深層強化学習を用いた自律サイバー推論システムの研究

A research on the autonomous cyber reasoning system based on deep reinforcement learning

藤本大輔・ネットワーク分科会・情報セキュリティ大学院大学

As threats in cyberspace increase, advanced and sophisticated cyber attacks are increasing. In recent years, security using machine learning has attracted attention. Conventional machine learning security mainly learns communications that occurred in the past, and compares them with actual communications to classify them. However, when learning from only past communications, problems such as classifying unusual legitimate communications as malicious communications and not being able to deal with completely new attacks can occur. Therefore, a method called reinforcement learning that maximizes reward by trial and error without ground truth is promising. In this research, we aim to implement the function of autonomous cyber reasoning system using deep reinforcement learning.

1. 研究背景

サイバー空間における驚異の増大に伴い，従来のセキュリティ機器では対応できない高度で巧妙なサイバー攻撃が増大している．そこで近年注目されているのが，機械学習を用いたセキュリティである．従来の機械学習セキュリティは，過去の通信から学習させ，実際の通信と比較し分類を行うものが主流であった．しかし，悪意の有無に関わらず，珍しい通信は十分に学習が行われず，間違った分類をしてしまう可能性がある．

機械学習セキュリティとしてはあまり注目されていないが，訓練データを与えずとも試行錯誤により報酬を最大化する強化学習という手法が期待される．これにより，未知のサイバー攻撃にも対抗できる可能性がある．

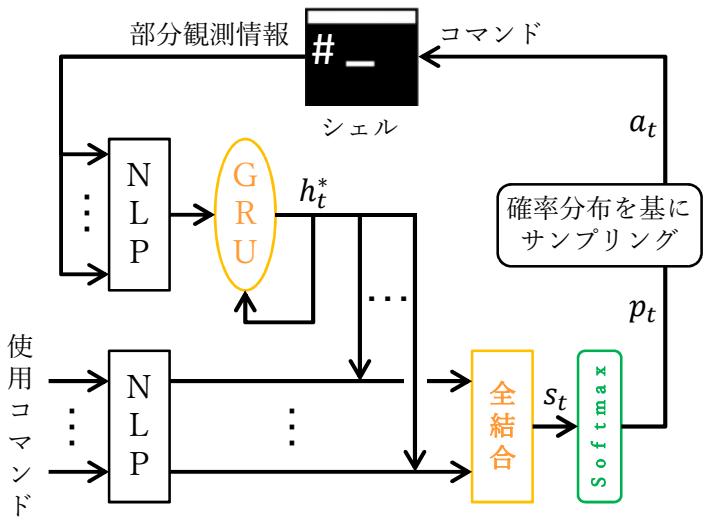
2. 研究目的

深層強化学習を用いて自律サイバー推論システムの機能を実装することである．サイバーセキュリティに関する行動をＡＩにより完全自動化することを究極の目標とし，まずはＣＴＦ解答を自動化する．

※サイバー推論システム：ソフトウェアの脆弱性検知，パッチ作成，エクスプロイト作成等を行うシステム

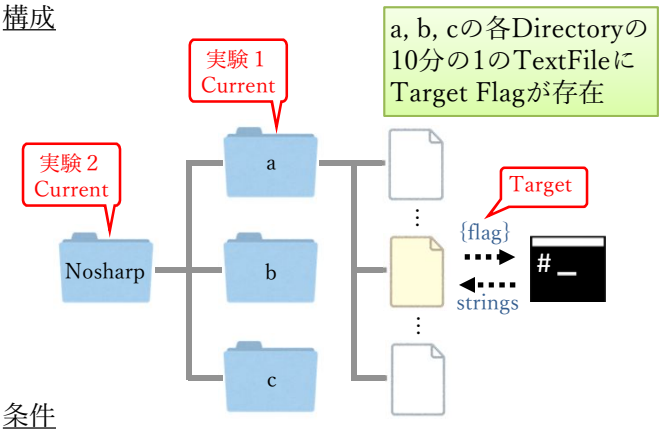
3. 提案手法

ＴｅｘｔＷｏｒｌｄをＣＴＦに適用する．具体的には，ＵＮＩＸのシェルと環境との対話をとおして得られた部分観測情報を状態として利用し，ニューラルエージェントに報酬を最大化する（最短経路でフラグを取得する）行動を学習させる．



4. 実験

構成

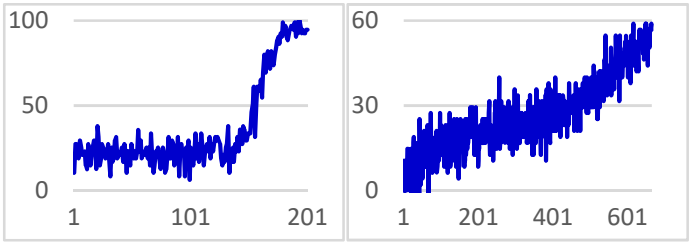


条件

使用コマンド	pwd，cd，strings
報酬	成功：＋100，失敗：-100
割引率 γ	実験１：0.01，実験２：0.8

結果

１００ステップ毎の平均期待報酬を下図に示す．実験１，実験２ともに過半数のエピソードにおいて最短経路でフラグを取得するようになった．



実験１　　実験２