MULTI-AGENT BASED REINFORCEMENT LEARNING FRAMEWORK FOR MULTI OBJECTIVE RESOURCE PROVISIONING IN CLOUD ENVIRONMENTS

P.S.L.A Jayaweera

218528C

MSc in Artificial Intelligence

Faculty of Information Technology

University of Moratuwa
Sri Lanka

July 2025

MULTI-AGENT BASED REINFORCEMENT LEARNING FRAMEWORK FOR MULTI OBJECTIVE RESOURCE PROVISIONING IN CLOUD ENVIRONMENTS

P.S.L.A Jayaweera

218528C

Thesis submitted in partial fulfillment of the requirements for the degree
MSc in Artificial Intelligence

Faculty of Information Technology

University of Moratuwa
Sri Lanka

July 2025

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 11/07/2025

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the MSc in Artificial Intelligence Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Professor (Mrs) Thushari Silva

Signature of the Supervisor: Date: 12/07/2025

DEDICATION

To my parents for their unconditional love and support

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor (Mrs) Thushari Silva, for her invaluable guidance, support, and encouragement throughout this research project. Her insights and feedback have been instrumental in shaping this thesis and helping me grow as a researcher.

I would also like to extend my sincere thanks to all the lecturers and academic staff who have shared their knowledge and provided support during my studies. Their dedication to teaching and mentorship has played a significant role in my academic journey.

To my fellow students and colleagues, thank you for the thought-provoking discussions, collaboration, and camaraderie. The shared experiences and encouragement made this journey more enjoyable and enriching.

To my beloved parents, no words can fully express my gratitude for your unconditional love, sacrifices, and unwavering faith in me. You are my foundation and my biggest inspiration.

To my loving husband, thank you for being my rock, always patient, endlessly supportive, and believing in me even when I doubted myself. Your love gave me the strength to keep going.

And to my best friend, Samodha Pallewatta, your constant encouragement, understanding heart, and timely words of comfort meant more to me than you'll ever know.

This thesis is not just a culmination of my work but a reflection of the love, support, and kindness I have received from all of you. Thank you from the bottom of my heart.

ABSTRACT

Cloud computing provides a powerful and flexible platform for executing large-scale and complex applications through a pay-as-you-go model. Due to its scalability, elasticity, and economic benefits, many enterprises now rely on cloud services for their business-critical operations. To meet growing demand, improve fault tolerance, and avoid vendor lock-in, organizations are increasingly adopting multi-cloud environments, leveraging the diverse capabilities of multiple cloud providers across different pricing models and geographic regions.

However, efficient and dynamic resource allocation in multi-cloud environments remains a significant challenge, particularly when aiming to balance conflicting objectives such as minimizing cost, meeting deadlines, and maintaining high quality of service. Traditional static scheduling methods lack the adaptability required for these dynamic, heterogeneous environments. Existing multi-objective provisioning techniques often struggle with scalability and responsiveness, especially under fluctuating and bursty workloads.

To address these limitations, this thesis introduces MARL4RP (Multi-Agent Reinforcement Learning for Resource Provisioning), a novel framework that leverages the power of reinforcement learning to tackle the dual challenges of resource provisioning and workflow scheduling in multi-cloud Infrastructure-as-a-Service (IaaS) environments. MARL4RP employs a decentralized, multi-agent reinforcement learning architecture that enables the system to learn optimal policies dynamically, with the goal of minimizing both operational costs and task execution latency.

The framework was evaluated using two state-of-the-art reinforcement learning algorithms, Proximal Policy Optimization (PPO) and Deep Q-Network (DQN), across varying task loads. Experimental results reveal that PPO consistently outperforms DQN, particularly under high and bursty workloads, achieving lower VM costs and improved execution times. While DQN demonstrates efficiency at lighter loads, PPO proves more robust and scalable as workload intensity increases.

These findings underscore the potential of MARL4RP as an intelligent and adaptive resource provisioning solution for dynamic and large-scale multi-cloud environments, paving the way for future research in autonomous cloud infrastructure management.

Keywords: Multi-Agent Reinforcement Learning, Cloud resource provisioning

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor i

Dedication ii

Acknowledgement iii

Abstract iv

Table of Contents v

List of Abbreviations viii

1 Introduction 1

1.1 Prolegomena 1

1.2 Aim and Objectives 3

1.3 Background and Motivation 3

1.4 Problem in Brief 5

1.5 Preamble 5

1.6 Resource Requirements 6

1.7 Thesis Structure 6

1.8 Summary 8

2 RECENT DEVELOPMENTS AND CHALLENGES IN CLOUD RESOURCE PROVISIONING 9

2.1 Introduction 9

2.2 Gestation of Cloud Resource Provisioning 9

2.3 Major Developments in Cloud Resource Provisioning 10

2.4 Future Directions of Cloud Resource Provisioning 11

2.5 Challenges of Cloud Resource Provisioning 12

2.6 Problem Definition 13

2.7 Summary 13

3 Technology Adapted 14

3.1 Introduction 14

3.2 Cloud Simulator 14

3.2.1 CloudSim 14
3.2.2 CloudSimPlus 16
3.3 Reinforcement Learning 17
3.3.1 Foundations of Reinforcement Learning in Agent Modeling 17
3.3.2 Proximal Policy Optimization 19
3.3.3 Deep Q-Network 19
3.3.4 Ray RLlib 20
3.3.5 Ray Tune 21
3.4 Summary 21
4 MARL system for cloud resource provisioning 22
4.1 Introduction 22
4.2 Hypothesis 22
4.3 Input 23
4.4 Output 23
4.5 Process 24
4.6 Features 25
4.7 Users 26
4.8 Summary 26
5 Design 27
5.1 Introduction 27
5.2 System Architecture 27
5.3 Agent Architecture 28
5.4 Orchestrator Agent 28
5.5 Datacenter Agents 28
5.5.1 Observation Space 29
5.5.2 Reward Structure 29
5.5.3 Action Space 31
5.5.4 Task scheduling Algorithm 31
5.6 Summary 32

6 Implementation of MARL4RP 33
6.1 Input Data Processing 33
6.2 Extension of CloudSimPlus for Multi-Cloud Simulation 33
6.2.1 VM Cost Calculation 34
6.2.2 Task execution time calculation 35
6.3 Reinforcement Learning Engine Implementation 35
6.3.1 Environment Setup 36
6.3.2 Environment Functions 37
6.4 Training with Ray Tune 37
6.5 Summary 38
7 Experiments and Evaluation 39
7.1 Introduction 39
7.2 Participants 39
7.3 Experimental Design 39
7.3.1 Experimental Setup 39
7.3.2 Evaluation Scenarios 40
7.3.3 Hyperparameter Tuning 40
7.4 Performance Comparison of DQN and PPO 42
7.4.1 Performance Under Increased Load 42
7.4.2 Adaptability To Bursty Load 43
8 Conclusion and Future Work 46
8.1 Introduction 46
8.2 Conclusion 46
8.3 Contributions 47
8.4 Future Work 47
References 49

LIST OF ABBREVIATIONS

CTDE Centralised Training with Decentralised Execution

DEA Differential Evolution Algorithm

DP Dynamic Programming

DQN Deep Q-Network

MARL Multi-Agent Reinforcement Learning

MAS Multi-Agent Systems

PE processing element

POMDP Partially Observable Markov Decision Process

PPO Proximal Policy Optimisation

QoS Quality of Service

RL Reinforcement Learning

RTP Real-Time Pricing

SLAs Service Level Agreements

TD Temporal Difference

TLB Targeted Load Balancer

TOU Time Of Use

VM virtual machine