DOI 10.37882/2223–2966.2020.11.26

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В ЗАДАЧЕ ТЕСТИРОВАНИЯ НА ПРОНИКНОВЕНИЕ

Мясников Алексей Владимирович
Санкт-Петербургский Политехнический Университет Петра Великого
myasnikovalexey@gmail.com

Аннотация. В рамках статьи рассмотрены вопросы применения машинного обучения с подкреплением к задаче тестирования на проникновение. Алгоритмы машинного обучения с подкреплением требуют определенного представления от среды, в которой они функционируют. В статье описан подход к представлению процесса тестирования на проникновение в терминах марковского процесса принятия решений, а также предложен подход к поиску оптимального пути атаки в рассмотренной модели с помощью методов машинного обучения.

Ключевые слова: машинное обучение, обучение с подкреплением, тестирование на проникновение, моделирование процесса тестирования на проникновение, марковский процесс принятия решений.

APPLICATION OF REINFOREMENT MACHINE LEARNING IN PENETRATION TESTING

A. Myasnikov

Summary. The article discusses the issues of applying reinforcement machine learning to the problem of penetration testing. Reinforcement machine learning algorithms require a specific representation of the environment in which they operate. The article describes an approach to representing the penetration testing process in terms of a Markov decision-making process, and also proposes an approach to finding the optimal attack path in the considered model using machine learning methods.

Keywords: machine learning, reinforcement learning, penetration testing, modeling of the penetration testing process, Markov decision making process.

Введение

Тестирование на проникновение — один из способов практической оценки безопасности цифровых активов путем проведения контролируемых атак на исследуемую систему. Процесс тестирования на проникновение связан с выполнением ряда технических задач, часть из которых поддаётся автоматизации.

Однако, на данный момент не существует подходов, которые бы автоматизировали тестирование на проникновение на всех шагах процесса. Этапы, связанные с проведением атак, требуют непосредственного контроля человека. Применение машинного обучения к задаче тестирования на проникновение является актуальной проблемой, так как позволяет автоматизировать шаги тестирования, связанные с продвижением в исследуемой сети.

Подходы с применением обучения с подкреплением

Системы, использующие искусственный интеллект, как правило подразделяются на два типа:

- Экспертные системы;
- Автоматизированные системы без учителя.

К экспертным системам принято относить такие продукты как антивирусы, файрволлы, IDPS и SIEM — системы. Они работают, опираясь на данные, подготовленные экспертами по безопасности (сигнатуры, правила распознавания и т.д.). Подобный подход приводит к высокому проценту ошибок в данных системах, а обучение с подкреплением позволяет создать автоматическую или полуавтоматическую контекстно-зависимую систему принятий решений.

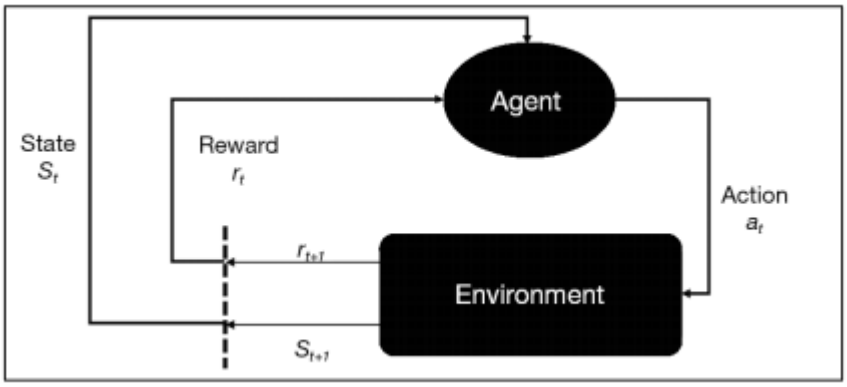

Рис. 1. Общая схема функционирования обучения с подкреплением.

Основными причинами, которые позволяют рассматривать обучение с подкреплением для задачи автоматизации тестирования на проникновение:

- Эффективное автономное обучение и улучшение результата через постоянное взаимодействие с окружающей средой;
- Обучение на базе функции вознаграждения. Контроль над функцией вознаграждения, который позволяет агенту обучения с подкреплением улучшать долгосрочные цели;
- Разнообразие сред для обучения с подкреплением позволяет отразить в модели такие свойства тестирования на проникновение как неопределенность и сложность.

Обучение с подкреплением один из подвидов машинного обучения. Оно позволяет программным агентам автоматически определить наиболее выгодную линию поведения в специфической среде с целью максимизировать производительность. Для обучения агенту достаточно получить значение функции награды. Таким образом, для построения и обучения модели достаточно иметь агента, среду и возможность контакта со средой. Взаимодействие агента со средой происходит через возможность агента получать состояние среды в каждый момент времени, затем в соответствии с некоторой стратегией выбирается некоторое взаимодействие на среду и ожидается обратный эффект от этого действия.[1]

На рисунке 1. изображено схематичное представление функционирования механизма обучения с подкреплением. Агент путем взаимодействия со средой, меняет состояние среды и получает некую награду. Таким образом на определенном шаге взаимодействий вырабатывается некая стратегия решения задачи, которая впоследствии может быть улучшена. Обучение с подкреплением имеет свойство сходимости к глобальному оптимальному значению, таким образом вырабатывается максимально эффективная стратегия принятия решений. [2]

Обучение с подкреплением исключает постоянный контроль эксперта-человека. Меньше временные затраты по сравнению с машинным обучением и экспертными системами. Кроме того, обучение с подкреплением активно развивающаяся область и новые алгоритмы для решения проблем обучения с подкреплением постоянно развиваются

Алгоритмы машинного обучения с подкреплением обучаются оптимальному решению через процесс взаимодействия со средой. Как правило процесс обучения начинается с некоторого начального состояния, которое для ряда задач может быть выбрано случайным образом. Взаимодействуя со средой из некоторого состояния s и совершая действие a, происходит переход и обновление весов Q(s, a).

Алгоритмы обучения с подкреплением отличаются выбором действий и обновлением весов. Существуют различные алгоритмы стратегии выбора действия, но наиболее применимы это UCB и e-greedy алгоритмы. Обе эти стратегии выбраны с целью соблюдения баланса между эксплуатацией уязвимостей и исследованием системы. Вариант, когда агент войдет в цикл бесконечной эксплуатации или сканирования неприемлем для системы. E-greedy алгоритм выбирает случайное действие с некоторой вероятностью ε, во всех остальных случаях выбираются оптимальные действия согласно текущей политики.

$$a_t = \begin{Bmatrix} \underset{a \epsilon A}{argmax}\, Q(a) \;\; with\, p(1-\varepsilon) \\ random\; a \epsilon A \; with\; p(\varepsilon) \end{Bmatrix}$$

Широко принятой практикой является использование ε-greedy алгоритмов с постепенным уменьшением значения ε.

Таблица 1. Представление модели ИС для использования с алгоритмами машинного обучения

Компонент	Определение
S	|M| x {скомпрометированные узлы} x {достижимые узлы} x |E| x {знание об уязвимых сервисах}
A	|M| x {сканирование, эксплуатация}
R(s`, a, s)	Награда(s`, s) — Стоимость_действия(a)
T(s`, a, s)	Неизвестно

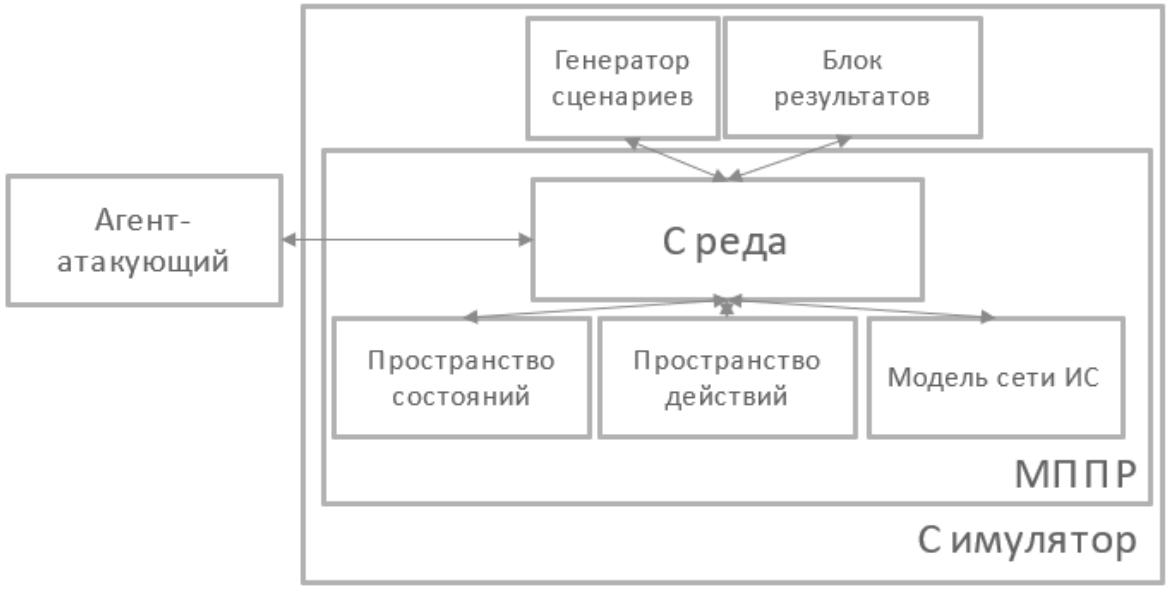

Рис. 2. Архитектура системы для обучения атакующих агентов

Моделирование сети

Одним из подходов к моделированию и планированию атак является использование марковского процесса принятия решений в средах симуляции. [3] Марковский процесс принятия решений — общий метод решения проблемы дискретного выбора в средах с неопределенностью. Марковский процесс задается следующим кортежем:

$\{S, A, R, T\}$, где S — множество состояний, A — множество действий, R — функция вознаграждения $R(s, a)$, T — множество функций перехода состояния $T(s, a, s`) = P(s`|s, a)$.

В любой момент времени система находится в некотором состоянии $s \in S$, а атакующий агент совершает некоторое действие a из множества A, которое приводит к двум исходам:

а) переход в некоторое состояние s`, определенный функцией перехода T;
б) получение награды R. (может быть отрицательным).

Цель агента-атакующего найти проекцию из множества состояний S на множество действий A, таким образом, чтобы максимизировать функцию вознаграждения. Данное решение называется оптимальной политикой решения π.

Для адаптации процесса тестирования на проникновение к данному представлению за множество состояний S можно взять все возможные конфигурации целевых машин в сети, множество действий — множество доступных атакующему действий (эксплуатация, сканирование), а функция вознаграждения — производная от стоимости действия и пользы, полученной в случае успешной компрометации системы.[4]

Существует несколько реализаций графов атаки с применением марковского процесса принятия решений. Один из подходов игнорирует конфигурацию рассматриваемой системы, неопределенность атакующего рассматривается в форме результата, который может получить атакующий. [5]

Построение плана атаки происходит за счет вероятности успеха применения эксплоита, основываясь на вероятности, полученной на предыдущих попытках применения рассматриваемого действия.[6]

Основным преимуществом данного подхода является возможность моделирования неопределенности атакующего с сохранением низкой вычислительной мощности. Однако, данный подход игнорирует конфигурацию системы, что является одним из ключевых знаний атакующего при проведении тестирования на проникновение. Так же данный подход предполагает, что мы заранее вычислили возможные исходы от применения атак (Модель переходов на графе), но данные вероятности напрямую зависят от типа атакуемой системы (например, от типа используемой ОС) и изменяются со временем.

Задача тестирования на проникновение с применением методов машинного обучения

Для того, чтобы иметь возможность использовать алгоритмы машинного обучения с подкреплением необходимо представить свойства процесса тестирования на проникновение в соответствии с марковским процессом принятий решений. Данная репрезентация задается в таблице 1. Так как методы обучения с подкреплением используются для поиска перехода с максимальной наградой, компонента T(s`, a, s) остается неизвестной.

На основе полученного отображения архитектура системы, позволяющей обучить агента для поиска оптимального атакующего пути, может быть представлена в виде архитектурной схемы (рис. 2).

Таким образом, процесс тестирования на проникновение может быть представлен в виде модели, на которой могут в дальнейшем быть использованы методы машинного обучения с подкреплением для поиска оптимального атакующего пути.

Заключение

В рамках данной статьи было рассмотрено применением методов машинного обучения с подкреплением к задаче тестирования на проникновение.

Процесс тестирования на проникновение был представлен в терминах марковского процесса принятия решений. Было предложена архитектура системы для обучения атакующих агентов на данной модели.

Дальнейшее направление исследований в данной области связано с обучением атакующих агентов и их связь с реальными действиями атакующего. Таким образом, агент-атакующий может выбирать какие из действий из данного состояния наиболее вероятно приведут к успеху и через интерфейс взаимодействия с ПО для тестирования на проникновения совершать действия в реальной системе.

ЛИТЕРАТУРА

1. M.G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, “The Arcade Learning Environment: An Evaluation Platform for General Agents” 1, vol. 47, pp. 253–279, Jun. 2013.
2. C. Szepesvari (2010). Algorithms of Reinforcement Learning. // Synthesis Lectures on Artificial Intelligence and Machine Learning.
3. R. Sutton, A. Barto Reinforcement Learning, second edition: An Introduction
4. Ghanem, M.C., & Chen, T. M. (2018). Reinforcement Learning for Intelligent Penetration Testing. // 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4).
5. Greenwald, L., & Shanley, R. (2009). Automated planning for remote penetration testing. // MILCOM 2009–2009 IEEE Military Communications Conference.
6. C. Sarraute, O. Buffet, and J. Hoffmann, “Penetration Testing == POMDP Solving?” [Электронный ресурс]. https://arxiv.org/pdf/1306.4714.pdf (Дата обращения: 10.10.2020)

© Мясников Алексей Владимирович (myasnikovalexey@gmail.com).
Журнал «Современная наука: актуальные проблемы теории и практики»