部分観測マルコフ決定過程によるニューラルエージェント強化学習を使用した自律型SQLインジェクション攻撃手法

The autonomous SQL injection exploitation using neural agent reinforcement learning by partial observation Markov decision process

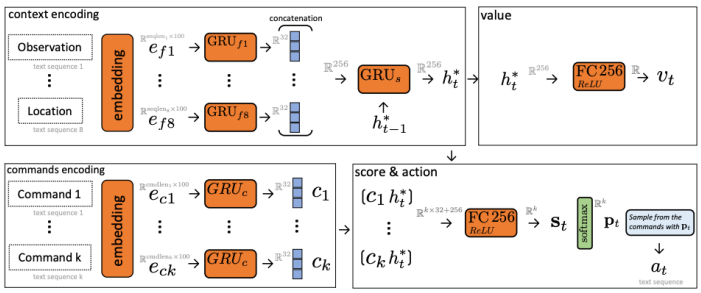
佐竹達也・ネットワーク分科会・情報セキュリティ大学院大学

In recent years, reinforcement learning approaches have been remarkable in various fields. Reinforcement Learning is one of machine learning methods but does not require teacher data which is a different point comparing with other machine learning methods. Agent repeat trial and error to maximize rewards by interacting with the environment and find the best policy. Reinforcement learning research trying to bring into the realm of cybersecurity is increasing. In this research, we constructed a reinforcement learning model to learn the optimal policy to solve SQL injection attack on CTF. The model is based on the Partially Observable Markov Decision Process (POMDP) and approximate state space and action space with distributed representation by deep reinforcement learning agent using state estimation by GRU and Attention. We successfully found out a better policy (mean steps is around 200/episode) to solve a SQL injection attack problem.

1.背景・目的

近年, 強化学習のアプローチは様々な分野で目覚ましい成果をもたらしてきた. 強化学習は,他の機械学習手法と異なり教師データを必要とせず, 環境との相互作用によってエージェントが試行錯誤を繰り返し報酬を最大化する最適な方策を発見する手法である.サイバーセキュリティの領域において,我々は強化学習を自律推論コンピューティングシステムとして応用する研究を進めている.その初期段階として,我々はCTF問題のSQLインジェクション問題を解く深層強化学習モデルを構築した.本研究の貢献は以下の3つである。1) CTFの実践環境(SECCON Beginners)に強化学習を適用させ実用性を示す. 2) POMDPに基づく深層強化学習エージェントが観測情報に応じて自律的に行動を生成する. 3) リカレントネットワークの一種であるGRUの自然言語処理技術を用いて実行可能行動の中から最適な行動を選択する.

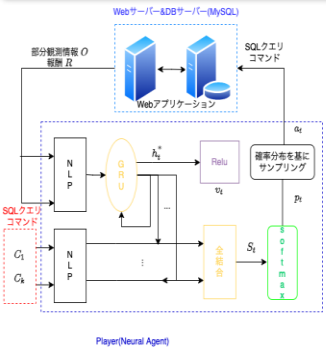
2. 先行研究(NLPニューラルエージェント)



テキストベースのダンジョンゲーム(Text-World)を自然言語文字列からなる観測情報をニューラル自然言語処理により分散表現に変換し, GRUで認識して推定状態を行う.

環境から与えられる行動集合も同様に分散表現の系列として表現し,推定状態と組み合わせてAttentionで最適な行動 を選択する.

3. 提案手法



部分観測情報

1	実行可能コマンド$A(s_t)$	状態s_t毎の実行可能行動の集合を表す行動空間
2	実行結果に関する描写	最後に行なったコマンドa_tの実行結果に関する具体的な描写
3	前のコマンド$a_{t-1} \in A(s_{t-1})$	最後にリクエストを送ったコマンド文字列
4	HTML	Webサーバからのレスポンス結果であるHTML要素の内,<>,(,),{,}, などの特殊文字以外を抽出した自然言語文字列
5	HTTPステータスコード	Webサーバからのレスポンス結果である3桁の文字コード
6	URL	WebアプリケーションのURL情報

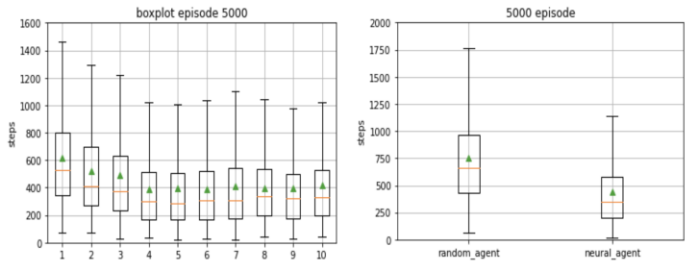
Webブラウザの出力結果(HTML)などに応じて新たな実行可能なSQLインジェクション攻撃コマンドを動的に生成し,行動空間を拡張する. phpファイルに記述されているFLAG文字列を取得することを目指す.

https://github.com/sataketatsuya/SQL_injection_pomdp

4. 評価実験・課題

Base SQL Query
SELECT login_id, id, auth_bit, last_time, comment
FROM users WHERE id = '{$_POST['id']}' AND password = '{$_POST['pw']}' ;

Injection Point



評価実験より,複雑なCTF問題に対してエージェントは,全てのエピソードにおいてもフラグを取得することができた.

ランダムエージェントより良い方策を獲得することができた.

2000エピソード以降,学習が停滞している.

我々が考える理論的な最適方策の期待値は138.5ステップに対して,訓練中の平均ステップ数は400ステップ周辺で停滞していることから最適方策 が獲得できていない.

5. まとめ

比較	先行研究	提案手法
確率過程	マルコフ決定過程	部分観測マルコフ決定過程
環境	モックデータベース	複雑なCTF問題の実践環境構築
状態S	既知	直接観測不可,部分的に観測
$\|A(s_t)\|$	51個	44→300個
モデル	Tabular Q-learning DQN	NLP技術を用いた深層強化学習