DOI: 10.15514/ISPRAS-2025-37(4)-2

Алгоритм оптимизации «черепаха и заяц» с адаптивной стратегией на основе взаимной информации для обнаружения сетевых вторжений

Т. Бхуванешвари, ORCID: 0000-0002-4403-307X <bhuvaneswarit@mepcoeng.ac.in>
К. Руба Соундар, ORCID: 0000-0003-1300-6519 <rubasoundar@mepcoeng.ac.in>
Р. Чандра Гуру Секар, ORCID: 0000-0002-1436-1785 <chandragurusekar@mepcoeng.ac.in>
Технический колледж Мепко Шленк,
Индия, 626005, Тамилнад, Сивакаси, Вирудхунагар.

Аннотация. В современную эпоху сильной взаимосвязанности данные и информация постоянно передаются по сетям. Обеспечение безопасности конфиденциальной информации и защиты компьютерных систем от сетевых угроз стало очень актуально. Поэтому важна разработка эффективной системы обнаружения вторжений в сеть (NIDS) с использованием оптимальных признаков. Эти признаки могут быть определены с помощью искусственного интеллекта путем изучения шаблонов и взаимосвязей методами машинного обучения (machine learning, ML). В статье представлена методика оптимизации типа «черепаха и заяц» для выбора оптимальных признаков. Для оценки используется набор данных UNSW-NB15. Результаты оптимизации достигают точности 94,12% для бинарной классификации и 93,92% для многоклассовой классификации, при этом из всего набора признаков выбираются 26 оптимальных. Чтобы улучшить подход, используется адаптивная стратегия на основе взаимной информации для управления количеством оптимальных признаков. Эта стратегия вместе с алгоритмом черепахи и зайца повышает точность, показывая 94,69% для бинарной классификации и 94,03% для многоклассовой классификации, при этом сокращая количество выбранных признаков до 9. Сравнительный анализ производительности показывает, что предлагаемый метод выбора признаков превосходит другие современные методы, обеспечивая более точные и надежные результаты при выявлении киберугроз. Кроме того, график связи количества оптимальных признаков и точности модели показывает, что выбор только 9 признаков является эффективным для достижения высокой точности обнаружения и прогнозирования кибератак.

Ключевые слова: система обнаружения сетевых вторжений (NIDS); алгоритм метаэвристической оптимизации; алгоритм оптимизации «черепаха и заяц» (RTOA); выбор признаков (FS); выбор признаков на основе оболочки; взаимная информация (MI).

Для цитирования: Бхуванешвари Т., Руба С. К., Чандра Г. С. Р. Алгоритм оптимизации «черепаха и заяц» с адаптивной стратегией на основе взаимной информации для обнаружения сетевых вторжений. Труды ИСП РАН, том 37, вып. 4, часть 1, 2025 г., стр. 31–50. DOI: 10.15514/ISPRAS–2025–37(4)–2.

Благодарности: Авторы благодарят анонимных рецензентов за их предложения и замечания по улучшению качества статьи.

Rabbit and Tortoise Optimization Algorithm with Mutual Information Based Adaptive Strategy for Network Intrusion Detection

T. Bhuvaneswari, ORCID: 0000-0002-4403-307X <bhuvaneswarit@mepcoeng.ac.in>
K. Ruba Soundar, ORCID: 0000-0003-1300-6519 <rubasoundar@mepcoeng.ac.in>
R. Chandra Guru Sekar, ORCID: 0000-0002-1436-1785 <chandragurusekar@mepcoeng.ac.in>
Mepco Schlenk Engineering College,
Virudhunagar, Sivakasi, Tamil Nadu, 626005, India.

Abstract. In the modern era of highly interconnectedness, data and information are constantly transmitted over networks. Ensuring the security of confidential information and protecting computer systems from network threats has become very important. Therefore, it is important to develop an effective network intrusion detection system (NIDS) using optimal features. These optimal features can be identified through computational intelligence by learning patterns and relationships among features using machine learning techniques. This paper presents a Rabbit and Tortoise optimization technique for selecting optimal features. For evaluation, the UNSW-NB15 dataset is utilized. The optimization results achieve an accuracy of 94.12% for binary classification and 93.92% for multi-class classification, with 26 optimal features selected from the entire feature set. To improve the approach, an adaptive strategy based on mutual information is used to control the number of optimal features. This strategy, together with the Rabbit and Tortoise algorithm, improves the accuracy, showing 94.69% for binary classification and 94.03% for multi-class classification, while reducing the number of selected features to 9 only. The comparative performance analysis shows that the proposed feature selection method outperforms other state-of-the-art methods, providing more accurate and reliable results in identifying cyber threats. In addition, the relationship plot between the number of optimal features and the accuracy of the model shows that selecting only 9 features is effective in achieving high accuracy in detecting and predicting cyber-attacks.

Keywords: network intrusion detection system (NIDS); metaheuristic optimization algorithm; rabbit and tortoise optimization algorithm (RTOA); feature selection (FS); wrapper-based feature selection; mutual information (MI).

For citation: Bhuvaneswari T., Ruba Soundar K., Chandra Guru Sekar R. Rabbit and Tortoise optimization algorithm with mutual information based adaptive strategy for network intrusion detection. Trudy ISP RAN/Proc. ISP RAS, vol. 37, issue 4, part 1, 2025, pp. 31-50 (in Russian). DOI: 10.15514/ISPRAS-2025-37(4)-2.

Acknowledgements. The authors thank anonymous reviewers for their suggestions and comments to improve the quality of the article.

1. Введение

В современном взаимосвязанном мире инновации стали жизненно важным компонентом повседневной жизни. Люди в значительной степени полагаются на Интернет, искусственный интеллект, пользуясь такими устройства, как смартфоны и умные гаджеты. Однако растущая зависимость от онлайн-сетей и устройств чревата растущим риском киберугроз. Эта уязвимость распространяется и на цифровые активы, принадлежащие предприятиям, включая сетевые системы. Со временем частота кибератак только растет, а их номенклатура постоянно расширяется [1]. Такие атаки приобрели международный характер, что привело к заметному снижению общей продуктивности мировой экономики. Чтобы преодолеть эти угрозы безопасности, компьютерные системы должны быть оснащены инструментами для проведения регулярного мониторинга и анализа поведенческих моделей, как для рутинных, так и для необычных действий. Концепция систем обнаружения вторжений (СОВ) играет важную роль в выявлении и устранении новых угроз безопасности и сетевых атак [2].

Функции обнаружения вторжений включают наблюдение и анализ действий пользователей и системы, оценку настроек и уязвимостей систем, оценку целостности систем и файлов, выявление семейств и шаблонов атак, изучение нестандартного поведения и мониторинг

нарушений политик пользователя. Сетевые вторжения обнаруживаются на основе анализа характеристик сетевого трафика.

Большинство современных систем обнаружения сетевых вторжений (network intrusion detection system, NIDS) идентифицируют атаки, анализируя все признаки, полученные из переданных по сети данных. При этом вычислительная стоимость анализа всех атрибутов очень высока. Однако для эффективного обнаружения атак не все признаки необходимы. Более того, показано, например, что избыток признаков в наборе данных UNSW-NB15 [1] приводит к переобучению и увеличению времени обнаружения вторжений.

Применение методов машинного обучения (machine learning, ML) в сочетании с отбором признаков в наборе данных UNSW-NB15 может облегчить идентификацию важных атрибутов для обнаружения вторжений. Уменьшение количества признаков может привести к сокращению времени обнаружения без ущерба для точности. Основная цель отбора признаков – идентификация и отбор признаков по степени их важности. Выбор признаков нашел применение в различных областях исследований, например, в таких, как обнаружение вторжений [3], прогнозирование болезни Паркинсона [4], прогнозирование рака [5] и многих других.

Методы отбора признаков можно разделить на три группы: фильтрация, обертка и гибридный метод. При фильтрации для определения критических признаков используют методы статистики, теории информации и теории множеств. Отбор признаков на основе оберток – это один из методов интеллектуального анализа данных, который оценивает и выбирает подмножества признаков с учетом их производительности в рамках выбранной модели машинного обучения. Процесс включает в себя итеративное обновление подмножеств признаков с помощью стратегии поиска и алгоритма оптимизации. Гибридный выбор признаков объединяет два или более методов выбора признаков, как правило, из методов фильтрации и обертки, чтобы использовать их сильные стороны и смягчить ограничения.

Последние исследования по этой теме используют методы мета-эвристической оптимизации как стратегию поиска для решения самых сложных задач. В этой статье мы представляем алгоритм оптимизации «черепаха и заяц» (Rabbit and Tortoise optimization algorithm, RTOA) и адаптивный алгоритм оптимизации «черепаха и заяц» (адаптивный RTOA, A-RTOA) для отбора признаков. Алгоритм RTOA рассматривается как стратегия поиска в методе обертки. Адаптивный алгоритм A-RTOA – это гибридный метод, который объединяет RTOA на основе обертки и взаимную информацию на основе фильтра в рамках адаптивной стратегии.

Статья организована следующим образом: в разделе 2 содержится обзор современных методов обнаружения вторжений и методов выбора признаков. В разделе 3 представлены и математически определены алгоритмы RTOA и A-RTOA. В разделе 4 обсуждается набор данных для NIDS и реализация предлагаемых для выбора признаков алгоритмов RTOA и A-RTOA. В разделе 5 экспериментальные результаты сравниваются с аналогичными методами в форме таблиц/столбчатых диаграмм, показывается и обсуждается важность адаптивного метода. В Заключении представлены основные выводы, намечена будущая работа.

2. Обзор литературы

Тахир и др. предложили восстанавливать пропущенные данные на основе глубокого обучения для улучшения качества данных в системах обнаружения вторжений (intrusion detection system, IDS). Метод использует технику Random Missing Value для имитации пропущенных данных, что позволяет проводить тщательную оценку различных вычислительных подходов. Эксперименты проводились на наборах данных NSL-KDD и UNSW-NB15 [6]. Сахид и др. представили метод ансамблевого обучения на основе обнаружения атакующих вторжений в сетях Интернета вещей (IoT) с помощью оптимизатора Gray Wolf Optimizer на наборах данных BoT IoT и UNSW-NB15 [7]. Чжу и др. интегрировали метод ансамблевого обучения и кластеризации подпространства для обнаружения вторжений

в сетях интернета вещей [8]. Халладжи и др. предложили структуру для классификации по нескольким меткам в СОВ, которая преодолевает ограничения традиционных методов с одной меткой [9].

Ли и др. представили SELSTM (Selective Encoding LSTM) – усовершенствованную систему IDS для Интернета вещей, основанную на сети сжатия и возбуждения [10] и объединяющую концепции нейронной сети семантического встраивания (neural semantic embedding network, NSENet) и долгой краткосрочной памяти (long short-term memory, LSTM). Диша и др. представили систему обнаружения вторжений, которая использует модели машинного обучения для защиты сетей и данных. Взвешенный случайный лес (Random Forest, RF) с примесью Джини используется для обработки многомерных данных для выбора признаков в наборе данных UNSW-NB15 [11]. Кесервани и др. представили всесторонний обзор систем СОВ, выделив их типы: основанные на неправильном использовании, основанные на аномалиях и гибридные, и обсудив проблемы, возникающие из-за увеличения сложности сети и объема данных. Рассматриваются шесть эталонных наборов данных и различные методы снижения размерности и классификации (в частности, машинное и глубокое обучение), способствующие повышению эффективности и надежности СОВ [12]. Юсефнежад и др. представили систему СОВ, использующую ансамблевые модели. Система использует опорные векторные машины (support vector machines, SVM) для обнаружения нормального трафика и метод k-ближайших соседей (k-nearest neighbors, kNN) для многоклассовой классификации [13].

Кабилан и др. предложили неконтролируемый метод обнаружения вторжений для сетей связи в автомобиле, использующий автокодировщики для извлечения признаков и нечеткую кластеризацию С-средних для точного обнаружения [14]. Алаззам и др. предложили оптимизацию процедуры выбора признаков, “вдохновленную голубем” (pigeon-inspired optimization) [15]. Фероз Хан и Анандхарадж представили многоуровневый подход к анализу проблем безопасности на основе пороговых контрмер против атак повторного воспроизведения на каждом уровне [16]. Для идентификации вторжений Мегантара и Тохари интегрировали отбор признаков с использованием контролируемого обучения и сокращение данных посредством неконтролируемого обучения на наборе данных NSL-KDD [17].

Альмиани и др. показали, как использовать глубокие рекуррентные нейронные сети (deep recurrent neural networks) и контролируемые методы управления устройствами для классификации и прогнозирования внезапных скоординированных атак для масштабирования устойчивых систем СОВ для медицинского интернета вещей. Затем была выполнена оптимизация функций для обнаружения вторжений с использованием “биологически вдохновленных” алгоритмов роя частиц [18].

Бхаттачарья и др. опубликовали гибридный анализ главных компонентов (principal component analysis, PCA) и метод машинного обучения, используемый в качестве основы для классификации записей IDS. В предлагаемом ими алгоритме проводится быстрое кодирование массивов данных, находящихся в распоряжении систем СОВ, а уменьшение размерности выполняется на основе гибридной модели “мотылька” (firefly model) PCA [19]. Moualla et al. использовали критерий примесей Джини для отбора важных признаков [20]. Сахид и Ароволо исследовали проблемы повышения безопасности медицинского интернета вещей с помощью контролируемых моделей машинного обучения, таких как случайный лес и дерево решений, а также глубокие рекуррентные нейронные сети [21].

Девпрасад и др. предложили систему обнаружения вторжений на основе аномалий. Они провели эксперименты на наборах данных NSL-KDD и UNSW-NB15, в которых использовали методику определения порядка предпочтения по сходству с идеальным решением, механизм ранжирования, включающий хи-квадрат и алгоритм летучей мыши для выбора признаков [22]. Ширавани и др. представили новый подход к выбору признаков с использованием оценки на основе корреляции и нечетких чисел (fuzzy numbers) для систем обнаружения вторжений. Определяя признаки как нечеткие числа и оптимизируя с помощью

эвристической функции, он эффективно уменьшает размер данных. Экспериментальные результаты на наборах данных KDD Cup, NSL-KDD и CICIDS демонстрируют более высокие показатели обнаружения и эффективность по сравнению с традиционными методами, такими как выбор признаков на основе корреляции [23].

Касонголо и др. проанализировали набор данных обнаружения вторжений UNSW-NB15 для обучения систем СОВ. Применяется фильтрующее сокращение признаков с использованием библиотеки XGBoost (extreme gradient boosting) [24]. Kumar et al. представили новую систему СОВ, разработанную на основе неправильного использования (misuse) для классификации сетевых угроз. Они применили набор данных UNSW-NB15 для обучения и оценки модели с новым набором данных в реальном времени RTNITP18, который был создан в NIT Patna [25]. Мефтах С. И др.. использовали двухэтапный процесс обнаружения аномалий с набором данных UNSW-NB15 и продемонстрировали оценку эффективности модели как в бинарной классификации, так и в мультиклассификации [26]. Альмомани представил метод выбора признаков с использованием

1. генетических алгоритмов;
2. оптимизации, вдохновлённой мигающим поведением светлячков (firefly);
3. оптимизатора Grey Wolf и
4. оптимизации роя частиц (Particle swarm optimization) для выбора атрибутов в наборе данных UNSW-NB15 с использованием классификаторов J48 и опорных векторных машин [27].

Ахмад и др. исследовали проблемы и достижения в области сетевых СОВ. Они рассмотрели методы машинного и глубокого обучения для улучшения производительности СОВ [28].

Белух и др. представили двухступенчатый классификатор с использованием алгоритма RepTree (reduced error pruning tree) и подмножеств протоколов, оцененных с наборами данных UNSW-NB15 и NSL-KDD для обнаружения сетевых вторжений [29]. Несколько авторов внесли вклад в методы выбора признаков в справочных статьях [30-39].

Описанные работы подчеркивают важность методов выбора признаков для повышения производительности классификатора. Однако адаптивным метаэвристическим подходам, нацеленным на обнаружение вторжений, в существующих исследованиях уделено мало внимания. В этой статье предлагается новый метаэвристический алгоритм оптимизации RTOA, интегрирующийся с адаптивными стратегиями (A-RTOA) для устранения этих пробелов, с упором на улучшенный отбор признаков для NIDS. Как часть реализации предлагаемого алгоритма авторы использовали набор данных UNSW-NB15, доступность данных подтверждена в [40]. Алгоритмы RTOA и A-RTOA использовались в процессе отбора признаков для обнаружения вторжений в сеть. Оптимальные признаки выбирались и анализировались с использованием пяти классификаторов: случайного леса [41], CatBoost [42], XGBoost [43], KNN [44] и дерева решений [45].

3. Предлагаемый метод

Этот раздел посвящен описанию алгоритмов RTOA и A-RTOA, предназначенных для отбора признаков. Демонстрируется важность адаптивной стратегии.

3.1 Алгоритм оптимизации «черепаха и заяц»

Как обсуждалось во введении, алгоритм оптимизации «черепаха и заяц» (RTOA) – это подражающий природе метаэвристический алгоритм оптимизации. Он вдохновлен известной историей о зайце и черепахе и имитирует характеристики этих животных: высокую скорость и непоследовательность зайца и медленное, но устойчивое движение черепахи. Эти характеристики используются для эффективного поиска оптимальных решений в задачах оптимизации (в основном в процессах выбора признаков).

Алгоритм RTOA моделирует поведение зайца и черепахи:

1. Быстрые движения зайца (высокая скорость).
2. Размеренное и неспешное движение черепахи (низкая скорость).

Объединяя эти две различные характеристики, RTOA эффективно перемещается по ландшафту оптимизации, балансируя между исследованием нового и эксплуатацией возможностей известного для достижения оптимальных решений. В литературе, помимо прочего, алгоритм используется для обнаружения цикла в связанном списке [46]. Однако предлагаемый в этой статье вариант RTOA специально разработан для обеспечения надежного и эффективного подхода к решению сложных задач оптимизации, особенно в контексте отбора признаков. В алгоритме шаги с 1 по 8, за исключением шага 5, представляют собой процесс RTOA.

3.2 Вклад и структура

RTOA имеет ряд существенных особенностей.

1. Это метаэвристический метод оптимизации, который снижает необходимость в обширной настройке параметров за счет использования только двух параметров, обозначаемых как α и β.
2. RTOA вдохновлен контрастным поведением зайца и черепахи. Он имитирует быстрый, но непоследовательный темп зайца и медленный, но уверенный ход черепахи.
3. Алгоритм использует динамику зайца и черепахи для управления процессом оптимизации, сочетая быстрое исследование с устойчивой эксплуатацией для поиска оптимальных решений.
4. RTOA используется для выбора соответствующих подмножеств признаков при обнаружении вторжений (UNSW-NB15), тем самым повышая производительность систем обнаружения за счет эффективного выбора признаков.
5. Предложенный метод RTOA был тщательно протестирован на известном наборе данных (UNSW-NB15) с использованием пяти популярных алгоритмов классификации (CatBoost, XGBoost, KNN, дерево решений и случайный лес), продемонстрировав свою эффективность и надежность в реальных сценариях.
6. Через контроль над количеством признаков предлагается адаптивная стратегия на основе взаимной информации с алгоритмом оптимизации «черепаха и заяц», которая обозначается как A-RTOA. Результаты A-RTOA сравниваются с другими методами, а также с обычным RTOA.

3.3 Адаптивная оптимизация зайца и черепахи на основе взаимной информации (A-RTOA)

Значение адаптивной стратегии заключается в преодолении ограничений неадаптивных подходов в выборе признаков. В неадаптивных стратегиях иногда возможно, что к концу максимальной итерации не будет выбрано ни одного признака (то есть двоичный вектор имеет значение (0, 0, 0, 0, …) в последней итерации). В этой ситуации он не только теряет время, но и не может дать эффективного решения. Адаптивная стратегия помогает обеспечить успех в выборе определенного числа оптимальных подмножеств признаков.

В области методов оптимизации адаптивная метаэвристика представляет собой метод, который использует самонастраивающиеся стратегии метаэвристической оптимизации для выбора подмножеств признаков. В подходе на основе обертки для выравнивания количества признаков в каждой итерации используются адаптивные стратегии, которые для конкретной задачи выбирают наиболее оптимальные признаки с помощью итерационного подхода.

Ключевым преимуществом адаптивной стратегии является ее способность обеспечивать контроль над количеством выбранных признаков при сохранении значения пригодности. В работе используется основанная на фильтре взаимная информация (ВИ) [41] для балансировки количества признаков между итерациями в RTOA. Этот метод особенно ценен для повышения эффективности моделей NIDS и снижения сложности набора данных за счет концентрации на наиболее информативных признаках. Такой подход является новым по сравнению с существующими исследованиями. В настоящем алгоритме A-RTOA представлен шагами от 1 до 8 (включая шаг 5).

3.4 Формулировка RTOA и A-RTOA

RTOA разработан в соответствии с вышеупомянутыми характеристиками зайца и черепахи. Предположим, что задача включает в себя D переменных, а популяция состоит из n_c выбранных решений-кандидатов. В нашей работе D представляет собой общее количество признаков в рассматриваемом наборе данных. Каждый кандидат представлен так:

$$v_k = (v_{k1,} v_{k2,} v_{k3,} \dots \dots v_{kD})$$

где k меняется от 1 до n_c.

3.5 Алгоритм RTOA и A-RTOA

Шаг 1: Инициализация

Число кандидатов в признаки (n_c) определено. Во время инициализации каждый кандидат случайным образом выбирает признаки следующим образом:

$$v_{kj}^0 = \text{Случайный}\{0,1\}$$

где:

- v_{kj}^0 – начальное значение (в момент времени 0) для j-того признака ($j = 1, 2, 3, \dots, D$) k-того кандидата ($k = 1, 2, 3, \dots, n_c$),
- Случайный {0, 1} означает случайное назначение 0 или 1.
- Если, $v_{kj}^0 = 1$, j-тый признак набора данных учитывается для процесса оценки, в противном случае – нет.

Шаг 2: Первоначальный расчет приспособленности и выбор глобально лучшего кандидата

Начальная пригодность каждого кандидата $v_k^0 = \left(v_{k1,}^0 v_{k2,}^0 v_{k3,}^0 \dots \dots v_{kD}^0\right)$ оценивается с помощью классификатора на основе выбранных им признаков v_{ki}^0. Кандидат с максимальной пригодностью определяет v_{best}^0 двоичный вектор, представляющий начальное наилучшее решение.

Шаг 3: Обновление позиций кандидатов

Как описано выше, алгоритм обновления позиций Зайца и Черепахи каждого кандидата v_k^{t+1} состоит из двух фаз: «Фаза Зайца» и «Фаза Черепахи» из предыдущей позиции v_k^t. Фазы определяются на основе случайности. Если $\text{Случайный}\{0,1\} < 0.5$, то будет выполнена Фаза Зайца, в противном случае будет выполнена Фаза Черепахи. Подробное описание выглядит следующим образом:

Шаг 3-1: Фаза Зайца. На этом этапе предпринимаются более масштабные шаги для исследования новых областей пространства решений (то есть параметра скорости движения $0.5 \leq \alpha \leq 1$).

$$\text{Шаг\_зайца} = \alpha * \left| v^t_{\text{лучший},j} - v^t_{k,j} \right|$$
$$v^{t+1}_{k,j} = v^t_{k,j} + \text{Шаг\_зайца}$$

Шаг 3-2: Фаза черепахи. На этом этапе делаются более мелкие и точные шаги, использующие известные хорошие решения (то есть параметр скорости движения $0 < \beta < 0.5$)

$$\text{Шаг\_черепахи} = \beta * \left| v^t_{\text{лучший},j} - v^t_{k,j} \right|$$
$$v^{t+1}_{k,j} = v^t_{k,j} + \text{Шаг\_черепахи}$$

где,

- $v^t_{k,j}$является текущим решением k-того отдельного признака в группе.
- $v^t_{\text{лучший},j}$ является наилучшим решением, которое в настоящее время было найдено.
- α— коэффициент, контролирующий размер шага исследования.
- β— коэффициент, контролирующий размер шага эксплуатации.
- Термин $\left| v^t_{\text{лучший},j} - v^t_{k,j} \right|$ представляет собой движение к лучшему кандидату.

Шаг 4: Обновление каждого кандидата для следующей итерации

Векторы решения v^{t+1}_k шага 3 могут не быть бинарными векторами. Следующее правило обновляет j-тую позицию k-того кандидата в итерации $t+1$, превращая v^{t+1}_k в бинарный вектор.

$$v^{t+1}_{k,j} = \begin{cases} 1, & P^t_{k,j} > \text{Случайный}(0,1) \\ 0, & Otherwise \end{cases}$$

Пространство поиска $P^t_{k,j}$ определяется как,

$$P^t_{k,j} = \frac{\left(e^{v^t_{k,j}} - e^{-v^t_{k,j}} \right)}{\left(e^{v^t_{k,j}} + e^{-v^t_{k,j}} \right)}$$

где,

- $v^{t+1}_{k,j}$, $v^t_{k,j}$ – обозначают текущее и предыдущее значение j-той позиции в k-том кандидате.
- $\text{Случайный}(0,1)$ обозначает случайное число, которое имеет равномерное распределение в диапазоне от 0 до 1.

Шаг 5: Выбор признаков с использованием концепции адаптивной стратегии на основе взаимной информации

В алгоритме RTOA шаг 5 пропускается, и процесс продолжается переходом к шагу 6. В алгоритме A-RTOA перед переходом к шагу 6 выполняется шаг 5.

Выбор кандидата был описан в предыдущих шагах 1-4. Количество выбранных признаков теперь определяется с использованием адаптивной стратегии на основе взаимной информации. Определение количества признаков, которые будут выбраны как 's' для каждого кандидата $k = 1, 2, \ldots n_c$.

Для каждого кандидата $\boldsymbol{k = 1, 2, \ldots n_c}$ выполняется расчет количества выбранных признаков:$x_k = |v^{t+1}_k|$

Если $x_k = s$, то:

никаких изменений v^{t+1}_k.

Если $x_k < s$, то:

a) Найти MI для невыбранных объектов (объектов j, которые имеют $v_{k,j}^{t+1} = 0$).
b) Добавить «$s - x_k$» признаки, имеющие наивысшее значение MI v_k^{t+1} (изменяя соответствующие $v_{kj} = 1$).

Если $x_k > s$, то:

a) Найти MI для выбранных признаков (признаков j, которые имеют $v_{k,j}^{t+1} = 1$).
b) Удалить «$x_k - s$ » признаки, имеющие самые низкие значения MI v_k^{t+1} (изменив соответствующие $v_k^{t+1} = 0$).

Шаг 6: Обновление глобального лучшего кандидата$(v_{\text{лучший}})$**.** На этом этапе можно обновить лучшего кандидата $v_{\text{лучший}}^t$.

Найти пригодность для v_k^{t+1}: пригодность $(v_k^{t+1}) = \text{Классификатор}(v_k^{t+1}), \quad \forall\, k = 1,2,3 \ldots n_c$

Определить кандидата $v_{k(\text{макс})}^{t+1}$, имеющего наибольшую точность среди $\{v_k^{t+1}: k = 1,2,3, \ldots n_c\}$, то есть пригодность $\left(v_{k(\text{макс})}^{t+1}\right) = \underset{1 \le k \le n_c}{\text{макс}}$ пригодность (v_k^{t+1})

Правило обновления:

$$v_{\text{лучший}}^{t+1} = \begin{cases} v_{k(\text{макс})}^{t+1}, & \text{Если пригодность}\left(v_{k(\text{макс})}^{t+1}\right) > \text{пригодность}\left(v_{\text{лучший}}^{t}\right) \\ v_{\text{лучший}}^{t}, & \text{В противном случае} \end{cases}$$

Шаг 7: Повторение шагов 3–6, пока не будет выполнен критерий остановки.

Шаг 8: Возвращение выбранных характеристик $v_{\text{лучший}}$ набора данных для подготовки модели NIDS.

4. Реализация RTOA и A-RTOA

Для реализации RTOA и A-RTOA используется набор данных UNSW-NB15, который широко применяется в исследованиях, связанных с обнаружением сетевых вторжений.

4.1 Описание набора данных UNSW-NB15

Набор данных UNSW-NB-15 [40] содержит в общей сложности сорок девять признаков. Каждая запись классифицируется как невредоносная и безопасная, либо как принадлежащая конкретному типу атак. Мустафа и др. [1] предоставили полное описание всех 49 признаков, включенных в набор. Признаки подразделяются на шесть основных групп: поток, базовые, содержимое, время, дополнительные признаки и помеченные признаки.

Такие столбцы, как порт назначения, порт источника, IP-адрес назначения и IP-адрес источника, были исключены из стандартных наборов данных для обучения и тестирования. Рассматриваемый набор данных UNSW-NB15 состоял из 45 столбцов (с 43 признаками, категорией атак и меткой класса), как указано в табл. 1. Кроме того, атаки в наборе данных классифицируются по 9 различным семействам атак: Exploit, Fuzzers, Reconnaissance, Backdoor, Analysis, Shellcode, Worm, DoS (отказ в обслуживании), Generic [1].

Табл. 1. Характеристики набора данных UNSW-NB15.
Table 1. Features of UNSW-NB15 dataset.

Этикетка	Название функции	Этикетка	Название функции	Этикетка	Название функции	Этикетка	Название функции
f_1	id	f_{13}	sload	f_{25}	tcprtt	f_{37}	ct_dst_src_ltm
f_2	dur	f_{14}	dload	f_{26}	synack	f_{38}	is_ftp_login

f_3	proto	f_{15}	sloss	f_{27}	ackdat	f_{39}	ct_ftp_cmd
f_4	service	f_{16}	dloss	f_{28}	smean	f_{40}	ct_flw_http_mthd
f_5	state	f_{17}	sinpkt	f_{29}	dmean	f_{41}	ct_src_ltm
f_6	spkts	f_{18}	dinpkt	f_{30}	trans_depth	f_{42}	ct_srv_dst
f_7	dpkts	f_{19}	sjit	f_{31}	response_body_len	f_{43}	is_sm_ips_ports
f_8	sbytes	f_{20}	djit	f_{32}	ct_srv_src	f_{44}	attack_cat
f_9	dbytes	f_{21}	swin	f_{33}	ct_state_ttl	f_{45}	Label
f_{10}	rate	f_{22}	stcpb	f_{34}	ct_dst_ltm	---	---
f_{11}	sttl	f_{23}	dtcpb	f_{35}	ct_src_dport_ltm	---	---
f_{12}	dttl	f_{24}	dwin	f_{36}	ct_dst_sport_ltm	---	---

После сбора данных они проходят фазу предварительной обработки, где из строятся репрезентативные выборки или образцы (samples) для обеспечения пригодности для последующего анализа. В любой модели машинного обучения предварительная обработка имеет решающее значение для удаления из необработанных данных ошибок и преобразования сетевого трафика в пригодный для использования готовый для обучения модели формат. Исходный набор данных UNSW-NB15 содержит в общей сложности 175341 обучающий образец. Этот набор, который включает девять различных семейств атак, демонстрирует значительный дисбаланс классов, как между атакующими посылками и не атакующими, так и внутри самих выделенных девяти категорий атак. Распределение обучающих и тестовых образцов показывает непропорциональное представительство по классам.

Для устранения этого дисбаланса применяется гибридная техника повторной выборки [48], балансирующая распределение классов перед передачей данных на фазы обучения и тестирования. Для бинарной классификации (атака/не атака) набор данных балансируется путем выбора 131742 образцов нормального поведения и 131742 образцов атаки, что дает в общей сложности 263484 образца, из которых затем 70% выделяются для обучения, а остальные 30% для проверки его проведения. Для многоклассовой классификации используется в общей сложности 146380 образцов, с равным количеством образцов (14638), выбранных для каждого типа атаки, чтобы обеспечить единообразное представление по классам, снова разделенных на 70% для обучения и 30% для проверки.

4.2 Внедрение механизма выбора признаков в NIDS

Следующим важным шагом является выбор признаков с помощью предлагаемых алгоритмов. Архитектура для интеграции предлагаемого метода выбора признаков в NIDS представлена на рис. 1.

Полученные выбранные признаки, извлеченные с помощью методов RTOA и A-RTOA, впоследствии оцениваются с помощью нескольких классификаторов, включая CatBoost, XGBoost, KNN, дерево решений и случайный лес, с использованием различных показателей производительности, таких как правильность (accuracy), точность (precision), полнота (recall) и F1-мера.

Метрики производительности предлагаемых методов сравниваются по разным классификаторам как для бинарных, так и для многомаркированных классов. Ниже сравнивается производительность предлагаемого подхода и других методов. Система NIDS готовится к использованию с учётом оптимальных функций, определенных с помощью лучшего предлагаемого метода A-RTOA. Наконец, чтобы различать нормальное поведение и определить специфическую природу атаки, система NIDS определяет оптимальные характеристики потоков пакетов и использует модель, созданную с учетом этих характеристик.

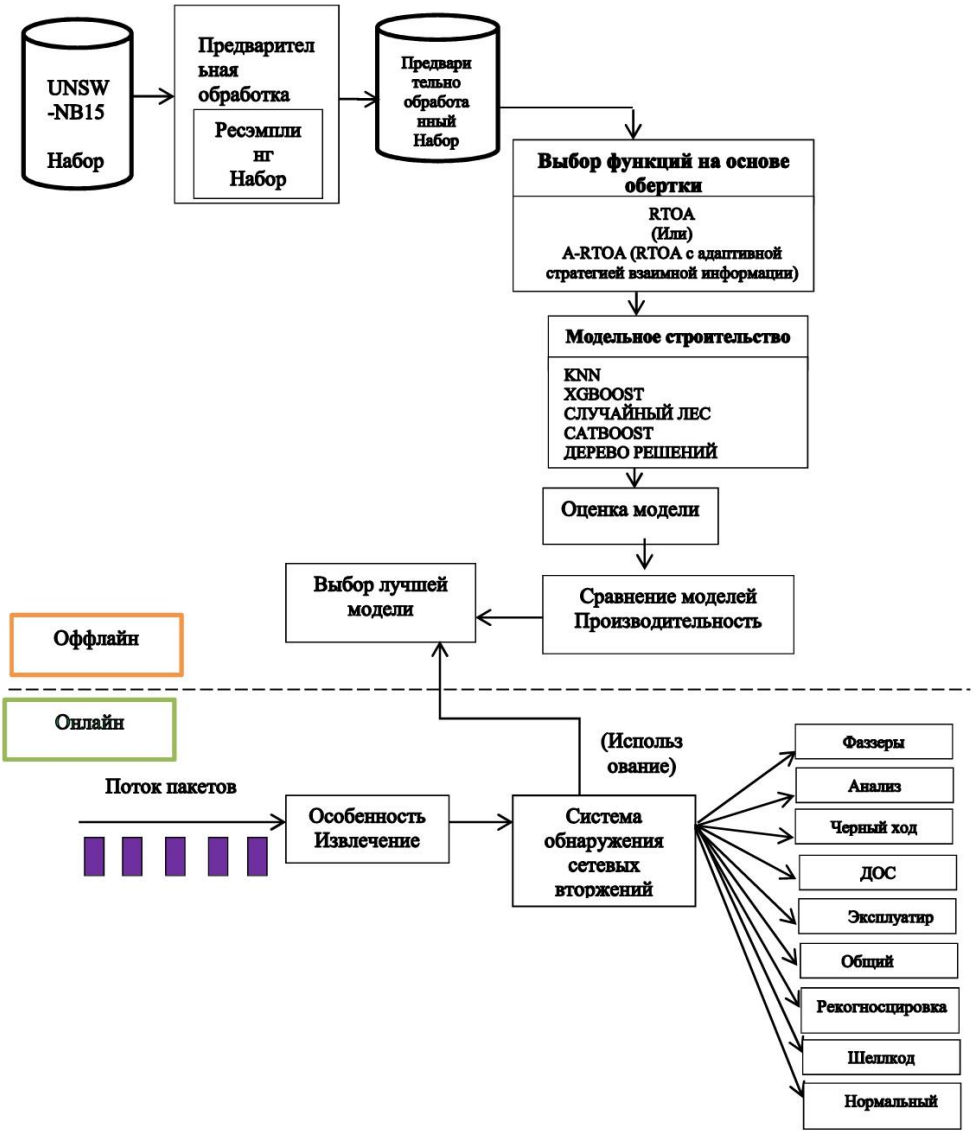


Рис. 1. Интеграция предлагаемого метода выбора функций в систему обнаружения сетевых вторжений (NIDS).
Fig. 1. Integration of the Proposed Feature Selection Method in the Network Intrusion Detection System (NIDS).

5. Результаты и обсуждение

Как обсуждалось ранее, набор данных UNSW-NB15 подвергается предварительной обработке, которая на основе предложенных алгоритмов позволяет выделить оптимальные признаки. Все эксперименты проводились на аппаратуре с 16,0 ГБ ОЗУ и процессором Intel® Core™ i7-9700F с тактовой частотой 3,00 ГГц.

Для обоих наших экспериментов в качестве классификатора используется случайный лес. Правильность модели определяется для функции пригодности, при этом значения параметров устанавливаются как $n_c = 10$, $\alpha = 0.7$, $\beta = 0.3$, а максимальное количество

итераций устанавливается равным 100. Первоначально ставится эксперимент, в котором представленный алгоритм RTOA выполняет шаги от первого до восьмого, пропуская шаг 5. Завершая шаг 8, RTOA генерирует 26 оптимальных признаков, показанных в табл. 2. Пояснения к индексам f_i (i = 1, 2, …, D) даны в табл. 1.

В эксперименте с алгоритмом A-RTOA количество выбираемых признаков, обозначенное как s, устанавливается разработчиком сети. Правильность модели с использованием случайного леса графически показана на рис. 2 для различных значений s от 1 до 43. Рис. 2 показывает эффективность A-RTOA, поскольку правильность для $s = 9$ почти такая же, как для случаев $s > 9$. Это означает, что при выборе только 9 признаков с использованием алгоритма A-RTOA вместо использования большего количества или даже всех признаков производительность сохраняется без значительной потери правильности, что позволяет быстрее проводить правильную детекцию. В конце шага 8 алгоритм A-RTOA выдает 9 оптимальных признаков, показанных в табл. 2.

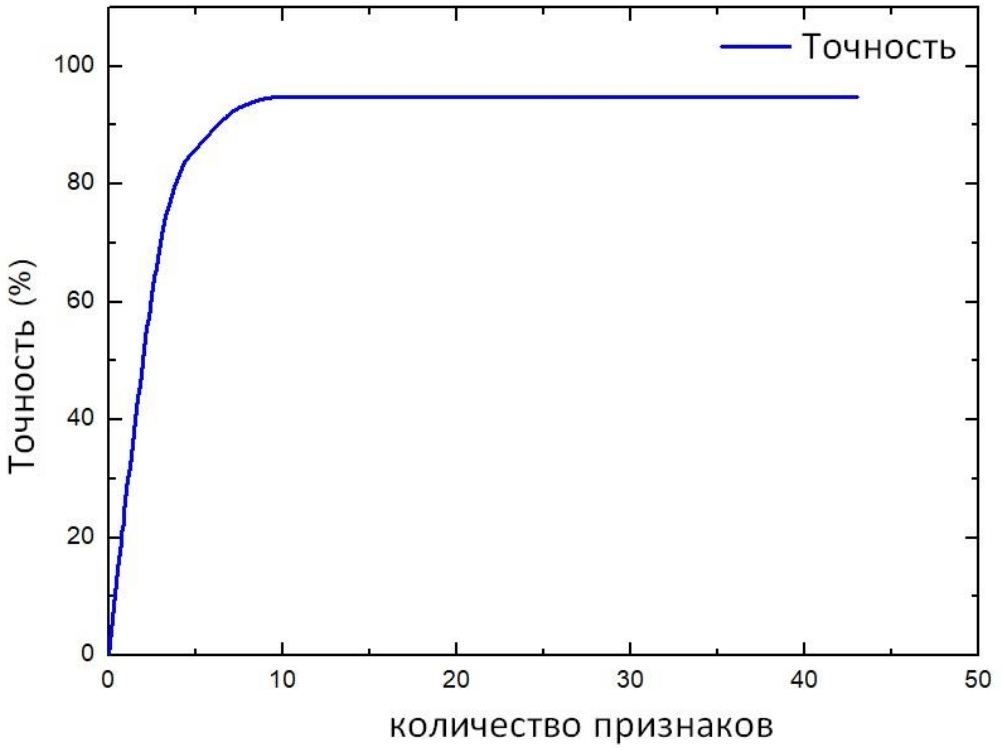
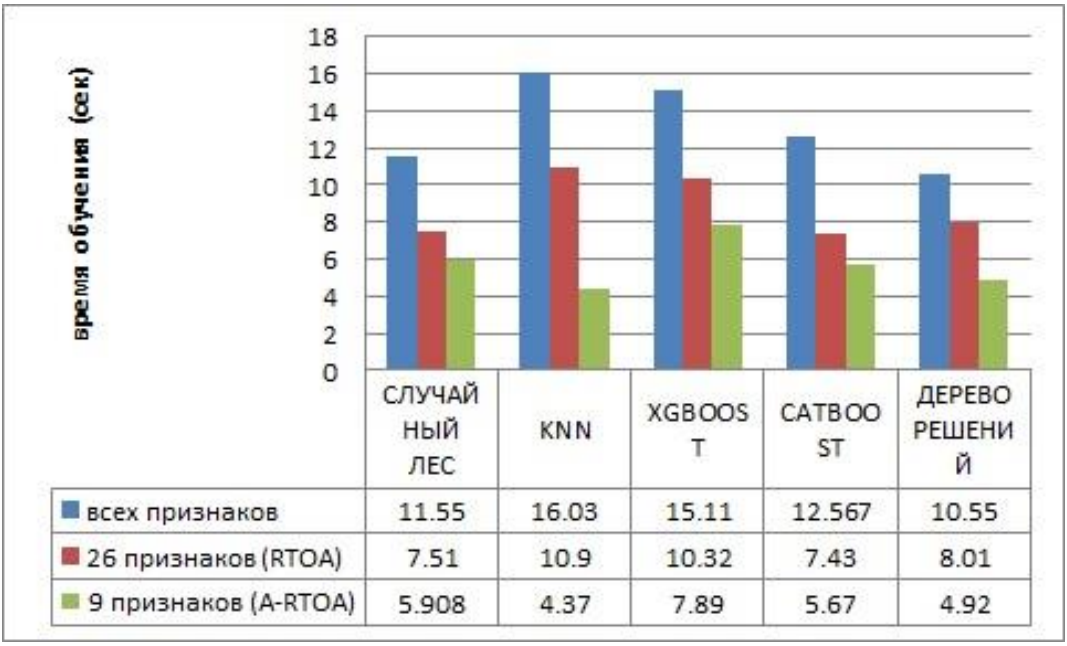


Рис. 2. Сравнение влияния количества признаков на правильность классификации.
Fig. 2. Comparison effect of feature selection on classification accuracy.

Табл. 2. Признаки, выбранные алгоритмами.
Table 2. Selected features by proposed feature selection algorithms.

Предлагаемые методы выбора признаков	Количество признаков	Признаки
RTOA	26	f_1, f_3, f_5, f_6, f_8, f_9, f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}, f_{17}, f_{18}, f_{19}, f_{20}, f_{21}, f_{22}, f_{23}, f_{24}, f_{25}, f_{26}, f_{28}, f_{29}, f_{30}, f_{31}
A-RTOA	9	f_1, f_5, f_{10}, f_{13}, f_{25}, f_{29}, f_{34}, f_{41}, f_{43}

На рис. 3 показано сравнение времени обучения при использовании всех признаков, подмножества из 26-ти признаков и 9-ти оптимальных признаков. Результаты показывают значительное сокращение времени обучения при меньшем количестве признаков, что подчеркивает эффективность выбора признаков.



	СЛУЧАЙНЫЙ ЛЕС	KNN	XGBOOST	CATBOOST	ДЕРЕВО РЕШЕНИЙ
всех признаков	11.55	16.03	15.11	12.567	10.55
26 признаков (RTOA)	7.51	10.9	10.32	7.43	8.01
9 признаков (A-RTOA)	5.908	4.37	7.89	5.67	4.92

Рис. 3: Сравнение времени обучения для всех признаков, для отобранных 26 признаков и 9 оптимальных признаков.
Fig. 3. Comparisons of training time for all features, 26 and optimal 9 features.

Для сравнения производительности, чтобы вычислить оценочные метрики, мы использовали классификаторы CATBoost, XGBoost, KNN, дерево решений и случайный лес. Метрики производительности бинарной классификации модели NIDS, построенные по выбранным признакам RTOA (26 признаков), A-RTOA (9 признаков) и всем признакам, показаны в табл. 3, табл. 4 и табл. 5 соответственно. Результаты демонстрируют более высокие показатели производительность алгоритма A-RTOA, особенно при сравнении с классификатором случайного леса.

Как указано в табл. 4 и 5, девять выбранных признаков имеют производительность, эквивалентную применению всех признаков в бинарной классификации, что достигается с помощью случайного леса с точностью 94,69% и 94,73%. Аналогично, приведенные выше результаты наглядно иллюстрируют эффективность выбора признаков в сохранении точности при уменьшении размерности.

Чтобы продемонстрировать эффективность предлагаемых алгоритмов RTOA и A-RTOA, мы сравниваем их с ныне используемыми подходами. Результаты эксперимента по бинарной классификации сравниваются с существующими методами классификации и кластеризации из литературы (такими как наивный байесовский, NSENet, AdaBoost, NB-SVM и т. д.) с использованием всего набора признаков, как показано в табл. 6. Она демонстрирует эффективность предлагаемого нами метода по сравнению с современными подходами, которые используют все признаки. Правильность, достигнутая с использованием 9 признаков, выше, чем с 26 признаками, и оба выше точности, полученной с использованием полного набора признаков. Это ясно подчеркивает эффективность нашей стратегии выбора

признаков. Кроме того, производительность алгоритмов RTOA и A-RTOA сравнивается с существующими подходами к оберткам и фильтрам (такими как алгоритмы PSO, GWO, FFA, GA и другими), предложенными разными авторами, как показано в табл. 7.

Табл. 3. Оценка производительности бинарной классификации с использованием оптимальных признаков (числом 26), выбранных алгоритмом RTOA.
Table 3. Performance evaluation for binary classification using optimal features (26) selected by RTOA.

Метрики	CATBOOST	XGBOOST	КНН	ДЕРЕВО РЕШЕНИЙ	СЛУЧАЙНЫЙ ЛЕС
Правильность (%)	87.26	88.22	87.31	93,55	**94.12**
Точность (%)	87,71	88.93	86.75	93.28	**93,87**
Отзывать (%)	87.26	88.22	87.31	93,55	**94.12**
Оценка F1 (%)	87.17	88.13	86.48	93.35	**93.92**

Табл. 4: Оценка производительности бинарной классификации с использованием оптимальных признаков (числом 9), выбранных адаптивным алгоритмом A-RTOA.
Table 4. Performance evaluation for binary classification using optimal features (9) selected by A-RTOA.

Метрики	CATBOOST	XGBOOST	КНН	ДЕРЕВО РЕШЕНИЙ	СЛУЧАЙНЫЙ ЛЕС
Правильность (%)	88.22	89.12	90,75	94.18	**94.69**
Точность (%)	88.73	89.97	90,69	93.97	**94.51**
Отзывать (%)	88.22	89.12	90,75	94.18	**94.69**
Оценка F1 (%)	88.18	89.05	90.26	94.01	**94,54**

Табл. 5: Оценка эффективности бинарной классификации с использованием всех признаков.
Table 5. Performance evaluation for binary classification using all features.

Метрики	CATBOOST	XGBOOST	КНН	ДЕРЕВО РЕШЕНИЙ	СЛУЧАЙНЫЙ ЛЕС
Правильность (%)	88.68	89.69	90.32	94.24	**94.73**
Точность (%)	89.16	90.44	90.21	94.02	**94,54**
Отзывать (%)	88.68	89.69	90.32	94.24	**94.73**
Оценка F1 (%)	88.69	89,68	89,75	94.07	**94.58**

Табл. 7 показывает эффективность предлагаемых нами методов RTOA и A-RTOA. Видно, что адаптивный алгоритм A-RTOA превосходит как алгоритм RTOA, так и другие современные методы с точки зрения как правильности, так и по количеству признаков.

Поскольку UNSW-NB15 представляет собой набор данных с многоклассовой разметкой, для многоклассовой классификации используется предварительно обработанный и нормализованный набор данных. Метрика производительности (правильность) предлагаемых нами RTOA (26 признаков) и A-RTOA (9 признаков) для многоклассовой классификации с использованием классификатора случайного леса сравнивается с методами из литературы [9, 26, 37 и 39], как показано в табл. 8. Видно, как предлагаемые методы повышают правильность и эффективность NIDS в сравнении с существующими подходами.

На основании результатов и обсуждений можно сделать вывод, что предлагаемый метод RTOA работает лучше с првильностью 94,12% по сравнению с существующими алгоритмами (68,18% против 93,01% с более чем 9 признаками) из литературы [11, 15, 17, 22, 24, 27, 29, 34-38]. Более того, адаптивная стратегия A-RTOA достигает наилучшей производительности с точностью 94,69%, превосходя RTOA (94,12%) и другие перечисленные современные методы для бинарной классификации. Этот вывод также распространяется на многоклассовую классификацию. Адаптивная стратегия A-RTOA достигает наилучшей производительности со средней правильностью 94,03%, превосходя RTOA со средней правильностью 93,92% для многоклассовой классификации.

6. Заключение

В этой статье предложена новый метаэвристическая оптимизация алгоритма «черепаха и заяц» (RTOA), а также адаптивная стратегия на основе взаимной информации с этим алгоритмом оптимизации, называемая A-RTOA. Эти методы выполняют отбор признаков при обнаружении сетевых вторжений. Методы реализованы с использованием набора данных UNSW-NB15. NIDS построены на двух оптимальных наборах признаков: 26 признаков, выбранных с помощью RTOA, и 9 признаков, выбранных с помощью A-RTOA. Производительность NIDS определяется и сравнивается с производительностью других алгоритмов NIDS. Сравнение бинарной классификации, многоклассовой классификации и важность сокращенных оптимальных признаков проведено в разделе результатов и показывает, что RTOA превосходит существующие методы отбора признаков, в то время как A-RTOA работает еще лучше, чем RTOA. Система NIDS, созданная с использованием A-RTOA, надежна и имеет большое значение для достижения надежного и эффективного обнаружения сетевых вторжений. Предложенные методы усиливают сетевую защиту и помогают справиться с проблемами, связанными с растущими киберугрозами. В будущем можно исследовать дальнейшие усовершенствования и методы для развития полученных результатов и улучшения возможностей многоклассовой классификации обнаружения сетевых вторжений.

Табл. 6. Сравнение бинарных классификаторов с несколькими методами классификации/кластеризации, а также RTOA и A-RTOA.
Table 6. Comparison of binary classifiers with a few Classification/Clustering methods and RTOA/A-RTOA.

	Методы классификации/кластеризации	Количество функций	Точность (%)
В литературе	Слияние NSENet и LSTM на основе SENet [12]	Все функции	82.14
	Дерево решений [14]		90.15
	АдаБуст [14]		90,51
	Дерево усиления градиента [14]		87,56
	Многослойный персептрон [14]		84.11
	AdaBoost, долго-кратковременная память [14]		87.90
	Закрытый рекуррентный блок [14]		82.87
	Модели ансамбля [17]		90.98
	Нечеткие C-средние [18]		73,62
	Дерево решений [45]		85,56
	Логистическая регрессия [45]		83.15
	Наивный Байес [45]		82.07
	Искусственная нейронная сеть [45]		81.34
	Кластеризация ожиданий-максимизации [45]		78.47
	Модель ансамбля Голосование [47]		89.29
	Дерево регрессии с градиентным усилением [48]		91.31
	NB-SVM [49]		93,75
Предлагаемый метод RTOA	CATBOOST	26	87.26
	XGBOOST		88.22
	KNN		87.31
	ДЕРЕВО РЕШЕНИЙ		93,55
	СЛУЧАЙНЫЙ ЛЕС		94.12
Предлагаемый метод A-RTOA	CATBOOST	9	88.22
	XGBOOST		89.12
	KNN		90,75
	ДЕРЕВО РЕШЕНИЙ		94.18
	СЛУЧАЙНЫЙ ЛЕС		94.69

Табл. 7. Сравнение различных существующих методов и предлагаемых методов RTOA и A-RTOA.
Table 7. Comparison between different existing methods and our proposed methods RTOA and A-RTOA.

№	Литература	Метод выбора характеристик	Число признаков	Используемые классификаторы	Точность (%)
1	Диша и др. [14]	Подход фильтра – взвешенный случайный лес на основе примесей Джини (ГИВРФ)	20	Дерево решений	93.01
				АдаБуст	90,51
				GBT	87.08
				MLP	87.26
				LSTM	88.99
				GRU	90.11
2	Алаззам и др. [19]	Подход обертки – оптимизатор, вдохновленный Pigeon (PIO): Sigmoid_PIO	14	Дерево решений (DT)	91.30
3	Мегантара и Ахмад [21]	Подход обертки: гибридный метод машинного обучения	11	Дерево решений	91.86
4	Девпрасад и др. [28]	Подход «Фильтр и обертка»: алгоритмы χ^2 и «летучая мышь»	9	Классификаторы DT и опорных векторных машин	89.43
5	Касонго и др. [30]	Нормализация функций с помощью XGBoost	19	Дерево решений	90,85
6	Альмомани и др. [35]	Подход обертки: PSO, GWO, FFA и GA, биоинспирированные алгоритмы и MI	13	Дерево решений	89,58
			30		90.48
7	Белух и др. [37]	Выбор признаков на основе фильтра (IG) с двухэтапным подходом к классификации	20	Уменьшение ошибок обрезки дерева (RepTree)	88.95
8	Мустафа и др. [50]	Фильтрационный подход (PCA) с геометрическим анализом площади (GAA)	15	Дерево решений	92.80
9	Мустафа и др. [51]	Гибридный выбор признаков с использованием ассоциативных правил	11	Кластеризация ожиданий-максимизации	77.2
				Логистическая регрессия	83.0
				Наивный Байес	79,5
10	Хаммасси и др. [52]	Подход обертка: Генетический алгоритм и логистическая регрессия	20	Классификатор C4.5	81.42
11	Сети и др. [53]	Подход, основанный на обучении с подкреплением	19	Случайный лес (RF)	82.12
				АдаБуст (ADB)	85,61
				Гауссовский наивный байесовский алгоритм (GNB),	70,62
				k-ближайших соседей (KNN)	78.47
				Квадратичный дискриминантный анализ (QDA)	68.18
				RF+ADB+QDA	85.09
12	Кумар и др. [54]	Метод, основанный на правилах	13	Дерево решений	84.83
				Дерево решений C5	90,74
13	**Предлагаемый подход к обертке: RTOA**		26	CATBOOST	87.26
				XGBOOST	88.22
				KNN	87.31
				ДЕРЕВО РЕШЕНИЙ	93,55
				СЛУЧАЙНЫЙ ЛЕС	**94.12**
14	**Предлагаемый подход к обертке: RTOA с адаптивной стратегией (A-RTOA)**		9	CATBOOST	88.22
				XGBOOST	89.12
				KNN	90,75
				ДЕРЕВО РЕШЕНИЙ	94.18
				СЛУЧАЙНЫЙ ЛЕС	**94.69**

Табл. 8: Сравнение показателей производительности (точность, %) для многоклассовой классификации: предлагаемые методы и данные из литературы.
Table 8. Comparison of Performance Metrics (Accuracy %) for Multi-Label Classification: Proposed Methods vs. Literature.

№	Характер атак	Халладжи и др., [9] Многокомпонентная нейронная сеть на основе ансамбля	Мефтах и др. [26] C5.0 Дерево решений	Сети и др., [37] Случайный лес	Мустафа и др., [39] Наивный Байес	Предлагаемые алгоритмы	
						RTOA (26 функций) Случайный лес	A-RTOA (9 функций) Случайный лес
1	Анализ	97.47	21.35	84.67	0	97.76	97.92
2	Черный ход	35.56	17.75	83.53	20	97.63	97.62
3	ДОС	94.40	10.38	92.12	71.1	91.74	92.33
4	Эксплуатация	70.17	96.68	79.21	54.6	86.48	89.34
5	Фаззеры	86.08	79.10	93.43	33.2	87.66	90.35
6	Общий	77.27	98.72	96.37	0	93.84	95.21
7	Нормальный	32.35	93.73	-	94.3	96.99	97.67
8	Рекогносцировка	36.88	76.29	89.45	69.9	95.78	96.46
9	Шеллкод	81.75	84.02	92.79	0	96.80	91.67
10	Черви	86.96	63.63	65.31	0	94.53	91.68

Список литературы / References

[1]. Moustafa, Nour, and Jill Slay. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In 2015 military communications and information systems conference (MilCIS), pp. 1-6. IEEE, 2015.
[2]. Choudhary, Sarika, and Nishtha Kesswani. Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 datasets using deep learning in IoT. Procedia Computer Science 167 (2020): 1561-1573.
[3]. Selvakumar, B., and Karuppiah Muneeswaran. Firefly algorithm based feature selection for network intrusion detection. Computers & Security 81 (2019): 148-155.
[4]. Bhuvaneswari, T., M. Chengathir Selvi, R. Naga Priyadarsini, U. Eswaran, and RK Ramesh Babu. Feature selection with mutual information based cuckoo search optimization for Parkinson's disease prediction. NeuroQuantology 20, no. 10 (2022): 1296.
[5]. Rana, Pratip, Phuc Thai, Thang Dinh, and Preetam Ghosh. Relevant and non-redundant feature selection for cancer classification and subtype detection. Cancers 13, no. 17 (2021): 4297.
[6]. Tahir, Mahjabeen, et al. A novel approach for handling missing data to enhance network intrusion detection system. Cyber Security and Applications 3 (2025): 100063.
[7]. Saheed, Yakub Kayode, and Sanjay Misra. A voting gray wolf optimizer-based ensemble learning models for intrusion detection in the Internet of Things. International Journal of Information Security 23.3 (2024): 1557-1581.
[8]. Zhu, Jingyi, and Xiufeng Liu. An integrated intrusion detection framework based on subspace clustering and ensemble learning. Computers and Electrical Engineering 115 (2024): 109113.
[9]. Hallaji, Ehsan, Roozbeh Razavi-Far, and Mehrdad Saif. Expanding analytical capabilities in intrusion detection through ensemble-based multi-label classification. Computers & Security 139 (2024): 103730.
[10]. Li, Shaoqin, Zhendong Wang, Shuxin Yang, Xiao Luo, Daojing He, and Sammy Chan. Internet of Things intrusion detection: Research and practice of NSENet and LSTM fusion models. Egyptian Informatics Journal 26 (2024): 100476.
[11]. Disha, Raisa Abedin, and Sajjad Waheed. Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity 5.1 (2022): 1.
[12]. Keserwani, Pankaj Kumar, Mahesh Chandra Govil, and Emmanuel S. Pilli. An effective NIDS framework based on a comprehensive survey of feature optimization and classification techniques. Neural Computing and Applications 35.7 (2023): 4993-5013.

[13]. Yousefnezhad, Maryam, Javad Hamidzadeh, and Mohammad Aliannejadi. Ensemble classification for intrusion detection via feature extraction based on deep Learning. Soft Computing 25.20 (2021): 12667- 12683.
[14]. Kabilan, N., Vinayakumar Ravi, and V. Sowmya. Unsupervised intrusion detection system for in-vehicle communication networks. Journal of Safety Science and Resilience 5.2 (2024): 119-129.
[15]. Alazzam, Hadeel, Ahmad Sharieh, and Khair Eddin Sabri. A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. Expert systems with applications 148 (2020): 113249.
[16]. Feroz Khan, A. B., and Anandharaj, G. A Multi-layer Security approach for DDoS detection in Internet of Things. International Journal of Intelligent Unmanned Systems 9, no. 3 (2021): 178-191.
[17]. Megantara, Achmad Akbar, and Tohari Ahmad. A hybrid machine learning method for increasing the performance of network intrusion detection systems. Journal of Big Data 8, no. 1 (2021): 142.
[18]. Almiani, Muder, Alia AbuGhazleh, Amer Al-Rahayfeh, Saleh Atiewi, and Abdul Razaque. Deep recurrent neural network for IoT intrusion detection system. Simulation Modelling Practice and Theory 101 (2020): 102031.
[19]. Bhattacharya, Sweta, Praveen Kumar Reddy Maddikunta, Rajesh Kaluri, Saurabh Singh, Thippa Reddy Gadekallu, Mamoun Alazab, and Usman Tariq. A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. Electronics 9, no. 2 (2020): 219.
[20]. Moualla, Soulaiman, Khaldoun Khorzom, and Assef Jafar. Improving the Performance of Machine Learning-Based Network Intrusion Detection Systems on the UNSW-NB15 Dataset. Computational Intelligence and Neuroscience 2021, no. 1 (2021): 5557577.
[21]. Saheed, Yakub Kayode., and Micheal Olaolu Arowolo. Efficient cyber attack detection on the internet of medical things-smart environment based on deep recurrent neural network and machine learning algorithms. IEEE Access 9 (2021): 161546-161554.
[22]. Devprasad, Kayathri Devi, Sukumar Ramanujam, and Suresh Babu Rajendran. Context adaptive ensemble classification mechanism with multi-criteria decision making for network intrusion detection. Concurrency and Computation: Practice and Experience 34, no. 21 (2022): e7110.
[23]. Shiravani, Anita, Mohammad Hadi Sadreddini, and Hassan Nosrati Nahook. Network intrusion detection using data dimensions reduction techniques. Journal of Big Data 10, no. 1 (2023): 27.
[24]. Kasongo, Sydney M., and Yanxia Sun. Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. Journal of Big Data 7, no. 1 (2020): 105.
[25]. Kumar, Vikash, Ditipriya Sinha, Ayan Kumar Das, Subhash Chandra Pandey, and Radha Tamal Goswami. An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset. Cluster Computing 23 (2020): 1397-1418.
[26]. Meftah, Souhail, Tajjeeddine Rachidi, and Nasser Assem. Network based intrusion detection using the UNSW-NB15 dataset. International Journal of Computing and Digital Systems 8, no. 5 (2019): 478-487.
[27]. Almomani, Omar. A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms. Symmetry 12, no. 6 (2020): 1046.
[28]. Ahmad, Zeeshan, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, and Farhan Ahmad. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies 32, no. 1 (2021): e4150.
[29]. Belouch, Mustapha, Salah El Hadaj, and Mohamed Idhammad. A two-stage classifier approach using reptree algorithm for network intrusion detection. International Journal of Advanced Computer Science and Applications 8, no. 6 (2017).
[30]. Moustafa, Nour, and Jill Slay. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Information Security Journal: A Global Perspective 25, no. 1-3 (2016): 18-31.
[31]. Swami, Rochak, Mayank Dave, and Virender Ranga. Voting-based intrusion detection framework for securing software-defined networks. Concurrency and computation: practice and experience 32.24 (2020): e5927.
[32]. Tama, Bayu Adhi, and Kyung-Hyune Rhee. An in-depth experimental study of anomaly detection using gradient boosted machine. Neural Computing and Applications 31 (2019): 955-965.
[33]. Gu, Jie, and Shan Lu. An effective intrusion detection approach using SVM with naïve Bayes feature embedding. Computers & Security 103 (2021): 102158.
[34]. Moustafa, Nour, Jill Slay, and Gideon Creech. Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks. IEEE Transactions on Big Data 5.4 (2017): 481-494.

[35]. Moustafa, Nour, and Jill Slay. A hybrid feature selection for network intrusion detection systems: Central points. arXiv preprint arXiv:1707.05505 (2017).
[36]. Khammassi, Chaouki, and Saoussen Krichen. A GA-LR wrapper approach for feature selection in network intrusion detection. computers & security 70 (2017): 255-277.
[37]. Sethi, Kamalakanta, E. Sai Rupesh, Rahul Kumar, Padmalochan Bera, and Y. Venu Madhav. A context-aware robust intrusion detection system: a reinforcement learning-based approach. International Journal of Information Security 19 (2020): 657-678.
[38]. Kumar, V., D. Sinha, A. K. Das, S. C. Pandey, and R. T. Goswami. An integrated rule based Intrusion Detection System: Analysis on UNSW-NB15 Data Set and the real time online dataset-cluster computing. SpringerLink (2019).
[39]. Moustafa, Nour, and Jill Slay. The significant features of the UNSW-NB15 and the KDD99 data sets for network intrusion detection systems. In 2015 4th international workshop on building analysis datasets and gathering experience returns for security (BADGERS), pp. 25-31. IEEE, 2015.
[40]. https://research.unsw.edu.au/projects/unsw-nb15-dataset.
[41]. Breiman, Leo. Random forests. Machine learning 45 (2001): 5-32.
[42]. Dorogush, Anna Veronika, Vasily Ershov, and Andrey Gulin. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363 (2018).
[43]. Chen, Tianqi, and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.
[44]. Wang, Hui. Nearest neighbours without k: a classification formalism based on probability. Faculty of Informatics, University of Ulster (2002).
[45]. Jung, Alexander. Machine learning: the basics. Springer Nature, 2022.
[46]. Peter Gammie, The Tortoise and Hare Algorithm, 2015, [online] Available: http://isa-afp.org/entries/TortoiseHare.html.
[47]. MI Hoque, Nazrul, Dhruba K. Bhattacharyya, and Jugal K. Kalita. MIFS-ND: A mutual information-based feature selection method. Expert systems with applications 41, no. 14 (2014): 6371-6385.
[48]. Srinivasan, Manohar, and Narayanan Chidambaram Senthil kumar. Class imbalance data handling with optimal deep learning-based intrusion detection in IoT environment. Soft Computing 28, no. 5 (2024): 4519-4529.

Информация об авторах / Information about authors

Тамиларасан БХУВАНЕШВАРИ – доцент кафедры компьютерных наук и инженерии инженерного колледжа Мерсо Schlenk с 2021 года. Ее научные интересы – машинное обучение, обнаружение вторжений, методы оптимизации.

Thamilarasan BHUVANESWARI – Assistant professor, Department of computer science and engineering of Mepco Schlenk engineering college since 2021. Her research interests are machine learning, intrusion detection, optimization techniques.

Саундар Катхавараян РУБА – Dr. Sci., доцент (старший класс) кафедры компьютерных наук и инженерии инженерного колледжа Мепко Шленк с 2021 года. Область научных интересов: классификация текстур, субтитры к видео, агрегация данных, обнаружение теней, DDoS-атаки и обнаружение вторжений.

Soundar Kathavarayan RUBA – Dr. Sci., Associate professor (Sr. Grade), Department of computer science and engineering of Mepco Schlenk engineering college since 2021. Research interests: texture classification, video captioning, data aggregation, shadow detection, DDoS attack and intrusion detection.

Гуру Секар Рамакришнан ЧАНДРА – доцент (старший класс) кафедры математики инженерного колледжа Мерсо Schlenk с 2017 года. Область научных интересов: численные методы, методы оптимизации и обработки изображений.

Guru Sekar Ramakrishnan CHANDRA – Assistant professor (Sr.Grade), Department of mathematics of Mepco Schlenk engineering college since 2017. Research interests: numerical methods, optimization techniques and image processing.