部分観測マルコフ決定過程に基づいたニューラルエージェントを使用したペネトレーションテスト手法の提案

Proposal of a penetration testing method using neural agents based on partially observable Markov decision process

米田智紀・ネットワーク分科会・情報セキュリティ大学院大学

abstract: Penetration testing is a method of testing for security vulnerabilities by attempting to penetrate devices and systems using various techniques. In particular, autonomous penetration testing technology based on machine learning is important for achieving offensive security. The autonomous penetration testing technology based on machine learning is expected to become an important method to realize offensive security and to cope with the increasing and sophisticated cyber attacks. In particular, autonomous penetration testing based on machine learning is an important method to realize offensive security and cope with cyberattacks' growing number and sophistication. Various machine learning-based autonomous penetration testing tools have already been developed, such as Deepexploit and Gyoison. In particular, penetration testers based on reinforcement learning, which can autonomously acquire attack methods without preparing training data in advance, are attracting attention. This paper focuses on a reinforcement learning model based on a partially observed Markov decision process. The system response obtained in the penetration testing process is interpreted by neural natural language processing techniques to estimate the state of the system and the subsequent optimal attack behavior. LeDeepChef[3], published in 2020, proposes a neural agent that can efficiently find the goal of Textworld, a text-based dungeon game, by reinforcement learning based on the Partial Observation Markov Decision Process (POMDP). This paper explores the applicability of this neural agent to penetration testing of real systems by making it work in OS environments such as Windows/Linux and having exploit commands in the action set.

研究の背景 ・ 目的

■背景

近年、強化学習を基にしたペネトレーションテスタが注目されており、ASAPやDeepExploitなど様々な手法が提案されている。

既存の研究の特徴点

・MDPを基にした研究がほとんど

・スキャン情報からIPやOS情報、脆弱性情報を抽出して、それら情報を状態として学習を行っている

■目標

本研究では、POMDPを基にしたニューラルエージェントを使用して、コマンド実行時の結果から自然言語処理によって解釈し、状態を推定しながら、次の最適な攻撃コマンドを選択するシステムを提案する

POMDPとは

・マルコフ決定過程(MDP)ではエージェントはマルコフ性のある状態を常に観測できるとしているが、部分観測マルコフ決定過程(POMDP)では、状態を部分的にしか観測できず、観測状態がマルコフ性を満たすとは限らない状況を扱う確率過程である。

・POMDPにおいて最も重要な点はMDPとは異なり、エージェントは状態を観測できず、代わりに観測(observation)と呼ばれる値を環境から観測して、次状態を推定する点である。

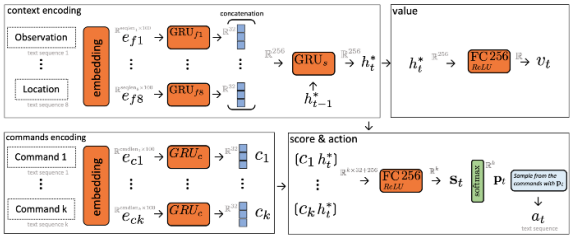
・POMDPを定義した式を下記に示す

$S = \{s^1 \dots s^N\}$: 状態(state)の有限集合
$A = \{a^1 \dots a^K\}$: 行動(action)の有限集合
$T: S \times A \times S \to [0,1]$: 条件付き状態遷移確率
$R: S \times A \to R$: 報酬関数
Ω: 観測の集合
$O: S \times A \times \Omega \to [0,1]$: 条件付き観測確率の集合

POMDPは状態更新関数$u: S \times A \times \Omega \to S$を用いて、次状態を次のように推定するプロセスとして定義される。この時、s_0は初期状態を表す。

$$s_{t+1} = u(s_t, a_t, a_{t+1}), \quad \text{for all } t \geq 0$$

ニューラルエージェント



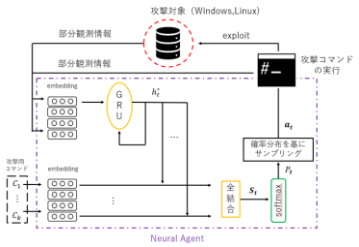
このシステムは実行結果等の部分観測情報と使用する攻撃コマンドを入力としている。embeddingによって分散表現を獲得し、GRUレイヤーにおいてベクトルを作成している。推定される次状態としては、context encodingの部分で出力されている隠れ状態h_t^*である。このh_t^*をFC層に渡すことで状態価値を算出している。また、 h_t^*とコマンドごとのベクトルを組み合わせてFC層に渡し、softmaxでコマンドのサンプリングを行い、行動となるa_tを決定している。

提案手法

提案手法のシステムはmetasploitのmsfconsole上で作用しており、nmapやsearchコマンドで抽出された攻撃コマンド（use exploit~やset rhostなど）とそれらコマンドの実行結果を入力としている。この提案手法の目標としては、windowsやlinuxシステムに対してexploitを成功させ、シェルやプロンプトを取得することである。

報酬設計

シェルの取得後、whoamiコマンドやgetidコマンドを実行し、権限としてrootを取得している場合、報酬を100与える。規定ステップ（2000）以内にシェルが取得できない場合に、報酬として-100を与えている。



今後の研究

今後の改善点としては、部分観測情報を増加させることあげられる。現在は実行結果が主な部分観測情報となっているが、OS情報やサービス情報やCVE情報も加えていきたい。また、スキャンコマンドも攻撃コマンドの一部として学習を行わせる機能を追加して、完全な自動ペンテスターを目指していきたい。