UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Deep Reinforcement Learning for Artificial Intelligence-enabled Autonomous Penetration Testing in Cyber Security

by

Hoang Khuong Tran

Thesis Submitted
in Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Supervisors: Prof. Chin-Teng Lin
Prof. Yu-Kai Wang

Sydney, Australia

December, 2022

Certificate of Authorship/Originality

I, Hoang Khuong Tran, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: December 30th, 2022 Place: Sydney, Australia

ABSTRACT

Deep Reinforcement Learning for Artificial Intelligence-enabled Autonomous Penetration Testing in Cyber Security

by

Hoang Khuong Tran

Penetration Testing (PT) is the set of methods used to enhance the security of a networked system by exploiting potential vulnerabilities. It is the practice of simulating attacks on computer systems, networks, or web applications to test their security and identify vulnerabilities that an attacker could exploit. Penetration testers use various tools and techniques to probe the defenses of a system and uncover weaknesses. These methods require significant time and resources for training and execution. There is a shortage of skilled professionals to deal with the increasingly complex cyber security landscape. Conventional approaches in penetration testing require threat model of the target network to gain context to the exploits and vulnerabilities. These requirements present intractable challenges for penetration testing tools in dealing with rapidly changing network software and attack dimensions.

Artificial intelligence (AI) and reinforcement learning (RL) can potentially be used in penetration testing to automate certain tasks and improve the efficiency of the testing process, such as identifying targets, generating attack strategies, and adapting to changes in the target system. AI-enabled PT has been under research in recent years due to the insurgence of new deep learning advances in which RL is considered an appropriate learning framework to develop such applications thanks to its capability in learning a sequential decision-making process without any labelled dataset through interacting with the environment. For example, an AI-powered tool could analyse the network architecture and system configurations and suggest

potential attack vectors based on this information. An RL system could learn from previous testing experiences and use this knowledge to adapt to new situations and improve its performance over time.

However, there are multitudes of challenges in using RL to develop an automatic penetration testing application. The main challenges are the large and structured configuration of the state space and action space which are unconventional in typical deep RL works. Other challenges include the partial observability, the scarcity of rewards and the stochastic dynamics of the environment.

This research aims at understanding the technical challenges presented by an autonomous penetration testing application and developing novel Deep Reinforcement Learning (DRL) frameworks to deal with two problems of scalable autonomous PT, which involve the complexity of the *action space* and the *state space*. By leveraging the recent advances in Multi-Agent Reinforcement Learning (MARL) paradigm, we re-formulate the conventional approach of using a single-agent DRL into a multi-agent learning framework, enabling the decomposition of the complex and structured *action space* into manageable sub-modules each of which is controlled by a DRL agent. The agents are trained cooperatively to develop PT policies under two different representations of the *action space*: a large and discrete action space and a multinomial parameterised action space. We introduced two new frameworks called *Cascaded reinforcement learning agents for large discrete action spaces* and *Multi-agent reinforcement learning for parameterised action spaces* for each of the aforementioned *action space* representations.

The complexity of the *state space* representation in autonomous PT consisting of the non-visual and binary-valued description of the cyber networks, the highly stochastic state transition probability coupled with the sparsity of the reward signals makes it challenging for DRL algorithms to learn in such application. We adapted Hierarchical Reinforcement Learning (HRL), a multi-layer learning approach wherein the high-level layer is trained to assign different sub-goals to the lower level, which in turn learns a primitive policy to achieve the given subgoals. This integration of HRL into the MARL training is innovative and can be devel-

oped into a more general framework for handling complex problem space in different domains. The subgoal learning is facilitated by using the Successor Representation (SR) as it enables the learning of a state abstraction under environment with sparse or no reward. All the proposed approaches can be integrated end-to-end to develop an AI-enabled autonomous penetration testing application.

Dissertation directed by Distinguished Professor Chin-Teng Lin
Australian Artificial Intelligence Institute (AAII)
School of Computer Science
University of Technology Sydney

To my loved ones

Acknowledgments

The 4-year PhD research is one of the toughest but worthwhile experiences in my life. It is a testimony to my perseverance, intellectual ability, and hard work. The Covid-19 pandemic has put a lot of constraints both mentally and physically to my research as well as my well-being. I experienced the passing of my grandpa and the birth of my first child during my PhD. I could not have done this without the unconditional support of my loved ones. There is no word that I can use to express my gratitude to my parents and my wife who have been standing with me throughout the journey.

I would like to sincerely thank Professor Chin-Teng Lin for giving me the opportunity to join the Brain Computer Interface lab at UTS to conduct my PhD research, and for continually supporting me in both work and personal life. I would also like to express my appreciation to Dr. Yu-Kai Wang for being my co-supervisor and his constant encouragement whenever we meet.

Last but not least, I extend my gratitude towards Dr. Junae Kim, Dr. Toby Richer and Max Standen from the Defence Science and Technology Group, Australia for supporting this work by having regular discussions with us and giving valuable feedbacks on the development of the involved researches.

This work was supported by the Australian Defence Science Technology Group (DSTG) under Agreement No: MyIP 10699.

Hoang Khuong Tran
Sydney, Australia, 2022.

List of Publications

Conference

C-1. **Tran, Khuong** and Akella, Ashlesha and Standen, Maxwell and Kim, Junae and Bowman, David and Richer, Toby and Lin, Chin-Teng, "Deep hierarchical reinforcement agents for automated penetration testing" *Proceedings of the 1st International Workshop on Adaptive Cyber Defense, IJCAI 2021*

Journal

J-1. **Tran, Khuong** and Standen, Maxwell and Kim, Junae and Bowman, David and Richer, Toby and Lin, Chin-Teng, "Cascaded reinforcement learning agents for large discrete action space" *Special Issue Machine Learning for Cybersecurity Threats, Challenges, and Opportunities II*

J-2. **Tran, Khuong** and Standen, Maxwell and Kim, Junae and Bowman, David and Richer, Toby and Lin, Chin-Teng, "A Multi-Agent Reinforcement Learning Approach for Multinomial Parameterised Action Space" *IEEE Transactions on Information Forensics and Security* (Under Review)

J-3. **Tran, Khuong**; Lin, Chin-Teng, "Subgoals Discovery Using Deep Successor Feature Representation in Autonomous Penetration Testing" *IEEE Transactions on Systems, Man, and Cybernetics* (In progress)

Contents

List of Figures

List of Tables

Abbreviation

RL - Reinforcement Learning

MARL - Multi-agent Reinforcement Learning

DRL - Deep Reinforcement Learning

MLP - Multi-layer perceptron

CybORG - Cyber Operation Reseach Gym

RNN - Recurrent Neural Network

GRU - Gated Recurrent Unit

DQN - Deep Q-Network

HRL - Hierarchical Reinforcement Learning

MDP - Markov Decision Process

SMDP - Semi-Markov Decision Process

PAMDP - Parameterised Action Markov Decision Process

POMDP - Partially Observable Markov Decision Process

CTF - Capture The Flag

SR - Successor Representation

SF - Successor Feature

DSR - Deep Successor Representation

ICT - Information Communication Technology