Enhancing Exfiltration Path Analysis Using Reinforcement Learning

Riddam Rishu[1], Akshay Kakkar[1], Cheng Wang[1], Abdul Rahman[1], Christopher Redino[1], Dhruv Nandakumar[1], Tyler Cody[2], Ryan Clark[1], Dan Radke[1], Edward Bowen[1]

[1] Deloitte & Touche LLP

[2] National Security Institute, Virginia Tech

OCT 20 2023

Agenda

1. Background - Penetration Testing & Reinforcement Learning (RL)
2. Approach - Incorporating Cyber Terrain into RL Models
3. Use Case – Protocol-based Exfiltration with Payload
4. Results & Discussions

Penetration Testing

Main concepts and challenges

ABOUT

A penetration test (pen test) is a simulated attack on a system to expose security weaknesses in a network.

Weaknesses can come from software **bugs**, network **misconfigurations**, design **flaws**, etc.

Pen-testers attempt to gain **access** to a particular host or asset by navigating through the network from some starting point.

CHALLENGES

1 – Requires deep understanding of **cyber tradecraft**

2 – **Manually driven** and **time consuming**

3 – **Limited in scope** due to the sheer number of options to explore

CONCEPTS

Network | Communication system of computers

Subnet| Subsystems of a network

Exploit | Program to take advantage of vulnerabilities or security flaws in software or hardware

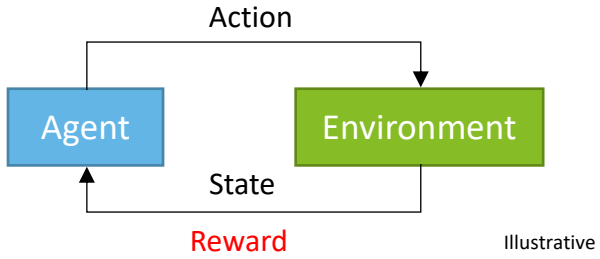
Reinforcement Learning (RL)

Main concepts and benefits

ABOUT

A type of machine learning that focuses on training an **agent** to make a **sequence of decisions** in order to maximize a cumulative reward through **trial-and-error**.

An agent **interacts** with an **environment** and learns from feedback in the form of **rewards** or **punishments**.



Illustrative

BENEFITS

Sequential Decision Making | RL can handle complex tasks that require making sequential decisions in uncertain environments.

Learning from Interaction | RL does not require explicit programming or domain-specific knowledge, or labelled data.

Autonomy and Adaptability | RL Enables agents to learn autonomously and adapt to changing circumstances. Once trained, RL agents can operate independently.

CONCEPTS

State | The current observation of the environment.

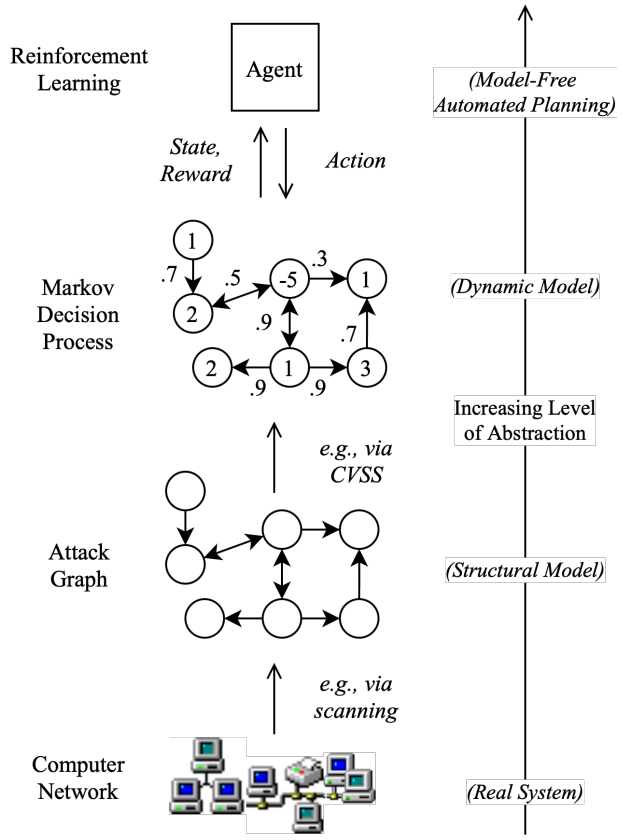
Action| The decision made by the agent based on the observed state.

Reward | The feedback signal the agent receives from the environment.

Policy | The strategy or set of rules that the agent uses to determine its actions in a given state.

Applying RL to Penetration Testing

Data-Driven Approach



CVSS: Common Vulnerability Scoring System

Illustrative

Related Works

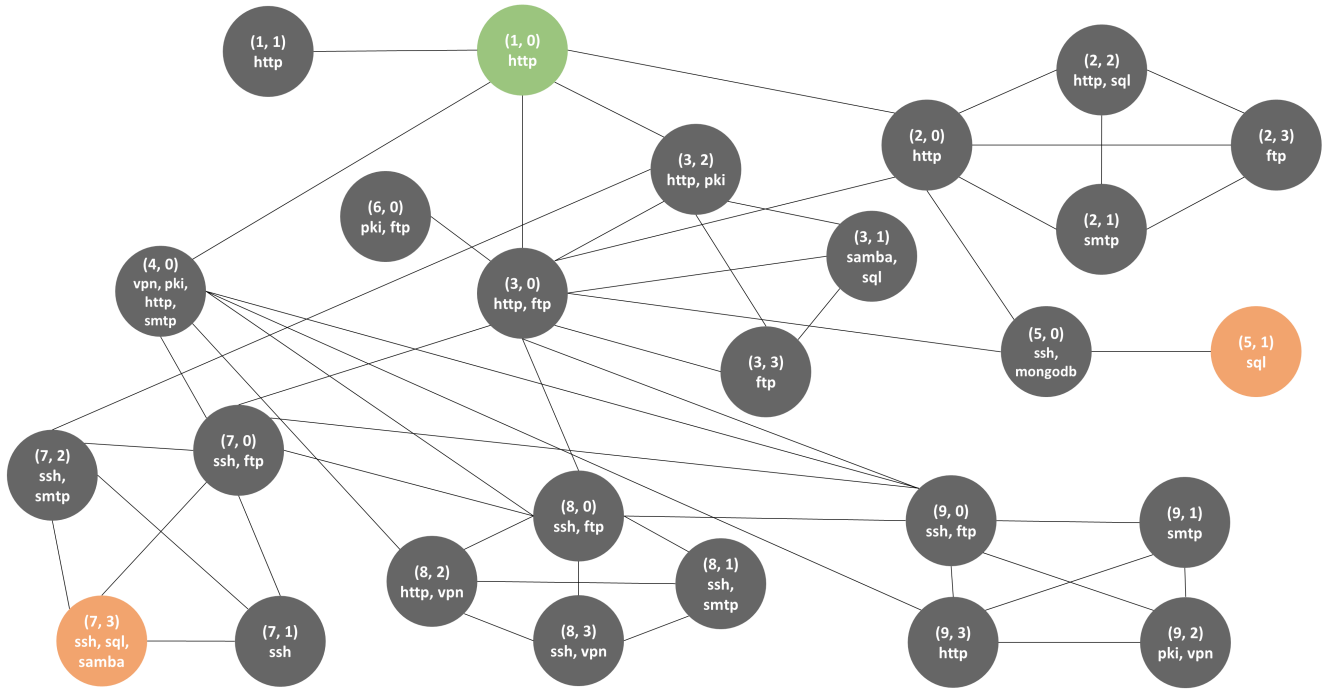
- Modeling Cyber Terrain (Gangupantulu 2021)
  - Build cyber terrain into Markov Decision Process (MDP) models by adding transition probabilities for traversing firewalls and negative rewards for different protocols.
- Performing Crown Jewel Analysis (Gangupantulu et al. 2021)
  - Identify effective entry points, pivot points, and footholds near crown jewels.
- **Discovering Exfiltration Paths** (Cody et al. 2022)
  - Find the path with the largest reward from a compromised node within a network to some exit nodes that are connected to the public internet.
- Exposing Surveillance Detection Routes (Huang et al. 2022)
  - Gain service information of the target node along with maximizing information gathering among other areas of the network while being cautious.
- Discovering Command and Control (C2) Channels (Wang et al. 2023)
  - Establish a communication channel and send data over it from the compromised host to the C2 server while evading firewall detections.

Reinforcement Learning Model for Pen Testing

States

The state at a given time step includes the following features for each host:

- Subnet ID and local ID
- Operating system
- Services
- Processes
- Discovered
- Compromised
- Isolated
- Time since infection
- Remaining payload size
- … …



ftp: File transfer protocol
http: Hypertext Transfer Protocol
pki: Public key infrastructure
smtp: Simple Mail Transfer Protocol

sql: Structured query language
ssh: Secure Shell Protocol
vpn: Virtual private network

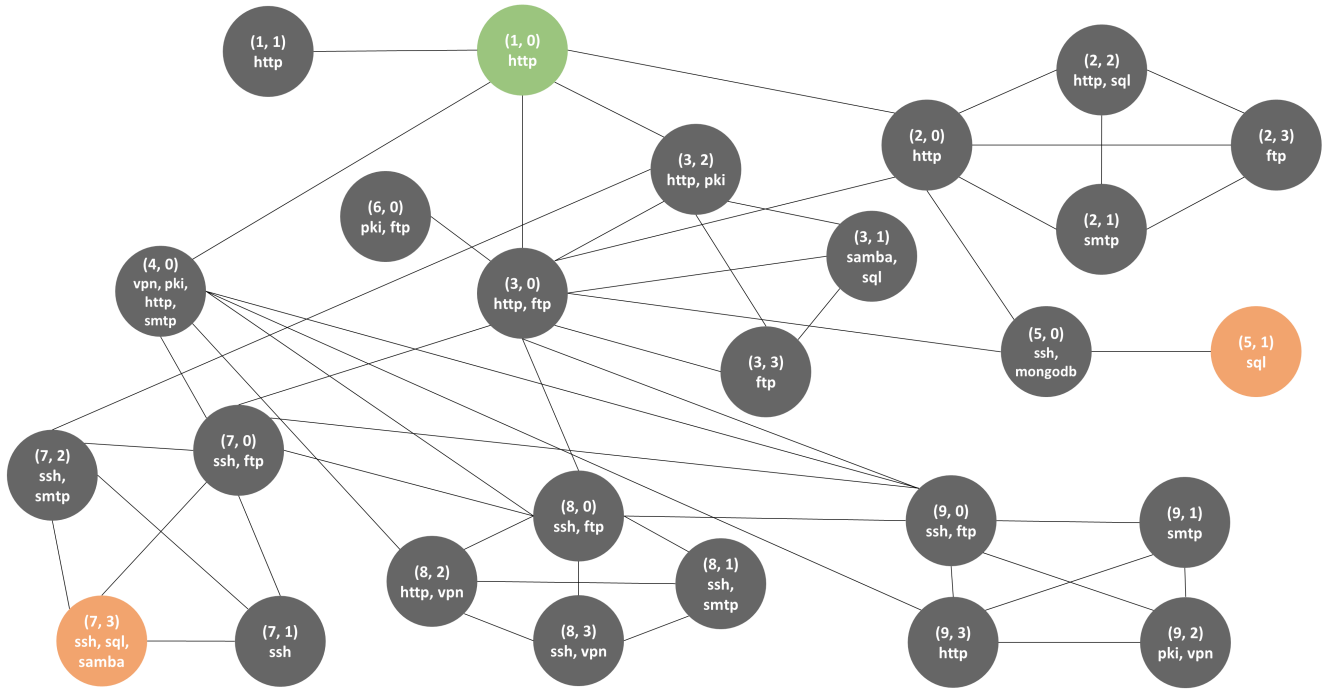
Illustrative

Reinforcement Learning Model for Pen Testing

Actions

The following types of actions are included:

- **Subnet Scan** at a compromised host
  - Discovers nearby hosts and their OS, services & processes information
- **Exploit** a nearby target
  - Gains user or root access based on target's vulnerability
- **Upload** a certain size of payload
  - Sends data from the compromised host to an exit node
- **Sleep** for a given period of time
  - Can be used to hide exfiltration activities and evade detections

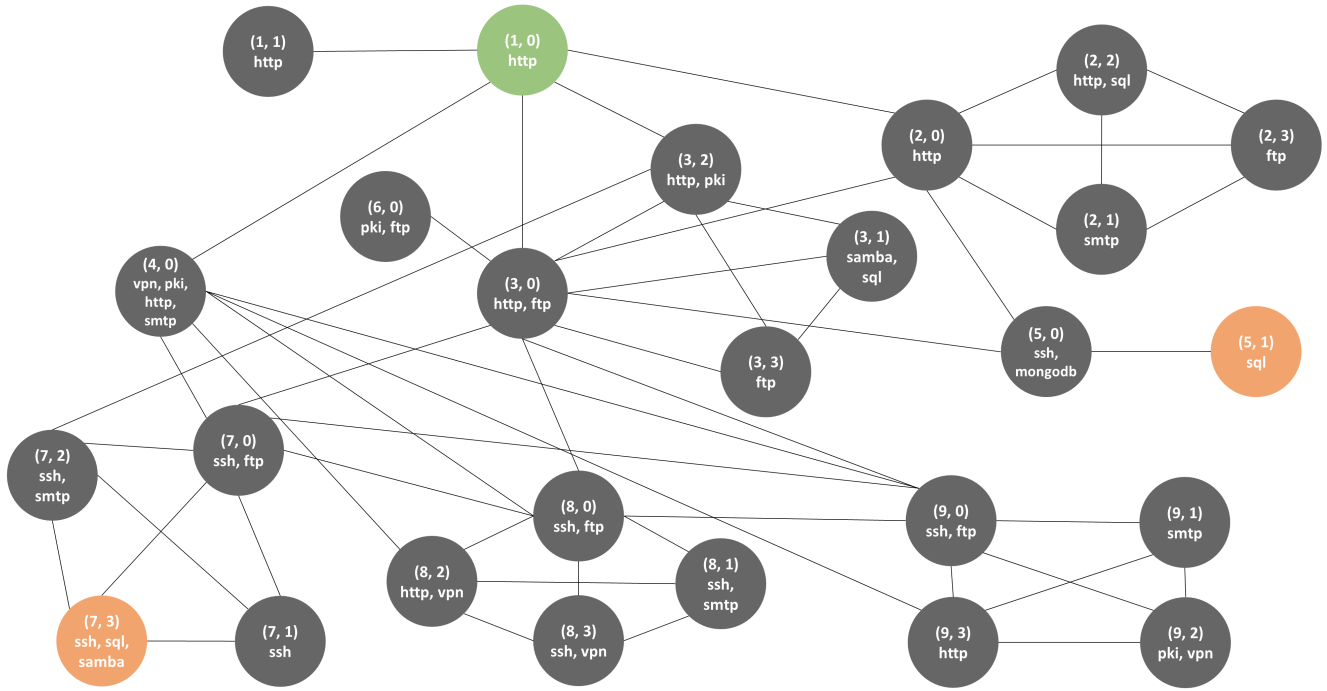


Illustrative

Reinforcement Learning Model for Pen Testing

Rewards

- A positive component for making progress towards the goal:
  - Discovers a target
  - Compromises a host
  - Uploads (partial) payload to an exit node
  - … …
- A negative term to penalize traversing defensive terrains:
  - Scanning/exploiting a host with sensitive services
  - Defensive measures vary by services (e.g., Telnet vs SSH)



Illustrative

Improvements over Previous Work[1]

- Considers utilization of common protocols (e.g., HTTP, DNS)
  - Widely available in many network environments
  - Not as closely monitored (e.g., data backup using FTP may align with security rules)
- Introduces payload and network firewalls
  - Different payload size can be specified to better simulate real-world exfiltration campaign
  - Network firewalls are modeled to monitor traffic within the network
  - In addition to finding effective paths, the agent should also learn to send data while evading firewall detections
- Experiments on larger networks
  - Previous work: a single network consists of 9 subnets and 26 hosts
  - This work: two networks
    - One with 10 subnets and 56 hosts, and
    - One with 101 subnets and 1444 hosts

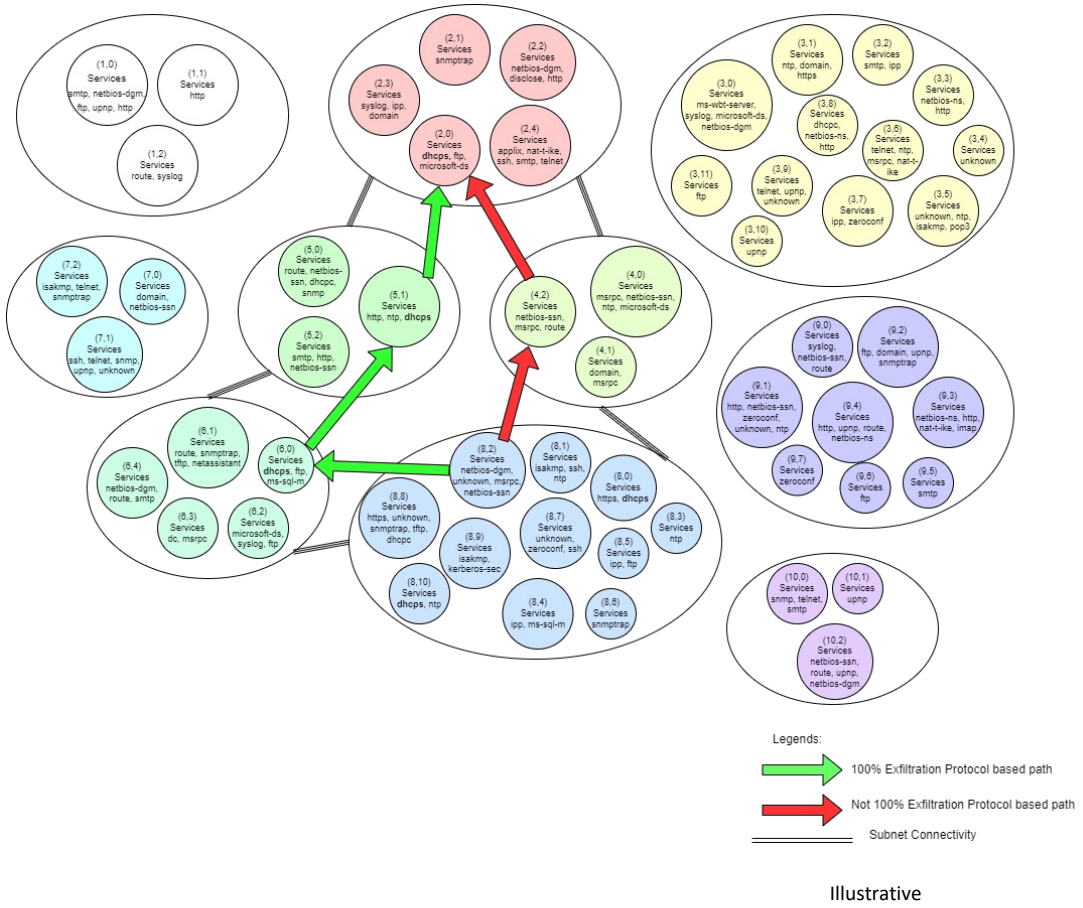
[1]Cody, Tyler, et al. "Discovering Exfiltration Paths Using Reinforcement Learning with Attack Graphs." *2022 IEEE Conference on Dependable and Secure Computing (DSC).* IEEE, 2022.

Protocol-Based Exfiltration with Payload

- Selecting path
  - The agent tries to maximize the utilization of a given protocol in its exfiltration path.
  - A path consists of 4 hosts running the same protocol (e.g., HTTP) is preferred over a path of 2 hosts running different protocols (e.g., HTTP and FTP).
  - If multiple paths are identified with the same protocol coverage, the shortest one is preferred.
  - After a new host is compromised, the agent will update candidate exfiltration paths.
  - If a better path is discovered, the agent will use it to exfiltrate the target payload.
- Sending payload
  - To evade firewall detection, the agent should learn to exfiltrate in a stealthy manner.
  - Frequent or large uploads will trigger alerts and incur a large penalty.
  - The task is complete if the entire payload is sent to the exit node.

Experiments and Results I

- An initial foothold is gained on host (8, 2) from subnet 8, which is not accessible to the Internet.
- The exit node (2, 0) from subnet 2 is directly connected to the Internet.
- DHCP is selected as the exfiltration protocol.
- The agent initially found path (8,2) -> (4,2) -> (2,0), but (4, 2) does not have DHCP.
- The agent later discovered and compromised host (6, 0) and (5, 1) subsequently.
- A new path (8, 2) -> (6,0) -> (5, 1) -> (2, 0) is selected for exfiltration since both (6, 0) and (5, 1) are running DHCP.

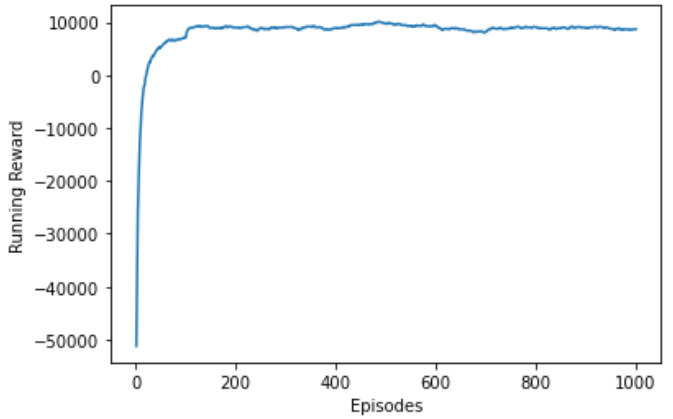


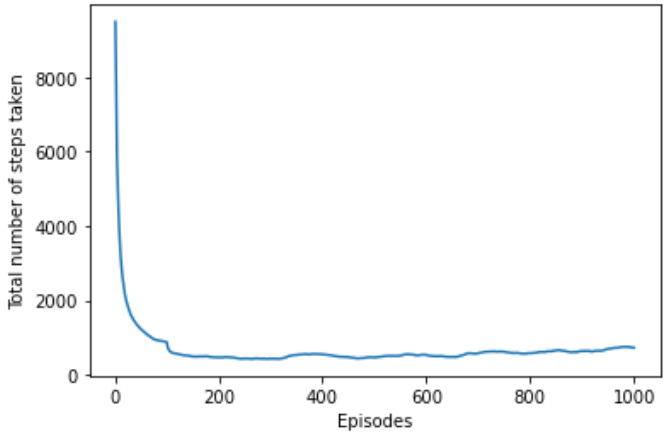
Illustrative

Experiments and Results II

- The second network has 101 subnets and 1444 hosts.
- The agent starts at host (44, 5) and tries to exfiltrate 10,000 MB data to the exit node (5, 10).
- One of host (5, 10)'s running services, HTTPS, is selected as the preferred exfiltration protocol.
- The RL policy converges in less than 200 episodes.
- The agent discovers and exploits host (24, 18), which is also running HTTPS, thus forming a single protocol-based path (44, 15) -> (24, 18) ->(5, 10).
- The entire payload is effectively exfiltrated without triggering alerts, as the agent has learned to take timely *sleep* actions between uploads.

Episode Rewards and Steps





Illustrative

Conclusion

Developed an RL-based method to automate the discovery process of exfiltration paths that incorporates protocol and payload considerations to account for nuances in adversarial behavior.

The strength of this approach is showcased through identification of intentional network misconfigurations that mimic real-world vulnerabilities.

Demonstrate that an RL agent can effectively discover an exfiltration path with maximum exfiltration protocol coverage and can perform exfiltration without being detected by firewalls.

The revealed attack paths provide insights for operators and analysts to assess present security measures, aiding them in crafting strategies for enhancing enterprise network security.

Deloitte.

This presentation contains general information only and Deloitte is not, by means of this presentation, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This presentation is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this presentation.

As used in this document, "Deloitte" means Deloitte & Touche LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.