

К.С. Григорян, Е.С. Басан**АЛГОРИТМ ПРОТИВОДЕЙСТВИЯ ПРЕДНАМЕРЕННЫМ ПОМЕХАМ
НА ОСНОВЕ ТЕОРИИ ИГР И ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ**

Преднамеренные помехи представляют собой одну из наиболее серьезных угроз информационной безопасности беспроводных систем связи, поскольку они могут приводить к нарушению доступности канала передачи данных и реализации атак отказа в обслуживании на физическом уровне. Особенно уязвимыми являются когнитивные радиосети, сенсорные сети и системы Интернета вещей, в которых злоумышленник способен динамически изменять стратегию постановки помех. Целью данной работы является разработка адаптивного алгоритма противодействия преднамеренным помехам, обеспечивающего устойчивость беспроводной системы связи в условиях неопределённости состояния канала и динамического поведения злоумышленника. Для достижения поставленной цели взаимодействие между легитимным узлом и постановщиком помех формализуется в виде марковской игровой модели Штаклберга, в которой передатчик выступает лидером, а злоумышленник – последователем. Обучение с подкреплением используется для адаптивного выбора стратегии легитимного узла в стохастической среде, тогда как робастная оптимизация SOCP (Second-order cone programming) обеспечивает выполнение требований к качеству обслуживания с учётом неопределённости оценки состояния канала. В качестве алгоритма обучения применяется метод SAC (Soft Actor-Critic), позволяющий эффективно работать в непрерывном пространстве действий. В ходе исследования решаются задачи формализации модели противодействия помехам, интеграции обучения с подкреплением с робастной оптимизацией и оценки эффективности предложенного подхода с использованием численного моделирования. Результаты моделирования показывают, что предложенный алгоритм RL-Stackelberg + SOCP обеспечивает снижение вероятности отказа соединения по сравнению с существующими методами противодействия помехам, включая ППРЧ (Псевдослучайная перестройка рабочей частоты) и игровые алгоритмы выбора частоты. Было проведено имитационное моделирование методом Монте-Карло предложенного алгоритма, а также нескольких алгоритмов на основе ППРЧ. Предложенный алгоритм противодействия преднамеренным помехам позволяет снизить вероятность обрыва связи на 0,14. Полученные результаты подтверждают повышение устойчивости беспроводных систем связи к атакам преднамеренных помех.

Беспроводные сети; робастная оптимизация; теория игр; защита каналов связи; обучение с подкреплением.

K.S. Grigoryan, E.S. Basan**GAME-THEORETIC AND REINFORCEMENT LEARNING-BASED ALGORITHM
FOR INTENTIONAL JAMMING MITIGATION**

Intentional jamming represents a serious threat to the security and availability of wireless communication systems. Modern wireless networks, including cognitive radio, sensor networks, and Internet of Things infrastructures, are particularly vulnerable because adversaries can dynamically adapt their jamming strategies. The objective of this study is to develop an adaptive anti-jamming algorithm capable of maintaining communication reliability under dynamic interference conditions. To achieve this goal, the interaction between the legitimate transmitter and the jammer is modeled as a Markov Stackelberg game, where the legitimate node acts as a leader and the jammer acts as a follower. Reinforcement learning is used to determine the optimal strategy of the leader in a stochastic environment, while robustness against channel uncertainty is ensured through a SOCP (Second-order cone programming) formulation that guarantees the required quality-of-service constraints. The learning process is implemented using the SAC (Soft Actor-Critic) algorithm, which enables stable policy optimization in continuous action spaces and stochastic environments. The research tasks include the formalization of the anti-jamming interaction as a Markov decision process, the integration of reinforcement learning with a robust SOCP optimization layer, and the evaluation of the proposed approach through simulation. A Monte Carlo simulation of the proposed algorithm, as well as several algorithms based on FHSS (Frequency-Hopping Spread Spectrum), was conducted. The proposed anti-jamming algorithm reduces the probability of communication outage by 0.14. The results indicate that the proposed approach improves the resilience of wireless communication systems and reduces the probability of successful denial-of-service attacks at the physical layer.

Wireless networks; robust optimization; game theory; communication channel protection; reinforcement learning.

Введение. Беспроводные системы связи широко применяются в критически важных инфраструктурах, сенсорных сетях, системах Интернета вещей и робототехнических комплексах. Однако, беспроводные каналы связи являются уязвимыми к целому спектру атак из-за открытой среды передачи. Поэтому, проблема защиты беспроводных каналов от деструктивного воздействия является актуальной проблемой защиты информации [1].

Основные пути повышения устойчивости связи в радиоканалах отражено на рис. 1. Классические методы защиты от помех основаны на использовании технологий расширения спектра, таких как прямое расширение спектра и частотная перестройка. Эти методы позволяют снизить вероятность подавления сигнала за счёт распределения энергии сигнала по широкому диапазону частот. Однако при наличии интеллектуального злоумышленника, способного адаптировать стратегию атаки, эффективность таких методов может существенно снижаться. Альтернативным способом повышения устойчивости радиоканалов является повышение пространственной селективности. Подобный способ позволяет исключить вклад сигналов, пришедших из направлений злоумышленника за счёт построения ДН (Диаграммы направленности) в направлении легитимного узла-приёмника, что достигается за счёт использования фазированных антенных решёток. В предложенной работе было принято решение исследовать повышение устойчивости именно через повышение пространственной селективности, поскольку предложенный метод позволяет добиться высоких показателей подавления преднамеренных помех [2, 3]. Кроме того, предложенный алгоритм не предполагает противопоставление традиционным алгоритмам на основе ППРЧ, а напротив, может дополнять уже существующие решения, повышая устойчивость беспроводных сетей.

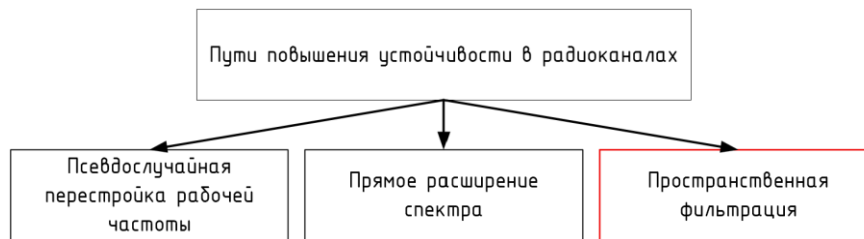


Рис. 1. Пути повышения устойчивости в радиоканалах

В последние годы активно развиваются методы противодействия помехам, основанные на использовании теории игр и методов машинного обучения. Игровые модели позволяют формализовать взаимодействие между легитимными пользователями и злоумышленниками как стратегическую игру с противоположными целями. В частности, игровые модели широко применяются в когнитивных радиосетях, где пользователи динамически выбирают каналы передачи данных в условиях присутствия помех и неопределённости среды [4].

В данной работе предлагается алгоритм противодействия преднамеренным помехам, основанный на комбинации методов теории игр, робастной оптимизации и обучения с подкреплением. Взаимодействие между легитимным узлом и злоумышленником формализуется в виде марковской игровой модели Штаклберга, в которой передатчик выступает лидером, а злоумышленник – последователем. Для обеспечения гарантированного уровня качества обслуживания используется робастная оптимизация SOCP (Second-order cone programming), учитывающая неопределённость оценки состояния канала. Оптимальная стратегия легитимного узла формируется с использованием алгоритма Soft Actor-Critic, позволяющего эффективно работать в непрерывном пространстве действий и стохастической среде.

1. Сравнительный анализ работ в области противодействия преднамеренным помехам. Теория игр представляет собой эффективный математический аппарат для моделирования взаимодействия рациональных агентов, стремящихся максимизировать собственную функцию полезности, что делает её особенно подходящей для исследования конфронтации между интеллектуальными глушителями и легитимными пользователями

[5]. В результате теоретико-игровые методы получили широкое распространение при разработке систем противодействия преднамеренным помехам, основанных на технологиях расширения спектра методом перестройки рабочей частоты и методом прямой последовательности [6, 7].

В последние годы значительное внимание уделяется применению методов обучения с подкреплением для защиты беспроводных систем связи. Несмотря на то что в литературе также рассматриваются альтернативные подходы, такие как методы ОО (Online convex optimization) [8–10], обучение с подкреплением обладает рядом преимуществ, среди которых способность адаптироваться к сложному и непредсказуемому поведению интеллектуальных глушителей [11–12]. В системах противодействия преднамеренным помехам методы обучения с подкреплением преимущественно применяются для оптимизации выбора рабочей частоты в ППРЧ-системах, выбора расширяющих последовательностей в системах с прямым расширением спектра, а также распределения мощности между несколькими подканалами.

Так как существенным недостатком традиционных систем с ППРЧ является фиксированная скорость перестройки рабочей частоты, для устранения данного недостатка в работе [13] было предложено расширить пространство действий агента, включив в него не только рабочую частоту, но и время пребывания на выбранной частоте. Для повышения эффективности обучения использовалась глубокая Q-сеть (Deep Q-Network), дополненная механизмом Pareto Prioritized Experience Replay (PPER). В данной работе наиболее значимыми считались те примеры обучения, которые одновременно характеризовались большой ошибкой временной разности (Temporal Difference Error) и высоким значением функции вознаграждения. Проведённое моделирование показало, что алгоритм PPER-DQN обеспечивает более быстрое обучение и более высокие значения SINR по сравнению с алгоритмами RC-DQN, TDC-DQN, Rank-DQN и классическим DQN.

В работе [14] предложен алгоритм Q-learning с инициализацией на основе априорных знаний о параметрах устройства постановки помех. Авторы сформировали граф знаний, содержащий информацию о типах устройств постановки помех, их характеристиках, стратегиях противодействия и начальных параметрах перестройки рабочей частоты. Использование экспертных знаний на этапе обучения позволило существенно ускорить сходимость алгоритма и повысить надёжность связи в условиях изменяющейся окружающей среды.

Похожий подход к организации перестройки рабочей частоты представлен в работе [15]. В отличие от традиционных методов обучения с подкреплением авторы модифицировали функцию вознаграждения, введя дополнительный штраф за смену рабочей частоты. Столкновение с помехой сопровождалось значительным штрафом, тогда как успешная передача данных приводила к меньшему штрафу, пропорциональному величине изменения рабочей частоты. Такая модификация позволила сократить число необоснованных перестроек частоты, уменьшить объём служебного обмена и повысить эффективность использования радиочастотного спектра.

В работе [16] предложена архитектура противодействия преднамеренным помехам, основанная на сочетании обучения с подкреплением и модели Transformer. Основная идея заключается в том, что действия интеллектуального устройства постановки помех обладают временной структурой и могут рассматриваться как последовательность событий. Использование слоя Transformer позволило учитывать долгосрочные зависимости в поведении нарушителя и повысить качество выбора рабочей частоты.

Для устройств с ограниченными вычислительными ресурсами в работе [17] предложен облегчённый алгоритм противодействия помехам на основе DQN. Метод ориентирован на борьбу с устройствами постановки помех, использующими динамически изменяющиеся шаблоны воздействия, а состояние среды формируется на основе измерений мощности сигналов в отдельных подканалах. Благодаря низкой вычислительной сложности предложенный подход пригоден для реализации на маломощных беспроводных устройствах.

В работе [18] разработана ППРЧ-система на основе DQN, в которой пространство действий включает одновременно выбор рабочей частоты и мощности передачи. Такой подход позволяет не только адаптировать используемый канал связи, но и одновременно минимизировать энергопотребление передатчика.

В отличие от рассмотренных выше методов, основанных на перестройке рабочей частоты, работа [19] посвящена оптимизации процесса пространственной селекции. Авторы использовали DQN сеть для оптимизации расположения подвижных антенных элементов с целью максимизации SINR. После определения оптимальной геометрии антенной решётки вычисление весового вектора пространственного фильтра выполнялось отдельно путём решения задачи максимизации отношения Рэлея.

Анализ существующих исследований показывает, что большинство методов противодействия преднамеренным помехам, основанных на обучении с подкреплением, ориентированы либо на выбор рабочей частоты, либо на распределение мощности передачи. Значительно меньше внимания уделяется совместной оптимизации робастного пространственного формирования луча и распределения мощности в условиях воздействия интеллектуального глушителя. Кроме того, существующие методы обучения с подкреплением редко сочетают механизмы адаптивного обучения с оптимизационными подходами, обеспечивающими детерминированные гарантии качества обслуживания. Указанные ограничения обуславливают необходимость разработки предлагаемой модели RL–Stackelberg–SOCP, объединяющей обучение с подкреплением, теорию игр и методы робастной выпуклой оптимизации в рамках единой архитектуры противодействия преднамеренным помехам.

2. Модель системы и постановка задачи. Рассматриваемый сценарий беспроводной связи представлен на рис. 2. Легитимные узлы соединяются по топологии РТР (Point-to-Point) либо РМР (Point-to-multiple-point). В случае применения топологии РМР связь ведомых узлов осуществляется через ведущий узел с использованием временного разделения каналов. В определённый квант времени обмен данными происходит между двумя легитимными узлами, один из которых является передатчиком, а другой приёмником. Передатчик выполняет двухуровневую процедуру оптимизации, позволяющую одновременно сформировать весовой вектор пространственного формирования луча в направлении легитимного приёмника и распределить мощность передачи между несколькими поднесущими. Цель злоумышленника заключается в нарушении связи между передатчиком и приёмником посредством создания преднамеренных помех. Принятый сигнал на i -й поднесущей может быть записан в следующем виде:

$$y_i = \sqrt{P_{s,i}} h_{s,i}^H \omega_i s_i + \sqrt{P_{j,i}} h_{j,i}^H \omega_i j_i + n_i,$$

где $P_{s,i}$ – мощность передачи легитимного передатчика на i -й поднесущей; $P_{j,i}$ – мощность помехового сигнала злоумышленника на i -й поднесущей; $h_{s,i}$ – вектор канала между легитимным передатчиком и приёмником; $h_{j,i}$ – вектор канала между злоумышленником и приёмником; ω_i – весовой вектор пространственного формирования луча; s_i – передаваемый информационный символ; j_i – помеховый сигнал злоумышленника; n_i – аддитивный белый гауссовский шум.

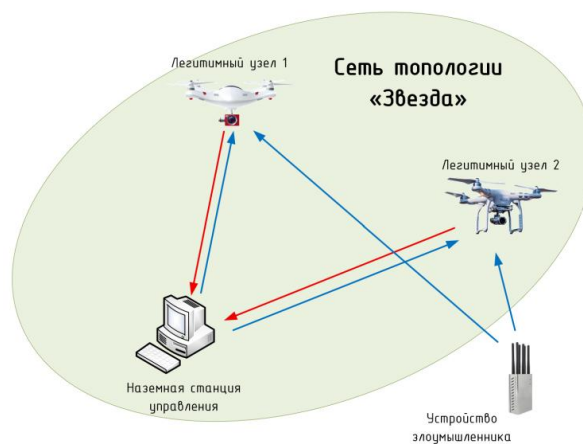


Рис. 2. Модель системы

3. Цель работы. Предлагается гибридный подход, в котором обучение с подкреплением оптимизирует стратегию лидера в марковской игровой модели Штаклберга, выбирая параметры робастности и порог SINR, а выполнение худшего случая ограничения SINR обеспечивается выпуклым робастным SOCP слоем. Таким образом достигается одновременно адаптивность к динамике злоумышленника и среды за счёт внедрения обучения с подкреплением и строгие гарантии по уровню SINR за счёт робастной конусной оптимизации. Игру можно сформулировать следующим образом:

$$\mathcal{G} = (\mathcal{S}, \mathcal{A}_L, \mathcal{A}_J, \mathcal{T}, r_L, r_J),$$

где $\mathcal{S}$ – пространство состояний игры; $\mathcal{A}_L, \mathcal{A}_J$ – пространства действий легитимного узла и злоумышленника соответственно; $\mathcal{T}$ – оператор перехода состояния; r_L, r_J – целевые функции легитимного узла и злоумышленника соответственно.

4. Неопределённость состояния канала и робастность к худшему случаю. Для подканала i SINR в момент t задаётся следующим образом:

$$SINR(t) = \frac{P_{s,i}(t) |\omega_i^H(t) h_{s,i}(t)|^2}{P_{j,i}(t) |\omega_i^H(t) h_{j,k}(t)|^2 + \sigma^2 \|\omega_i(t)\|_2^2},$$

где σ^2 – мощность аддитивного белого гауссовского шума. Принимается что полезный сигнал известен с некоторой погрешностью:

$$h_{s,i}(t) = h'_{s,i}(t) + \delta h_{s,i}(t), \quad \|\delta h_{s,i}(t)\|_2 \leq \varepsilon(t),$$

где $\varepsilon(t)$ – радиус множества неопределённости. Таким образом, радиус множества неопределённости является одним из параметров робастности, который может быть подобран в ходе обучения предложенного алгоритма.

Задача распределения мощности осуществляется за счёт максимизации пропускной способности с учётом ограничения по энергетическому бюджету передатчика:

$$\begin{aligned} \max_P \quad & \sum_{k=1}^K \log(1 + SINR_k) - \lambda \sum_{i=1}^K P_{s,k}, \\ \text{s. t.} \quad & \sum_{k=1}^K P_{s,k} \leq P_s^{max}, \\ & P_{s,k} \geq 0, \end{aligned}$$

где λ – коэффициент штрафа за избыточную мощность передачи. Таким образом, параметр λ является вторым параметром робастности, подбираемым в ходе обучения предложенного алгоритма.

5. Действия легитимного узла. Легитимный узел, как лидер в иерархической игре, выбирает управляемые параметры:

$$a_t(\lambda_t, \varepsilon_t),$$

где ε_t – предполагаемая неопределённость в измерении параметров канала легитимного узла; $\lambda_t \geq 0$ – вес штрафа за использование избыточной мощности в игре. Так как моделирование конфронтации злоумышленника и легитимного узла происходит по модели Штаклберга, поиск равновесия происходит при помощи метода обратной индукции. В рассмотренной модели злоумышленник моделируется как последователь, а легитимный узел как лидер. По этой причине, на первой итерации происходит поиск решения для лучшего ответа злоумышленника, а затем лидер учитывает эту реакцию.

После выбора a_t легитимным узлом, злоумышленник выбирает распределение мощности как лучший ответ:

$$P'_j(t) \in \arg \max_{P_j \geq 0, \sum_i P_{j,i} \leq P_j^{max}} U_j(s_t, a_t, P_j),$$

где U_j – целевая функция злоумышленника. В дальнейшем легитимный пользователь учитывает $P'_j(t)$ для расчёта лучшего ответа.

Если целевая функция U_j имеет вид суммы вогнутых функций по подканалам и линейное ограничение мощности, то решение $P'_{j,i}$ может быть получено из ККТ (Каруша-Куна-Такера) с учётом ограничений на максимальную мощность злоумышленника P_j^{max} и имеет следующую форму:

$$P'_{j,i}(t) = \left[\frac{\beta_i}{\mu} - \frac{1}{k_i(t)} \right], \quad \sum_i P'_{j,i}(t) = P_j^{max},$$

где μ – множитель Лагранжа; $k(t)$ – эффективность помехи. Лучший ответ легитимного узла в соответствии с моделью Штаклберга может быть вычислен следующим образом:

$$P'_{s,i} = \left[\frac{1}{\mu \ln 2} - \frac{P'_{j,i} |\omega_i^H h_{j,i}|^2 + \sigma^2 |\omega_i|^2}{|\omega_i^H h_{s,i}|^2} \right]_+,$$

где μ – множитель Лагранжа; σ – уровень фоновых шума на один приёмный тракт.

6. Слой поиска весовых коэффициентов антенной решетки. После получения $P'_{j,i}(t)$ легитимный узел вычисляет вектор весов ω . Предположив, что известна ковариация помех $R_{j,i}$ для каждого подканала такая что:

$$R_{j,k}(t) = P'_{j,i}(t) h_{j,k}(t) h_{j,k}^H(t).$$

Следующим шагом необходимо найти вектор весовых коэффициентов. Предлагается искать весовые коэффициенты при фиксированных параметрах мощности для подканалов $P_{s,i}$. Тогда требование для SINR можно сформулировать следующим образом:

$$\min_{\|\delta h_{s,i}\| \leq \varepsilon(t)} \frac{P_{s,i}(t) |\omega_i^H (h'_{s,i} + \delta h_{s,i})|^2}{\omega_i^H R_{j,k} \omega_i + \sigma^2 \|\omega_i\|_2^2} \geq \gamma_t.$$

При этом сохраняется стандартная робастная нижняя оценка, представленная формулой ниже:

$$\min_{\|\delta h\| \leq \varepsilon} |\omega^H (h' + \delta h)| \geq |\omega^H h'| - \varepsilon \|\omega\|_2.$$

В итоге задача определения ω_i будет иметь следующий вид:

$$\begin{aligned} \min_{\omega_i \in \mathbb{C}^M} & \|\omega_i\|_2^2 \\ \Re\{\omega_i^H h'_{s,i}\} - \varepsilon(t) \|\omega_i\|_2 & \geq \sqrt{\gamma_t} \left\| \frac{R_{j,i}^{\frac{1}{2}} \omega_i}{\sqrt{\sigma^2}} \right\|_2, \\ \Im\{\omega_i^H h'_{s,i}\} & = 0. \end{aligned}$$

где $R_{j,i}$ – ковариационная матрица i -го канала злоумышленника. Вышеописанные ограничения позволяет найти такое ω_i , которое удовлетворяет гарантии $SINR \geq \gamma_t$ при $\|\delta h_{s,i}\| \leq \varepsilon(t)$.

7. Марковская динамика и состояние. Марковская игра требует, чтобы состояние содержало всю информацию, влияющую на динамику среды, т.е.:

$$\Pr(s_{t+1} | s_t, a_t) = \Pr(s_{t+1} | s_0, a_0, \dots, s_t, a_t).$$

Таким образом, состояние должно включать следующие составляющие:

- ◆ Текущая геометрия каналов.
- ◆ Текущий уровень неопределённости ε .
- ◆ Текущая мощность злоумышленника.

Предлагаемый вектор состояния:

$$s_t = (g_s(t), g_j(t), \varepsilon'(t), \lambda(t), P_j^{max}),$$

где $g_s(t) = [\|h'_{s,1}(t)\|_2^2, \dots, \|h'_{s,K}(t)\|_2^2] \in \mathbb{R}^K$ – эффективность полезного канала; $g_j(t) = [\|h'_{j,1}(t)\|_2^2, \dots, \|h'_{j,K}(t)\|_2^2] \in \mathbb{R}^K$ – эффективность канала злоумышленника; $\varepsilon'(t) = \sqrt{\frac{1}{L} \sum_{i=t-L}^t \|h'_s(i) - h'_s(i-1)\|_2^2}$ – текущая оценка неопределённости; $\lambda(t)$ – размер штрафа за использование избыточной мощности передачи; P_j^{max} – ограничение по мощности злоумышленника.

Для моделирования подвижности легитимного узла необходим учёт временной динамики. Подобные воздействия на канал могут быть учтены при помощи модели Джейка [20]. С учётом дискретного времени t динамика каналов для моделирования задаётся следующим образом:

$$\begin{aligned} h_{s,i}(t) &= a(\theta_s) \odot g_{s,i}(t), \\ h_{j,i}(t) &= a(\theta_j) \odot g_{j,i}(t), \\ g_{j,i,m}(t) &= \frac{1}{\sqrt{N}} \sum_{n=1}^N e^{j(2\pi f_{D,i} \cos \theta_{n,m} t + \phi_{n,i})}, \end{aligned}$$

где $g_{s,i,m}$, $g_{j,i,m}$ – векторы случайных процессов в соответствии с моделью замирания Джейка для i -го подканала -ой антенны для легитимного узла и злоумышленника соответственно, $f_D = \frac{v f_c}{c}$ – доплеровская частота.

8. Описание функции награды лидера. Награда лидера определяется как компромисс между пропускной способностью, энергозатратами и устойчивостью к резкому изменению вектора ω :

$$\begin{aligned} r_t &= \sum_{i=1}^I \log_2(1 + SINR_i(t)) - q_1 \lambda_t \sum_{i=1}^I p_{s,i} + \\ &+ q_2 \max(0, \Delta_{SINR}) - q_3 \cdot \max(0, -\Delta_{SINR}) - q_4 \cdot I_{outage}, \end{aligned}$$

где $\log_2(1 + SINR_i(t))$ – награда, пропорциональная пропускной способности каждого подканала; $\sum_{i=1}^I p_{s,i}$ – штраф за уровень мощности передатчика; $\max(0, \Delta_{SINR})$ – награда за прирост SINR; $-\max(0, \Delta_{SINR})$ – штраф за деградацию SINR; $I_{outage} = \begin{cases} 1, SINR < \gamma \\ 0, SINR > \gamma \end{cases}$ – штраф за отказ связи; q_i – весовые коэффициенты каждого -го слагаемого награды.

9. Обучение лидера и RL-политика. Так как легитимный узел выбирает детерминированный лучший ответ, можно определить эффективные для этого узла функции перехода и награды, которые могут быть представлены формулой ниже:

$$\bar{r}(s, a) = r_L(s, a, BR(s, a)),$$

где $BR(\cdot)$ – Лучший ответ. Оператор перехода может быть описан следующим образом:

$$\bar{T}(s'|s, a) = T(s'|s, a, BR(s, a)).$$

В итоге задача для легитимного узла является обычным Марковским процессом принятия решений:

$$\mathcal{M}_L(\mathcal{S}, \mathcal{A}_L, \bar{T}, \bar{r}, \gamma_{RL}),$$

где $\gamma_{RL}^t \in (0,1)$ – коэффициент дисконтирования. Задаётся параметризованная политика $\pi_\theta(a|s)$. Данная политика максимизирует математическое ожидание от суммарной пропускной способности за время t :

$$\max_{\theta} \mathbb{E} \left[\sum_{t=0}^T \gamma_{RL}^t \bar{r}(s_t, a_t) \right],$$

где $\gamma_{RL}^t \in (0,1)$ – коэффициент дисконтирования. Функция ценности может быть описана формулой ниже:

$$Q^\pi(s, a) = \mathbb{E}_\pi \left[\sum_{t=0}^T \gamma_{RL}^t \bar{r}(s_t, a_t) | s_0 = s, a_0 = a \right].$$

Тогда функция Беллмана для оптимальной политики $\pi(s) \in \arg \max_a Q^\pi(s, a)$:

$$Q'(s, a) = \bar{r}(s, a) + \gamma_{RL} \mathbb{E}_{s' \sim \bar{\mathcal{T}}(\cdot|s, a)} \left[\max_{a' \in \mathcal{A}_L} Q'(s_{t+1}, a_{t+1}) \right].$$

10. Обновление параметров сети. В процессе обучения обновляются следующие параметры:

- ◆ параметры сети актора θ ;
- ◆ параметры первой сети критика φ_1 ;
- ◆ параметры второй сети критика φ_2 ;
- ◆ коэффициент энтропийной регуляризации α ;
- ◆ параметры целевых сетей критика $\bar{\varphi}_1$ и $\bar{\varphi}_2$.

После каждого перехода вычисляется целевое значение:

$$y_t = r_t + \gamma \left(\min_{1,2} Q_{\bar{\varphi}_i}(s_{t+1}, a_{t+1}) - a \log \pi_\theta(a_{t+1}|s_{t+1}) \right),$$

где $Q_{\bar{\varphi}_i}$ – целевая сеть критика; γ – коэффициент дисконтирования. Каждая сеть критика минимизирует среднеквадратичную ошибку Беллмана:

$$J_Q(\varphi_i) = \mathbb{E} \left[(Q_{\bar{\varphi}_i}(s_t, a_t) - y_t)^2 \right].$$

Параметры критика обновляются методом градиентного спуска:

$$\varphi_i \leftarrow \varphi_i - \eta_Q \nabla_{\varphi_i} J_Q(\varphi_i),$$

где η_Q – скорость обучения сети критика. После обновления критиков производится обновление сети актора посредством максимизации ожидаемой функции $J_\pi(\theta)$:

$$J_\pi(\theta) = \mathbb{E} \left[a \log \pi_\theta(a_t|s_t) - \min_{1,2} Q_{\bar{\varphi}_i}(s_t, a_t) \right].$$

Параметры актора обновляются по правилу, схожему с обновлением параметров критика:

$$\theta \leftarrow \theta - \eta_\pi \nabla_\theta J_\pi(\theta),$$

где η_π – скорость обучения сети актора. Для автоматического регулирования компромисса между исследованием среды и использованием накопленного опыта применяется энтропийная регуляризация. Коэффициент энтропии обновляется путём минимизации $J(a)$:

$$J(a) = \mathbb{E} \left[-a (\log \pi_\theta(a_t|s_t) + \mathcal{H}_{target}) \right],$$

где $\mathcal{H}_{target}$ – целевое значение энтропии. Обновление коэффициента энтропии осуществляется по правилу:

$$a \leftarrow a - \eta_a \nabla_a J(a),$$

где η_a – скорость обучения коэффициента энтропии. Следующим шагом происходит обновление целевых сетей критика:

$$\bar{\varphi}_i \leftarrow \tau \varphi_i + (1 - \tau) \bar{\varphi}_i,$$

где τ – коэффициент мягкого обновления, удовлетворяющий условию $0 < \tau \ll 1$.

11. Алгоритм обучения с подкреплением. Итоговый алгоритм машинного обучения изображён на рис. 3.

Результаты. Для сравнения эффективности алгоритмов защиты от преднамеренных помех использовалась метрика ОР (Outage Probability) рис. 4. Данный показатель характеризует вероятность успешного нарушения доступности канала связи злоумышленником. Моделирование проводилось методом Монте-Карло с числом итераций равным 5000. В качестве механизма исследования использовалась энтропийная регуляризация SAC. В алгоритме используется архитектура Soft Actor-Critic, включающая одну сеть актора и две сети критика. Актор представляет собой полносвязную нейронную сеть с входным сло-

ем размерности $2K + 3 = 13$ (K – количество подканалов, 3 дополнительных элемента – значения параметров робастности ε , λ , а также мощность злоумышленника), двумя скрытыми слоями по 128 нейронов с функцией активации ReLU и выходным слоем, формирующим средние значения и стандартные отклонения гауссовского распределения для двух управляющих параметров ε и λ . Коэффициент дисконтирования установлен равным 0,99, коэффициент мягкого обновления целевых сетей – 0,005, целевая энтропия – -2 . Обучение сети происходило за 250 эпизодов, в каждом из которых было 20 шагов.

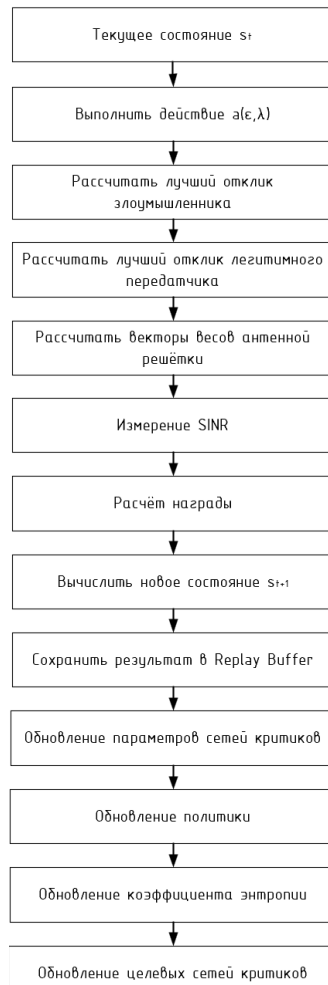


Рис. 3. Схема предложенного алгоритма

Для оценки эффективности предложенного алгоритма использовалась многоканальная система связи с передачей по нескольким подканалам OFDM. Канал связи моделировался на основе геометрической модели с использованием направляющих векторов, которые определяются стандартной моделью для равномерной линейной антенной решётки:

$$a(\theta) = [1, e^{-j\pi \sin(\theta)}, \dots, e^{-j\pi(M-1) \sin(\theta)}]^T,$$

где θ – угол, M – номер элемента антенной решётки. Для каждого подканала независимо генерировались два комплексных вектора канала:

- ♦ канал легитимного передатчика $h_{s,i}$;
- ♦ канал злоумышленника $h_{j,i}$.

Для моделирования временной изменчивости беспроводного канала использовался коррелированный процесс эволюции канала. На каждом временном шаге генерировались новые реализации каналов, после чего состояние канала обновлялось с использованием коэффициента временной корреляции $\rho = 0,92$. Такой подход позволял моделировать плавное изменение характеристик канала во времени, характерное для квазистационарных беспроводных каналов связи. В моделировании использовалась линейная антенная решетка с 4 равномерно расположенными элементами. Источник преднамеренных помех и легитимный узел находятся равноудалённо от приёмника, а угол между двумя источниками составляет 8 градусов.

Для анализа использовалась система с 5 различными каналами, на каждом из которых передаётся сигнал с модуляцией QPSK (Quadrature Phase-Shift Keying). Мощность передатчика взята равной 1 Вт, в то время как мощность злоумышленника бралась в диапазоне от 1 Вт до 10 Вт. Отношение ОСШ было взято равным 8дБ. В качестве моделирования случайных процессов в канале была взята модель Джейка. Для сравнения использовались алгоритмы: ППРЧ с использованием игровой модели Штаклберга [21], ППРЧ с использованием алгоритма обучения с подкреплением Stackelberg Q-learning [22] и RL-Stackelberg + SOCP. По графику можно увидеть что предложенный алгоритм позволяет обеспечить низкую вероятность отказа. До уровня 10дБ вероятность отказа у предложенного алгоритма близка к нулю, далее медленно возрастает до 0.2 на SINR 20дБ.

Повышение устойчивости системы обеспечивается использованием робастной оптимизации при формировании передающих весов. Учет неопределённости оценки состояния канала снижает влияние ошибок оценки параметров среды и предотвращает возможное усиление эффективности атаки за счёт неточностей измерений.

Таким образом, результаты моделирования подтверждают, что предложенный алгоритм позволяет повысить устойчивость беспроводной системы связи к атакам, направленным на нарушение доступности канала передачи данных, что означает снижение вероятности успешной реализации атаки отказа в обслуживании на физическом уровне беспроводной сети.

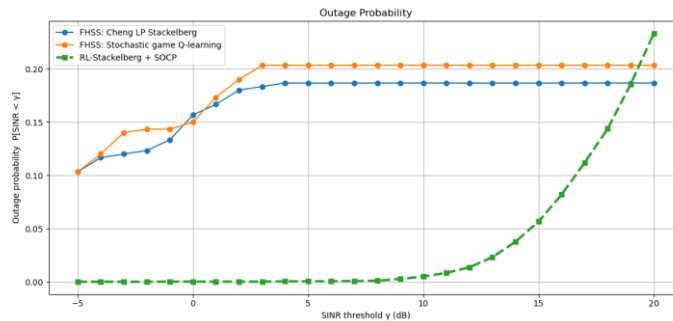


Рис. 4. Сравнение вероятности отказа связи относительно требуемого порога γ

Заключение. В работе предложен алгоритм противодействия преднамеренным помехам в беспроводных системах связи, основанный на комбинации методов теории игр, робастной оптимизации и обучения с подкреплением. Для обеспечения гарантированного уровня качества обслуживания задача формирования передающих весов решается с использованием робастной оптимизации второго порядка (SOCP), позволяющей учитывать неопределённость оценки состояния канала. Результаты моделирования показали, что предложенный алгоритм обеспечивает более низкую вероятность отказа соединения по сравнению с традиционными методами противодействия помехам, такими как ППРЧ. Это достигается за счёт адаптивной настройки параметров передачи и учёта динамического поведения злоумышленника.

Полученные результаты могут быть использованы при разработке устойчивых к преднамеренным помехам систем беспроводной связи, включая когнитивные радиосети, сенсорные сети и системы Интернета вещей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Xiao Y., Ren H., Wu S., Liu L., Meng X., Ding P. Anti-Jamming Method of Cognitive Radio Based on Q-Learning, *Proceedings of the 12th International Conference on Electronics, Communications and Networks (CECNet 2022)*. *Frontiers in Artificial Intelligence and Applications*, 2022, Vol. 363, pp. 97-104. DOI: 10.3233/FAIA220523 (accessed 09 March 2026).
2. Zumegen F.J., Jain I.K., Bharadia D. BeamArmor: Seamless Anti-Jamming in 5G Cellular Networks with MIMO Null-Steering, *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications (HotMobile '24)*. New York: Association for Computing Machinery, 2024, pp. 121-126. DOI: 10.1145/3638550.3641138 (accessed 09 March 2026).
3. Andrei V.C., Djuhera A., Li X., Mönich U.J., Boche H., Saad W. Resilient-by-Design Framework for MIMO-OFDM Communications under Smart Jamming. – 2024. – URL: <https://rosap.nrl.bts.gov/view/dot/80055> (accessed 09 March 2026).
4. Jia L., Qi N., Chu F., Fang S., Wang X., Ma S., Feng S. Game-Theoretic Learning Anti-Jamming Approaches in Wireless Networks, *IEEE Communications Magazine*, 2022, Vol. 60, pp. 60-66. DOI: 10.1109/MCOM.001.00496.
5. Han C., Huo L., Tong X., Wang H., Liu X. Spatial Anti-Jamming Scheme for Internet of Satellites Based on the Deep Reinforcement Learning and Stackelberg Game, *IEEE Transactions on Vehicular Technology*, 2020, Vol. 69, pp. 5331-5342. DOI: 10.1109/TVT.2020.2982672.
6. Aref M., Jayaweera S., Yezpe E. A Survey on Cognitive Anti-Jamming Communications, *IET Communications*, 2020, Vol. 14, No. 18, pp. 3111-3127. DOI: 10.1049/iet-com.2020.0024.
7. Imran M., Pan Z., Liu N., Sajjad M., Butt F. Anti-Jamming for Cognitive Radio Networks with Stackelberg Game-Assisted DSSS Approach, *EURASIP Journal on Wireless Communications and Networking*, 2024, Vol. 2024, pp. 1-18. DOI: 10.1186/s13638-024-02397-x.
8. Liu L., Pu W., Li Y., Luo Z.-Q. Learning an Opponent-Aware Anti-Jamming Strategy via Online Convex Optimization, *arXiv*, 2026. arXiv:2604.11016. DOI: 10.48550/arXiv.2604.11016.
9. Liu L., Pu W., Li Y., Jiu B., Luo Z.-Q. Radar Anti-Jamming Strategy Learning via Domain-Knowledge Enhanced Online Convex Optimization, *arXiv*, 2024. arXiv:2402.16274. DOI: 10.48550/arXiv.2402.16274.
10. Chen S., Tu W.-W., Zhao P., Zhang L. Optimistic Online Mirror Descent for Bridging Stochastic and Adversarial Online Convex Optimization, *Proceedings of the 40th International Conference on Machine Learning (ICML)*. *Proceedings of Machine Learning Research*, 2023, Vol. 202, pp. 5002-5035. Available at: <https://proceedings.mlr.press/v202/chen23aa.html> (accessed 22 June 2026).
11. Yao F., Jia L. A Collaborative Multi-Agent Reinforcement Learning Anti-Jamming Algorithm in Wireless Networks, *arXiv*, 2018. arXiv:1809.04374. DOI: 10.48550/arXiv.1809.04374.
12. Joulani P., Gyorgy A., Szepesvari C. Delay-Tolerant Online Convex Optimization: Unified Analysis and Adaptive-Gradient Algorithms, *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016, Vol. 30, pp. 1744-1750. DOI: 10.1609/aaai.v30i1.10320.
13. Chen Y., Zhao Z. Intelligent Anti-Jamming Decision Algorithm of Bivariate Frequency Hopping Pattern Based on ET-PPO, *Telecommunications Science*, 2022, Vol. 38, pp. 1-13. DOI: 10.11959/j.issn.1000-0801.2022264.
14. Feng X., Niu Y., Liu Q., Zhou Q. Decision Making for Communication Anti-Jamming Tasks with Knowledge-Graph-Based Q-Learning, *Electronics*, 2024, Vol. 13, Art. 4757. DOI: 10.3390/electronics13234757.
15. Chang X., Li Y., Zhao Y., Du Y., Liu D. An Improved Anti-Jamming Method Based on Deep Reinforcement Learning and Feature Engineering, *IEEE Access*, 2022, Vol. 10, pp. 69992-70000. DOI: 10.1109/ACCESS.2022.3187030.
16. Xu J., Lou H., Zhang W., Sang G. An Intelligent Anti-Jamming Scheme for Cognitive Radio Based on Deep Reinforcement Learning, *IEEE Access*, 2020, Vol. 8, pp. 202563-202572. DOI: 10.1109/ACCESS.2020.3036027.
17. Ali A., Naser S., Muhaidat S. Defeating Proactive Jammers Using Deep Reinforcement Learning for Resource-Constrained IoT Networks, *arXiv*, 2023. arXiv:2307.06796. DOI: 10.48550/arXiv.2307.06796.
18. Li X., Chen J., Ling X., Wu T. Deep Reinforcement Learning-Based Anti-Jamming Algorithm Using Dual Action Network, *IEEE Transactions on Wireless Communications*, 2023, Vol. 22, pp. 5403-5418. DOI: 10.1109/TWC.2022.3227575.
19. Tang X., Jiang Y., Liu J., Du Q., Niyato D., Han Z. Deep Learning-Assisted Jamming Mitigation with Movable Antenna Array, *arXiv*, 2024. arXiv:2410.20344. DOI: 10.48550/arXiv.2410.20344.
20. Goldsmith A. *Wireless Communications*. Cambridge: Cambridge University Press, 2005, 584 p.
21. Cheng L., Xu Z., Zhou J., Tian D., Duan X., Qu K., Zhao D. Adaptive Spectrum Anti-Jamming in UAV-Enabled Air-to-Ground Networks: A Bimatrix Stackelberg Game Approach, *Electronics*, 2023, Vol. 12, Art. 4344. DOI: 10.3390/electronics12204344.
22. Noori H., Vilni S. Jamming and Anti-Jamming in Interference Channels: A Stochastic Game Approach, *IET Communications*, 2020, Vol. 14, pp. 682-692. DOI: 10.1049/iet-com.2019.0637.

Григорян Карен Сергеевич – Южный федеральный университет; e-mail: kgr@sfedu.ru; г. Таганрог, Россия; тел.: +79889490693; кафедра безопасности информационных технологий им. О.Б. Макаревича; аспирант.

Басан Елена Сергеевна – Южный федеральный университет; e-mail: ebasan@sfedu.com; г. Таганрог, Россия; тел.: +79880158697; кафедра безопасности информационных технологий им. О.Б. Макаревича; доцент.

Grigoryan Karen Sergeevich – Southern Federal University; e-mail: kgr@sfedu.ru; Taganrog, Russia; phone: +79889490693; the Department of Information Technology Security named after O.B. Makarevich; postgraduate student.

Basan Elena Sergeevna – Southern Federal University; e-mail: ebasan@sfedu.com; Taganrog, Russia; phone: +79880158697; the Department of Information Technology Security named after O.B. Makarevich; associate professor.

УДК 004.932

DOI 10.18522/2311-3103-2026-4-77-90

С.В. Ермоленко, И.Л. Каширина

УЛУЧШЕНИЕ СЕГМЕНТАЦИИ НА МУЛЬТИФАЗНЫХ КТ СНИМКАХ С ПОМОЩЬЮ ОБУЧАЕМОГО ФАЗОВОГО СОВМЕЩЕНИЯ

Автоматизированный анализ мультифазных КТ снимков, получаемых в разные моменты времени после введения контрастного вещества, крайне важен для точного выявления патологий, но связан с фундаментальной проблемой пространственного несовпадения изображений на разных фазах из-за дыхания и движения пациента, что приводит к значительному снижению точности автоматической сегментации. Существующие методы предварительной регистрации (совмещения фаз) требуют ручной настройки параметров и не интегрируются в обучаемые конвейеры, что препятствует полной автоматизации процесса сегментации. Цель данного исследования заключалась в разработке дифференцируемого метода совмещения мультифазных КТ-изображений на основе обучаемых линейных аффинных преобразований, полностью встроенного в конвейер обучения модели сегментации. В отличие от традиционных подходов, предложенный метод реализует дифференцируемый модуль аффинной регистрации (включающий перенос, поворот и масштабирование), параметры которого оптимизируются градиентным спуском без ручной настройки и встроены непосредственно в вычислительный граф модели nnU-Net. В работе на открытом датасете КТ органов брюшной полости WAW-TACE проведено сравнение предлагаемого метода с базовым подходом сегментации мультифазных изображений (основанном на простой конкатенации фаз без регистрации) и методом аффинной регистрации, реализованным в библиотеке ITK.Elastix. Относительно сегментации без регистрации достигнут существенный прирост метрик качества: коэффициента Dice на 43,19% и метрики ROC-AUC на 13,55%. По сравнению с Elastix улучшение составило 13,43% по Dice, 6,86% по ROC-AUC и 19,6% по точности определения количества патологий на КТ-снимке. Практическая значимость исследования заключается в разработке готового модуля на PyTorch для интеграции в конвейеры анализа КТ-изображений, обеспечивающего полную автоматизацию регистрации без необходимости в ручной настройке гиперпараметров и с высокой вычислительной эффективностью.

Мультифазная компьютерная томография; сегментация медицинских изображений; дифференцируемая регистрация; глубокое обучение; nnU-Net; аффинное преобразование.

S.V. Ermolenko, I.L. Kashirina

IMPROVING SEGMENTATION IN MULTIPHASE CT IMAGES USING TRAINABLE PHASE SUPERIMAGING

Joint Automated analysis of multiphase CT scans, acquired at different time points after contrast agent administration, is highly important for accurate pathology detection. However, it faces a fundamental problem of spatial misalignment between phases due to patient breathing and movement, which leads to a significant reduction in the accuracy of automatic segmentation. Existing pre-registration (phase alignment) methods require manual parameter tuning and are not integrated into trainable pipelines, hindering the full au-