DOI 10.36622/1682-7813.2024.27.3.007
УДК 004.056.5:378.3

ИССЛЕДОВАНИЕ СПЕЦИФИКИ МОДЕЛИРОВАНИЯ КОМПЬЮТЕРНЫХ АТАК С ИСПОЛЬЗОВАНИЕМ МАРКОВСКИХ ПРОЦЕССОВ ПРИНЯТИЙ РЕШЕНИЙ И Q-ОБУЧЕНИЯ

В.В. Подтопельный

Рассматриваются проблемы, возникающие при решении задачи построения вектора атаки в сетевой инфраструктуре. Приведены и охарактеризованы разновидности различных методов моделирования атак с использованием методологии ФСТЭК, применяемых при построении вектора сетевой атаки, а также рассмотрена специфика их взаимосвязей с использованием марковских процессов принятия решений при моделировании атакующих воздействий. Рассмотрена их пригодность для различных процедур определения параметров вектора. При построении вектора сетевой атаки приводятся особенности определения стратегий по Q-обучению. Изучается формирование вектора атаки в контексте эксплуатации уязвимостей информационной системы в реальном времени. Определяются особенности построения вектора атаки с учетом специфики связей тактик (состояний).

Ключевые слова: сетевая атака, уязвимость, марковские процессы моделирование, стратегия, политика, метод обучения

Введение

При исследовании аспектов безопасности информационных систем используют разные способы определения уровня защищенности. Особое место занимает анализ защищенности систем с применением моделирования компьютерных атак как совокупности атакующих воздействий. Следует отметить, что не все из существующих методов одинаково применимы к задачам анализа атак в процессе эксплуатации информационных систем. Существует системы, которые не всегда позволяют реализовать приёмы пен-теста. Отсюда следует, что при аудите информационной безопасности определение возможных путей (векторов в терминологии ФСТЭК) развертывания атак предполагает построение сценариев, которые могут создаваться и анализироваться как на формальном уровне, в виде моделирования, так и в практической форме (пен-тест) [1]. При этом следует отметить, что определение сценариев атакующих воздействий возможно без учёта конкретного периода времени, которое требуется на проведение атакующих мероприятий, на противодействие им или для блокировки каналов связи. Подобное необходимо в том случае, если моделирование атаки, построение сценария нападения реализуется для выявления эффективных путей атакующих воздействий злоумышленника с целью их превентивной блокировки. Кроме того, анализ атакующих воздействий в режиме реального времени не обязательно предполагает определение наиболее эффективного пути развёртывания атаки по критерию наименьших затрат времени, поскольку действия злоумышленника не всегда подчиняются логике поиска наилучшего способа нападения в силу ограниченности квалификации злоумышленника и доступных ему средств нападения.

Таким образом, предполагается наличие двух типов моделирования с точки зрения потребности аудита:

1. Первый тип, позволяет рассматривать последовательность атакующих воздействий во времени с целью определения порядка блокирования атакующих воздействий как мер, которые разворачиваются либо параллельно действиям злоумышленника, либо в процессе выявления последовательности атакующих воздействий.

2. Второй тип аудита предполагает рассмотрение сценария вне контекста противодействия в режиме реального времени. Такой сценарий ориентируется на выявление эффективного пути достижения

цели злоумышленника с последующим превентивным интегрированием мер защиты с целью изоляцией или уничтожения найденного пути нападения.

Существует разные способы моделирования, которые позволяют построить и проанализировать сценарии атаки:

1. Регрессионные модели имеют среднюю сложность построения и дают количественную форму прогноза, основанную на математическом анализе данных. Такие модели эффективны при наличии большого количества данных и четких зависимостей между переменными.

2. Авторегрессионные модели характеризуются высокой сложностью построения, но также дают количественную форму прогноза. Они подходят для прогнозирования временных рядов с четко выраженными закономерностями.

3.Экспертные оценки характеризуются низкой сложностью построения модели, что делает их доступными для применения. Они дают качественную форму прогноза, основанную на логическом анализе и обработке экспертной информации. Однако они могут быть субъективными и не всегда учитывать все факторы, влияющие на прогноз.

В целом можно выделить множество методов (деревья атак, различные способы, связанные с использованием машинного обучения, использование байесовских сетей доверия, сети Петри Маркова, нечетких множеств; теория игр, теория графов, теория случайных процессов), которые использовались для построения сценариев нападения (векторов атак).

Для задачи определения вектора атаки наиболее часто сейчас используют модель на базе Марковских процессов принятия решений (МППР) поскольку она позволяет учесть фактор неопределенности: марковские процессы позволяют моделировать среду, где результаты действий не всегда предсказуемы. Это особенно важно при анализе компьютерных атак, где атакующие могут изменять свои тактики и стратегии. Следует отметить, что учитываются в данном случае следующие факторы, влияющие на анализ атак:

1. Динамический характер процесса. Атаки представляют собой динамический процесс, где существует зависимость между состоянием системы и воздействием на нее. Марковские модели, обладая высокой степенью адаптивности, позволяют эффективно учитывать эту зависимость. Изменения в модели атаки при смене стратегии поведения злоумышленника во время нападения можно отслеживать. Модели могут адаптироваться к новой информации по мере ее поступления, что позволяет постоянно изучать и обновлять политику на основе результатов предыдущих решений. Это особенно полезно в контексте растущих киберугроз, появлении новых атак (моделирование вероятностных переходов, оптимизация политик действий, структурированное представление состояний).

2. Использование вероятностного подхода. В сфере ИБ часто присутствует неопределенность, связанная с поведением злоумышленников. Он позволяет учитывать вероятности различных событий и оценивать риски при различных стратегиях принятия решений. Это помогает оптимизировать защиту от компьютерных атак и минимизировать потенциальные угрозы. С другой стороны, позволяет найти оптимальную последовательность атакующих воздействий. При это МППР не теряет свойства структуриванности в описании последовательностей действий злоумышленника. Кроме того, в этом случае можно отметить адаптивность МППР в задачах построения модели атак на системы ИС, которая проявляется через обновление моделей на основе новых данных, учет вероятностных переходов, интеграцию с методами обучения, моделирование многоэтапных атак и оптимизацию стратегий реагирования.

3. Возможность учета ранее неизвестных факторов. Марковские модели позволяют включать в анализ широкий спектр факторов, влияющих на вектор атаки, такие как характеристики системы, действия злоумышленника, защитные меры.

4. Четкая структура. В рамках МППР четко определены состояния, действия и вознаграждения, что упрощает анализ и

информирование о процессе принятия решений. Такой структурированный подход помогает понять сложные взаимодействия внутри системы. Это важно для эксперта, составляющего сценарий атак тем, что позволяет описать предметную область и одновременно учитывать фактор неопределённости (облегчает процесс понимания и анализа новых угроз, позволяя более эффективно разрабатывать стратегии защиты.).

Другие методы также обладают своими отрицательными и положительными сторонами в решении рассматриваемых задач:

1. Деревья позволяют представить атаку как иерархическую структуру (каждый узел соответствует определенной стадии атаки, а ветви — возможным действиям злоумышленника). Это позволяет реализовать визуализацию последовательности действий злоумышленника и упростить анализ атак. Однако при многоуровневых атаках в отличии от МППР возникают сложности в учете всех состояний и переходов для многоуровневых атак, могут не учитывать вероятностные аспекты и неопределенности.

2. Методы машинного обучения способны обрабатывать большие объемы данных и выявлять аномалии, что может быть полезно для предсказания атак, кроме того позволяют автоматизировать процесс обнаружения угроз. Однако машинное обучение требует больших объемов данных для обучения и сильно зависимо от их качества, что может быть проблемой в условиях ограниченных ресурсов, массивов доступных данных, также результаты могут быть трудными для интерпретации, в то время как MDP предлагают более понятную структуру.

3. Байесовские сети учитывают вероятностные зависимости и неопределенности, что делает их мощным инструментом для анализа сложных сценариев. Позволяют оценивать влияние различных факторов на уровень риска. Однако по сравнении с МППР формирование и настройка байесовских сетей может быть сложным и требовать значительных ресурсов, в том числе вычислительных ресурсов ограничить их применение по сравнению с более простыми моделями MDP.

4. Сети Петри подходят для анализа систем с параллельными процессами, что может быть полезно при моделировании атак.и кроме того, позволяют исследовать временные аспекты атак, что может улучшить понимание динамики угроз. Однако они могут быть сложны для интерпретации, тогда как МППР имеют более интуитивно понятную структуру. Кроме того, при работе с моделями на основе Сетей Петри присутствует необходимость тщательного описания всех возможных состояний системы.

5. Нечеткие множества позволяют учитывать неопределенности и субъективные оценки рисков, что может быть полезно в условиях недостатка информации. Однако, могут потребовать больше вычислительных ресурсов для обработки нечетких данных, а результаты могут быть трудными для интерпретации, в отличие от более четких вероятностных выводов в MDP.

6. Теория игр позволяет учитывать поведение как атакующих, так и защищающихся, что может дать более полное представление о сценарии атаки. Однако модели теории игр могут быть сложными при реализации и, самое важное в этом случае, требуют точных данных о поведении противника, чего может не хватать при прогнозировании действий потенциального злоумышленника, тем более, что рассчитывать на точные данные сложно при вероятностном анализе и построения надежной модели

7. Теория графов позволяет визуализировать сложные взаимосвязи между компонентами системы и действиями злоумышленников. При этом модели на основе графов могут сталкиваться с проблемами масштабируемости при анализе больших систем и не всегда учитывает временные аспекты атак, что ограничивает ее применение в динамичных сценариях.

8. Модели случайных процессов хорошо подходят для анализа непредсказуемых изменений в системе. Но подобные модели могут быть сложными для построения и интерпретации при анализе действий злоумышленника, что затрудняет их

использование на практике, и также требуют значительных вычислительных ресурсов для анализа.

Необходимо отметить, что несмотря на явные преимущества в задачах определения сценария атак, МППР имеют некоторую сложность применения: настройка модели подразумевает детальное понимания системы, включая определение состояний, действий и вероятностей переходов. Эта сложность может быть препятствием для реализации, особенно в больших и динамичных системах. Однако в контексте моделирования атак на ИС и составления сценариев атак специалистом в области ИБ при аудите этот недостаток нивелируется конкретизацией задач и введением уровней детализации ИС (по примеру нотаций IDEF0): аудитор сосредотачивается на системных компонентах и уязвимостях отдельного сегмента ИС или на определённых типах угроз (предполагаемых типах атак), при более высоком уровне абстрагирования каждый сегмент представляет собой компонент ИС более высокого порядка. Это органично вписывается в методику составления сценариев атак, учитывая сегментированность современных информационных систем и систем. Этим же компенсирует и другой недостаток метода МППР, который заключается в том, что по мере увеличения числа состояний и действий вычислительные ресурсы, необходимые для решения задач МППР, могут значительно возрасти.

Таким образом, в задачах аудита и составления моделей атак (в частности, формирования сценариев атак потенциальных злоумышленников как последовательности атакующих воздействий посредством эксплуатации уязвимостей, когда требуется выявить этапы продвижения злоумышленника по степени компрометации системы (состояниями, при которых нарушитель получает возможности осуществить действия для повышения степени компрометации целевой системы)), такие особенности метода МППР, как четкая структурированность (МППР обеспечивают четкую структуру для описания состояний системы и возможных действий), простота реализации, учет вероятности атакующих воздействий (позволяет моделировать неопределенности и риски, связанные с атаками), возможность определения оптимальной стратегии злоумышленника являются выгодными.

Вопрос использования Марковских процессов для исследования компьютерных атак рассматривается в ряде научных работ [2-4], что доказывает их применимость для решения задач информационной безопасности.

При этом следует отметить специфику анализа. Построение модели предполагает изучение особенностей уязвимостей программно-аппаратного обеспечения, топологические особенности информационных систем (если они распределённого типа), модели нарушителей, существующие способы и механизмы эксплуатации уязвимостей. Таким образом, прежде чем определить порядок реализации атак, предварительно исследуются целевые информационные системы (ЦИС), специфика взаимосвязей объектов в ней. Поскольку ЦИС бывают разного типа, то атакуемые действия рассматриваются как последовательность эксплуатации уязвимостей, то есть как этапы работы злоумышленника для достижения целевого состояния, которое означает достижения успеха атаки. Таким образом, атаки могут быть представлены как перемещение злоумышленника, наблюдаемого через фиксацию атакующих воздействий, между узлами сети, на которых были обнаружены уязвимости и которые были использованы для достижения целевого объекта. Так формируется вектор атаки.

Однако уязвимости разного типа, позволяющие продвинуться злоумышленнику на этапе компрометации или захвата ресурсов, при эксплуатации могут предполагать, как локальные, так и межузловые действия злоумышленников. Учитывая эти факторы, сценарий атаки (последовательность действий злоумышленника) в современной методологии построения вектора атак (по Методика ФСТЭК, далее - Методика) предлагается рассматривать не как перемещение между узлами, а как перемещения между состояниями, которые

маркируют повышение уровня компрометации (появление новых возможностей у злоумышленника) системы (объекта системы) от начального состояния атаки (Тактика 1(Разведка)) до конечного состояния, которое определяется как достижение целей злоумышленника (Тактика 10) [1]. Таким образом, действия при атаке соответствуют техникам и их тактикам. Реализация тактики рассматривается как успех (наступление некого состояния) в последовательности этапов повышения степени компрометации. Таким образом анализ защищённости информационной системы представлен как поиск наилучшей последовательности атакующих воздействий (наилучшего вектора атаки) злоумышленника направленных на повышение компрометации ЦИС. Следует отметить, что на сегодняшний момент есть несколько программных решений, которые позволяют построить вектор атак, например, MulVal. Однако в основном, эти решения используют на основе логики описания, например, логические выражения на языке Datalog. Подобные подходы не учитывают вероятностные показатели атак в полной мере в сопряжении с методологий их описания по Методике ФСТЭК.

Определение ограничений при моделировании атак

Следует отметить, что в произвольный момент времени атакуемая система может находиться в любом из десяти состояний, которые образуют пространство состояний [1]. Подобный подход удобен при рассмотрении посредством марковских процессов принятия решений (МППР) последовательности атакующих воздействий как череды смены состояний и, одновременно, позволяет отследить их сопряжённость (для этого достаточно применения системы линейных уравнений). Необходимо подчеркнуть сложную природу состояний, поскольку они учитывают с одной стороны, тип Тактики (который приведён в методологии ФСТЭК), а, с другой стороны, поскольку данные состояния достигаются при успешной эксплуатации уязвимости (узлов сети, их программного обеспечения), в пространстве состояний также должны учитываться специфика процесса достижения успеха, то есть действия, которое позволяет эксплуатировать уязвимость, а также сама уязвимость.

Следует учитывать то, что атакующие воздействия могут принадлежать к различным последовательностям, реализуемым разными злоумышленниками в один и тот же период времени. Воздействие на один и тот же узел (объект системы) повышает уровень его компрометации (уязвимости), и, следовательно, вероятность использования узла в сети как промежуточного объекта в последовательности атак. Таким образом, вероятность поражения какого-либо узла может являться суммой результатов различных действий, направленных на эксплуатацию уязвимости или множество уязвимостей, принадлежащих одному объекту информационной системы. Также следует учитывать то, что состояние компрометации в последовательности сценария атаки может достигаться с помощью атакующих воздействий, результаты которых, между собой не всегда сопрягаются. В этом случае суммирование результатов эксплуатации уязвимостей при одновременном наблюдении атак, принадлежащих разным классам, недопустимо. При рассмотрении сценариев атак следует учитывать природу эксплуатируемых уязвимостей, требуется определить их класс и уже на основе совпадения классов эксплуатируемых уязвимостей решать: стоит ли суммировать или нет результаты эксплуатации уязвимости.

Таким образом для рассмотрения атакующих воздействий нужно вести некоторые правила и некоторые ограничения:

1.Логика описания атак до начала реальных нападений по методологии ФСТЭК подразумевает, что если злоумышленники не достигают цели промежуточной компрометации, то есть отсутствует фиксация успешного достижения состояния, соответствующего тактике, которая учтена в сценарии, то ветвь пути, на котором встретилась эта тактика, отбрасывается, потому что состояние недостижимо по предложенному варианту сопряжения

состояния, уязвимости и действия, которое обеспечивает реализацию этого состояния, как этапа атаки, на котором появляется возможность осуществить набор новых компрометирующих действий, советующих достигнутому состоянию (этапу) системы. Следуя этой логике, возвраты в графе в предыдущее состояние не обязательны (это допустимо при отсутствии реального режима времени, когда фиксируется начало развёртывания атаки).

2. Рассматриваются состояния как некое комплексное описание всей системы, в которых она может пребывать. При этом надо учитывать, что это комплексная система состоит из множества различных отдельных узлов сети, которые могут включать как множество уязвимостей, так и одинаковые уязвимости, то есть те, которые классифицируются одинаково: допустим, такие уязвимости могут присутствовать в операционной системе которые были изначально закуплены по одной лицензии и установлены в компьютерах определённой организации.

3. Нужно классифицировать уязвимости не только на уровне переходов от одной тактики к другой, но ещё нужно классифицировать те тактики, которые между собой могут быть сопряжены в рамках повышения мощности одного состояния. С вниманием следует относится к вопросу накопления мощности состояния компрометации в каком-то определённом диапазоне, если при этом существует множество различных уязвимостей, которые ведут к одному и тому же состоянию в локальной или сетевой системе: суммарно они могут дать высокую вероятность промежуточных компрометации (реализации какой-то промежуточной тактики по Методике ФСТЭК).

Если аудитор рассматривает каждую уязвимость отдельно, маркируя её принадлежность к какой-либо тактике (состоянию), то фиксируется сопряжение уязвимостей через определённые действия без учёта их общей суммарной мощности и без общего состояния, которое описывает в целом всю систему (то есть система может побывать в каком-либо определённом состоянии). Это удобно при возникновении потребности выявления отдельного пути по конкретным уязвимостям принадлежащим отдельным узлам и программному обеспечению.

Таким образом возникает два представления о переходах в МППР при моделировании атак:

1. Передвижение злоумышленника с помощью действий по уязвимостям, которые рассматриваются в режиме реального времени развертывания атаки (on-line). При этом переходы по уязвимостям (то есть их эксплуатация) будут многочисленны: они будут иметь возвратные состояния при возникновении контрмер или иных препятствий при развертывании вектора атаки (рис. 1). Тогда теряется последовательность переходов между тактиками, как состояниями компрометации, методологии ФСТЭК, если эти переходы рассматривать с точки зрения потребности достижения конкретной цели (состояния системы), поскольку цели злоумышленника могут зависеть от индивидуальных предпочтений злоумышленника. Кроме того, целью, злоумышленника может быть промежуточное состояние. В режиме реального времени развертывания атаки (on-line) следующие состояния должны учитываться при наличии возвратных переходах, что усложняет порядок вычислений.

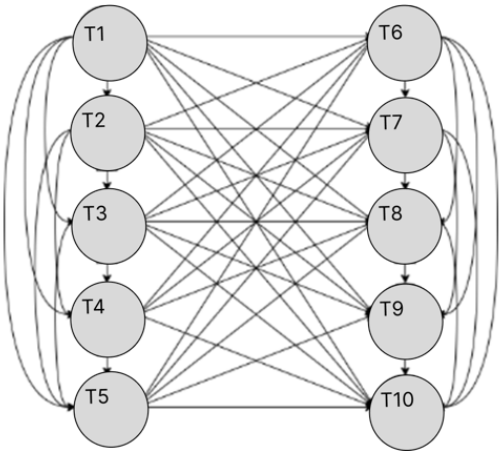


Рис. 1. Граф возможных состояний системы

2. Второй способ описания переходов предполагает, что используется методология ФСТЭК при плановом аудите или в режиме off-line. Методика позволяет определить

более эффективный путь, исходя из того, что при построении сценария (вектора) атаки не повторяется уже прошедшие этапы в силу того, что возвратные состояния отдаляют злоумышленника от целевого состояния или объекта системы [1]. При этом требуется конкретизировать состояние узла в распределённой системе или системы в целом для определения специфики компрометации для определения последовательности состояний вектора атаки (могут учитываться при построении сценария как общесистемные состояния, так состояния отдельных узлов сети). Следует учесть: длительность времени, затрачиваемого на переход не важна, поскольку не влияет на успешность поиска эффективного пути при появлении контрмер защищающейся стороны. Однако длительность каждого перехода может быть параметром метрики эксплуатируемой уязвимости. Кроме того, целью злоумышленника может быть промежуточное состояние (успех тактики), тогда следующие состояния отбрасываются, что упрощает порядок вычислений (на рис. 2 отброшены состояния с Т2 по Т5).

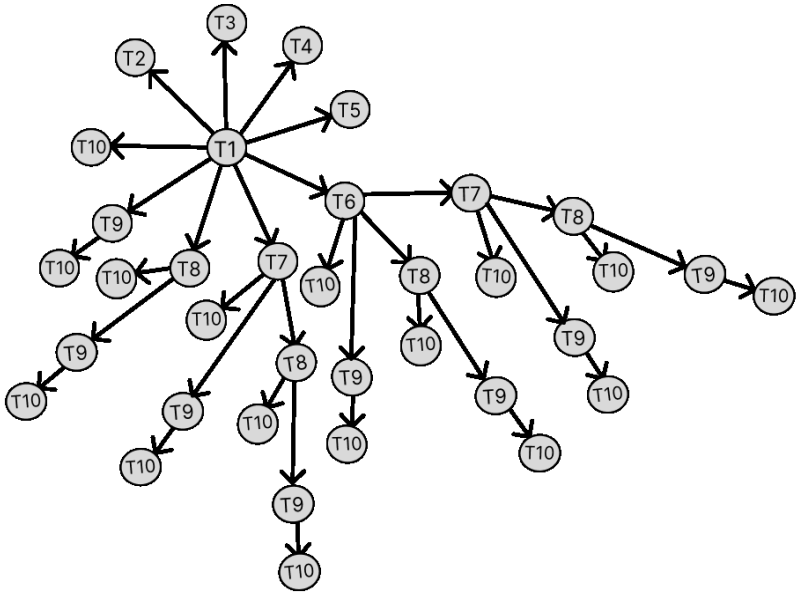


Рис. 2. Граф без возвратных связей между возможными состояниями системы

Марковские модели, использующие только уязвимости без определения классификационной принадлежности их состояний к тактикам Методики, не упорядочены и поэтому сложно определять качество состояния как повышения степени компрометации. При отсутствии поглощающих состояний задача определения наилучших атакующих последовательностей усложняется в еще большой степени, поскольку возможно зацикливание в эргодической последовательности. Определение специфических особенностей поиска наилучших последовательностей действий атакующего в режиме on-line является целью данной работы. Достижение этой цели дает возможность сформировать более полный сценарий потенциальной атаки, построенный на основе последовательности состояний, интерпретируемых как наступление возможности реализовать тактики злоумышленника.

Таким образом, достижение этой цели позволяет:

- выявить наиболее опасный сценарий атак (группу сценариев атак), описываемый тактиками (рассматриваются в приведенном исследовании как этапы атак) и принадлежащими им техниками (конкретная реализация этапа посредством эксплуатации уязвимостей) по Методике ФСТЭК, что позволяет сосредоточится стороне защиты на перекрытии уязвимостей (первоначально с помощью пакетов обновления), которые используются при реализации техник сценария;

- дает исследовательский инструмент – модель позволяет оценивать различные стратегии злоумышленника (последовательности действий с учетом

наград) для максимизации ожидаемых выгод, исследовать динамику изменений состояний модели при изменении вознаграждений;

- позволяет облегчить работу эксперта–аналитика (сузить область поиска), составляющего и анализирующего сценарии атак (в плане определения наиболее опасных состояний и последовательностей) при разработке Модели угроз организации (использование экспертного подхода регламентировано Методикой ФСТЭК);

- помогает найти оптимальную последовательности действий специалисту по активному аудиту ИБ, проводящего оценку защищенности информационных систем путем тестирования на проникновение.

Соответственно, следует:

- более подробно рассмотреть специфику методов моделирования атак (в качестве основного метода была выбран МППР);

- изучить специфику поиска наилучших для злоумышленника последовательностей атакующих воздействий;

- рассмотреть особенности использования алгоритмов оценки стратегий злоумышленника при нападении;

- изучить порядок получения исходных данных для моделирования;

- выявить особенности определения стратегий в задачах моделирования сетевых атак в режиме on-line.

Специфика определения модели поиска наилучшего вектора атаки

При использовании марковских процессов принятия решения (МППР) сценарий атаки (вектор атаки) описывается направленным графом. Предполагается, что злоумышленник получает вознаграждение *r*, которое зависит от действия *a* и состояния *s*. Требуется найти функцию, называемую стратегией, или политикой, которая определяет, какое действие предпринять в каждом состоянии, чтобы максимизировать некоторую другую функцию (например, среднюю или ожидаемую дисконтированную сумму) последовательности вознаграждений.

В общем случае применяемый метод прогнозирования включает следующие ключевые этапы:

1. Определение пространства состояний системы и возможных действий по отношению к ним.

2. Назначение вероятностей перехода состояний (постоянных во времени).

3. Определение характеристик выхода (политика действий злоумышленника).

4. Разработка математической модели (формирование матриц вероятностей переходов, вознаграждений) и решение задачи МППР.

5. Анализ результатов.

Марковский процесс принятия решений – это кортеж

$$\langle S, A, P, R, \gamma \rangle,$$

для которого случайные состояния из последовательности $\{s1, s2, \dots\}$ обладают Марковским свойством, где

$S = \{s1, s2, \dots, s10\}$ – множество вершин-состояний системы, соответствующих тактикам T1 - T10. Количество состояний и переходов между ними определяется на основе технического исследования целевой инфраструктуры (состояния как совокупность возможных действий (техник тактики по Методике ФСТЭК), получаемых нарушителем на достигнутом уровне компрометации, воспринимаемом как этап атак),

$A = \{(s_i, s_j) | s_i, s_j \in S, a\epsilon\{1, \dots, 11\}\}$ – множество ребер, отражающих способы эксплуатации уязвимостей (действия), которые могут быть применены атакующим для перехода между состояниями. Необходимо учитывать, что в зависимости от следующего состояния системы (тактики), количество доступных действий для следующего перехода может варьироваться из-за различий в способах эксплуатации уязвимостей,

$R = \{r_{ij} \vee (s_i, s_j) \in S\}$ – это награды за переход в определенное состояние, характеризующие эффективность или сложность эксплуатации уязвимостей (действия$a \in A$) для перехода к следующему состоянию (успешной реализации тактики),

$P(s, a; s')$ – вероятность перехода из состояния $s\epsilon S$ при действии $a\epsilon A$ в состояние $s'\epsilon S$,

γ – коэффициент дисконтирования (оценивает ценность будущих наград) $\gamma \in [0; 1]$.

Стратегией, или политикой π в данном случае – распределение действий злоумышленника, которые связаны с эксплуатацией уязвимостей при заданных состояниях. Она целиком описывает поведение атакующего зависит только от текущего состояния. Если нарушитель следует стратегии π на шаге i то $\pi(a|s)$ – вероятность того, что будет предпринято действие $A_i = a$, если планируется переход в состояние $S_i = s$ (1):

$$\pi(a|s) = P[A_i = a \vee S_i = s] \quad (1)$$

Функция ценности определяется следующим образом вознаграждений на каждом i-ом шаге (2).

$$G_i = R_{i+1} + yR_{i+2} + y^2R_{i+3} + \cdots = \sum_{k=0}^{\infty} y^k R_{i+k+1} \quad (2)$$

Ценность состояния – это ожидаемое дисконтированное вознаграждение R злоумышленника, начинающего эксплуатировать уязвимость из состояния $s_i = S$ при соблюдении стратегии атаки π. Функция ценности состояния s при стратегии π определяется как математическое ожидание дисконтированной суммы будущих вознаграждений R (описано уравнением оптимальности Беллмана (3)). Для поглощающего состояния графа атаки ценность равна нулю.

$$V_\pi(s) = M[R_{i+1} + \gamma V_\pi(S_{i+1})|S_i = s], s \in S \quad (3)$$

где $V_\pi(s)$ – функция ценности состояний (важность достижения состояния успеха эксплуатации уязвимостей атакующим в векторе атаки) при стратегии атаки π;

M – математическое ожидание случайной величины;

γ – коэффициент дисконтирования.

Графически стратегия может быть представлена как граф, узлами которого являются состояния, в которых находится система на каждом из этапов атаки, а дуги – действиями атакующего.

Решить Марковский процесс принятия решений означает найти наилучшую стратегию поведения нарушителя в заданной системе. Наилучшая стратегия π – это такая стратегия, при соблюдении которой достигается максимальная ожидаемая совокупная награда, которая в дальнейшем будет определятся как параметр компрометации ЦИС [5]. Оптимальность стратегии может пониматься в различных смыслах в зависимости от выбранного критерия.

Политика выгоды (в Марковских процессах принятия решений) связана с оптимизацией стратегий принятия решений в последовательности атакующих воздействий при условии неопределенности среды. Как упоминалось ранее, наилучший вектор атаки обеспечивает максимальную ожидаемую ценность состояния (4):

$$\pi_*(s) = \underset{\pi}{argmax} V_\pi(s) \quad (4)$$

Функция ценности, обеспечивающая максимальную ожидаемую величину компрометации относительно всех стратегий, называется оптимальной функцией ценности, представлена как важнейшая величина компрометации (эксплуатируемости уязвимости) и обозначается как $V_*(s)$ (5):

$$V_*(s) = \underset{\pi}{max} V_\pi(s). \quad (5)$$

Для сравнения стратегий нападения и определения наилучшей атаки необходимо ввести понятие оценки стратегии как последовательности атакующих воздействий, где вычисляется функция ценности V_π. С этой целью вводится отношение порядка на множестве стратегий π (6). Оно означает, что при использовании стратегии как последовательности действий атакующего π нарушитель будет получать большую или равную ожидаемую величину компрометации по сравнению со стратегией π' во всех состояниях:

$$\pi \geq \pi' \Leftrightarrow V_\pi(s) \geq V'_\pi(s), \forall s \quad (6)$$

Иным способом определения стратегии является использование функции ценности действия a, представляемого как эксплуатация уязвимости. Ценность применения действия в состоянии s при стратегии π формируется как к ожидаемый доход, когда нарушитель начинает реализацию атаки в состоянии s, выполняет действие a и затем следует стратегии π (7) [5]:

$$Q_\pi(s,a) = M[G_i|S_i = s, A_i = a], \quad (7)$$

где $Q_\pi(s,a)$ – функция ценности действия (как эксплуатации уязвимости) при атакующей последовательности (стратегии) π.

Эта функция ценности действия позволяет представить функцию ценности состояния, а также стратегию как (8) и (9) соответственно [5]:

$$V(s) = \max_a Q(s,a) \quad (8)$$

$$\pi(s) = \arg\max_a Q(s,a) \quad (9)$$

Для вычислений, связанных со стратегией принятия решений, и для поиска оптимальной стратегии, применяются различные алгоритмы, в том числе относящиеся к методам динамического программирования, Q-обучения. Используемые методы рассматриваются далее. При этом каждая стратегия по-разному проявляет свою эффективность. Ее выбор определяется особенностями описания проблемной области:

1. Динамичность или статичность состояний;

2. Полная или неполная известность модели (не точные знания о количестве состояний, то есть о количестве эксплуатируемых уязвимостей). При поиске стратегий в реальном режиме времени злоумышленник может обнаружить ранее неучтенные уязвимости.

3. Наличие обратных связей при использовании времени достижения состояния.

Особенности использования алгоритмов оценки политики

При построении марковских моделей часто используется динамическое программирование. Это семейство алгоритмов, которые применяются для поиска оптимальных стратегий в условиях модели окружающей среды, представленной в виде Марковского процесса принятия решений [6]. Чаще всего применяются следующие алгоритмы:

1. policy iteration – итерация по стратегиям;

2. value iteration – итерация по значениям.

Итерация по стратегиям – это метод нахождения оптимальной стратегии путем последовательного чередования оценки текущей стратегии и ее улучшения с помощью жадного алгоритма. Этот подход гарантированно сходится к оптимальному решению независимо от начальной стратегии. Сходимость достигается за конечное число итераций, так как существует ограниченное количество возможных стратегий, и каждая итерация улучшает стратегию, если это возможно. Несмотря на то, что число допустимых стратегий растет экспоненциально с увеличением количества состояний, итерация по стратегиям на практике сходится достаточно быстро [7]. Однако, это означает, что исследуема ЦИС должна быть полностью описана, а все возможные пути атакующих последовательностей учтены.

В результате работы этого алгоритма создается цепочка политик, где каждая из которых является улучшением по сравнению с предыдущей (10):

$$\pi_0 \xrightarrow{E} v_{\pi_0} \xrightarrow{I} \pi_1 \xrightarrow{E} v_{\pi_1} \xrightarrow{I} \pi_2 \xrightarrow{E} \ldots \xrightarrow{I} \pi_* \xrightarrow{E} v_* \quad (10)$$

Оценка политики (E – с англ. evaluation – оценка) и шаги по ее совершенствованию (I – с англ. improvement – улучшение) проводятся до тех пор, пока политика больше не перестанет улучшаться [5].

Итерация по стратегиям часто бывает вычислительно затратной, потому что в каждой итерации необходимо вычислять оценку стратегии. Вариант итерационного поиска стратегии, называемый

модифицированной итерацией по стратегиям, аппроксимирует функцию полезности, используя итеративную оценку стратегии вместо точной оценки [6]. Можно выбрать количество итераций оценки стратегии между шагами улучшения. Если использовать только одну итерацию между шагами, то этот подход идентичен итерации по значениям.

Используя алгоритм итерации по значениям, можно снизить сложность вычислений за счет отсутствия необходимости в оценке стратегии. Выбор действия для оптимальной политики происходит напрямую на основе получаемой награды.

Когда разница значений между двумя последовательными итерационными шагами будет ниже заданного значения, итерация будет прекращена.

Основное отличие алгоритма итерации по значениям от итерации по стратегиям заключается в том, что итерация по значениям фокусируется на вычислении оптимальной функции полезности для каждого состояния, в то время как итерация по стратегиям направлена на поиск оптимальной стратегии действий.

Алгоритм итерации по значениям обновляет функцию полезности за один шаг, учитывая все будущие возможные вознаграждения, в то время как итерация по стратегиям обновляет стратегию за один шаг, выбирая действие с максимальной полезностью [8]. Итерация по стратегиям начинает с некоторой начальной стратегии и итеративно улучшает ее, вычисляя значения функции полезности для каждого состояния при текущей стратегии, а затем находя новую стратегию, максимизирующую ожидаемую полезность каждого состояния

Алгоритм итерации по стратегиям, поскольку, несмотря на свою вычислительную сложность, работает быстрее и использует меньшее количество итераций для обеспечения сходимости алгоритма.

Поскольку приведенные методы определения стратегий требует предварительного описания исследуемой среды, их целесообразно применять в режиме off-line.

В случаях исследования атакующих последовательностей без предварительного знания модели среды часто используется Q-обучение - алгоритм обучения с подкреплением, который позволяет агенту находить оптимальную стратегию поведения в не полностью описанной среде.

Кроме того, того существует возможность обучения в режиме реального времени: Q-обучение позволяет агенту обучаться и улучшать свою стратегию по мере взаимодействия со средой. Методы подобные Q-обучение целесообразно применять при определении наилучшего вектора атаки в режиме on-line.

Порядок получения исходных данных для моделирования

Формально вектор атаки можно представить в виде графа, генерирующегося с помощью разрабатываемого инструмента, и принимающего в качестве входных данных следующую информацию о компонентах анализируемой системы:

- IP-адрес хоста;
- идентификатор (ID) уязвимости, при этом хосты могут иметь более одной уязвимости;
- CVSS оценка эксплуатируемости уязвимости по стандарту 2.0, значение которой варьируется в пределах от 0 до 10: чем выше значение, тем более опасной является уязвимость. Это число будет использовано как один из коэффициентов в расчете вероятностей и наград математической модели для перехода между состояниями.

Получение входных данных является необходимым этапом работы разрабатываемого инструмента, так как модель на основе Марковских процессов инициализируется с некоторым условным начальным распределением состояний и вероятностями переходов между ними.

Другим требованием для формирования модели является выбор и категорирование состояний системы, чтобы связать с каждой вершиной графа набор уязвимостей, делающих переход к следующей вершине возможным. В качестве набора возможных состояний были приняты десять тактик из

Методики оценки угроз безопасности информации ФСТЭК России.

Тактическая задача: достижение нарушителем конечной цели (состояния T1), приводящее к реализации моделируемой угрозы и причинению недопустимых негативных последствий [3].

Таким образом, данные тактики позволяют определить все возможные пути атаки на информационную систему и разработать меры по ее защите. В зависимости от заданных условий, сценарий реализации атаки может включать не все тактики, то есть их количество может варьироваться.

Злоумышленник переходит из одного состояния в другое, когда эксплуатируется хотя бы одна из уязвимостей, относящихся к следующему состоянию исследуемой системы.

Целевой вершиной графа (поглощающее состояние) является состояние тактики T10 – при нахождении одного из компонентов системы в данном состоянии нарушается состояние безопасности всей системы в целом. Как только злоумышленник достигает этого состояния, система считается скомпрометированной, то есть злоумышленник достиг своей цели. Таким образом, система будет оставаться в этом состоянии до тех пор, пока специалисты по безопасности не примут превентивные меры по удалению присутствия злоумышленника в системе.

Проблемой является отсутствие прямого сопоставления тактик с уязвимостями. В соответствии с утвержденной формой описания уязвимостей в БДУ ФСТЭК, каждая уязвимость может быть классифицирована по одному из двенадцати способов эксплуатации, которые были выведены путем синтеза шаблонов атак из международного стандарта CAPEC [9]. Способы эксплуатации уязвимостей, в свою очередь, могут быть соотнесены с техниками, которые применяются для реализации тактик. В результате анализа перечня техник получена табл. 1, отражающая соотношения тактик и способов эксплуатации, что позволит в дальнейшем отнести уязвимости к состояниям системы. В процессе совершения атаки у злоумышленника имеется возможность выполнять некоторые действия (способы эксплуатации) по отношению к уязвимостям компонентов системы, посредством которых он меняет ее состояние. При этом вероятности перехода для каждого действия принимают разные значения при наличии статистической информации от датчиков (поставщиков данных о событиях в системе), и равномерное распределение вероятностей с учетом соблюдения нормировки, исходя из количества действий от состояния к состоянию.

Таблица 1

Соотношение тактик и способов эксплуатации уязвимостей

	Тактика	**Способы эксплуатации уязвимости**
1	Сбор информации о системах и сетях (T1)	Несанкционированный сбор информации Анализ целевого объекта
2	Получение первоначального доступа к компонентам систем и сетей (T3)	Нарушение аутентификации Нарушение авторизации Инъекция
3	Внедрение и исполнение вредоносного программного обеспечения в системах и сетях(T3)	Инъекция Манипулирование структурами данных
4	Закрепление (сохранение доступа) в системе или сети (T4)	Манипулирование структурами данных Манипулирование ресурсами
5	Управление вредоносным программным обеспечением и (или) компонентами, к которым ранее был получен доступ (T5)	Злоупотребление функционалом
6	Повышение привилегий по доступу к компонентам систем и сетей(T6)	Нарушение аутентификации Нарушение авторизации

Продолжение табл. 1

	Тактика	Способы эксплуатации уязвимости
7	Сокрытие действий и применяемых при этом средств от обнаружения (Т7)	Манипулирование сроками и состоянием
8	Получение доступа (распространение доступа) к другим компонентам систем и сетей или смежным системам и сетям (Т8)	Подмена при взаимодействии
9	Сбор и вывод из системы или сети информации, необходимой для дальнейших действий при реализации угроз безопасности информации или реализации новых угроз(Т9)	Несанкционированный сбор информации
10	Несанкционированный доступ и (или) воздействие на информационные ресурсы или компоненты систем и сетей, приводящие к негативным последствиям(Т10)	Манипулирование структурами данных Манипулирование ресурсами Исчерпание ресурсов

В отношении уязвимостей может применяться не весь набор способов эксплуатации, что отражено в табл. 1. Возможность соотнесения состояний системы, выраженных в виде фиксации успеха реализации тактик ФСТЭК, и способов эксплуатации уязвимостей, которые позволяют достичь состояние успеха, говорит о целесообразности их использования в качестве действий. Переходы между состояния вероятностны и зависят от особенностей эксплуатации (сложности) уязвимостей.

Оценка потенциала эксплуатации уязвимостей (в диапазоне от 0 до 10) осуществляется в соответствии со стандартом CVSS 2.0 (Common Vulnerability Scoring System). Пусть $r(v)$ – мера сложности эксплуатации уязвимости v. Стандарт CVSS использует формулу (11) для расчета данного параметра, в котором константа 20 представляет собой коэффициент серьезности уязвимости [10].

$$r(v) = 20 * AV * AC * Au, \qquad (11)$$

где *Access Vector (AV)* – параметр, показывающий, каким путем может быть внедрена уязвимость;

Access Complexity (AC) – сложность атаки, оценивающая, насколько легко или сложно использовать данную уязвимость;

Authentication (Au) – оценивает количество аутентификаций, которые атакующий должен произвести прежде чем воспользоваться уязвимостью.

Учитывая оценки эксплуатируемости для каждой из уязвимостей в графе атак, возможно оценить вероятности перехода путем нормализации оценок уязвимостей по всем ребрам, начиная с исходного состояния системы. Пусть p_{ij} – вероятность того, что злоумышленник, находящийся в состоянии i, воспользуется уязвимостью в состоянии j. Соответственно можно, формально определить вероятность перехода, используя формулу (12):

$$p_{ij} = 1 - \frac{\sum rr_j}{\sum_{k=1}^{n} r_k}, \qquad (12)$$

где n – количество уязвимостей, доступных из состояния i;

r_j – сумма оценок эксплуатируемости уязвимостей в состоянии j;

r_k – сумма оценок эксплуатируемости всех уязвимостей всех состояний, доступных из состояния i.

Исходя из описанных выше условий, формируется матрица переходных вероятностей при выполнении действия a_i в отношении набора состояний.

Следующим важным параметром модели является вознаграждение, которое отражает успех злоумышленника в случае реализации атакуемого воздействия или затраты в случае неудачи.

Ключевым фактором для данного параметра является базовая метрика уязвимости. Чем выше значение данной метрики, тем большее вознаграждение должен получать атакующий при переходе в соответствующее состояние в модели Марковского процесса.

Коэффициент дисконтирования задает горизонт задачи для нарушителя и

определяется в диапазоне от 0 до 1. Он показывает, насколько весомыми станут будущие вознаграждения в сравнении с полученными на текущем этапе атаки. Для определения значения коэффициента дисконтирования используются следующие подходы [11]:

1. Учет временных предпочтений нарушителя: коэффициент дисконтирования может отражать, насколько важны для нарушителя будущие вознаграждения по сравнению с текущими. Если предполагается, что нарушитель ориентирован на быстрый результат, коэффициент будет ближе к 0, если на долгосрочный – ближе к 1.

2. Оценка рисков срыва атаки на каждом шаге: чем выше риски прерывания атаки на промежуточных шагах, тем ниже должен быть коэффициент дисконтирования. Это отражает неопределенность получения будущих вознаграждений.

Особенности определения стратегий в задачах анализа сетевых атак в режиме on-line

Политики задают правила принятия решений в определённых условиях, а стратегии ориентированы на достижение максимальной выгоды в долгосрочной перспективе, выбирая наиболее подходящую политику.

Поскольку действия связывают состояния, выявления оптимального набора действий позволяют определить наиболее вероятные состояния (совокупность допустимых к осуществлению действий (техник тактики по Методике ФСТЭК)), в которые попадет злоумышленник при развертывании атаки.

Каждое состояние имеет соответствующую политику, которая указывает, какое действие следует предпринять в этом состоянии. Стратегии помогают выбрать оптимальную политику для максимизации вознаграждений. Эти концепции тесно связаны с состояниями системы и активно используются для анализа и оптимизации поведения злоумышленника в условиях неопределённости.

Для изучения особенностей определения стратегий как последовательности атакующих воздействий в задачах анализа сетевых атак в режиме on-line рассмотрим атаку ARP-спуффинг. Инфраструктура описывается диапазоном IP-адресов 130.10.1.1 – 130.10.1.100. В заданном диапазоне идентифицируются уязвимости, соответствующие при их успешной эксплуатации (посредством реализации действий) достигаемым состояниям (фиксации успешной реализации тактики, согласно таб.1). При это в режиме реального времени исследуемая система изменчива, поэтому целесообразно применять алгоритмы, которые минимально требовательны к исходным данным и предполагает изначально неполную известность исходных данных политик.

```
{'T1': 'A1_2', 'T2': 'A2_3', 'T3': 'A3_6', 'T6': 'A6_10', 'T8': 'A8_10', 'T10': 'Terminate'}
{'T1': 'A1_2', 'T2': 'A2_3', 'T3': 'A3_6', 'T6': 'A6_8', 'T8': 'A8_10', 'T10': 'Terminate'}
{'T1': 'A1_2', 'T2': 'A2_6', 'T3': 'A3_6', 'T6': 'A6_10', 'T8': 'A2_6', 'T10': 'Terminate'}
```

Рис 3. Список состояний и действий моделируемой атаки

Список состояний моделируемой атаки приведен на рис.3, а рис. 4 иллюстрирует визуализацию последовательности атакующих воздействий на компонент системы в виде графа. Круглые вершины отражают тактики ФСТЭК, которые последовательно описывают ход выполнения атаки. Переход между вершинами осуществляется через действие (ребра графа), которое вероятнее всего предпримет злоумышленник в отношении системы, для того чтобы перейти к следующему этапу атаки. В случае наличия или отсутствия уязвимостей, подходящих под тот или иной способ эксплуатации, количество доступных действий может варьироваться. Ребра, выделенные жирным, указывают на наиболее вероятный вектор атаки.

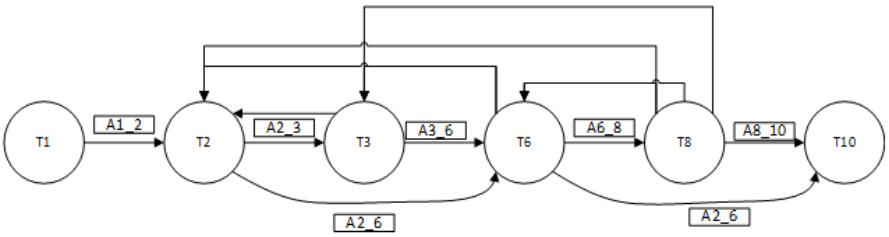


Рис. 4. Список состояний и действий моделируемой атаки

Множество действий описывается следующим набором (рис.5).

```
Предполагаемый вектор атаки:
[сбор информации-->(130.10.1.11)
инъекция-->(130.10.1.12)
манипулирование структурами данных-->(130.10.1.12)
манипулирование ресурсами->(130.10.1.21)
злоупотребление функционалом-->(130.10.1.21)
нарушение аутентификации-->(130.10.1.24)
манипулирование сроками и состоянием-->(130.10.1.24)
подмена при взаимодействии-->(130.10.1.24)
исчерпание ресурсов-->(130.10.1.24)]
```

Рис. 5. Предполагаемый вектор атаки

Описанная атака характеризуется следующим: реализацией в реальном режиме времени, динамичностью состояний, неполнотой описания модели, при поиске стратегий в реальном режиме времени злоумышленник может обнаружить ранее неучтенные уязвимости, наличием обратных связей. В данном случае для определения стратегий были выбраны методы классификационно принадлежащие к методу определения наилучших стратегий Q-обучение.

Изучаемые стратегии выбора действий (горизонт их планирования конечный) следующие:

1. Стратегия *ε-greedy*. Выбирает действие, максимизирующее текущую оценку функции полезности (greedy action), с вероятностью 1-ε, выбирает случайное действие с вероятностью ε, обеспечивает баланс между эксплуатацией (выбор лучшего известного действия) и исследованием (выбор случайного действия).

2. Стратегия greedy. Всегда выбирает действие, максимизирующее текущую оценку функции полезности. Не обеспечивает исследование альтернативных действий, что может привести к застреванию в локальном оптимуме

3. Стратегия softmax стратегия. Выбирает действия пропорционально их оценкам функции полезности. Вероятность выбора действия *a* пропорциональна $exp(Q(s,a)/\tau)$, где τ - параметр времени. Обеспечивает более плавное распределение вероятностей между действиями, чем ε-greedy. Позволяет сфокусировать поиск на наиболее перспективных действиях, но может пренебрегать менее привлекательными действиями.

Стратегия USB (Upper Confidence Bound) является еще одним подходом, который выбирает действия, основываясь на верхней границе доверительного интервала для оценок функции полезности. Она направлена на оптимальное исследование пространства действий.

Псевдокод ниже демонстрирует процедуру применение поиска стратегии (повторяется до сходимости или конечного, или получения терминального состояния):

Листинг 1. Псевдокод процедуры поиска стратегии.

```
    # Наблюдение:
    s_пред = s
    a_пред = a
    s = Получить текущее состояние сенсора
    r = Получить вознаграждение за предыдущее действие
    # Обновление полезности:
    Q[s_пред, a_пред] = Q[s_пред, a_пред] + LF * (r + DF * MAX(Q, s) - Q[s_пред, a_пред])
    # Выбор действия:
    a = ARGMAX(Q, s)
    Передать действие a активатору
  Функция MAX(Q, s):
    max = minValue
    Для каждого a из возможных действий для состояния s:
      Если Q[s, a] > max, то max = Q[s, a]
    Вернуть max
  Функция ARGMAX(Q, s):
    amax = Первое возможное действие для состояния s
    Для каждого a из возможных действий для состояния s:
      Если Q[s, a] > Q[s, amax], то amax = a
    Вернуть amax
```

В листинге 1 использованы обозначения:

LF — это коэффициент обучения α. Чем он выше, тем сильнее агент доверяет новой информации;

DF — это коэффициент обесценивания (дисконтирования) γ. Чем он меньше, тем меньше агент предпочитает выгоду от своих будущих действий в пользу настоящего.

Проведенные подготовительные иследования выявили следующие:

При 100 000 итераций по поиску лучшей стратегии (методы e-greedy, greedy, softmax) было определено:

- стратегия greedy работает быстрее остальных, но является наименее точной;

- Softmax имеет хорошую точность на базовых параметрах (но поиск требует больше итераций в сравнении с остальными методами (медленная работа)).

При 100 раундах стратегия e-greedy показывает ту же точность, что и ранее. Это означает, что ей не требуется в данной задаче много раундов для обучения (что экономит вычислительные ресурсы).

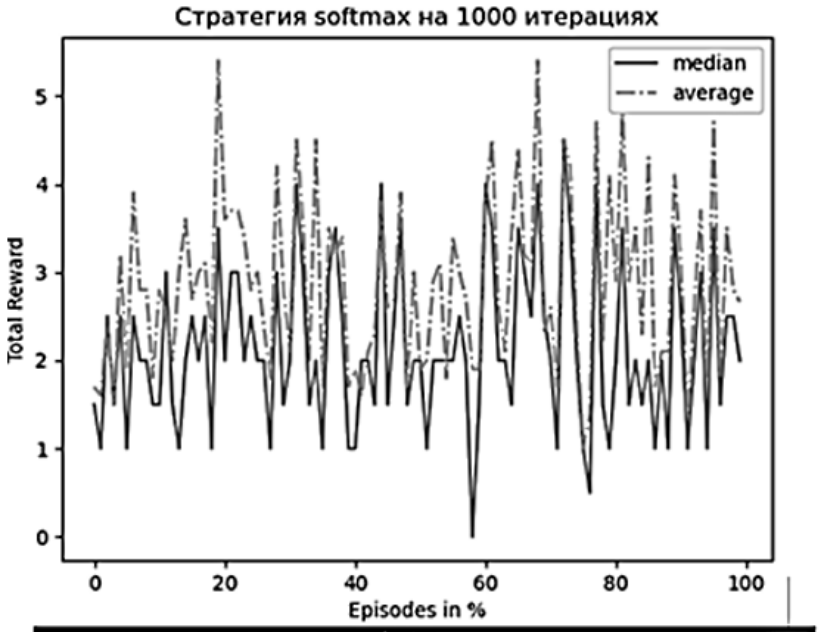
Далее (рис. 6-9) приводятся результаты исследования (полученные стратегии по E-greedy, greedy, usb (медианные значения по 1% от итераций (общая награда за раунд)). При этом следует отметить, что некоторые стратегии (greedy и usb) для данной задачи опираясь на собственные результаты зацикливались и не выходили из цикла при 1000 и более итераций. Соответственно был введён параметр дальности обзора 3000, так же внутри раунда обучения проводилась проверка на сходимость, чтобы при условии, что дальность обзора превышена досрочно прогон не завершался.


```
Медианная награда: 337.0
Средняя награда: 343.1631631631632
Максимальная награда: 612
Последняя награда: 593
Q table:
{'AD': 6.150240977355153, 'AC': 5.288324750157858}
{'AC': 5.396841756262222, 'AD': 6.106125490565373, 'AR': 5.521260791930304}
{'AC': 1.7022345201178892, 'AD': -5.026274723471368, 'AR': 5.505856473910464}
{'Terminate': 0}
```

Рис. 6. Реализация e-greedy стратегии

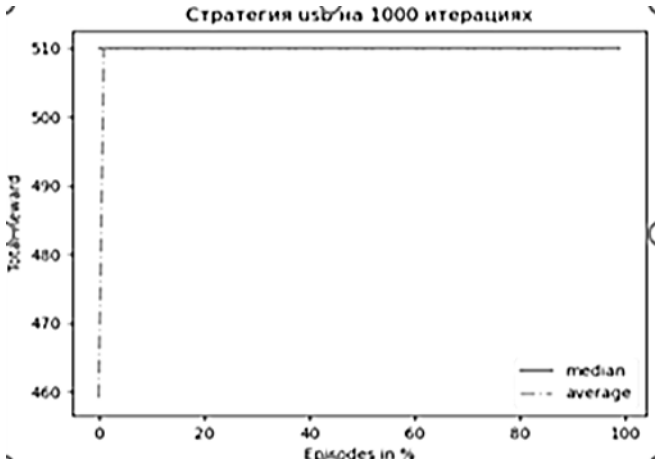

```
Медианная награда: 296.0
Средняя награда: 296.0610610610611
Максимальная награда: 345
Последняя награда: 325
Q table:
{'AD': 3.668273324681356, 'AC': 0}
{'AC': 2.2532373904434326, 'AD': 0, 'AR': 0}
{'AC': -5.773502691896258, 'AD': -5.0, 'AR': 3.3137422449888194}
{'Terminate': 0}
```

Рис. 7. Реализация greedy стратегии




```
Политика: N-AD; T-AD; C-AR; B-Terminate;
Медианная награда: 2.0
Средняя награда: 2.87987987987988
Максимальная награда: 18
Последняя награда: 0
Q table:
{'AD': 6.131555928867277, 'AC': 5.283353034122967}
{'AC': 5.381628173116712, 'AD': 6.0590020690523225, 'AR': 5.503180081502875}
{'AC': 3.102822485298449, 'AD': -5.962420952780851, 'AR': 5.492870302990163}
{'Terminate': 0}
```

Рис. 8. Реализация стратегии softmax




```
Медианная награда: 510.0
Средняя награда: 509.4904904904905
Максимальная награда: 510
Последняя награда: 510
Q table:
{'AD': 5.263157894736832, 'AC': 0}
{'AC': 0.0, 'AD': 0, 'AR': 4.736842105263148}
{'AC': -5.773502691896258, 'AD': 0, 'AR': 4.736842105263148}
{'Terminate': 0}
```

Рис. 9. Реализация e-greedy стратегии usb

Результаты моделирования атаки

Стратегии greedy и usb для данной задачи, опираясь на свои алгоритмически получаемые данные (специфика их алгоритма) попадали в бесконечный цикл и не выходили из него, при этом был введён параметр дальности обзора 3000 итераций, так же внутри раунда обучения проводилась проверка на сходимость, чтобы при условии, что дальность обзора превышена досрочно прогон не завершался.

1. Стратегия e-greedy проявила себя лучше всех, поскольку исследовала большее количество состояний, аналогично softmax (в результате дали одинаковые политики).

2. Стратегии usb, greedy плохо исследовали среду, однако гарантированно отводит агента от терминального состояния (выделяются состояния challenged – reset).

3. Стратегия softmax проявила себя лучше всех, так как исследовала большее количество состояний.

При рассмотрении модели атаки с возвратными состояниями в предложенных стратегиях при условии противодействия атакующим воздействиям задача определения эффективной стратегии атаки усложняется. В данном случае целесообразно использовать Multi-Agent Reinforcement Learning (MARL), который расширяет концепцию *RL* на случай, когда в среде действует несколько автономных агентов, каждый из которых пытается максимизировать свою собственную награду. Так же следует отметить следующее: возникает необходимость оптимизации обучения, поскольку можно использовать нейронную сеть, которая должна оценивать значения для неизвестных состояний.

Заключение

При определении особенностей поиска наилучших последовательностей действий атакующего были рассмотрены специфические особенности применения МППР как метода моделирования. При этом в процесс моделирования были интегрированы методы описания атак Методики ФСТЭК. На основании методических ограничений и режимов аудита были определены режимы моделирования. Было выявлено, что в режиме off-line целесообразно применять поиск наилучших последовательностей атакующих воздействий с помощью стратегии по значениям, в режиме on-line использовать методы Q-обучения, или классификационно схожие с ними.

При построении моделей в качестве входных данных использовались метрики уязвимостей, которые были классифицированы по методам эксплуатации, и, соответственно, по принадлежности к тактикам Методики (следует учесть, что оценка уязвимости может носить экспертный характер, если четкого соотнесения с тактиками не наблюдается). При исследовании аспектов моделирования на примере атаки ARP-спуффинг в режиме on-line была определена наилучшая стратегия из выбранных – softmax.

Список литературы

1. Методический документ «Методика оценки угроз безопасности информации» (утв. Федеральной службой по техническому и экспортному контролю 5 февраля 2021 г.). 83 с.

2. Zhou X, Decision Process For Automatic Cyber Defense (Accepted) / Xiaofan Zhou, Simon Yusuf Enoch, Dan Dong Seong Kim. Markov // In the 23rd World Conference on Information Security Applications (WISA 2023).

3. Booker, L.B. A model-based, decision-theoretic perspective on automated cyber response. / L.B. Booker, S.A. Musman // arXiv preprint arXiv:2002.08957 (2020)

4. Zheng J. Defending sdn-based iot networks against ddos attacks using markov decision process. / J. Zheng, A.S. Namin. // In: 2018 IEEE International Conference on Big Data (Big Data). IEEE (2018)

5. Кохендерфер М. Алгоритмы принятия решений / М. Кохендерфер, Т. Уилер, К. Рэй. // пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2023. 684 с.

6. Саттон Р. С., Барто Э. Дж. Обучение с подкреплением: Введение. 2-е изд. / пер. с англ. А. А. Слинкина. М.: ДМК Пресс, 2020. 552 с.

7. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 2nd edition. Melbourne, Australia : OTexts, 2018. 382 p.

8. Mazengia D.H. Forecasting Spot Electricity Market Prices Using Time Series Models: Thesis for the degree of Master of Science in Electric Power Engineering. Gothenburg, Chalmers University of Technology, 2008. 89 p.

9. Common Attack Pattern Enumerations and Classifications // URL: https://capec.mitre.org/ (дата обращения: 12.05.2024).

10. Common Vulnerability Scoring System Calculator // URL: https://nvd.nist.gov/vuln-metrics/cvss/v2-calculator (дата обращения: 13.05.2024).

11. Щеглов А.Ю. Защита компьютерной информации от несанкционированного доступа. СПб.: Наука и Техника, 2004. 384 с.

Калининградский государственный технический университет
Kaliningrad State Technical University

Поступила в редакцию 28.08.24

Информация об авторе

Подтопельный Владислав Владимирович – старший преподаватель, Институт цифровых технологий (ИЦТ), Калининградский государственный технический университет (КГТУ), e-mail: ionpvv@mail.ru

INVESTIGATION OF THE SPECIFICS OF MODELING COMPUTER ATTACKS USING MARKOV DECISION-MAKING PROCESSES AND Q-LEARNING

V.V. Podtopelny

The problems arising in solving the problem of constructing an attack vector in a network infrastructure are considered. The varieties of various methods of attack modeling using the FSTEC methodology used in constructing a network attack vector are presented and characterized, as well as the specifics of their interrelationships using Markov decision-making processes in modeling attacking influences are considered. Their suitability for various procedures for determining vector parameters is considered. When constructing a network attack vector, the features of defining Q-learning strategies are given. The formation of an attack vector in the context of exploiting vulnerabilities of an information system in real time is studied. The features of the attack vector construction are determined, taking into account the specifics of the relations of tactics (states).

Keywords: network attack, vulnerability, Markov processes modeling, strategy, policy, learning method.

Submitted 28.08.24

Information about the author

Vladislav V. Podtopelny – senior Lecturer, Institute of Digital Technologies, Kaliningrad State Technical University, e-mail: ionpvv@mail.ru