ПРОТИВОДЕЙСТВИЕ ВНЕШНИМ ВТОРЖЕНИЯМ В ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СЕТИ

Задбоев В. А.[1], Абрамова Н. И.[2], Москалев В. С.[3]

DOI:10.21681/3034-4050-2025-5-18-23

Ключевые слова: *DDoS, машинное обучение, классификация, Logistic Regression, Random Forest, Gradient Boosting, синтетические данные, анализ трафика, метрики качества, ROC-кривая, PR-кривая, матрица ошибок.*

Аннотация

Цель: *на основе общих теоретических положений и актуальных релевантных работ раскрыть содержательную (вербальную) модель системы управления робототехническими комплексами (РТК).*

Метод: *на основе системного подхода выполнен анализ методов, моделей и технологий управления современных и перспективных РТК. Методом классификации и методом аналогий определены основные категории предметной области управления РТК специального назначения (СН).*

Результат: *приведены результаты анализа известных моделей, разработанных при проектировании и исследовании систем управления РТК СН. Определены основные принципы управления и технологии, применяемые для обеспечения управления РТК СН.*

Научная новизна: *заключается в обобщении и систематизации известных моделей и методов, а также технологий управления РТК СН. Сформирована вербальная модель системы управления РТК СН. Известные в настоящее время модели СУ РТК СН не в полной мере характеризуют свойства системы управления РТК как интеллектуальной системы со свойствами суперсистемы.*

Введение

Современные информационные системы подвергаются множеству угроз, среди которых атаки типа DDoS занимают одно из ведущих мест. В последние годы наблюдается значительный рост частоты и мощности подобных атак. Они способны полностью парализовать работу организаций, нарушить предоставление сервисов и нанести существенный экономический ущерб. В условиях цифровой трансформации и зависимости бизнеса от доступности сетевых ресурсов проблема своевременного обнаружения DDoS-атак приобретает критическое значение.

Существующие решения можно условно разделить на три группы: сигнатурные методы, эвристические алгоритмы и методы машинного обучения. Сигнатурные методы, хотя и эффективны против известных атак, плохо справляются с новыми сценариями. Эвристические алгоритмы позволяют выявлять аномалии, но часто страдают от высокого числа ложных срабатываний. В этих условиях использование машинного обучения становится естественным шагом, позволяющим автоматизировать процесс анализа и обнаружения аномалий

Постановка задачи в данной работе заключается в исследовании эффективности трёх алгоритмов классификации: Logistic Regression, Random Forest и Gradient Boosting. Каждый из этих методов имеет свои особенности и применяется в различных областях анализа данных. Задача исследования – выяснить, какой из алгоритмов лучше подходит для выявления DDoS-атак при работе с синтетически сгенерированными данными, имитирующими реальный сетевой трафик. Современные информационные системы подвергаются множеству угроз, среди которых атаки типа DDoS занимают одно из ведущих мест. В последние годы наблюдается значительный рост частоты и мощности подобных атак. Они способны полностью парализовать работу организаций, нарушить предоставление сервисов и нанести существенный экономический ущерб. В условиях цифровой трансформации и зависимости бизнеса от доступности сетевых ресурсов проблема своевременного обнаружения DDoS-атак приобретает критическое значение.

1 Задбоев Вадим Александрович, младший научный сотрудник Военной академии связи, г. Санкт-Петербург, Россия. E-mail: zadboev89@mail.ru

2 Абрамова Нина Ивановна, научный сотрудник Военной академии связи, г. Санкт-Петербург, Россия. E mail: HIAbram@yandex.ru

3 Москалев Владимир Сергеевич, слушатель Военной академии связи, г. Санкт-Петербург, г. Санкт-Петербург, Россия. E-mail: gsg.1991@mail.ru

Таблица 1

Пример синтетического датасета

packets_per_sec	bytes_per_sec	unique_src_ips	syn_rate	rst_rate	fin_rate	udp_qps	is_ddos
18981.326836783574	743102.9635478084	29.20733073396296	4790.482140191341	258.8130172384032	526.0855838712544	0.0	1
dns_qps	**http_req_rate**	**failed_conn_ratio**	**entropy_dst_port**	**Inbound_out_ratio**	**avg_ttl**	**pct_small_packets**	
63.589542598623325	908.7611284701245	0.017875279932732382	3.703961805562273	2.8167719454935187	77.67448470466215	0.3635508654165375	

Решение поставленной задачи

Для проведения эксперимента был разработан программный код, реализующий полный цикл: от генерации данных до анализа результатов. Синтетический датасет формировался с учётом таких характеристик, как количество пакетов в секунду, объём байт, число уникальных IP-адресов, доля неудачных соединений и прочие признаки, традиционно используемые для анализа сетевой активности (табл. 1).

Общее количество строк в датасете составило 20 тысяч. Данные разделялись на обучающую и тестовую выборки в пропорции 75 % к 25 %. Для сравнения алгоритмов классификации для всех трёх моделей определялся общий набор из всех признаков.

Logistic Regression использовалась для базового анализа линейных зависимостей. Random Forest позволял выявлять нелинейные закономерности и устойчиво работать с шумными данными. Gradient Boosting обеспечивал наилучшие результаты за счёт последовательного исправления ошибок предыдущих итераций.

В ходе экспериментов выяснили, что все три модели справляются примерно с одинаковой точностью с небольшим перевесом в пользу Logistic Regression, которая показала средний уровень точности около 0,94, при этом метрика Recall у всех оставалась на уровне 0,80, что указывает на возможность пропуска части атак.

Важным этапом исследования стал анализ визуализаций.

ROC-кривые (рис. 1) позволили убедиться, что ансамблевые методы (Random Forest и Gradient Boosting) обеспечивают лучшую отделимость классов по сравнению с линейной моделью Logistic Regression.

PR-кривые (рис. 2) подтвердили устойчивость Random Forest к несбалансированным данным.

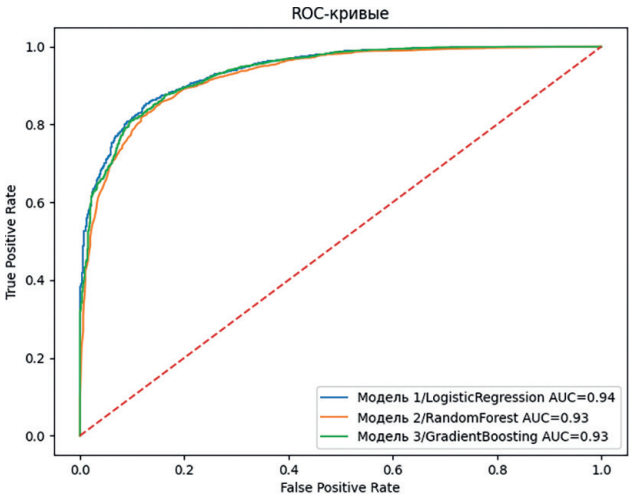

Рис. 1. График ROC-кривой

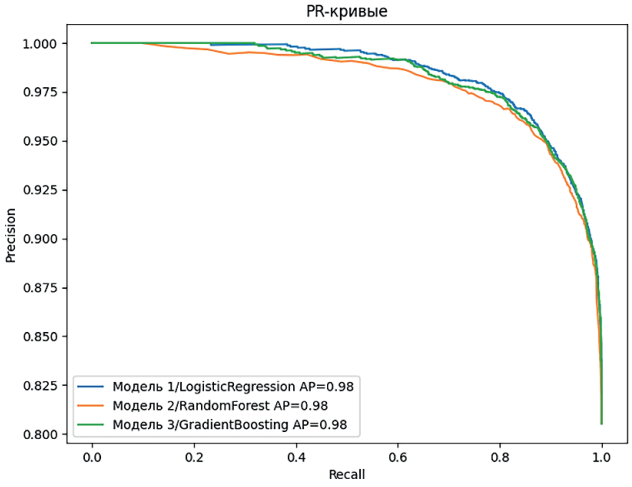

Рис. 2. График PR-кривой

Однако несмотря на итоговую оценку точности, матрицы ошибок (рис. 3) наглядно показали, что Logistic Regression чаще допускает ошибки второго рода (пропуск атак), в то время как ансамблевые методы обладают меньшим числом ложных срабатываний. Для последующих сравнений выбран метод

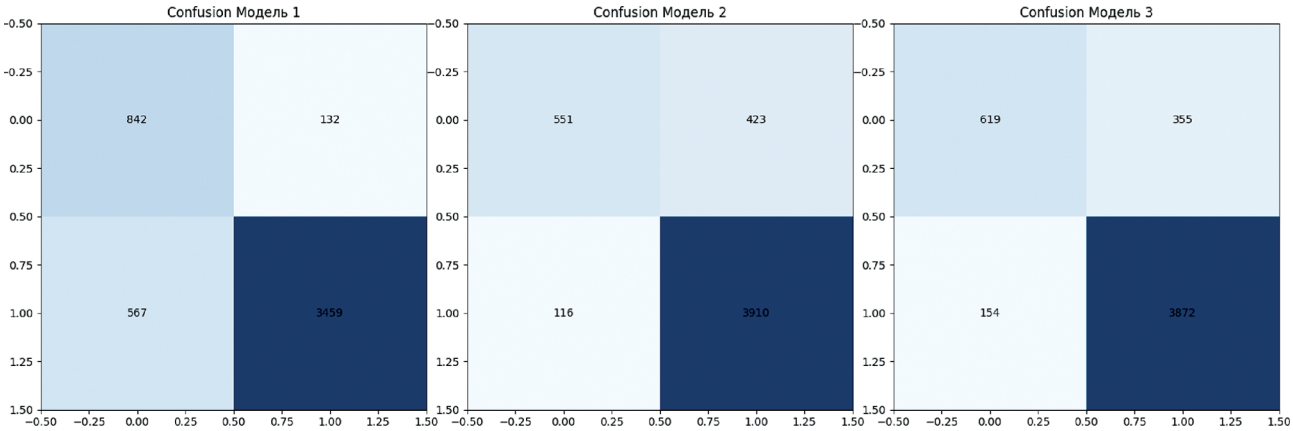

Рис. 3. Матрицы ошибок для каждой модели

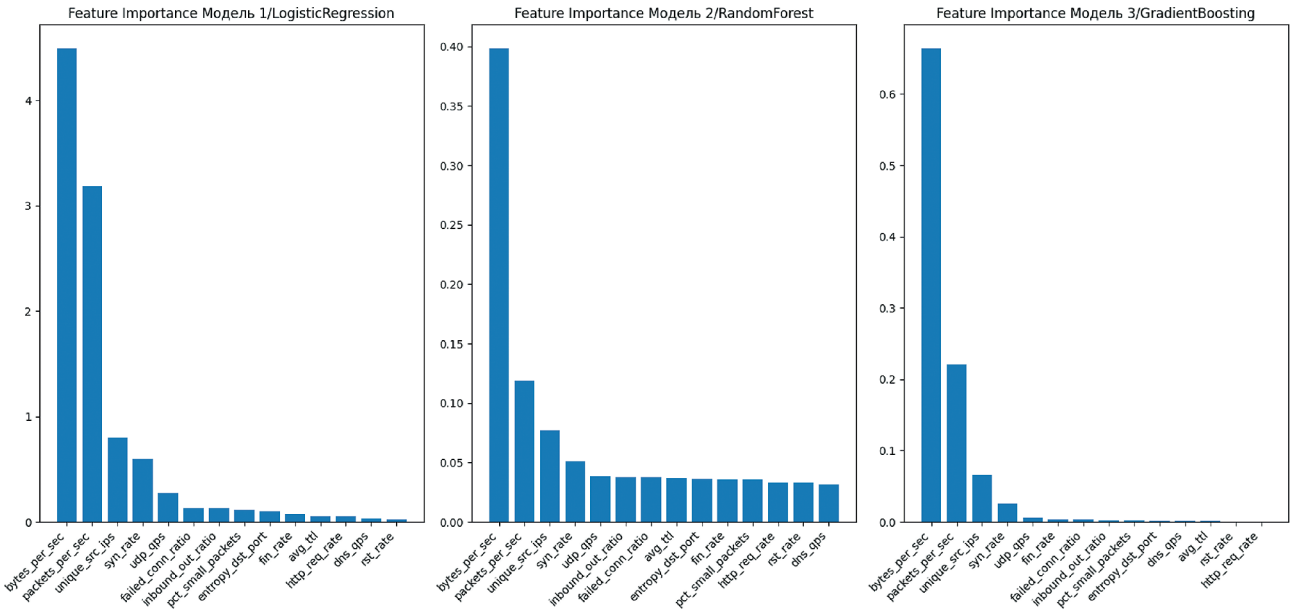

Рис.4. Графики с ключевыми индикаторами

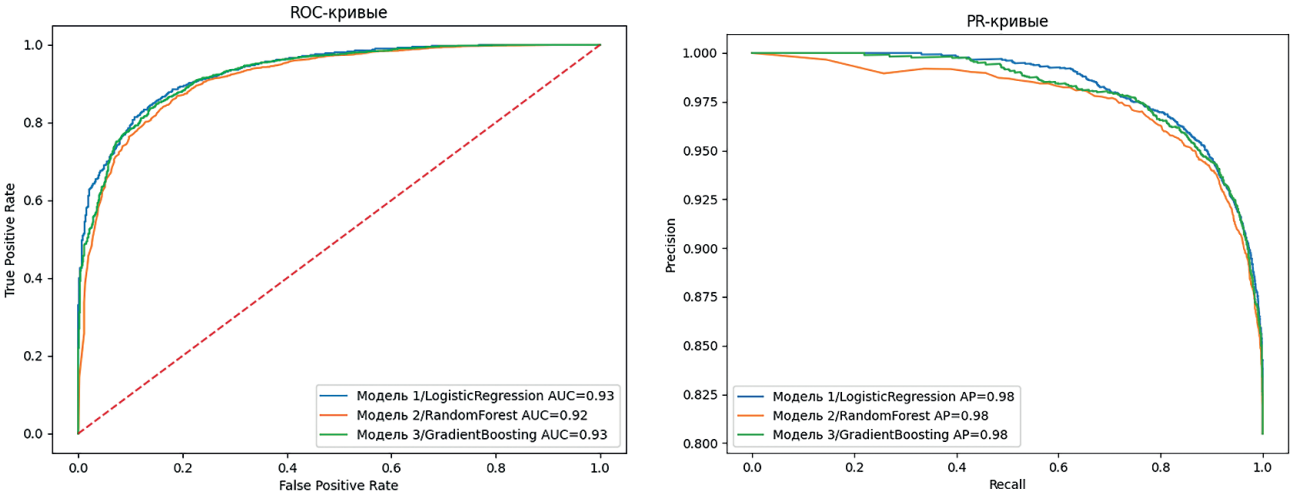

Рис. 5. Графики ROC-кривой и PR-кривой

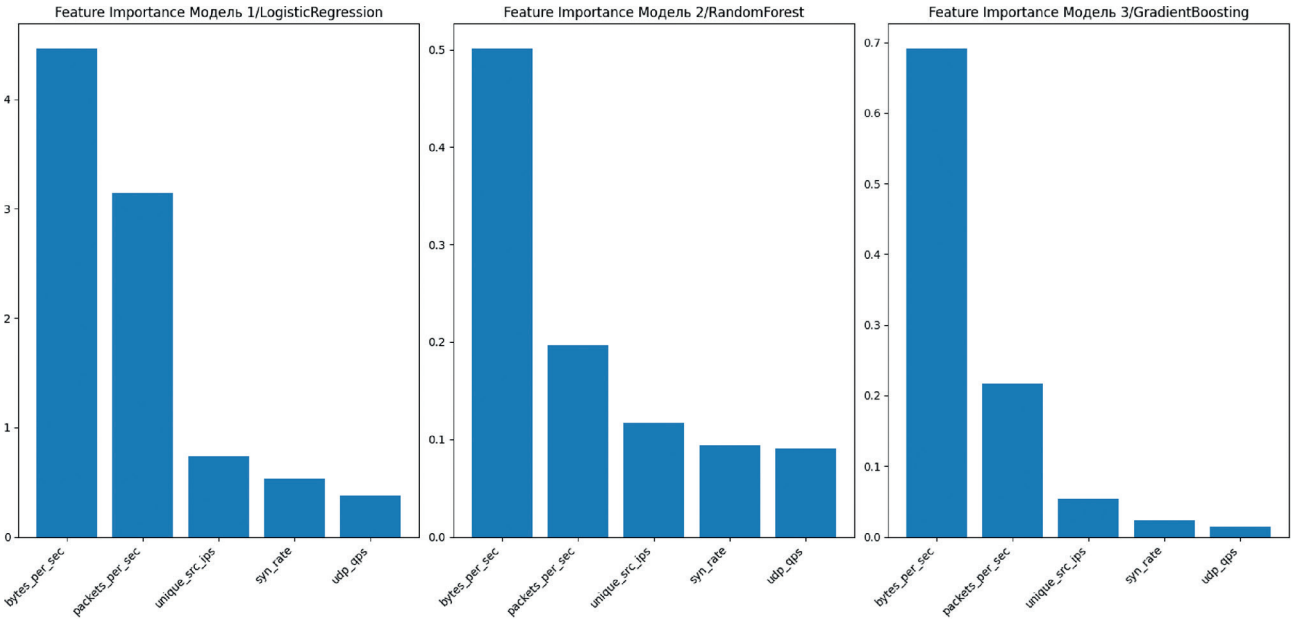

Рис. 6. Графики с ключевыми индикаторами

Gradient Boosting, так как он является наиболее сбалансированным при определении True Negative и True Positive характеристик.

Анализ важности признаков (рис. 4) уже на первом этапе выявил, что ключевыми индикаторами атак являются количество пакетов в секунду, количество бит в секунду, количество уникальных IP-адресов, скорость SYN-запросов и количество UDP-запросов. Поэтому для дальнейшего изучения выбраны вышеупомянутые признаки.

Следующим шагом запущен аналогичный предыдущему процесс обучения и тестирования моделей с применением только 5 наиболее важных признаков. Как видно на основании ROC- и PR-кривой (рис. 5) при наименьшем количестве вводных данных результаты почти остались на том же уровне с погрешностью в меньшую сторону.

Также на основании (рис. 6) видно, что в данной ситуации метод Gradient Boosting, меньше полагается на все 5 признаков, выдавая при этом схожий результат с аналогичными методами на основе чего можно сделать вывод о высокой эффективности данного метода, что делает его приоритетным при выборе обучающей модели.

Таким образом, проведённые эксперименты продемонстрировали преимущество в сравнении с аналогами ансамблевого метода Gradient Boosting в задаче обнаружения DDoS-атак.

Выводы

По результатам исследования достигнуты следующие результаты:

1. На основе сгенерированного синтетического датасета на 20 тысяч записей проведено обучение трёх моделей: Logistic Regression, Random Forest и Gradient Boosting, каждая из которых продемонстрировала свои сильные и слабые стороны.
2. С применением графиков (ROC, PR, матрицы ошибок, важности признаков) получены количественные оценки качества, подтверждающие, что ансамблевые методы обеспечивают более высокую точность и полноту.
3. Выявлены ключевые индикаторы атак при обнаружении DDoS-атак, которыми являются количество пакетов в секунду, количество бит в секунду, количество уникальных IP-адресов, скорость SYN-запросов и количество UDP-запросов.
4. Подтвержденная эффективность обучения моделей на основе выявленных ключевых признаков, а также отдельно модели Gradient Boosting в сравнении с аналогами.

Достигнутый эффект исследования заключается в том, что была подтверждена эффективность ансамблевых методов классификации в задачах обнаружения атак. Random Forest и Gradient Boosting показали наилучшие результаты по метрикам ROC-AUC и F1.

Gradient Boosting при использовании меньших входных данных для вычислений, обладает аналогичными результатами в обнаружении проводимых DDoS-атак и может быть рекомендована для систем обнаружения вторжений в информационно-вычислительных системах.

Практическая ценность работы состоит в том, что результаты работы могут быть использованы при проектировании систем обнаружения атак в реальных условиях. Они позволяют выстраивать комбинированные подходы, сочетающие простоту линейных и мощь ансамблевых моделей, а также более точно подбирать алгоритмы под конкретные сценарии, таким образом внося вклад в развитие методов информационной безопасности и открывая перспективы для дальнейших исследований в области применения гибридных систем и методов глубокого обучения для анализа сетевого трафика.

Литература

1. Волостных, В. А. Организация защиты конфиденциальной информации на предприятии средствами криптографической защиты / В. А. Волостных, П. А. Воробьев, В. А. Задбоев // Перспективы безопасности-2024: Сборник материалов II научно-технической конференции, посвященной информационной безопасности, Санкт-Петербург, 19-20 июня 2024 года. – Санкт-Петербург: ООО «Специальный Технологический Центр», 2024. – С. 9–13. – EDN FKFHVG.
2. Свидетельство о государственной регистрации программы для ЭВМ № 2024661259, Российская Федерация. Программа расчета вероятностно-временных характеристик средств сетевого контроля в условиях многоэтапных атак: № 2024617992: заявл. 11.04.2024: опубл. 16.05.2024 / В. А. Робак, В. А. Липатников, В. А. Парфиров [и др.]. – EDN WTNJLT.
3. Задбоев В. А. Определение цикла атаки противника в информационно-вычислительной сети / В. А. Задбоев, М. Д. Беседин, А. С. Антонов // Проблемы стандартизации, унификации и метрологии систем и средств связи специального назначения в аспекте их цифровой трансформации: Сборник материалов научно-практической конференции, Санкт-Петербург, 10 апреля 2023 года. – Санкт-Петербург: Военная академия связи им. Маршала Советского Союза С. М. Буденного, 2024. – С. 47–51. – EDN ZMTGNS.
4. Задбоев В. А. Способ защиты базовых сервисов, работающих с сетью интернет, серверов специального назначения / В. А. Задбоев, М. А. Магера, А. И. Скреблюков // Военная связь будущего. Квантовый скачок как неизбежность: Сборник материалов международной научно-практической конференции, Санкт-Петербург, 10-11 ноября 2023 года. – Санкт-Петербург: Федеральное государственное казенное военное образовательное учреждение высшего образования «Военная Академия Связи Имени Маршала Советского Союза С. М. Буденного» Министерства Обороны Российской Федерации, 2023. – С. 195–199. – EDN URMMWN.
5. Задбоев В. А. Алгоритм сканирования сетевой инфраструктуры для выявления внешних злоумышленников / В. А. Задбоев, В. А. Липатников // Инновационные достижения и результаты научной деятельности операторов научных рот Вооруженных Сил Российской Федерации: сборник научных статей по материалам круглого стола, Санкт-Петербург, 29 ноября 2022 года / Военная академия связи. – Санкт-Петербург: Федеральное государственное казенное военное образовательное учреждение высшего образования «Военная академия связи имени Маршала Советского Союза С. М. Буденного» Министерства обороны Российской Федерации, 2022. – С. 89–96. – EDN KQNXRV.

COUNTERACTING EXTERNAL INTRUSIONS IN THE INFORMATION AND COMPUTING NETWORK

Zadboev V. A.[4], Abramova N. I.[5], Moskalev V. S.[6]

Keywords: *DDoS, machine learning, classification, Logistic Regression, Random Forest, Gradient Boosting, synthetic data, traffic analysis, quality metrics, ROC curve, PR curve, error matrix**Abstract. The article presents the results of a large-scale study of the effectiveness of various machine learning algorithms in the problem of detecting distributed DDoS attacks.*

4 Vadim A. Zadboev, Junior Researcher, Military Academy of Communications, St. Petersburg, Russia. E-mail: zadboev89@mail.ru

5 Nina I. Abramova, Researcher, Military Academy of Communications, St. Petersburg, Russia. E-mail: HIAbram@yandex.ru

6 Vladimir S. Moskalev, Student of the Military Academy of Communications, St. Petersburg, St. Petersburg, Russia. E-mail: gsg.1991@mail.ru

Abstract

The purpose of the work *is to conduct a comparative analysis of three classification algorithms using a synthetic dataset that includes network traffic parameters typical for distributed attacks. It is necessary to identify the most appropriate method for the tasks of early detection of attacks and ensuring the stability of information systems.*

Method: *based on synthetically generated data reflecting the characteristics of network traffic, three approaches are compared: Logistic Regression, Random Forest, and Gradient Boosting.*

Results of the study: *the results of a large-scale study of the effectiveness of various machine learning algorithms in the task of detecting distributed DDoS attacks are presented. The paper shows differences in accuracy, completeness, F1-mer, as well as in resistance to unbalanced data. ROC curves, PR curves, error matrices are constructed, and the importance of the features on the basis of which the attack is determined is determined, key indicators of attacks in the detection of DDoS attacks are identified. which are the number of packets per second, the number of bits per second, the number of unique IP addresses, the rate of SYN requests, and the number of UDP requests. The effectiveness of training models based on the identified key features was confirmed. The effectiveness of the Gradient Boosting ensemble model in comparison with the above analogues has been confirmed.*

The scientific novelty *lies in the fact that within the framework of a single study, a systematic comparison of three different models is carried out with an emphasis on the analysis of their strengths and weaknesses in the context of information security tasks. As a result, it is possible to identify patterns that can be used in the construction of hybrid attack detection systems.*

References

1. Volostnyh, V. A. Organizacija zashhity konfidencial'noj informacii na predprijatii sredstvami kriptograficheskoj zashhity / V. A. Volostnyh, P. A. Vorob'ev, V. A. Zadboev // Perspektivy bezopasnosti-2024: Sbornik materialov II nauchno-tehnicheskoj konferencii, posvjashhennoj informacionnoj bezopasnosti, Sankt-Peterburg, 19-20 ijunja 2024 goda. – Sankt-Peterburg: OOO «Special'nyj Tehnologicheskij Centr», 2024. – S. 9–13. – EDN FKFHVG.
2. Svidetel'stvo o gosudarstvennoj registracii programmy dlja JeVM № 2024661259 Rossijskaja Federacija. Programma rascheta verojatnostno-vremennyh harakteristik sredstv setevogo kontrolja v uslovijah mnogojetapnyh atak: № 2024617992: zajavl. 11.04.2024: opubl. 16.05.2024 / V. A. Robak, V. A. Lipatnikov, V. A. Parfirov [i dr.]. – EDN WTNJLT.
3. Zadboev V. A. Opredelenie cikla ataki protivnika v informacionno-vychislitel'noj seti / V. A. Zadboev, M. D. Besedin, A. S. Antonov // Problemy standartizacii, unifikacii i metrologii sistem i sredstv svjazi special'nogo naznachenija v aspekte ih cifrovoj transformacii: Sbornik materialov nauchno-prakticheskoj konferencii, Sankt-Peterburg, 10 aprelja 2023 goda. – Sankt-Peterburg: Voennaja akademija svjazi im. Marshala Sovetskogo Sojuza S. M. Budennogo, 2024. – S. 47–51. – EDN ZMTGNS.
4. Zadboev V. A. Sposob zashhity bazovyh servisov, rabotajushhih s set'ju internet, serverov special'nogo naznachenija / V. A. Zadboev, M. A. Magera, A. I. Skrebljukov // Voennaja svjaz' budushhego. Kvantovyj skachok kak neizbezhnost': Sbornik materialov mezhdunarodnoj nauchno-prakticheskoj konferencii, Sankt-Peterburg, 10-11 nojabrja 2023 goda. – Sankt-Peterburg: Federal'noe gosudarstvennoe kazennoe voennoe obrazovatel'noe uchrezhdenie vysshego obrazovanija «Voennaja Akademija Svjazi Imeni Marshala Sovetskogo Sojuza S. M. Budennogo» Ministerstva Oborony Rossijskoj Federacii, 2023. – S. 195–199. – EDN URMMWN.
5. Zadboev V. A. Algoritm skanirovanija setevoj infrastruktury dlja vyjavlenija vneshnih zloumyshlennikov / V. A. Zadboev, V. A. Lipatnikov // Innovacionnye dostizhenija i rezul'taty nauchnoj dejatel'nosti operatorov nauchnyh rot Vooruzhennyh Sil Rossijskoj Federacii: sbornik nauchnyh statej po materialam kruglogo stola, Sankt-Peterburg, 29 nojabrja 2022 goda / Voennaja akademija svjazi. – Sankt-Peterburg: Federal'noe gosudarstvennoe kazennoe voennoe obrazovatel'noe uchrezhdenie vysshego obrazovanija «Voennaja akademija svjazi imeni Marshala Sovetskogo Sojuza S. M. Budennogo» Ministerstva oborony Rossijskoj Federacii, 2022. – S. 89–96. – EDN KQNXRV.