad-hoc cognitive radio - development to frequency sharing system by using multi-hop network”, IEEE International Symposium on DySPAN: 589-592.

Gao Cunhao, Shi Yi, Hou Y.T., Sherali H.D. and Zhou Huaibei (2011), “Multicast Communications in Multi-Hop Cognitive Radio Networks”, IEEE Journal on Selected Areas in Communications 29(4): 784-793.

GNU Radio Website: http://www.gnuradio.org/redmine/projects/gnuradio.

Guan Q., Yu F. R., Jiang S. and Wei G. (2010), “Prediction-based topology control and routing in cognitive radio mobile ad hoc networks”, IEEE Transaction on Vehicular Technology 59(9): 4443-4452.

Hinojosa W. M., Nefti, S. and Kaymak U., “System Control with generalized probabilistic fuzzy-reinforcement learning”, IEEE Transaction on Fuzzy Systems 19(1), 51-64.

Hou Y. Thomas., Shi Yiand and Sherali Hanif D. (2008), “Spectrum sharing for multi-hop networking with cognitive radios”, IEEE Journal on Selected Areas in Communication 26(1): 146-155.

How K. C., Ma M. and Qin Y. (2011), “Routing and QoS provisioning in cognitive radio networks”, Elsevier Journal on Computer Network 55(1): 330-342.

Huang Xin-Lin, Wang Gang, Hu Fei and Sunil Kumar (2011), “Stability-capacity-adaptive routing for high mobility multihop cognitive radio networks”, IEEE Transaction on Vehicular Technology 60(6): 2714-2729.

Huang Zhengxing, Wil MP van der Aalst, Xudong Lu and Huilong Duan (2011), “Reinforcement learning based resource allocation in business process management”, International Journal on Data and Knowledge Engineering 70: 127-145.

Huisheng, Ma, Lili Zheng, Xiao Ma and Yongjian luo (2008), “Spectrum aware routing for multi-hop cognitive radio networks with a single transceiver”, 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications CrownCom: 1-6.

Jian Tang, Roberto Hincapie, Guoliang Xue, Weiyi Zhang, and Roberto Bustamante (2010), "Fair Bandwidth Allocation in Wireless Mesh Networks With Cognitive Radios", IEEE Transactions on Vehicular Technology 59(3): 1487-1496.

Jiao Wang and Yuqing Huang (2010), "A cross-layer design of channel assignment and routing in Cognitive Radio Networks", 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT) 9-11 July 2010 doi: 10.1109/ICCSIT.2010.5564800, 7: 542-547.

Jin Xiaocong, Zhang Rui, Sun Jingchao and Zhang Yanchao (2014), "TIGHT: A Geographic Routing Protocol for Cognitive Radio Mobile Ad Hoc Networks", IEEE Transactions on Wireless Communications 13(8): 4670-4681.

Juncheng Jia, Jin Zhang and Qian Zhang (2009), "Relay-Assisted Routing in Cognitive Radio Networks", IEEE International Conference on Communications doi: 10.1109/ICC.2009.5199406: 1-5.

Kaebling L. P., Littman M. L., Moore A. W. (1996), "Reinforcement learning: a survey", Journal on Artificial Intelligence Research 4: 237-285,

Khalife H., Malouch N., Fdida S. (2009), "Multi-hop cognitive radio networks: to route or not to route", IEEE Network Magazine doi: 10.1109/MNET.2009.5191142, 23(4): 20-25.

Khan Athar Ali, Mubashir Husain Rehmani and Yasir Saleem (2015), "Neighbour discovery in traditional wireless networks and cognitive radio networks: Basics, taxonomy, challenges and future research directions", Elsevier Journal on Network and Computer Applications 52: 173-190.

Kiam Cheng How, Maode Ma and Yang Qin (2011), "Routing and QoS provisioning in cognitive radio networks", Elsevier Journal on Computer Network 55(1): 330-342.

Kok-Lim Alvin Yau, Peter Komisarczukb and paul D.Teal (2012), "Reinforcement learning for context awareness and intelligence in wireless network: Review, new features and open issues", Elsevier Journal on Network and Computer Application 35: 253-267.

Kulkarni Parag (2012), "Reinforcement and Systemic Machine Learning for Decision Making", IEEE series on systems science and engineering, Published by John Wiley and Sons Inc, ISBN 978-0-470-91999-6.

Kyasanur Pradeep and Vaidya Nitin (2006), "Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks", ACM Mobile Computing and Communications Review 1(2): 1-13.

Kyounghwan L. and Yener A. (2006), "Outage Performance of Cognitive Wireless Relay Networks", IEEE Global Telecommunications Conference GLOBECOM: 1-5.

Lee Jae-Joon and Jaesung Lim (2014), "Cognitive routing for multi-hop mobile cognitive radio ad hoc networks", IEEE Journal on Communications and Networks 16(2): 155-161.

Lei Ding, Tommaso Melodia, Stella N. Batalama, John. D. Matyjas and Micheal. J. Medley (2010), "Cross-Layer Routing and Dynamic Spectrum Allocation in Cognitive Radio Ad Hoc Networks", IEEE Transaction On Vehicular Technology 59(4): 1969-1979.

Li Xiukui and Zekavat Seyed A. R.(2009), "Cognitive Radio Based Spectrum Sharing: Evaluating Channel Availability via Traffic Pattern Prediction", Journal on Communications and Networking 11(2): 104-114.

Liang J. and Chen J. (2013), "Resource Allocation in Cognitive Radio Relay Networks", IEEE Journal on Selected Areas In Communications 31(3): 476-488.

Liu Bing-Hong, Min-Lun Chen, Ming-Jer Tsai (2011), "Message-efficient location Prediction for mobile objects in wireless sensor networks using a maximum likelihood technique", IEEE Transaction on Computers 60(6): 865-878.

Liu Yongkang, Lin X. Cai and Xuemin Shen (2012), "Spectrum-aware opportunistic routing in multi-hop cognitive radio networks", IEEE Journal on Selected Areas in Communication 30(10): 1958-1968.

Lucian Busoniu, Robert Babuska and Bart De Schutter (2008), "A Comprehensive survey of Multiagent Reinforcement Learning", IEEE Transaction on Systems, Man and Cybernetics - Part C: Applications and Review 38(2): 156-172.

Lunden, J., Kulkarni, S. R., Koivunen, V. and Vincent P.(2013), "Multi-agent Reinforcement Learning Based Spectrum Sensing Policies for Cogntive Radio Networks", IEEE Journal on Selected Topics in Signal Processing 7(5): 858-867.

Ma Miao and Tsang Danny H.K. (2009), "Joint Design of Spectrum Sharing and Routing with Channel Heterogeneity in Cognitive Radio Networks", Elsevier Journal on Physical Communication 2: 127-137.

Macaluso Irene, Finn Danny, Ozgul Baris and DaSilva Luiz (2013), "Complexity of Spectrum Activity and Benefits of Reinforcement Learning for Dynamic Channel Selection", IEEE Journal on Selected Areas in Communications, 31(11): 2237-2248.

Mankar Praful S., Das Goutam, Pathak S. S. and Rajkumar R. V. (2015), "A method for Accessing Spatial Spectrum Holes for Relay Based Cognitive Cellular Networks", IEEE Wireless Communication Letters, 4(3): 245-248.

Marco Di Felice, Kaushik Roy Chowdhury, Wooseong Kim Andreas Kassler and Luciano Bononi (2010), "End-to-end protocol for cognitive radio ad hoc networks: an evaluation study", International Journal on Performance Evaluation 68(9): 859-875.

Martin Mario (2011): http://www.cs.upc.edu/ mmartin/Ag5-4x.pdf.

Matt Ettus (2015), "Universal software radio peripheral", http://www.ettus.com.

Mitola Joseph (2001), "Cognitive radio for flexible mobile multimedia communications", Mobile Networks and Applications 6: 435-441.

Mitola J. and Maguire G.Q. (1999), “Cognitive radio: making software radios more personal”, IEEE Personal Communications 6(4): 13-18.

Mumey B., Tang J., Judson I.R. and Stevens D.(2012), “On Routing and Channel Selection in Cognitive Radio Mesh Networks”, IEEE Transactions on Vehicular Technology 61(9): 4118-4128.

Nie J, Haykin S (1999), “A Dynamic channel assignment policy through Q-learning”, IEEE Transaction on Neural Network 10(6): 1443-1455.

Parker L.E.(2002), “Distributed algorithms for multi-robot observation of multiple moving targets”, Autonomous Robots 12(3): 231-255.

Parunak H.V.D. (1999), “Industrial and practical applications of DAI”, Multi-Agent Systems: A Modern Approach to Distributed Artificial Intelligence, MIT Press, Ch. 9: 377-412.

Pefkianakis I, Wong S.H.and Lu S. (2008), “SAMER: spectrum aware mesh routing in cognitive radio networks”, 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks DySPAN: 1-5.

Ping Shuyu, Aijaz Adnan, Holland O and Aghvami A., “SACRP: A Spectrum aggregation based cooperative routing protocol for cognitive radio ad hoc networks.”, IEEE Transaction on Communications 63(6): 2015-2030.

Pruyt E. (2006), “System Dynamics and Decision-making in the context of dynamically complex Multi-dimensional Societal Issues”, 24^{th} International Conference of the System Dynamics Society: 1-19

Pynadath D.V. and Tambe M. (2002), “The communicative multi-agent team decision problem: analyzing teamwork theories and models”, Journal on Artificial Intelligence Research 16: 389-423.

Qin Yang, Zhong Xiaoxiong, Yang Yuanyuan, Lia Li and Yea Yibin (2015), “Combined channel assignment and network coded opportunistic routing in cognitive radio networks”, Elsevier Journal on Computers and Electrical Engineering (In Press).

Rabiner L. R. (1989), "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", In Proceeding of IEEE, 77(2): 257-286.

Rehmani Mubashir Husain, Viana Aline Carneiro, Khalife Hicham , Fdida Serge (2013), "SURF: A distributed channel selection strategy for data dissemination in multi-hop cognitive radio networks", Elsevier Journal on Computer Communications 36(10-11): 1172-1185.

RFC (2003): http://www.ietf.org/rfc/rfc3561.txt.

Royer E. M. and C. K. Toh (1999), "A review of current routing protocols for ad hoc mobile wireless networks", IEEE Personal Communication 6(2): 46-55.

Saleem Yasir, Kok-Lim Alvin Yau, Hafizal Mohamad, Nordin Ramli and Mubashir Husain Rehmani (2015), "SMART: A SpectruM-Aware ClusteR-based rouTing scheme for distributed cognitive radio networks", Elsevier Journal on Computer Networks 91(14): 196-224.

Saleem Yasir, Salim Farrukh and Rehmani Mubashir Husain (2015), "Routing and channel selection from cognitive radio network's perspective: A survey", Elsevier Journal on Computers and Electrical Engineering 42: 117-134.

Saleem Yasir, and Rehmani Mubashir Husain (2014), "Primary radio user activity models for cognitive radio networks:A survey", Elsevier Journal on Network and Computer Applications 43: 1-16.

Samah E., Baher A., and Hossam A. (2013), "Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto", IEEE Transactions on Intelligent Transportation Systems 14(3): 1140-1150.

Schoch Elmar, Feiri Michael, Kargl Frank and Weber Michael (2008), "Simulation of ad hoc networks: ns-2 compared to JiST/SWANS", In Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks, systems and workshops (Simutools '08), ICST, Brussels, Belgium, 36: 1-8.

Sengupta S, Subbalakshmi KP (2013), "Open research issues in multihop cognitive radio networks", IEEE Communication Magazine 51(4): 168-176.

Shah G.A., Gungor V.C. and Akan O.B.(2013), "A Cross-Layer QoS-Aware Communication Framework in Cognitive Radio Sensor Networks for Smart Grid Applications", IEEE Transactions on Industrial Informatics 9(3): 1477-1485.

Sharma Manuj, Sahoo Anirudha and Nayak K. D. (2008), "Channel modeling based on interference temperature in underlay cognitive wireless networks", IEEE International Symposium on Wireless Communication System: 224-228.

Sharma M., Sahoo A., Nayak K.D. (2007), "Channel selection under interference temperature model in multi-hop cognitive mesh networks", 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks DySPAN doi: 10.1109/DYSPAN.2007.25 : 133-136.

Sharma R, Gopal M (2010), "Synergizing reinforcement learning and game theory-a new direction for control", Elsevier International Journal of Applied Soft Computing 10: 675-688.

Shih Chao-Fang, Liao Wanjiun and Chao Hsi-LU (2011), "Joint routing and spectrum allocation for multi-Hop Cognitive Radio Networks with Route Robustness Consideration", IEEE Transaction on Wireless Communication 10(9): 2940-2949.

Stone P. and Veloso M. (2000), "Multi-agent systems: A Survey from the machine Learning perspective", Springer Journal Autonomous Robots 8(3): 345-383.

Sutton Richard S., Barto Andrew G. (1998), "Reinforcement learning: an introduction", MIT Press, Cambridge.

Spachos P. and Hantzinakos D.(2014), "Scalable Dynamic Routing Protocol for Cognitive Radio Sensor Networks", IEEE Sensors Journal 14(7): 2257-2266.

Sobron Iker, Diniz Paulo S. R., Martins Wallace A. and Velez Manuel (2015), “Energy Detection Technique for Adaptive Spectrum Sensing”, IEEE Transactions on Communications 63(3): 617-627.

Talay A.C. and Altilar D.T.(2009), “ROPCORN: Routing protocol for cognitive radio ad hoc networks”, International Conference on Ultra Modern Telecommunications : 1-6.

Talay A. C. and Altilar D. T.(2012), “Self adaptive routing for dynamic spectrum access in cognitive radio networks”, Elsevier Journal on Network and Computer Applications 36(4): 1140-1151.

Tan Xiaobo, Zhang Hang, Chem Quin and Hu Jian (2013), “Opportunistic channel selection based on time series prediction in cognitive radio networks”, Transaction on Emerging Telecommunication Technologies 25(11): 1126-1136.

Troxel G. D. et. al.(2008), “Enabling open source cognitively controlled collaboration among software-defined radio nodes”, Elsevier Journal on Computer Networks 52: 898-911.

Tsagkaris K. (2008), “Neural network-bsed learning schemes for cognitive radio systems”, Elsevier Journal on Computer Communications, 31: 3394-3404.

Tumuluru Vamsi Krishna, Wang Ping and Niyato Dusit (2010), “Channel status prediction for cognitive radio networks”, Wireless Communications and Mobile Computing 10: 1-13.

Usaha W, Barria JA (2007), “Reinforcement learning for resource allocation in LEO satellite networks”, IEEE Transaction on Systems, Man Cybern-Part B: Cybern 37(3): 515-527.

Wang Xingwei, Cheng Hui and Huang Min(2014), “QoS multicast routing protocol oriented to cognitive network using competitive coevolutionary algorithm”, Elsevier Journal on Expert Systems with Applications 41(10): 4513-4528.

Wei-bing LIU, Xian-jia WANG (2009), “Dynamic decision model in evolutionary games based on reinforcement learning”, Elsevier Journal of Systems Engineering - Theory and Practice 2(3): 28-33.

Weiss G., 1999. Learning in multi-agent systems: A Modern Approach to Distributed Artificial Intelligence, MIT Press, Ch. 6: 259-298.

Wu C., Chowdhury K., Felice M. D. and Meleis W. (2010), “Spectrum management of cognitive radio using multi-agent reinforcement learning”, 9th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2010): 1705-1712.

Wu C., Ohzahata S. and Kato, T.(2013), “A Routing Protocol for Cognitive Radio Ad Hoc Networks Giving Consideration to Future Channel Assignment”, First International Symposium on Computing and Networking: 227-232.

Xia Bing, M Wahab, Yang Y, Z Fan, Zhong F and Mahesh S (2009), “Reinforcement learning based spectrum-aware routing in multi-hop cognitive radio networks”, 4th IEEE International conference on Cognitive Radio Oriented Wireless Networks and Communications CROWNCOM: 1-5.

Xie Min, Zhang Wei and Kai-Kit Wong (2010), “A Geometric approach to improve spectrum efficiency for cognitive relay networks”, IEEE Transaction on Wireless Communications 9(1): 268-281.

Xu Y. et.al. (2012), “Opportunistic Spectrum Access in Cognitive Radio Networks: Global Optimization Using Local Interaction Games”, IEEE Journal on Selected Topics In Signal Processing 6(2): 180-193.

Yong Duana, Qiang Liub and XinHe Xub (2010), “Cross-layer routing and dynamic spectrum allocation in cognitive radio ad hoc network”, IEEE Transaction on Vehicular Technology 59(4): 1969-1979.

Yuan Guangxiang, Ryan C. Grammenos, Yang Yang, and Wang Wenbo (2010), “Performance Analysis of Selective Opportunistic Spectrum Access With Traffic Prediction”, IEEE Transactions on Vehicular Technology, 59(4): 1949-1959.

Zeeshan, M., Manzoor, M.F., Qadir, J.(2010), “Backup channel and cooperative channel switching on-demand routing protocol for multi-hop cognitive radio ad hoc networks (BCCCS)”, 6th International Conference on Emerging Technologies: 394-399

Zhou Yuan, Bin Song Ju and Zhu Han (2013), “Interference Aware Routing Using Network Formation Game in Cognitive Radio Mesh Networks”, IEEE Journal on Selected Areas in Communications 31(11): 2494-2503.

Publications

And

Research Grants

Publications and Research Grants

1. Barve Sunita S. and Kulkarni Parag A., “Multi-agent reinforcement learning based opportunistic routing and channel assignment for mobile cognitive radio ad hoc network”, ACM Springer International Journal on Mobile Networks and Applications, Volume 19, No 6, pp. 720-730, December 2014, DOI 10.1007/s11036-014-0551-6, ISSN: 1383-469X (print version), ISSN: 1572-8153 (electronic version), Impact Factor: 1.496.

2. Barve Sunita S. and Kulkarni Parag A., “A Performance Based Routing Classification in Cognitive Radio Networks”, International Journal on Computer Applications (IJCA), Volume 44, No. 19, pp. 11-20, April 2012, DOI:10.5120/ 6370-8762, ISSN: 0975-8887, Impact Factor: 0.791.

3. Barve Sunita S. and Kulkarni Parag A., “Dynamic Channel Selection and Routing Through Reinforcement Learning in Cognitive Radio Networks”, In proceeding of IEEE International Conference on Computational Intelligence and Computing Research, December 2012, IEEE Xplore Catalog No.CFP1220J-ART, ISBN:978-1-4673-1244-5, pp. 6-12, DOI: 10.1109/ICCIC.2012.6510175.

4. Barve Sunita S., “Cognitive Radio Networks: A Tutorial”, IJACKD Journal of Research, Volume 3, Issue 1, pp. 32-37, September 2014, ISSN (Print): 2278-5698.

5. Barve Sunita S., Kanchan H. and Kulkarni Parag A., “Link Prediction-Based Topology Control and Adaptive Routing for Cognitive Radio Mobile Ad-Hoc Networks”, International Journal on Emerging Technologies in Computational

and Applied Sciences (IJETCAS), Volume 9, No. 3, pp. 319-325, June-August 2014, ISSN: 2279-0055 (print version), ISSN: 2279-0047 (electronic version), Impact Factor: 1.237.

Research Grants Received

Project No. 1	
Title of the Project	Efficient and enhanced spectrum management using cognitive radio networking
Funding Agency	Research Promotion Scheme of All India Council for Technical Education, New Delhi
Grant-in-aid Sanctioned	Rs. 16,38,000/-
Duration	2013-2016
Project No. 2	
Title of the Project	Cognitive Radio Network using GNU Radio and Universal Software Radio Peripherals
Funding Agency	BCUD, Savitribai Phule Pune University, Pune
Grant-in-aid Sanctioned	Rs. 1,40,000/-
Duration	2013-2015

ALL INDIA COUNCIL FOR TECHNICAL EDUCATION
4th Floor, East Tower, N.B.C.C. Place, Bhisham Pitamah Marg,
Pragati Vihar, New Delhi-110003

Ref. No.: **20/AICTE/RIFD/RPS(POLICY-III)78/2013-14** **July 29, 2013**

The Drawing and Disbursing Officer
All India Council for Technical Education
7TH Floor, Chandralok Building,
Connaught Place, New Delhi – 110 001.

Sub: **Release of Grants under Research Promotion Scheme (RPS) Scheme during the financial year 2013-14.**

Sir,

This is to convey the sanction of the Council for payment of **Rs. 1501334 (Rupees Fifteen Lakh One Thousand Three Hundred ThirtyFour Only)** during 2013-14 under the Research Promotion Scheme (RPS) as Grant-in-aid for meeting the expenditure for implementing the Scheme in order to create & update general research capabilities, as per details given below:

1.	Name of the Beneficiary Institution (University / College / Institution)	:	**MIT ACADEMY OF ENGINEERING (OLD NAME - MAHARASHTRA ACADEMY OF ENGINEERING, ALANDI(D))** **DEHU PHATA , ALANDI(D), TAL:KHED, DIST: PUNE-412105, MAHARASHTRA., Maharashtra**
2.	Principal Investigator's Name & Deptt.	:	**Mrs. SUNITA BARVE** **COMPUTER SCIENCE AND ENGINEERING**
3.	Grant-in-aid Sanctioned	:	**Rs. 1638000/-**
4.	Amount to be Released during the year 2013-14 (100% of non-recurring and recurring of year 1 remaining recurring in year 2 and 3)	:	**Non Recurring : Rs. 1433000/-** **Recurring (1st Year): Rs. 68334/-** **Total : Rs. 1501334/-**
5.	Approved Duration	:	**3 Years (Three Years)**
6.	Title of the Project	:	**EFFICIENT AND ENHANCED SPECTRUM MANAGEMENT USING COGNITIVE RADIO NETWORKING**

1. The sanctioned grant-in-aid is debatable to the major "**601.1 RPS** " and is valid for payment during the financial year 2013-14

2. The grant-in-aid of the grant shall be drawn by the Drawing and Disbursing Officer (DDO), All India Council for Technical Education, New Delhi on the Grants-in-aid bill and shall be disbursed to and credited to the account of **Director/Registrar/ Principal, MIT ACADEMY OF ENGINEERING (OLD NAME - MAHARASHTRA ACADEMY OF ENGINEERING, ALANDI(D)) DEHU PHATA , ALANDI(D), TAL:KHED, DIST: PUNE-412105, MAHARASHTRA., Maharashtra** through **RTGS**.

3. The date of release of the grant by AICTE shall be taken as the date of commencement of the project. The **Principal/Director/Registrar** shall intimate about the receipt of the grant to AICTE. Any Expenditure incurred prior the issuance of the approval letter is not allowed to be adjusted in the grant and if the University/Institution do not take the project work within 6 months of the receipt of the grant, approval shall ipso facto lapse.

30-7-13

Contd…2/-

University of Pune

Pune: - 411 007

Board of Colleges and University Development

Ref. No:- OSD /BCUD/360/71 Date:- 27/11/2013

To,
The Principal,
Maharashtra Academy of Engineering
And Educational Research
MIT Academy of Engineering
Addr: Devu Phata Alandi Devachi Pune
Ta: Khed Dist: Pune

Subject:- Sanction BCUD research Proposals for 2013-14 to 2014-15.

Dear Sir /Madam,

With reference to the acceptance letter and revised budget of the Research Proposals received from Principle Investigators, the University authorities are pleased to approve the projects submitted by the following Principle Investigators along with the sanctioned amount show against their names for year 2013-14 to 2014-15.

Sr. No.	Full Name	Sanction Amount
1	Satyajit Abhinandan Pangaonkar	175000/-
2	Sunilkumar Madhusudan Bhagat	140000/-
3	Sunita Shivprakash Barve	140000/-
4	Shitalkumar Arvind Jain	140000/-
5	Dipti Yashodhan Sakhare	100000/-
	Total	6,95,000/-

The 1st Installment of the Sanction research project has been released. You are requested to inform concerned teachers.

Details about the Norms and Guidelines can be download from <u>www.unipune.ac.in</u>

Dr. Ravindra G. Jaybhaye
OSD/BCUD

Dr. V. B. Gaikwad
Director BCUD