Temasek Defence Systems Institute

# Multi – Armed Bandit Models for Exploitation of Cyber Networks

**Author:** Chan Baixian, Alvin (ST Engineering)

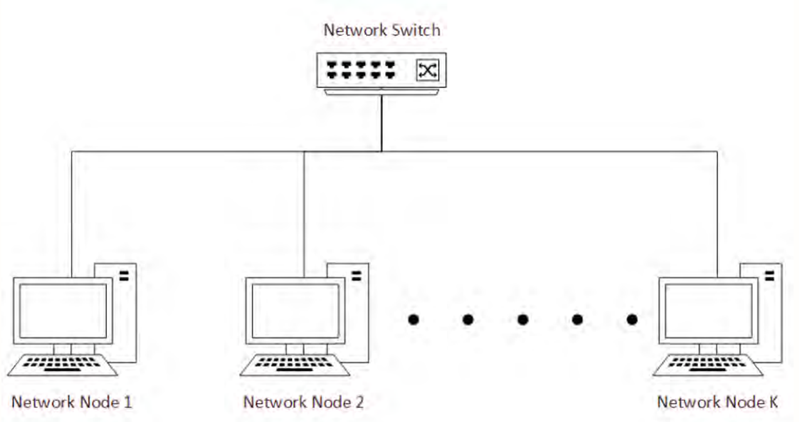
**Thesis Advisors:** Dr. Roberto Szechtman (Naval Postgraduate School)
Dr. James Grant (Lancaster University)

## Introduction

Computer networks are often the target of cyber-attacks carried out by malevolent agents, to either disable critical system operations or to surreptitiously gain access to sensitive data. Simulated network attacks allow identification of potential attack surfaces, and enable bolstering of network defences.

## Objectives

- Use multi – armed bandit models as a framework to formulate possible network attack strategies that maximizes expected reward earned.
- Evaluate regret performance of relevant multi – armed bandit models to defined problem.
- Develop our own stylized model for network exploitation with improved regret performance.




## Main Research Idea

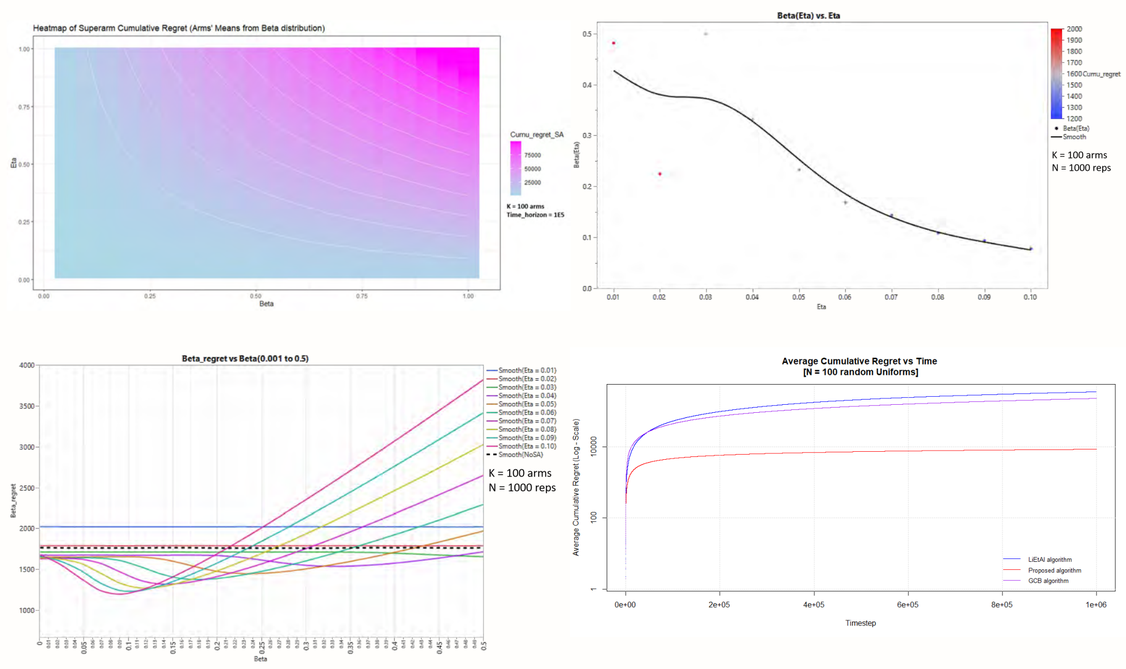
- Consider a ‘star’ LAN topology with single network switch and K connected nodes. In each time step, the attacker samples either the super – arm or the individual arms.

- Define super – arm: network switch.
- Define arms: connected network nodes.
- Attacker has no knowledge of the underlying distributions and rewards of the respective nodes.

- Sampling the super – arm earns no reward, but earns a regret equal to reward value of best arm. Attacker gains observations on random subset of arms, which helps identify promising arms to exploit over time.

- Sampling an individual arm earns the attacker corresponding node reward, or regret equal to reward difference between best node and current node.

- Goal: Identify (via exploration) and exploit best possible node, with minimal cumulative regret earned over time.




## Key Research Results

1. Evaluated cumulative regret performance of:
- Global Confidence Bound Learning Algorithm
- LiEtAl Algorithm

2. Developed own stylized model for defined problem, using multi – armed bandit model framework. Cumulative regret performance achieved was significantly lower over time, in comparison with other algorithms evaluated.

## Benefits / Potential Applications

Formulation of optimal network attack strategies to evaluate network defenses, based on specific topologies and reward feedback mechanisms.

## Follow – Up Research

Optimization of the number of arms that get at least a single sample during initialization phase, and also balancing of the initialization effort done by super – arm in terms of number of arms.

NUS National University of Singapore