БЕЗОПАСНОСТЬ ЦИФРОВЫХ ТЕХНОЛОГИЙ. – 2024. – № 3 (114). – 21–33

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 004.056.5 DOI: 10.17212/2782-2230-2024-3-21-33

МУЛЬТИАГЕНТНОЕ ТЕСТИРОВАНИЕ НА ПРОНИКНОВЕНИЕ НА ОСНОВЕ AIRL*

М.М. ГРЕКОВ[1], А.А. СЫЧУГОВ[2]

[1] *300012, РФ, г. Тула, пр. Ленина, 92, Тульский государственный университет, ассистент кафедры информационной безопасности. E-mail: grekov.web@yandex.ru*
[2] *300012, РФ, г. Тула, пр. Ленина, 92, Тульский государственный университет, заведующий кафедрой информационной безопасности. E-mail: xru2003@list.ru*

В статье рассматривается применение мультиагентного подхода на основе метода Adversarial Inverse Reinforcement Learning (AIRL) для тестирования на проникновение в информационные системы. Описаны теоретические аспекты мультиагентного AIRL, включая возможности моделирования сложных и многоступенчатых атак, координации действий агентов, а также обучения с частичным наблюдением, что позволяет учитывать ограничения в доступе к информации. Практическое применение такого подхода продемонстрирует его эффективность в выявлении уязвимостей, обеспечивая более глубокий и точный анализ безопасности.

Ключевые слова: состязательное обучение с обратным подкреплением, информационная безопасность, тестирование на проникновение, автоматизация, мультиагентное обучение, частичное наблюдение, машинное обучение, нейронные сети

ВВЕДЕНИЕ

Информационная безопасность становится одной из ключевых задач в современном мире, где данные и цифровые ресурсы играют критическую роль в функционировании организаций, правительств и общества в целом. Угроза кибератак растет с каждым годом, и злоумышленники продолжают разрабатывать всё более сложные и изощренные методы компрометации систем. В связи с этим обеспечение безопасности информационных систем требует использования продвинутых методов и подходов, которые способны эффективно выявлять и устранять уязвимости до того, как они будут использованы в атаках.

* Статья получена 10 августа 2024 г.

Тестирование на проникновение (Penetration Testing, PT) – один из важнейших инструментов обеспечения информационной безопасности, направленный на обнаружение слабых мест в системах путем имитации реальных атак. Однако традиционные методы тестирования на проникновение сталкиваются с рядом ограничений. Во-первых, они зачастую предполагают сценарии атак, которые не учитывают сложность и динамичность современных кибератак, включающих координацию между несколькими участниками или использованием разнообразных тактик и стратегий. Во-вторых, традиционные методы требуют значительных затрат ресурсов и времени, особенно при тестировании крупных и комплексных сетевых инфраструктур. Эти ограничения требуют разработки новых методов и подходов, которые позволят повысить эффективность и глубину анализа систем безопасности.

Одним из перспективных направлений, способных решить указанные проблемы, является применение нейросетевых алгоритмов и методов машинного обучения. В частности, применение метода обратного подкрепления (Adversarial Inverse Reinforcement Learning, AIRL), который позволяет агентам обучаться на основе демонстраций, создавая оптимальные стратегии для достижения заданных целей [1, 2]. Мультиагентный подход в AIRL представляет собой дальнейшее развитие этого метода, при котором несколько агентов взаимодействуют друг с другом и с окружающей средой для моделирования сложных сценариев атак и тестирования на проникновение [3].

Таким образом, мультиагентный подход в AIRL представляет собой мощный и гибкий инструмент для тестирования на проникновение, который позволяет моделировать сложные и многоступенчатые атаки, координировать действия агентов, обучать их на основе частичной информации и использовать ансамбли агентов для повышения эффективности анализа. Этот подход имеет потенциал значительно улучшить безопасность информационных систем, позволяя обнаруживать и устранять уязвимости на более ранних этапах и с большей точностью. В настоящей статье рассматриваются теоретические аспекты мультиагентного AIRL, а также его практическое применение для тестирования на проникновение, что позволяет расширить возможности такого тестирования и сделать его более адаптивным, масштабируемым и эффективным в условиях современных угроз.

1. МУЛЬТИАГЕНТНЫЙ ПОДХОД AIRL

Применение мультиагентного AIRL для тестирования на проникновение открывает новые горизонты для анализа и выявления уязвимостей в информационных системах. Такой подход позволяет моделировать много-

ступенчатые атаки, в которых действия одного агента зависят от решений и поведения других агентов. Это особенно важно в условиях, когда злоумышленники координируют свои действия для преодоления защитных мер и достижения своих целей. Мультиагентные системы также способны адаптироваться к изменяющимся условиям, что позволяет им находить уязвимости, которые могут быть упущены при использовании традиционных методов.

Важной особенностью мультиагентного AIRL является возможность моделирования как кооперативных, так и конкурентных сценариев, что делает его особенно полезным для тестирования на проникновение. Кооперативные сценарии могут включать взаимодействие агентов для выявления и использования уязвимостей, тогда как конкурентные сценарии позволяют моделировать противодействие между несколькими атакующими агентами для формирования наиболее эффективных траекторий атак.

Многоуровневое обучение, являющееся одной из ключевых составляющих мультиагентного AIRL, позволяет разбивать сложные задачи на несколько уровней абстракции, что упрощает координацию действий агентов и позволяет более эффективно решать сложные задачи. Это особенно актуально в условиях тестирования на проникновение, когда необходимо моделировать атаки на различных уровнях системы – от сети до приложений и данных.

Еще одной важной характеристикой мультиагентного AIRL является возможность обучения с частичным наблюдением, что позволяет агентам принимать решения на основе неполной или неточной информации о системе. Это приближает модель к реальным условиям тестирования на проникновение, когда атакующие не всегда имеют полный доступ к информации о цели и должны действовать на основе ограниченных данных.

2. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

2.1. ФОРМАЛИЗАЦИЯ ВЗАИМОДЕЙСТВИЯ МЕЖДУ НЕСКОЛЬКИМИ АГЕНТАМИ

Мультиагентный подход в Adversarial Inverse Reinforcement Learning (AIRL) является расширением классического алгоритма AIRL, при котором вместо одного агента существует несколько агентов, взаимодействующих друг с другом и с окружающей средой. Такой подход особенно важен для моделирования сложных систем, в которых поведение одного агента зависит от действий других агентов.

В мультиагентных системах каждый агент принимает решения исходя из текущего состояния среды и действий других агентов. Это взаимодействие может быть для кооперативных, конкурентных или смешанных систем.

- Кооперативные системы. Агенты работают вместе для достижения общей цели. Примеры включают командные виды спорта и координированные задачи в робототехнике.
- Конкурентные системы. Агенты преследуют противоположные цели. Примеры включают игры, такие как шахматы или го, где выигрыш одного агента означает проигрыш другого.
- Смешанные системы. Содержат элементы как кооперации, так и конкуренции. Например, экономические модели, когда агенты могут сотрудничать для достижения некоторых целей, но конкурируют за ресурсы.

В мультиагентном AIRL каждый агент обучает свою политику, учитывая действия и стратегии других агентов. Это требует более сложного подхода к обучению, так как агенты должны адаптироваться к изменяющимся стратегиям своих коллег.

Каждый агент имеет свой дискриминатор, который помогает оценивать, насколько действия агента соответствуют демонстрациям. Дискриминаторы также помогают агентам различать истинные демонстрации от сгенерированных траекторий, способствуя обучению более реалистичных и эффективных стратегий.

Состояния и действия:

- S – множество всех возможных состояний среды;
- A_i – множество действий агента i;
- $A = (A_1, A_2, \ldots, A_n)$ – совместное множество действий всех агентов.

В мультиагентной системе состояние среды $s \in S$ является общим для всех агентов, но каждое действие $a_i \in A_i$ выполняется отдельным агентом i. Совместное действие всех агентов представляется вектором

$$a = (a_1, a_2, \ldots, a_n). \quad (1)$$

Каждый агент i имеет свою собственную политику $\pi_{\theta i}(a_i | s)$, которая определяет выбор действия a_i в зависимости от текущего состояния s. Политики всех агентов вместе образуют совокупную политику системы

$$\Pi = \{\pi_{\theta_1}, \pi_{\theta_2}, \ldots, \pi_{\theta_n}\}. \quad (2)$$

$P(s' \mid s, a)$ – функция перехода состояний, определяющая вероятность перехода из состояния s в состояние s' при выполнении совместного действия $a = (a_1, a_2, \ldots, a_n)$.

В мультиагентной системе каждый агент i имеет следующие компоненты.

1. Политика $\pi_{\theta_i}(a_i \mid s)$. Определяет вероятности выбора действия a_i в состоянии s для агента i. Параметризована вектором θ_i.

2. Функция вознаграждений $R_i(s, a_i, a_{-i})$. Вознаграждение агента i в состоянии s за действие a_i, учитывающее действия остальных агентов $a_{-i} = (a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$.

3. Дискриминатор $D_{\phi_i}(s, a_i)$. Оценивает вероятность того, что пара «состояние, действие» (s, a_i) агента i взята из истинных демонстраций.

Дискриминатор агента i обучается различать истинные демонстрации и траектории, сгенерированные текущей политикой агента:

$$D_{\phi_i}(s, a_i) = \frac{\exp\left(f_{\phi_i}(s, a_i)\right)}{\exp\left(f_{\phi_i}(s, a_i)\right) + \pi_{\theta_i}(a_i \mid s)}. \tag{3}$$

Функция потерь для дискриминатора:

$$L_{D_{\phi_i}} = E_{(s,a_i)\sim X_i}\left[\log D_{\phi_i}(s, a_i)\right] + E_{(s,a_i)\sim \pi_{\theta_i}}\left[\log\left(1 - D_{\phi_i}(s, a_i)\right)\right]. \tag{4}$$

Политика агента i обучается минимизировать функцию потерь, чтобы действия агента соответствовали истинным демонстрациям.

Функция вознаграждения для агента i:

$$R_i(s, a_i) = f_{\phi_i}(s, a_i) - \log \pi_{\theta_i}(a_i \mid s). \tag{5}$$

Функция потерь для политики:

$$L_G(\theta_i) = E_{\tau_i \sim \pi_{\theta_i}}\left[\sum_{t=0}^{T}\left(f_{\phi_i}(s_t, a_{i,t}) - \log \pi_{\theta_i}(a_{i,t} \mid s_t)\right)\right]. \tag{6}$$

Обновление параметров происходит итеративно.

1. Обновление дискриминатора:

$$\phi_i \leftarrow \phi_i + \eta_{\phi_i} \nabla_{\phi_i} L_{D_{\phi_i}} . \quad (7)$$

2. Обновление политики:

$$\theta_i \leftarrow \theta_i - \eta_{\theta_i} \nabla_{\theta_i} L_G(\theta_i) . \quad (8)$$

Мультиагентный подход в состязательном обучении с обратным подкреплением позволяет моделировать и обучать агентов, взаимодействующих друг с другом в сложных системах. Это расширяет возможности традиционного AIRL, делая его применимым к более широкому кругу задач, где поведение одного агента зависит от действий других агентов. В результате агенты могут обучаться более сложным стратегиям и взаимодействиям, что ведет к более реалистичному и эффективному поведению в реальных системах.

Применение мультиагентного подхода в контексте тестирования на проникновение может значительно усилить эффективность и глубину анализа безопасности сетевых систем. В этом контексте каждый агент может специализироваться на определенных аспектах безопасности, работая совместно для более комплексного и всестороннего тестирования.

2.2. ФОРМАЛИЗАЦИЯ МНОГОУРОВНЕВОГО ОБУЧЕНИЯ (HIERARCHICAL LEARNING)

Многоуровневое обучение (Hierarchical Learning) – это подход, который разбивает сложные задачи на несколько уровней абстракции или подзадач, что позволяет агентам решать их более эффективно. В контексте мультиагентного Adversarial Inverse Reinforcement Learning (AIRL) многоуровневое обучение может быть применено для улучшения координации и взаимодействия агентов [4, 5].

1. *Иерархические модели* делят задачу на высокоуровневые цели и низкоуровневые подзадачи. В мультиагентных системах это может означать разделение на стратегические (высокоуровневые) и тактические (низкоуровневые) задачи.

2. *Менеджеры и рабочие*. Высокоуровневые агенты (менеджеры) отвечают за определение целей и стратегий, в то время как низкоуровневые агенты (рабочие) выполняют конкретные действия для достижения этих целей.

В многоуровневом мультиагентном AIRL каждый агент имеет две политики:

- *высокоуровневая политика* $\pi_{\theta_i^H}(g \mid s)$: определяет цели или подзадачи g в состоянии s;
- *низкоуровневая политика* $\pi_{\theta_i^L}(a_i \mid s, g)$: определяет конкретные действия a_i для выполнения цели g в состоянии s.

Для каждого уровня политики существует свой дискриминатор:

- *дискриминатор высокоуровневой политики* $D_{\phi_i^H}(s, g)$: оценивает вероятность того, что цель g в состоянии s является частью истинных демонстраций;
- *дискриминатор низкоуровневой политики* $D_{\phi_i^L}(s, a_i \mid g)$: оценивает вероятность того, что действие a_i в состоянии s для цели g является частью истинных демонстраций.

Высокоуровневая политика:

$$\pi_{\theta_i^H}(g \mid s) = Pr\left(g \mid s;\ \theta_i^H\right). \tag{9}$$

Низкоуровневая политика:

$$\pi_{\theta_i^L}(a_i \mid s, g) = Pr\left(a_i \mid s, g;\ \theta_i^L\right). \tag{10}$$

Дискриминатор высокоуровневой политики:

$$D_{\phi_i^H}(s, g) = \frac{\exp\left(f_{\phi_i^H}(s, g)\right)}{\exp\left(f_{\phi_i^H}(s, g)\right) + \pi_{\theta_i^H}(g \mid s)}. \tag{11}$$

Дискриминатор низкоуровневой политики:

$$D_{\phi_i^L}(s, a_i \mid g) = \frac{\exp\left(f_{\phi_i^L}(s, a_i \mid g)\right)}{\exp\left(f_{\phi_i^L}(s, a_i \mid g)\right) + \pi_{\theta_i^L}(a_i \mid s, g)}. \tag{12}$$

Функция потерь для дискриминатора высокоуровневой политики:

$$L_{D_{\phi_i^H}} = E_{(s,g)\sim X_i}\left[\log D_{\phi_i^H}(s,g)\right] + E_{(s,g)\sim\pi_{\theta_i^H}}\left[\log\left(1 - D_{\phi_i^H}(s,g)\right)\right]. \quad (13)$$

Функция потерь для дискриминатора низкоуровневой политики:

$$L_{D_{\phi_i^L}} = E_{(s,a_i,g)\sim X_i}\left[\log D_{\phi_i^L}(s,a_i|\, g)\right] +$$

$$+ E_{(s,a_i,g)\sim\pi_{\theta_i^L}}\left[\log\left(1 - D_{\phi_i^L}(s,a_i|\, g)\right)\right]. \quad (14)$$

Функция вознаграждения для высокоуровневой политики:

$$R_i^H(s,g) = f_{\phi_i^H}(s,g) - \log\pi_{\theta_i^H}(g \mid s). \quad (15)$$

Функция потерь для высокоуровневой политики:

$$L_G^H\left(\theta_i^H\right) = E_{\tau_i\sim\pi_{\theta_i^H}}\left[\sum_{t=0}^{T}\left(f_{\phi_i^H}(s_t,g_t) - \log\pi_{\theta_i^H}(g_t|\, s_t)\right)\right]. \quad (16)$$

Функция вознаграждения для низкоуровневой политики:

$$R_i^L(s,a_i|\, g) = f_{\phi_i^L}(s,a_i|\, g) - \log\pi_{\theta_i^L}(a_i|\, s,g). \quad (17)$$

Функция потерь для низкоуровневой политики:

$$L_G^L\left(\theta_i^L\right) = E_{\tau_i\sim\pi_{\theta_i^L}}\left[\sum_{t=0}^{T}\left(f_{\phi_i^L}(s_t,a_{i,t}|\, g_t) - \log\pi_{\theta_i^L}(a_{i,t}|\, s_t,g_t)\right)\right]. \quad (18)$$

2.3. ФОРМАЛИЗАЦИЯ ОБУЧЕНИЯ С ЧАСТИЧНЫМ НАБЛЮДЕНИЕМ

Обучение с частичным наблюдением (Partially Observable Learning) относится к случаям, когда агенты не имеют полного доступа к состоянию среды и могут принимать решения только на основе частичной информации [6, 7]. В мультиагентных системах это добавляет дополнительный уровень слож-

ности, так как каждый агент должен принимать решения, учитывая как ограниченную информацию о среде, так и действия других агентов.

Частичные наблюдения o_i – наблюдения, которые доступны агенту *i*, зависят от истинного состояния среды *s* и могут содержать шум.

Политики агентов $\pi_{\theta_i}(a_i|\, o_i)$ – политики, которые агенты используют для принятия решений на основе своих наблюдений.

Дискриминаторы $D_{\phi_i}(o_i, a_i)$ – дискриминаторы, которые оценивают вероятность того, что пара «наблюдение, действие» является частью истинных демонстраций.

Агент *i* получает частичное наблюдение $o_i \in O_i$, которое связано с истинным состоянием среды $s \in S$ через наблюдательную модель *Z*:

$$o_i \sim Z(o_i|\, s)\,. \tag{19}$$

Политика агента *i* определяется как вероятностное распределение действий a_i на основе частичного наблюдения o_i:

$$\pi_{\theta_i}(a_i|\, o_i) = Pr(a_i|\, o_i;\ \theta_i)\,. \tag{20}$$

Дискриминатор агента *i* оценивает вероятность того, что пара «наблюдение, действие» $(a_i|\, o_i)$ взята из истинных демонстраций. Функция дискриминатора:

$$D_{\phi_i}(o_i, a_i) = \frac{\exp\left(f_{\phi_i}(o_i, a_i)\right)}{e\exp\left(f_{\phi_i}(o_i, a_i)\right) + \pi_{\theta_i}(a_i|\, o_i)}\,. \tag{21}$$

Функция потерь для дискриминатора:

$$L_{D_{\phi_i}} = E_{(o_i,a_i)\sim X_i}\left[\log D_{\phi_i}(o_i, a_i)\right] + E_{(o_i,a_i)\sim\pi_{\theta_i}}\left[\log\left(1 - D_{\phi_i}(o_i, a_i)\right)\right]. \tag{22}$$

Функция вознаграждения агента *i* основана на дискриминаторе:

$$R_i(o_i, a_i) = f_{\phi_i}(o_i, a_i) - \log\pi_{\theta_i}(a_i|\, o_i)\,. \tag{23}$$

Функция потерь для политики агента *i*:

$$L_G(\theta_i) = E_{\tau_i\sim\pi_{\theta_i}}\left[\sum_{t=0}^{T}\left(f_{\phi_i}(o_{i,t}, a_{i,t}) - \log\pi_{\theta_i}(a_{i,t}|\, o_{i,t})\right)\right]. \tag{24}$$

ЗАКЛЮЧЕНИЕ

Мультиагентный AIRL позволяет моделировать сложные и многоступенчатые атаки, где несколько агентов взаимодействуют друг с другом и с окружающей средой. Этот подход значительно расширяет возможности тестирования на проникновение, предоставляя более реалистичные сценарии и позволяя учитывать динамику атак, которая часто остается незамеченной при использовании традиционных методов. Благодаря возможности координации действий между агентами мультиагентные системы могут выявлять уязвимости, которые были бы недоступны для одного агента, действующего в изоляции.

Одной из ключевых особенностей мультиагентного AIRL является использование многоуровневого обучения, что позволяет делить сложные задачи на иерархические уровни и обеспечивать более эффективную координацию действий агентов. Это особенно важно в условиях тестирования на проникновение, где необходимо учитывать разнообразные аспекты системы – от сети до приложений и данных. Многоуровневое обучение также упрощает решение задач, связанных с моделированием атак на разных уровнях абстракции, что делает тестирование на проникновение более комплексным и точным.

Обучение с частичным наблюдением, также реализуемое в рамках мультиагентного AIRL, позволяет агентам принимать решения на основе неполной информации о среде. Это приближает процесс тестирования на проникновение к реальным условиям, в которых злоумышленники не всегда имеют полный доступ к данным о цели. Такой подход делает анализ более реалистичным и эффективным, поскольку агенты могут адаптироваться к неопределенности и принимать обоснованные решения даже при наличии ограниченной информации.

Таким образом, мультиагентный подход в AIRL представляет собой мощный и гибкий инструмент для тестирования на проникновение, который значительно расширяет возможности анализа и выявления уязвимостей в современных информационных системах. Применение этого подхода позволяет повысить точность, адаптивность и глубину анализа, что в конечном итоге способствует улучшению безопасности систем и снижению рисков кибератак. В условиях быстро меняющегося ландшафта угроз мультиагентный AIRL является перспективным направлением, которое может существенно изменить подходы к обеспечению информационной безопасности и стать основой для будущих исследований и разработок в области тестирования на проникновение.

СПИСОК ЛИТЕРАТУРЫ

1. *Fu J., Luo K., Levine S.* Learning robust rewards with adversarial inverse reinforcement learning // arXiv e-prints. – 2017. – arXiv: 1710.11248.

2. *Sychugov A., Grekov M.* Automated penetration testing based on adversarial inverse reinforcement learning // 2024 International Russian Smart Industry Conference (SmartIndustryCon). – IEEE, 2024. – P. 373–377.

3. *Yu L., Song J., Ermon S.* Multi-agent adversarial inverse reinforcement learning // Proceedings of Machine Learning Research. – 2019. – Vol. 97: Proceedings of the 36th International Conference on Machine Learning, Long Beach, California. – P. 7194–7201.

4. *Chen J., Lan T., Aggarwal V.* Hierarchical adversarial inverse reinforcement learning // IEEE Transactions on Neural Networks and Learning Systems. – 2023. – DOI: 10.1109/TNNLS.2023.3305983.

5. Multi-task hierarchical adversarial inverse reinforcement learning / J. Chen, D. Tamboli, T. Lan, V. Aggarwal // Proceedings of Machine Learning Research. – 2023. – Vol. 202: Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii. – P. 4895–4920.

6. Adversarial reinforcement learning under partial observability in autonomous computer network defence / Y. Han, D. Hubczenko, P. Montague, O. De Vel, T. Abraham, B.I.P. Rubinstein, C. Leckie, T. Alpcan, S. Erfani // 2020 International Joint Conference on Neural Networks (IJCNN). – IEEE, 2020. – P. 1–8. – DOI: 10.1109/IJCNN48605.2020.9206634.

7. *Choi J.D., Kim K.E.* Inverse reinforcement learning in partially observable environments // Journal of Machine Learning Research. – 2011. – Vol. 12 (1). – P. 691–730.

Греков Михаил Михайлович, ассистент кафедры информационной безопасности Тульского государственного университета. Основное направление научных исследований – применение машинного обучения и нейронных сетей в области информационной безопасности, тестирование на проникновение. E-mail: grekov.web@yandex.ru

Сычугов Алексей Алексеевич, заведующий кафедрой информационной безопасности Тульского государственного университета. Область научных интересов – методы и алгоритмы оперативного обнаружения опасных состояний промышленных объектов, информационная безопасность. E-mail: xru2003@list.ru

DOI: 10.17212/2782-2230-2024-3-21-33

Multiagent penetration testing based on AIRL*

M.M. Grekov[1], A.A. Sychugov[2]

[1] *Tula State University, 92 Lenina Avenue, Tula, 300012, Russian Federation, assistant at the Department of Information Security. E-mail: grekov.web@yandex.ru*
[2] *Tula State University, 92 Lenina Avenue, Tula, 300012, Russian Federation, Head of the Department of Information Security. E-mail: xru2003@list.ru*

This paper explores the application of a multi-agent approach based on the Adversarial Inverse Reinforcement Learning (AIRL) method for penetration testing in information systems. Theoretical aspects of multi-agent AIRL are discussed, including the ability to model complex, multi-stage attacks, coordinate agent actions, and learn with partial observability, which accounts for limitations in information access. The practical application of this approach will demonstrate its effectiveness in identifying vulnerabilities, providing a deeper and more accurate security analysis.

Keywords: Adversarial Inverse Reinforcement Learning, information security, penetration testing, automation, multi-agent learning, partial observation, machine learning, neural networks

REFERENCES

1. Fu J., Luo K., Levine S. Learning robust rewards with adversarial inverse reinforcement learning. *arXiv e-prints*, 2017, arXiv: 1710.11248.

2. Sychugov A., Grekov M. Automated penetration testing based on adversarial inverse reinforcement learning. *2024 International Russian Smart Industry Conference (SmartIndustryCon)*. IEEE, 2024, pp. 373–377.

3. Yu L., Song J., Ermon S. Multi-agent adversarial inverse reinforcement learning. *Proceedings of Machine Learning Research*, 2019, vol. 97: *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, California, pp. 7194–7201.

4. Chen J., Lan T., Aggarwal V. Hierarchical adversarial inverse reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. DOI: 10.1109/TNNLS.2023.3305983.

5. Chen J., Tamboli D., Lan T., Aggarwal V. Multi-task hierarchical adversarial inverse reinforcement learning. *Proceedings of Machine Learning Research*, 2023, vol. 202: *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, Hawaii, pp. 4895–4920.

* *Received 10 August 2024.*

6. Han Y., Hubczenko D., Montague P., De Vel O., Abraham T., Rubinstein B.I.P., Leckie C., Alpcan T., Erfani S. Adversarial reinforcement learning under partial observability in autonomous computer network defense. *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9206634.

7. Choi J.D., Kim K.E. Inverse reinforcement learning in partially observable environments. *Journal of Machine Learning Research*, 2011, vol. 12 (1), pp. 691–730.

Для цитирования:

Греков М.М., Сычугов А.А. Мультиагентное тестирование на проникновение на основе AIRL // Безопасность цифровых технологий. – 2024. – № 3 (114). – С. 21–33. – DOI: 10.17212/2782-2230-2024-3-21-33.

For citation:

Grekov M.M., Sychugov A.A. Mul'tiagentnoe testirovanie na proniknovenie na osnove AIRL [Multi-agent penetration testing based on AIRL]. *Bezopasnost' tsifrovykh tekhnologii = Digital Technology Security*, 2024, no. 3 (114), pp. 21–33. DOI: 10.17212/2782-2230-2024-3-21-33.