НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2026 Том 26 № 4 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
July–August 2026 Vol. 26 No 4 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-4-783-792
УДК 004.056.5

Обнаружение атак на веб-приложения на основе комбинирования трансформеров, сверточной нейронной сети и многослойного перцептрона

Павел Сергеевич Соболев[1], Игорь Витальевич Котенко[2]✉

[1] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[2] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

[1] pavels21@bk.ru, https://orcid.org/0009-0002-8091-8024
[2] ivkote@comsec.spb.ru✉, https://orcid.org/0000-0001-6859-7120

Аннотация

Введение. В связи с постоянным ростом количества атак на веб-приложения задача автоматического обнаружения таких атак остается одной из ключевых проблем информационной безопасности. Представлена разработка нового гибридного подхода к обнаружению атак на веб-приложения, в котором этап извлечения признаков переносится на уровень предварительно обученных трансформеров. **Метод.** Предлагается гибридный метод обнаружения атак на веб-приложения, основанный на совместном использовании специализированных трансформерных моделей и нейронных сетей различной архитектуры. Разработанный подход предполагает, что Uniform Resource Locator для запроса обрабатывается моделью URLBERT, обученной на структурных особенностях веб-адресов, тогда как содержимое запроса анализируется моделью SecRoBERT, ориентированной на задачи кибербезопасности. Полученные контекстные представления объединяются и передаются в параллельные блоки сверточной нейронной сети и многослойного перцептрона, обеспечивающие извлечение локальных и глобальных признаков соответственно. Предложенный подход позволяет объединить возможности трансформерных моделей обработки данных с эффективным выявлением статистических и локальных зависимостей на основе сверточной нейронной сети и многослойного перцептрона, что повышает точность классификации вредоносных запросов. **Основные результаты.** Экспериментальная оценка проведена на трех наборах данных, два из которых служат для бинарной классификации и один для многоклассовой. Тестирование проводилось с использованием *k*-fold кросс-валидации и посредством измерения задержки вывода модели, для демонстрации применимости метода в системах обнаружения атак в реальном времени. Результаты экспериментальной оценки предлагаемой модели продемонстрировали достаточно высокие показатели точности обнаружения атак на веб-приложения, а также показали существенный недостаток, связанный со значительным временем обработки запросов. **Обсуждение.** Анализ результатов тестирования и сравнение с релевантными исследованиями подтвердил целесообразность использования данной модели в качестве второго этапа анализа запросов в системах обнаружения атак на веб-приложения.

Ключевые слова

обнаружение атак на веб-приложения, информационная безопасность, глубокое обучение, трансформерные модели, SQL-инъекции, XSS-атаки

Благодарности

Работа выполнена при поддержке гранта Российского научного фонда и Санкт-Петербургского научного фонда № 25-11-20028 (https://rscf.ru/project/25-11-20028/) в СПб ФИЦ РАН.

Ссылка для цитирования: Соболев П.С., Котенко И.В. Обнаружение атак на веб-приложения на основе комбинирования трансформеров, сверточной нейронной сети и многослойного перцептрона // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 4. С. 783–792. doi: 10.17586/2226-1494-2026-26-4-783-792

Web application attack detection based on a combination of transformers, convolutional neural network, and multilayer perceptron

Pavel S. Sobolev[1], Igor V. Kotenko[2]✉

[1] ITMO University, Saint Petersburg, 197101, Russian Federation
[2] St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation

[1] pavels21@bk.ru, https://orcid.org/0009-0002-8091-8024
[2] ivkote@comsec.spb.ru✉, https://orcid.org/0000-0001-6859-7120

Abstract

Given the steady increase in the number of attacks on web applications, the task of automatically detecting such attacks requests remains one of the key challenges in information security. The aim of this work is to develop a new hybrid approach to detecting attacks on web applications in which the feature extraction stage is shifted to the level of pre-trained transformers. To achieve this goal, the paper proposes a hybrid method for detecting attacks on web applications based on the combined use of specialized transformer models and neural networks of various architectures. The suggested approach offers that the request URL is processed by a URLBERT model trained on the structural features of web addresses, while the request content is analyzed by a SecRoBERT model tailored for cybersecurity tasks. The resulting contextual representations are combined and fed into parallel blocks of a convolutional neural network and a multilayer perceptron, which extract local and global features, respectively. The proposed approach combines the capabilities of transformer data processing models with the efficient identification of statistical and local dependencies based on a convolutional neural network and a multilayer perceptron, which improves the accuracy of malicious query classification. Experimental evaluation is conducted on three datasets, two of which are used for binary classification and one for multi-class classification. Testing was performed using *k*-fold cross-validation and by measuring the model output latency to demonstrate the method applicability in real-time attack detection systems. The results of the experimental evaluation of the proposed model demonstrate high accuracy in detecting attacks on web applications, but also reveal a significant drawback related to the considerable time required to process requests. Analysis of the test results and comparison with relevant studies confirm the feasibility of using this model as the second stage of request analysis in web application attack detection systems.

Keywords

web application attack detection, information security, deep learning, transformer models, SQL injection, XSS attacks

Acknowledgements

This study was supported by the Russian Science Foundation and the Saint Petersburg Science Foundation (grant No. 25-11-20028, https://rscf.ru/project/25-11-20028/) in SPC RAS.

For citation: Sobolev P.S., Kotenko I.V. Web application attack detection based on a combination of transformers, convolutional neural network, and multilayer perceptron. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 4, pp. 783–792 (in Russian). doi: 10.17586/2226-1494-2026-26-4-783-792

Введение

Веб-приложения являются основой современной цифровой инфраструктуры и активно используются в бизнесе, государственных системах и облачных сервисах. Рост их количества сопровождается увеличением числа атак, направленных на эксплуатацию уязвимостей прикладного уровня. Наиболее распространенными остаются Structured Query Language (SQL) инъекции и Cross-Site Scripting (XSS) атаки, позволяющие злоумышленнику выполнять произвольный код или получать несанкционированный доступ к данным [1].

Существующие системы защиты часто используют сигнатурный анализ, который плохо адаптируется к новым и модифицированным атакам [2]. В связи с этим все большее распространение получают методы глубокого обучения [3], способные выявлять скрытые закономерности в HyperText Transfer Protocol (HTTP) запросах.

В работе [4] установлено, что гибридные архитектуры, объединяющие сверточные нейронные сети (Convolutional Neural Networks, CNN) и многослойные перцептроны (Multilayer perceptron, MLP), эффективно извлекают локальные и статистические признаки текстовых последовательностей, обеспечивая высокую точность обнаружения атак. Однако такие подходы обычно используют простые методы предварительной обработки, например статистическую меру Term Frequency-Inverse Document Frequency (TF-IDF), и символьную токенизацию, что ограничивает способность модели учитывать контекст и семантику запросов.

Современные трансформерные модели, основанные на механизме внимания, позволяют получать контекстные представления текста и успешно применяются в задачах анализа естественного языка и кибербезопасности [5]. Специализированные модели, обученные на Uniform Resource Locator (URL) структурах и данных информационной безопасности, способны выявлять сложные зависимости.

Целью работы является разработка нового гибридного подхода к обнаружению атак на веб-приложения, в котором этап извлечения признаков переносится на уровень предварительно обученных трансформеров. Предлагаемый метод объединяет преимущества моделей Bidirectional Encoder Representations from Transformers (BERT) семейства с архитектурами CNN и MLP, сохраняя эффективность гибридной схемы и повышая качество представления входных данных.

В настоящей работе исследовано применение специализированных трансформеров для раздельной

обработки URL и содержимого запроса; интеграция контекстных эмбеддингов в гибридную архитектуру CNN+MLP; экспериментальная оценка основных метрик эффективности обнаружения.

Релевантные работы

В современных исследованиях, посвященных обнаружению и классификации веб-атак, все более широкое применение находят методы глубокого обучения, позволяющие эффективно извлекать информативные признаки из данных и повышать качество распознавания атакующих воздействий. Рассмотрим подробнее ряд значимых релевантных исследований в данной сфере.

В [6] предложена гибридная модель глубокого обучения Long Short-Term Memory (CNN-LSTM) для обнаружения SQL-инъекций в веб-приложениях. Методология работы включает многоэтапный пайплайн обработки SQL-запросов: представление текста запросов с помощью TF-IDF, последующий отбор признаков методом χ^2 (хи-квадрат), усечение предложений с сохранением только статистически значимых токенов и обучение нейросетевых архитектур. Эксперименты проводились на объединенном наборе данных, включающем реальные SQL-запросы информационной системы организации, и публичный набор SQLiV3 из Kaggle. Результаты показывают, что предложенная модель CNN-LSTM достигает значение Accuracy, Recall и F1-score 99,85 %, а Precision — 99,86 %.

В [7] предложен метод автоматического обнаружения SQL-инъекций на основе автоэнкодерной нейросети Autoencoder Network (AE-Net) для автоматического извлечения признаков из текстовых SQL-запросов. Экспериментальные исследования проведены на наборе данных из 46 392 SQL-запросов и показали, что классификатор XGBoost в сочетании с признаками AE-Net достигает точности до 99 % при 10-кратной кросс-валидации.

В [8] представлен фреймворк обнаружения веб-атак, основанный на комбинации BERT и MLP, предназначенный для бинарной классификации HTTP-запросов на нормальные и аномальные. Векторные представления размерности 80 × 768 (матрицы из 80 токенов, где каждый элемент представлен вектором скрытого пространства размером 768) формируются с помощью BERT-токенизатора и используются в качестве входа для MLP из 6 полносвязных слоев. Эксперименты проводились на объединенном наборе данных CSIC 2010, FWAF и HttpParams. Полученные результаты демонстрируют точность до 99,98 % и F1-score — 98,7 %, при среднем времени классификации около 0,4 мс на запрос.

В [9] рассмотрен подход к автоматическому обнаружению инъекционных атак на основе методов глубокого обучения. В работе сравнивают глубокую нейронную сеть (DNN) и CNN, обученных на наборах данных ECML/PKDD 2007, CSIC 2010, а также Hybrid 2020, сформированном с использованием реальных вредоносных payloads. Экспериментальные результаты показывают, что CNN стабильно превосходит DNN по точности и уровню абстракции признаков, особенно при классификации легитимных запросов, а DNN демонстрирует лучшую переносимость при обнаружении вредоносных запросов на данных, отличных от обучающего набора.

В [10] предложен метод аномального обнаружения атак на веб-приложения с использованием ошибки реконструкции HTTP GET-запросов. Модель обучается только на нормальных запросах. Ключевой особенностью подхода является сегментация запросов на числовые значения, пути, параметры и специальные символы. Реконструкционная нейросетевая архитектура, основанная на объединении многоголового самовнимания и управляемой сверточной сети, применяется для моделирования нормального поведения запросов. Средняя кросс-энтропийная ошибка реконструкции служит для определения аномальности запроса. Результаты тестирования на реальных журналах веб-серверов продемонстрировали высокую эффективность подхода с показателем Area Under the Curve (AUC) до 0,99.

В [11] представлен подход для обнаружения SQL-инъекций в веб-приложениях на основе двух специализированных аритектур — CNN-SIDNet-1 и SIDNet-2. Особенность подхода — автоматическое извлечение признаков. Данные модели предназначены для задачи бинарной классификации. Предлагаемая архитектура SIDNet-2 является расширением базовой модели посредством применения dropout-регуляризации, что позволяет снизить переобучение и улучшить обобщающую способность модели. Результаты экспериментов на публичных наборах данных SQLi и SQLiV2 продемонстрировали высокую точность классификации, достигающую 98 %.

В [12] предложен метод обнаружения атак на веб-приложения, основанный на концепции обучения без примеров и CNN. Обучение модели выполняется исключительно на легитимных HTTP-запросах. В качестве входных данных используется Uniform Resource Identifier (URI) путь запроса, описанный через векторные представления символов, сформированные на основе ASCII-кодирования. CNN применяется для извлечения признаков, после чего классификация выполняется посредством сравнения полученных представлений с векторами классов по евклидову расстоянию. Метод был протестирован на нескольких наборах данных, включая CSIC 2010 и реальные журналы веб-трафика. Результаты экспериментов показали высокий уровень полноты обнаружения.

В [13] представлена модель SecurityBERT, предназначенная для обнаружения сетевых атак в Internet of Things (IoT) и IIoT-средах. Суть подхода заключается в адаптации специализированной вариации BERT к анализу сетевых данных с обеспечением конфиденциальности трафика. В рамках работы предложен метод кодирования фиксированной длины с сохранением конфиденциальности, который преобразует числовые и категориальные характеристики сетевого трафика в текстовые представления, посредством объединения наименований признаков с их значениями и последующим хэшированием. После чего последовательности токенов обрабатываются токенизатором кодирования

пар байтов. Результаты экспериментального тестирования на наборе данных Edge-IIoTset продемонстрировали высокие показатели точности (до 98,2 %).

Анализ современных исследований в рамках задачи обнаружения атак на веб-приложения демонстрирует эффективность методов глубокого обучения, в частности архитектур на основе MLP [14, 15] и на основе CNN [16, 17], которые характеризуются эффективным выявлением признаков и высокой точностью классификации запросов. В последние годы одним из перспективных направлений является использование моделей обработки естественного языка, основанных на трансформенных архитектурах, в частности BERT [18], которые позволяют использовать контекстные зависимости HTTP-запросов. Также одним из перспективных направлений исследований, является добавление объяснимого искусственного интеллекта в существующие подходы для улучшения интерпретируемости результатов [19, 20]. Данное направление является важным аспектом для внедрения разрабатываемых систем в практические среды эксплуатации, в связи с тем, что интерпретация принятых решений способна упростить задачу обнаружения при анализе срабатываний.

Предлагаемый подход

Предлагаемый метод основан на гибридной архитектуре глубокого обучения, предназначенной для бинарной классификации HTTP-запросов на нормальные и вредоносные. Основная идея заключается в многоуровневом извлечении признаков, объединяющем контекстное представление данных и классические методы нейронной обработки последовательностей.

В отличие от подхода [4], где признаки формировались на основе TF-IDF и символьных последовательностей, в новом методе начальный этап полностью основан на трансформерных моделях BERT.

Ключевым компонентом архитектуры BERT является механизм самовнимания, который обеспечивает моделирование взаимосвязей между всеми элементами входной последовательности. Механизм самовнимания позволяет учитывать глобальный контекст входной строки. Это особенно важно при анализе веб-запросов, поскольку вредоносные шаблоны часто формируются не отдельными символами или токенами, а их комбинациями, распределенными по всей строке запроса. В основе механизма самовнимания лежит преобразование входной последовательности токенов в три набора векторных представлений, включающих в себя запросы, ключи и значения. Для каждого токена входной последовательности вычисляется степень его значимости относительно всех остальных токенов. Дополнительным важным свойством архитектуры BERT является использование двунаправленного контекстного кодирования входной последовательности, благодаря которому BERT учитывает информацию одновременно из левой и правой частей последовательности.

Блок-схема предлагаемого подхода представлена на рис. 1.

URL-запроса передается в специализированную модель URLBERT, обученную на корпусах веб-адресов. Модель выполняет токенизацию и формирует контекстные эмбеддинги, учитывающие структуру URL, взаимосвязь сегментов пути, характерные паттерны атак и позиционные зависимости символов. Длина входной последовательности составляет 64 токена.

Содержимое HTTP-запроса обрабатывается моделью SecRoBERT — трансформером, адаптированным для задач информационной безопасности. Данная модель обучена выявлять вредоносные конструкции, характерные для XSS, SQL-инъекций и других атак, на основе контекстных признаков запросов. Длина входной последовательности составляет 128 токенов.

Полученные эмбеддинги передаются в сверточный блок. Модель CNN служит для выявления локальных зависимостей в рамках последовательностей признаков. В результате можно обнаружить характерные шаблоны атак. Архитектура CNN состоит из двух одномерных сверточных слоев. Первый содержит 128 фильтров с размером ядра 3, а второй 128 фильтров с размером ядра 5. Далее происходит агрегирование и преобразование последовательности признаков в один итоговый вектор.

Параллельно с архитектурой CNN применяется MLP, предназначенный для обработки статистических признаков запроса. На данном этапе анализируются агрегированные представления из трансформеров и дополнительные числовые параметры запроса, в число которых входит длина запроса и HTTP-метод. В результате усиливается статистическая составляющая анализа. Числовые признаки запроса передаются в полносвязный слой на 64 нейрона с функцией активации

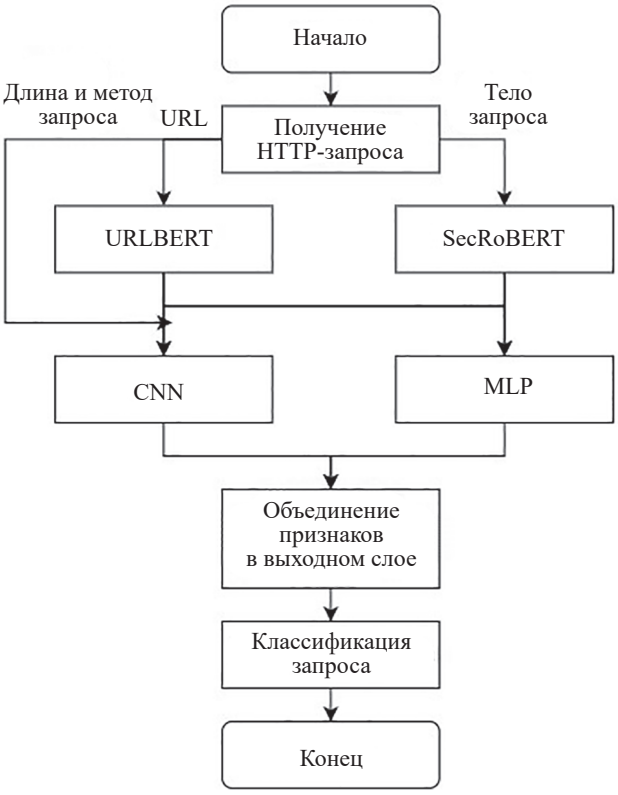

Рис. 1. Блок-схема предлагаемого подхода

Fig. 1. Block diagram of the proposed adapter

Rectified Linear Unit (ReLU), усредненные векторные представления адресной строки и содержимого запроса подаются в два разных полносвязных слоя на 128 нейронов с функцией активации ReLU.

Признаки, полученные из URLBERT, SecRoBERT, CNN и MLP, конкатенируются в единый вектор признаков. Итоговый классификатор является полносвязной нейронной сетью, которая завершается функцией softmax, служащей для определения вероятности отнесения запроса к определенному классу. Также предлагается вариация данного подхода для задачи бинарной классификации. Структура остается прежней, но завершается не функцией softmax, а сигмоидной функцией.

Таким образом, предложенная архитектура объединяет контекстные признаки трансформерных моделей, локальные структурные признаки CNN и глобальные статистические признаки MLP.

Эксперименты и результаты

На основании проведенного анализа современных исследований в области обнаружения атак на веб-приложения [21], был выбран набор данных CSIC 2010 [22], который наиболее часто используется для обучения и тестирования предлагаемых подходов. Набор данных CSIC 2010 содержит 25 тыс. вредоносных запросов и 36 тыс. нормальных запросов, при этом охватывает широкий спектр угроз, таких как SQL-инъекции, XSS-атаки, выполнение команд на стороне сервера и ряда других атак. Важной особенностью CSIC 2010 является наличие дополнительных параметров запросов, помимо текстового содержимого или URL-строки. Набор данных имеет характеристики реальных сетевых взаимодействий, что позволяет оценивать эффективность моделей в условиях, приближенных к эксплуатационной среде веб-приложений. В настоящей работе были также использованы два набора данных, размещенных на платформе Kaggle, содержащие HTTP-запросы с различными типами веб-атак. Первый набор данных включает 7373 вредоносных запроса, содержащих SQL-инъекции, а также 6313 легитимных запросов [23]. Второй набор данных содержит 11 450 вредоносных запросов с XSS-атаками и 19 150 нормальных запросов [24].

На этапе предварительной обработки данных из исходных наборов были удалены повторяющиеся записи и были сформированы два производных набора данных путем их объединения и балансировки классов.

В наборе данных представлены три класса: легитимные запросы (non_attack); запросы, содержащие SQL-инъекции (sql_injection); запросы, содержащие XSS-атаки (xss-attack). Распределение классов после объединения данных представлено на рис. 2, *a*.

Первый набор данных предназначен для решения задачи бинарной классификации. При его формировании все вредоносные запросы были объединены в

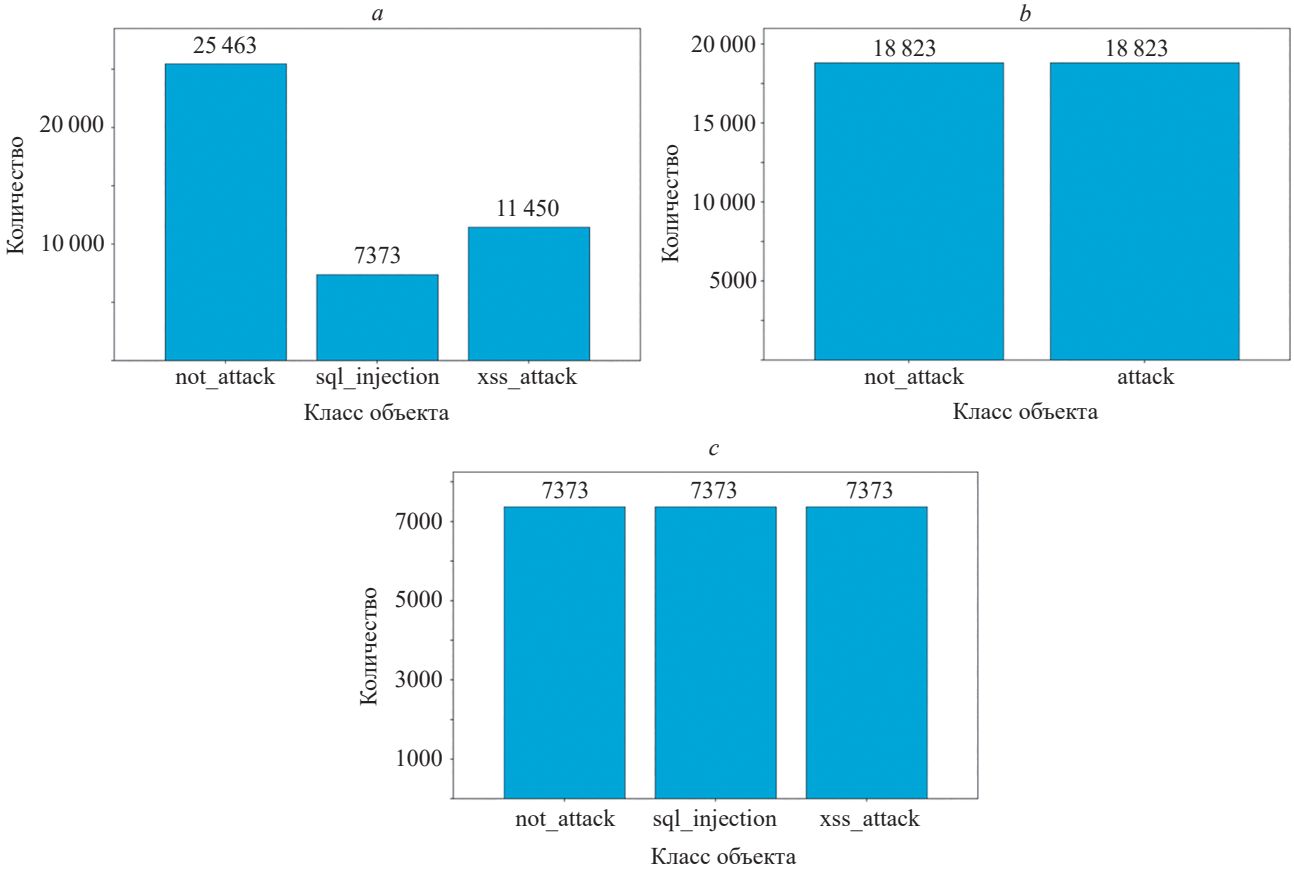

Рис. 2. Распределение классов: в объединенном наборе данных (*a*); после балансировки для бинарной (*b*) и для мультиклассовой (*c*) классификаций

Fig. 2. Class distribution: in the combined dataset (*a*); after balancing for binary (*b*) and multiclass (*c*) classifications

один класс, а нормальные в другой. Совокупное число вредоносных запросов составило 18 823, тогда как число нормальных запросов — 25 463. Для устранения дисбаланса классов из множества нормальных запросов было случайным образом отобрано 18 823 экземпляра, что позволило сформировать сбалансированный набор данных. В результате итоговый набор для бинарной классификации содержит 37 646 запросов, из которых 18 823 являются вредоносными и 18 823 нормальными (рис. 2, *b*).

Второй набор данных использовался для задачи многоклассовой классификации. В данном случае рассматривались три класса: non_attack; sql_injection; xss-attack. Для обеспечения балансировки классов за основу принимался наименьший по объему класс — SQL-инъекции sql_injection, содержащий 7373 запроса. Из остальных классов случайным образом отбиралось такое же количество запросов. Итоговый набор данных включает 22 119 запросов, по 7373 экземпляра в каждом из трех классов (рис. 2, *c*).

Тестирование проводилось на персональном компьютере с процессором AMD Ryzen 5 2600 (6 ядер, 3,4 ГГц), 16 ГБ оперативной памяти с частотой 2400 МГц и видеоадаптером NVIDIA GeForce RTX 3060.

Для предотвращения переобучения в процессе обучения использовались методы регуляризации и ранней остановки обучения, с целью ограничения избыточный адаптации модели к обучающей выборке. Дополнительно оценка обобщающей способности модели проводилась с применением 5-кратной кросс-валидации, что обеспечило более устойчивую и статистически корректную оценку качества классификации.

Результаты тестирования на наборе данных CSIC 2010 приведены в табл. 1.

Результаты бинарной классификации приведены в табл. 2.

Поскольку распределение классов в исследуемом наборе данных является сбалансированным, для оценки качества многоклассовой классификации использовались метрики с усреднением macro, при котором значения показателей вычисляются независимо для каждого класса и затем усредняются без учета их численности. Такой подход обеспечивает равный вклад каждого класса в итоговую оценку модели и позволяет получить объективную характеристику ее обобщающей способности.

Результаты многоклассовой классификации приведены в табл. 3.

В рамках тестирования были сформированы матрицы ошибок для всех тестируемых моделей. Матрицы ошибок для методов на основе BERT представлены на рис. 3.

Матрицы ошибок для методов на основе SecRoBERT, представлены на рис. 4.

Сравнение полученных результатов предлагаемого подхода на основе SecRoBERT + CNN + MLP с результатами из других работ в контексте мультиклассовой классификации показано в табл. 4. Следует отметить, что используемые наборы данных не являются идентичными по составу и характеристикам, вследствие чего их прямое сопоставление не позволяет сделать однозначные выводы.

Сравнение с аналогами в задаче мультиклассовой классификации демонстрирует, что предлагаемый подход имеет высокую эффективность в рамках обнаружения атак на веб-приложения и превосходит суще-

Таблица 1. Результаты тестирования на наборе данных CSIC 2010

Table 1. Test results on the CSIC 2010 dataset

Модель	Accuracy	Precision	Recall	F1-score	Время обработки одного запроса, мс
BERT + CNN	0,952	0,997	0,883	0,937	8,47
BERT + MLP	0,926	0,938	0,879	0,907	7,51
BERT + MLP + CNN	0,965	0,985	0,931	0,957	7,45
SecRoBERT + URLBERT + CNN	0,964	0,998	0,914	0,954	7,46
SecRoBERT + URLBERT + MLP	0,965	0,992	0,924	0,956	7,31
SecRoBERT + URLBERT + MLP + CNN	0,968	0,991	0,932	0,961	7,23

Таблица 2. Результаты бинарной классификации

Table 2. Binary classification results

Модель	Accuracy	Precision	Recall	F1-score	Время обработки одного запроса, мс
BERT + CNN	0,993	0,997	0,989	0,993	5,68
BERT + MLP	0,995	0,998	0,990	0,994	5,42
BERT + MLP + CNN	0,996	0,999	0,993	0,996	5,31
SecRoBERT + CNN	0,996	0,998	0,993	0,995	5,27
SecRoBERT + MLP	0,997	0,999	0,995	0,997	5,14
SecRoBERT + MLP + CNN	0,998	0,999	0,997	0,998	5,05

Таблица 3. Результаты многоклассовой классификации
Table 3. Results of multi-class classification

Модель	Accuracy	Precision	Recall	F1-score	Время обработки одного запроса, мс
BERT + CNN	0,991	0,995	0,987	0,991	5,71
BERT + MLP	0,993	0,996	0,988	0,992	5,49
BERT + MLP + CNN	0,995	0,996	0,991	0,994	5,38
SecRoBERT + CNN	0,995	0,996	0,992	0,994	5,36
SecRoBERT + MLP	0,996	0,996	0,993	0,995	5,18
SecRoBERT + MLP + CNN	0,997	0,997	0,997	0,997	5,05

ствующие подходы. Однако, из-за различий в наборах данных, показатели могут отличаться при тестировании всех представленных моделей на одном общем наборе данных.

В результате, на основе проведенного тестирования, можно сделать общий вывод, что использование BERT, а в особенности специализированных версий SecRoBERT и URLBERT, повышает общее качество модели. Предлагаемый подход с использованием SecRoBERT + URLBERT + MLP + CNN демонстрирует высокие показатели на несбалансированном наборе данных CSIC 2010, который содержит множество атак, однако Precision выше у моделей на основе CNN + BERT и его вариациях. В то же время, предлагаемый подход показывает лучшие результаты на сбалансированном наборе данных в рамках бинарной и многоклассовой классификации, что говорит о его высоком потенциале применимости в системах обнаружения

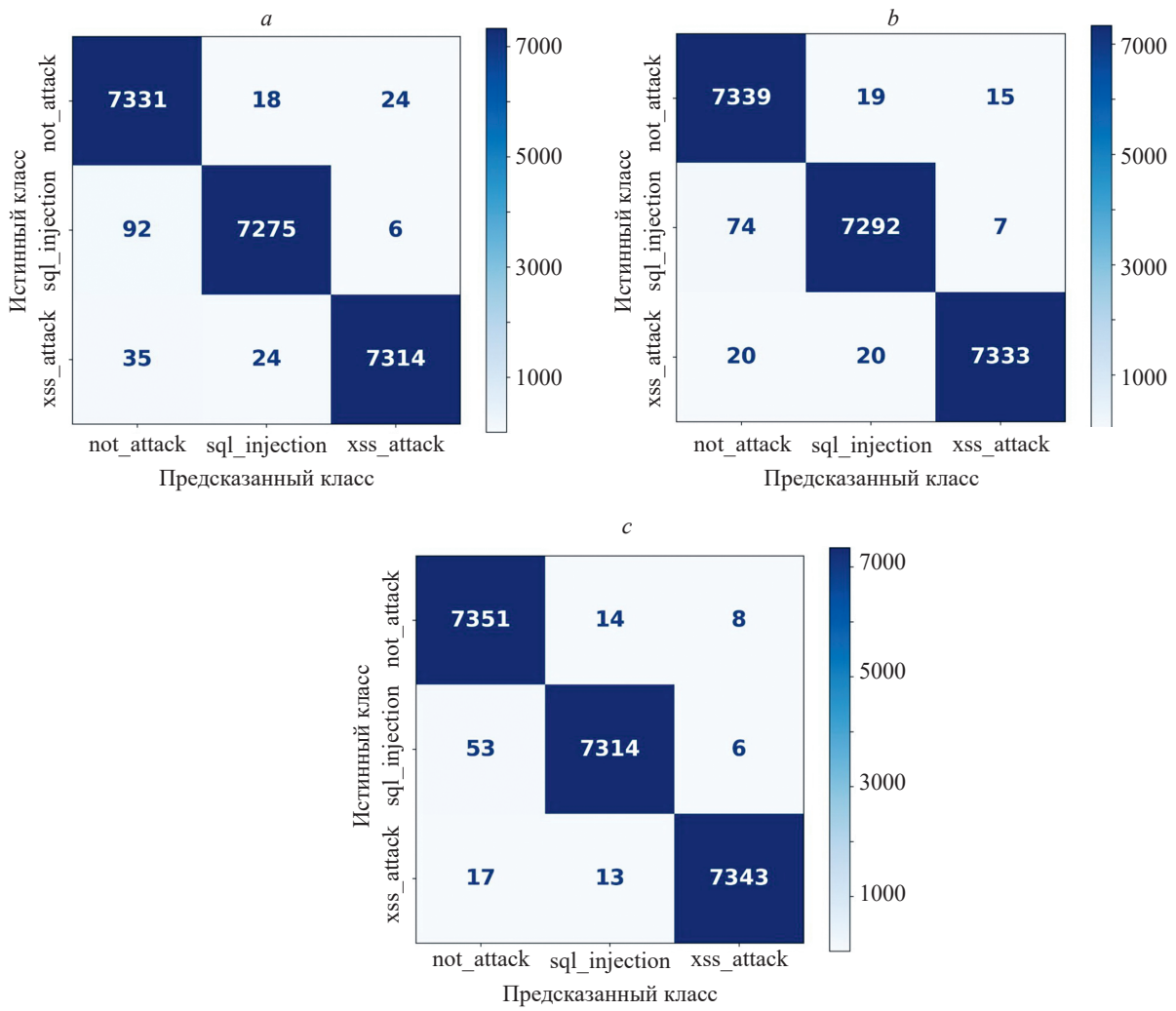

Рис. 3. Матрицы ошибок методов на основе BERT: BERT + CNN (*a*); BERT + MLP (*b*); BERT + MLP + CNN (*c*)
Fig. 3. Confusion matrices of BERT-based methods: BERT + CNN (*a*); BERT + MLP (*b*); BERT + MLP + CNN (*c*)

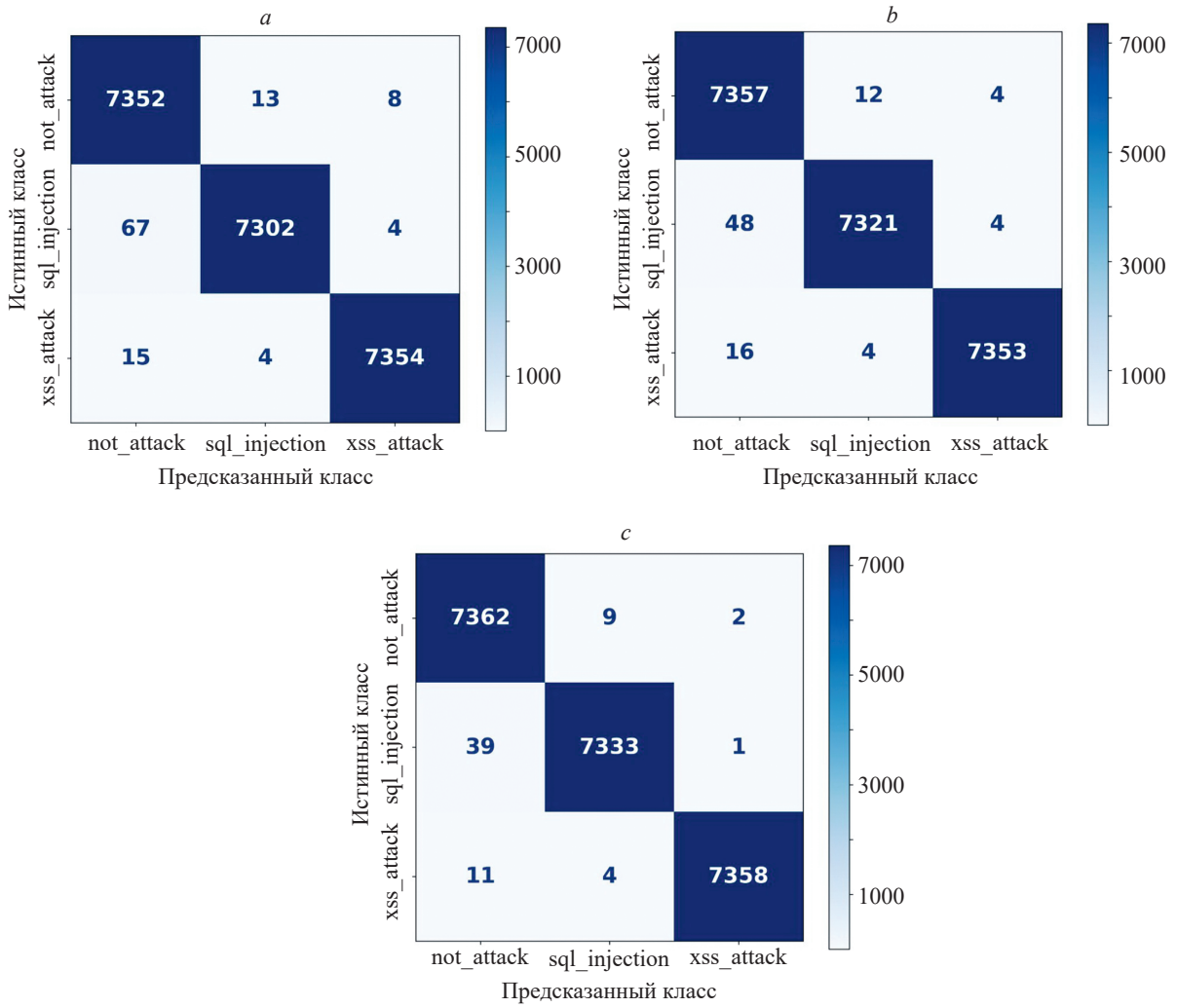

Рис. 4. Матрицы ошибок методов на основе SecRoBERT: SecRoBERT + CNN (*a*); SecRoBERT + MLP (*b*); SecRoBERT + MLP + CNN (*c*)

Fig. 4. Confusion matrices of SecRoBERT-based methods: SecRoBERT + CNN (*a*); SecRoBERT + MLP (*b*); SecRoBERT + MLP + CNN (*c*)

Таблица 4. Сравнение предлагаемого подхода с аналогами

Table 4. Comparison of the proposed approach with analogues

Ссылка на источник	Модель	Набор данных	Accuracy	Precision	Recall	F1-score
[25]	DA-SANA	Собственный набор данных	0,966	0,965	0,965	0,966
[26]	CNN + GRU	Собственный набор данных	0,996	0,996	0,995	0,996
Представленная работа	SecRoBERT + MLP + CNN	Объединение наборов данных из платформы Kaggle	0,997	0,997	0,997	0,997

атак. Однако все подходы, основанные на трансформерах BERT и его вариациях, имеют общие недостатки. Модели требуют много времени на обработку запросов, что затрудняет их внедрение в системы обнаружения атак на веб-приложения. С учетом среднего времени обработки одного запроса получается около 192 запросов в минуту, что явно указывает на неприменимость в реальных системах. Также, основываясь на результатах матриц ошибок — все модели плохо разделяют SQL-инъекции от обычных запросов.

Заключение

Предложен подход к обнаружению атак на веб-приложения на основе гибридной архитектуры, объединяющей специализированные трансформерные модели

URLBERT и SecRoBERT, а также сверточные нейронные сети и перцептроны. Сущность подхода заключается в объединении контекстных представлений Uniform Resource Locator и содержимого HTTP-запроса, локальных структурных признаков и глобальных статистических характеристик, что позволяет повысить полноту анализа признаков для выявления атак и их классификации. Таким образом, научная новизна подхода состоит в построении единой архитектуры, интегрирующей специализированные трансформерные представления адресной строки и содержимого запроса с методами выделения локальных и глобальных признаков. В отличие от подходов, основанных только на текстовом анализе или на статистических характеристиках, предложенное решение обеспечивает комплексное описание запроса за счет объединения контекстных, структурных и статистических признаков. Разработанный подход демонстрирует высокие показатели при работе со сбалансированными данными, а также демонстрирует высокую эффективность в рамках многоклассовой классификации. Однако при применении данной модели к несбалансированным данным наблюдается снижение показателей качества. К недостаткам подхода также можно отнести высокую длительность обработки запросов и наличие проблем с отнесением Structured Query Language-инъекций к классу легитимных запросов. Представленную модель можно применять в реальных системах по обнаружению атак на веб-приложения в качестве дополнительной проверки запросов с целью отнесения их к конкретным видам атак, но необходимо реализовать отдельный этап для первичной бинарной классификации, с целью снижения нагрузки на систему и обеспечения работы в реальном времени. В качестве дальнейшего направления развития модели рассматривается переход к двухступенчатой архитектуре обнаружения атак на веб-приложения. На первом этапе планируется реализовать облегченную модель с бинарной классификацией запросов и передачей всех потенциально вредоносных в разработанную модель для многоклассовой классификации, что должно повысить практическую применимость разработанного метода за счет возможности обработки множества запросов в реальном времени. Также перспективным направлением развития является добавление моделей объяснимого искусственного интеллекта для интерпретируемости результатов.

Литература

1. Kaur D., Kaur P. Empirical analysis of Web attacks // Procedia Computer Science. 2016. V. 78. P. 298–306. doi: 10.1016/j.procs.2016.02.057
2. Desnitsky V.A., Kotenko I.V., Nogin S.B. Detection of anomalies in data for monitoring of security components in the Internet of Things // Proc. of the XVIII International Conference on Soft Computing and Measurements (SCM). 2015. P. 189–192. doi: 10.1109/SCM.2015.7190452
3. Dong H., Kotenko I. Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection // Knowledge and Information Systems. 2025. V. 67. N 5. P. 3915–3966. doi: 10.1007/s10115-025-02366-w
4. Котенко И.В., Соболев П.С. Обнаружение атак на веб-приложения на основе комбинирования сверточных нейронных сетей и многослойных перцептронов // Информатизация и связь. 2025. № 2. С. 90–96. doi: 10.34219/2078-8320-2025-16-2-90-96
5. Kotenko I., Abramenko G. Ontology-guided heterogeneous graph retrieval for large-language-model interpretation of Suricata events // Lecture Notes in Networks and Systems. 2026. V. 1762. P. 261–272. doi: 10.1007/978-3-032-13615-2_22
6. Casmiry E.N., Sinde R.S., Mduma N.M. A Hybrid deep learning model for SQL injection attack detection // IEEE Access. 2026. V. 14. P. 6450–6463. doi: 10.1109/ACCESS.2026.3651991
7. Thalji N., Raza A., Islam M.S., Samee N.A., Jamjoom M.M. AE-Net: Novel autoencoder-based deep features for SQL injection attack detection // IEEE Access. 2023. V. 11. P. 135507–135516. doi: 10.1109/ACCESS.2023.3337645
8. Seyyar Y.E., Yavuz A.G., Ünver H.M. An attack detection framework based on BERT and deep learning // IEEE Access. 2022. V. 10. P. 68633–68644. doi: 10.1109/ACCESS.2022.3185748
9. Odumuyiwa V., Chibueze A. Automatic detection of HTTP injection attacks using convolutional neural network and deep neural network // Journal of Cyber Security and Mobility. 2021. V. 9. N 4. P. 489–514. doi: 10.13052/jcsm2245-1439.941
10. Li J., Fu Y., Xu J., Ren C., Xiang X., Guo J. Web application attack detection based on attention and gated convolution networks // IEEE Access. 2020. V. 8. P. 20717–20724. doi: 10.1109/ACCESS.2019.2955674
11. Muduli D., Shookdeb S., Zamani A.T., Saxena S., Kanade A.S., Parveen N., Shameem M. SIDNet: A SQL Injection detection network for enhancing cybersecurity // IEEE Access. 2024. V. 12. P. 176511–176526. doi: 10.1109/ACCESS.2024.3502293
12. Demirel D.Y., Sandikkaya M.T. Web based anomaly detection using zero-shot learning with CNN // IEEE Access. 2023. V. 11. P. 91511–91525. doi: 10.1109/ACCESS.2023.3303845

References

1. Kaur D., Kaur P. Empirical analysis of Web attacks. *Procedia Computer Science*, 2016, vol. 78, pp. 298–306. doi: 10.1016/j.procs.2016.02.057
2. Desnitsky V.A., Kotenko I.V., Nogin S.B. Detection of anomalies in data for monitoring of security components in the Internet of Things. *Proc. of the XVIII International Conference on Soft Computing and Measurements (SCM)*, 2015, pp. 189–192. doi: 10.1109/SCM.2015.7190452
3. Dong H., Kotenko I. Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection. *Knowledge and Information Systems*, 2025, vol. 67, no. 5, pp. 3915–3966. doi: 10.1007/s10115-025-02366-w
4. Kotenko I.V., Sobolev P.S. Detection of attacks on Web applications based on a combination of convolutional neural networks and multilayer perceptrons. *Informatization and communication*, 2025, no. 2, pp. 90–96. (in Russian). doi: 10.34219/2078-8320-2025-16-2-90-96
5. Kotenko I., Abramenko G. Ontology-guided heterogeneous graph retrieval for large-language-model interpretation of Suricata events. *Lecture Notes in Networks and Systems*, 2026, vol. 1762, pp. 261–272. doi: 10.1007/978-3-032-13615-2_22
6. Casmiry E.N., Sinde R.S., Mduma N.M. A Hybrid deep learning model for SQL injection attack detection. *IEEE Access*, 2026, vol. 14, pp. 6450–6463. doi: 10.1109/ACCESS.2026.3651991
7. Thalji N., Raza A., Islam M.S., Samee N.A., Jamjoom M.M. AE-Net: Novel autoencoder-based deep features for SQL injection attack detection. *IEEE Access*, 2023, vol. 11, pp. 135507–135516. doi: 10.1109/ACCESS.2023.3337645
8. Seyyar Y.E., Yavuz A.G., Ünver H.M. An attack detection framework based on BERT and deep learning. *IEEE Access*, 2022, vol. 10, pp. 68633–68644. doi: 10.1109/ACCESS.2022.3185748
9. Odumuyiwa V., Chibueze A. Automatic detection of HTTP injection attacks using convolutional neural network and deep neural network. *Journal of Cyber Security and Mobility*, 2021, vol. 9, no. 4, pp. 489–514. doi: 10.13052/jcsm2245-1439.941
10. Li J., Fu Y., Xu J., Ren C., Xiang X., Guo J. Web application attack detection based on attention and gated convolution networks. *IEEE Access*, 2020, vol. 8, pp. 20717–20724. doi: 10.1109/ACCESS.2019.2955674
11. Muduli D., Shookdeb S., Zamani A.T., Saxena S., Kanade A.S., Parveen N., Shameem M. SIDNet: A SQL Injection detection network for enhancing cybersecurity. *IEEE Access*, 2024, vol. 12, pp. 176511–176526. doi: 10.1109/ACCESS.2024.3502293
12. Demirel D.Y., Sandikkaya M.T. Web based anomaly detection using zero-shot learning with CNN. *IEEE Access*, 2023, vol. 11, pp. 91511–91525. doi: 10.1109/ACCESS.2023.3303845

13. Ferrag M.A., Ndhlovu M., Tihanyi N., Cordeiro L.C., Debbah M., Lestable T., et al. Revolutionizing cyber threat detection with large language models: a privacy-preserving BERT-based lightweight model for IoT/IIoT devices // IEEE Access. 2024. V. 12. P. 23733–23750. doi: 10.1109/ACCESS.2024.3363469
14. Tang P., Qiu W., Huang Z., Lian H., Liu G. Detection of SQL injection based on artificial neural network // Knowledge-Based Systems. 2020. V. 190. P. 105528. doi: 10.1016/j.knosys.2020.105528
15. Mokbal F.M.M., Wang D., Imran A., Lin J., Akhtar F., Wang X. MLPXSS: An Integrated XSS-based attack detection scheme in Web applications using multilayer perceptron technique // IEEE Access. 2019. V. 7. P. 100567–100580. doi: 10.1109/ACCESS.2019.2927417
16. Xie X., Ren C., Fu Y., Xu J., Guo J. SQL injection detection for Web applications based on elastic-pooling CNN // IEEE Access. 2019. V. 7. P. 151475–151481. doi: 10.1109/ACCESS.2019.2947527
17. Tian Z., Luo C., Qiu J., Du X., Guizani M. A Distributed deep learning system for Web attack detection on edge devices // IEEE Transactions on Industrial Informatics. 2020. V. 16. N 3. P. 1963–1971. doi: 10.1109/TII.2019.2938778
18. El Mahdaouy A., Lamsiyah S., Idrissi M.J., Alami H., Yartaoui Z., Berrada I. DomURLs_BERT: Pre-trained BERT-based model for malicious domains and URLs detection and classification // Journal of Network and Systems Management. 2026. V. 34. N 2. P. 36. doi: 10.1007/s10922-025-10010-9
19. Bacevicius M., Paulauskaite-Taraseviciene A., Zokaityte G., Kersys L., Moleikaityte A. Comparative analysis of perturbation techniques in LIME for intrusion detection enhancement // Machine Learning and Knowledge Extraction. 2025. V. 7. N 1. P. 21. doi: 10.3390/make7010021
20. Bacevicius M., Paulauskaite-Taraseviciene A. Machine learning algorithms for raw and unbalanced intrusion detection data in a multi-class classification problem // Applied Sciences. 2023. V. 13. N 12. P. 7328. doi: 10.3390/app13127328
21. Котенко И.В., Соболев П.С. Обнаружение атак на веб-приложения: анализ современных подходов // Актуальные проблемы инфотелекоммуникаций в науке и образовании: Сборник научных статей XIII Международной научно-технической и научно-методической конференции. 2024. Т. 1. Р. 497–501.
22. Еремин Е.О. Обзор открытых наборов данных для выявления атак на веб-приложения // International Journal of Open Information Technologies. 2024. Т. 12. № 3. С. 106–113.
23. Venkatramulu S., Waseem Md.S., Taneem A., Thoutam S.Y., Apuri S., Nachiketh. Research on SQL injection attacks using word embedding techniques and machine learning // Journal of Sensors, IoT & Health Sciences. 2024. V. 2. N 1. P. 55–64. doi: 10.69996/jsihs.2024005
24. Bakır R. UniEmbed: A Novel approach to detect XSS and SQL injection attacks leveraging multiple feature fusion with machine learning techniques // Arabian Journal for Science and Engineering. 2025. V. 50. N 19. P. 15591–15604. doi: 10.1007/s13369-024-09916-4
25. Karacan H., Sevri M. A Novel data augmentation technique and deep learning model for Web application security // IEEE Access. 2021. V. 9. P. 150781–150797. doi: 10.1109/ACCESS.2021.3125785
26. Yang W., Zuo W., Cui B. Detecting malicious URLs via a keyword-based convolutional gated-recurrent-unit neural network // IEEE Access. 2019. V. 7. P. 29891–29900. doi: 10.1109/ACCESS.2019.2895751

13. Ferrag M.A., Ndhlovu M., Tihanyi N., Cordeiro L.C., Debbah M., Lestable T., et al. Revolutionizing cyber threat detection with large language models: a privacy-preserving BERT-based lightweight model for IoT/IIoT devices. *IEEE Access*, 2024, vol. 12, pp. 23733–23750. doi: 10.1109/ACCESS.2024.3363469
14. Tang P., Qiu W., Huang Z., Lian H., Liu G. Detection of SQL injection based on artificial neural network. *Knowledge-Based Systems*, 2020, vol. 190, pp. 105528. doi: 10.1016/j.knosys.2020.105528
15. Mokbal F.M.M., Wang D., Imran A., Lin J., Akhtar F., Wang X. MLPXSS: An Integrated XSS-based attack detection scheme in Web applications using multilayer perceptron technique. *IEEE Access*, 2019, vol. 7, pp. 100567–100580. doi: 10.1109/ACCESS.2019.2927417
16. Xie X., Ren C., Fu Y., Xu J., Guo J. SQL injection detection for Web applications based on elastic-pooling CNN. *IEEE Access*, 2019, vol. 7, pp. 151475–151481. doi: 10.1109/ACCESS.2019.2947527
17. Tian Z., Luo C., Qiu J., Du X., Guizani M. A Distributed deep learning system for Web attack detection on edge devices. *IEEE Transactions on Industrial Informatics*, 2020, vol. 16, no. 3, pp. 1963–1971. doi: 10.1109/TII.2019.2938778
18. El Mahdaouy A., Lamsiyah S., Idrissi M.J., Alami H., Yartaoui Z., Berrada I. DomURLs_BERT: Pre-trained BERT-based model for malicious domains and URLs detection and classification. *Journal of Network and Systems Management*, 2026, vol. 34, no. 2, pp. 36. doi: 10.1007/s10922-025-10010-9
19. Bacevicius M., Paulauskaite-Taraseviciene A., Zokaityte G., Kersys L., Moleikaityte A. Comparative analysis of perturbation techniques in LIME for intrusion detection enhancement. *Machine Learning and Knowledge Extraction*, 2025, vol. 7, no. 1, pp. 21. doi: 10.3390/make7010021
20. Bacevicius M., Paulauskaite-Taraseviciene A. Machine learning algorithms for raw and unbalanced intrusion detection data in a multi-class classification problem. *Applied Sciences*, 2023, vol. 13, no. 12, pp. 7328. doi: 10.3390/app13127328
21. Kotenko I.V., Sobolev P.S. Detection of attacks on Web applications: analysis of modern approaches. *Proc. of the Actual Problems of Infotelecommunications in Science and Education (APINO 2024)*, 2024, vol. 1, pp. 497–501. (in Russian)
22. Eremin E. Verview of public datasets for Web application attack detecting. *International Journal of Open Information Technologies*, 2024, vol. 12, no. 3, pp. 106–113. (in Russian)
23. Venkatramulu S., Waseem Md.S., Taneem A., Thoutam S.Y., Apuri S., Nachiketh. Research on SQL injection attacks using word embedding techniques and machine learning. *Journal of Sensors, IoT & Health Sciences*, 2024, vol. 2, no. 1, pp. 55–64. doi: 10.69996/jsihs.2024005
24. Bakır R. UniEmbed: A Novel approach to detect XSS and SQL injection attacks leveraging multiple feature fusion with machine learning techniques. *Arabian Journal for Science and Engineering*, 2025, vol. 50, no. 19, pp. 15591–15604. doi: 10.1007/s13369-024-09916-4
25. Karacan H., Sevri M. A Novel data augmentation technique and deep learning model for Web application security. *IEEE Access*, 2021, vol. 9, pp. 150781–150797. doi: 10.1109/ACCESS.2021.3125785
26. Yang W., Zuo W., Cui B. Detecting malicious URLs via a keyword-based convolutional gated-recurrent-unit neural network. *IEEE Access*, 2019, vol. 7, pp. 29891–29900. doi: 10.1109/ACCESS.2019.2895751

Авторы

Соболев Павел Сергеевич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 59499265100, https://orcid.org/0009-0002-8091-8024, Pavels21@bk.ru
Котенко Игорь Витальевич — доктор технических наук, профессор, главный научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, SC 15925268000, https://orcid.org/0000-0001-6859-7120, ivkote@comsec.spb.ru

Authors

Pavel S. Sobolev — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 59499265100, https://orcid.org/0009-0002-8091-8024, Pavels21@bk.ru
Igor V. Kotenko — D.Sc., Professor, Chief Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation, SC 15925268000, https://orcid.org/0000-0001-6859-7120, ivkote@comsec.spb.ru

Статья поступила в редакцию 14.04.2026
Одобрена после рецензирования 05.06.2026
Принята к печати 22.07.2026

Received 14.04.2026
Approved after reviewing 05.06.2026
Accepted 22.07.2026