МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра Информационных технологий и систем безопасности

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

(Дипломная работа)

На тему «Интеграция искусственного интеллекта в системы защиты информации**»**

Исполнитель _____ Загрядцкий Арсений Игоревич

(подпись) (фамилия, имя, отчество)

Руководитель _____ Козлов Юрий Викторович

(подпись) (фамилия, имя, отчество)

«К защите допускаю»

Заведующий кафедрой _____ Лепешкин Олег Михайлович

(подпись) (фамилия, имя, отчество)

«____»____ 20___ г.

Санкт-Петербург

МИНИСТЕРСТВО НАУКИ И ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра Информационных технологий и систем безопасности

«УТВЕРЖДАЮ»

Заведующий кафедрой

_____ _____
(подпись) (фамилия, имя, отчество)

«___»_____20___ года

Задание

на выпускную квалификационную работу

студенту _____
(фамилия, имя, отчество)

1. Тема Интеграция искусственного интеллекта в системы защиты информации

закреплена приказом ректора Университета от «____» _____20___ года, № _____

2. Срок сдачи законченной работы «___» _____20___ года

3. Исходные данные к выпускной квалификационной работе:

4. Перечень вопросов, подлежащих разработке (краткое содержание работы:

Введение. Актуальность темы, цели и задачи ВКР

Глава 1 Основы обеспечения безопасности в образовательной организации
(наименование главы)

Глава 2 Разработка аналитико-математической модели принятия решений в системе обеспечения информационной безопасности образовательной организации
(наименование главы)

Глава 3 Методика применения аналитико-математической модели в системе обеспечения информационной безопасности образовательной организации
(наименование главы)

Глава 4 Научно-экономическое обоснование методики построения системы управления информационной безопасностью образовательной организации
(наименование главы)

Заключение. Выводы по работе в целом. Оценка степени решения поставленных задач. Практические рекомендации.

5. Перечень материалов, представляемых к защите:

– Пояснительная записка;

– Схема _____
(наименование схемы)

– Диаграмма _____
(наименование диаграммы)

6. Консультанты по работе

6.1. _____

6.2. _____

…

7. Дата выдачи задания: «__» _____ 20_____ года Руководитель выпускной квалификационной работы

_____ _____
(должность, ученая степень, ученое звание, фамилия, имя, отчество) (подпись)

Задание принял к исполнению «___»_____20___ года

Студент _____ _____
(фамилия, имя, отчество, учебная группа) (подпись)

РЕФЕРАТ

Дипломная работа: с., ____рис., _табл., __ приложения, ____ источников литературы.

СИСТЕМА УПРАВЛЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТЬЮ, СТАНДАРТИЗАЦИЯ В ОБЛАСТИ СИСТЕМ УПРАВЛЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТЬЮ, ПОЛИТИКА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ, АНАЛИЗ РИСКОВ.

Объект исследования:

Предмет исследования:

Цель работы:

В дипломной работе проводится анализ…

Разработан …

СОДЕРЖАНИЕ

Оглавление

ВВЕДЕНИЕ

В условиях цифровизации бизнеса и активного развития веб-технологий веб-приложения становятся основным инструментом взаимодействия пользователей с информационными ресурсами. Одновременно с этим возрастает количество и сложность атак на веб-приложения, что делает вопросы информационной безопасности особенно актуальными.

В последние годы характер киберугроз существенно изменился. Всё чаще злоумышленники используют автоматизированные инструменты и технологии машинного обучения для проведения разведки, анализа инфраструктуры и поиска уязвимостей. Такие атаки отличаются высокой скоростью, масштабируемостью и адаптивностью, что значительно усложняет их обнаружение и предотвращение традиционными средствами защиты информации.

Современные системы защиты информации (СЗИ) обладают высокой функциональностью и способны обеспечивать комплексную защиту информационных ресурсов. Однако их внедрение и эксплуатация требуют значительных финансовых затрат и наличия квалифицированных специалистов по информационной безопасности. Большинство коммерческих решений ориентированы на крупные организации и недоступны для малого и среднего бизнеса.

В результате многие организации вынуждены использовать упрощённые механизмы защиты или полностью отказываться от специализированных средств безопасности. Это приводит к росту рисков утечки данных, компрометации систем и финансовых потерь. Особенно уязвимыми оказываются веб-приложения, которые являются основным каналом взаимодействия с пользователями и внешними сервисами.

Практический опыт подтверждает актуальность рассматриваемой проблемы. В процессе тестирования инфраструктуры была развёрнута база данных с открытым сетевым доступом и белым IP-адресом. Спустя непродолжительное время после публикации ресурс подвергся атаке: данные

были зашифрованы, а злоумышленники потребовали выкуп за их восстановление. Несмотря на то, что база данных носила тестовый характер и не содержала критически важной информации, данный инцидент продемонстрировал высокую скорость обнаружения уязвимых ресурсов и эффективность автоматизированных средств атакующих.

Данный случай стал отправной точкой для анализа существующих подходов к защите информационных систем. Он показал, что современные угрозы требуют не только наличия средств защиты, но и их адаптивности, доступности и простоты внедрения. В условиях ограниченного бюджета и нехватки квалифицированных специалистов особую значимость приобретают решения, способные сочетать эффективность защиты с относительно низкой стоимостью внедрения и эксплуатации.

Одним из перспективных направлений повышения эффективности защиты информационных систем является применение методов машинного обучения. Интеллектуальные алгоритмы позволяют анализировать поведение пользователей, выявлять аномалии в сетевом трафике и адаптироваться к изменяющимся условиям угроз, что делает их особенно актуальными в условиях роста автоматизированных атак.

Целью данной дипломной работы является разработка и внедрение интеллектуальной системы защиты веб-приложений на основе методов машинного обучения, ориентированной на использование в условиях ограниченного бюджета и минимального участия специалистов по информационной безопасности.

Для достижения поставленной цели в работе необходимо решить следующие задачи, проанализировать современные угрозы безопасности веб-приложений;

- исследовать существующие подходы к защите веб-приложений и информационных систем;
- разработать архитектуру веб-приложения с системой аутентификации;

- реализовать интеллектуальный механизм анализа и фильтрации сетевого трафика на основе машинного обучения;
- сформировать датасеты легитимного и вредоносного трафика;
- обучить и протестировать модели машинного обучения;
- оценить эффективность и экономическую целесообразность предложенного решения.

Объектом исследования является система защиты веб-приложений.

Предметом исследования являются методы и алгоритмы машинного обучения, применяемые для анализа HTTP-трафика и выявления атак.

Практическая значимость работы заключается в возможности использования разработанного решения в реальных условиях эксплуатации веб-приложений, особенно в организациях с ограниченными ресурсами. Предложенный подход позволяет снизить порог входа для внедрения современных средств защиты информации за счёт упрощения архитектуры, автоматизации процессов анализа трафика и использования методов машинного обучения.

Основным концептуальным посылом данной дипломной работы является разработка системы защиты веб-приложений, которая сочетает в себе простоту внедрения и эксплуатации с высоким уровнем эффективности. Предлагаемая система ориентирована на практическое применение в условиях ограниченного бюджета и дефицита квалифицированных специалистов по информационной безопасности, при этом она должна обеспечивать уровень защиты, сопоставимый или превосходящий возможности распространённых бесплатных решений.

Таким образом, работа направлена на решение актуальной научно-практической задачи - создание доступной, масштабируемой и интеллектуальной системы защиты веб-приложений, способной адаптироваться к современным угрозам и требованиям цифровой среды.

1.ТЕОРЕТИЧЕСКИЕ И ИНЖЕНЕРНЫЕ ОСНОВЫ СИСТЕМ ЗАЩИТЫ ИНФОРМАЦИИ И ИНТЕЛЛЕКТУАЛЬНЫХ МЕТОДОВ АНАЛИЗА

1.1. Система защиты информации как объект научного исследования и инженерного проектирования

Развитие цифровых технологий привело к тому, что информационные системы стали ключевым элементом функционирования организаций, государственных структур и общества в целом. Современные веб-приложения и сетевые сервисы обеспечивают хранение, обработку и передачу значительных объёмов данных, включая персональную, коммерческую и служебную информацию. При этом рост функциональности и доступности информационных ресурсов неизбежно сопровождается увеличением числа угроз и усложнением методов их реализации.

Система защиты информации в научной интерпретации рассматривается как целостный комплекс взаимосвязанных средств, методов и процессов, направленных на обеспечение устойчивости информационной системы к воздействию угроз. Данный комплекс включает технические и программные средства защиты, криптографические механизмы, организационные регламенты, процедуры мониторинга и реагирования на инциденты, а также методы управления рисками. В инженерном аспекте система защиты информации представляет собой функционально интегрированную подсистему, встроенную в архитектуру информационной системы и функционирующую в режиме реального времени.

Информационная безопасность определяется как состояние защищенности информации и информационных ресурсов, при котором обеспечиваются ее конфиденциальность, целостность и доступность. Конфиденциальность отражает способность системы предотвращать несанкционированный доступ к данным, целостность характеризует устойчивость данных к несанкционированным изменениям, а доступность определяет возможность

легитимного использования информации в требуемый момент времени. Данные свойства образуют фундаментальную модель безопасности, используемую в международных стандартах и научных исследованиях.

С инженерной точки зрения система защиты информации должна рассматриваться как динамическая и адаптивная система. Это означает, что она не может быть сведена к набору статических правил или фиксированных механизмов контроля. Современные угрозы характеризуются высокой степенью изменчивости, автоматизации и интеллектуализации, что требует применения методов анализа, способных учитывать контекст, поведение пользователей и статистические характеристики трафика. В этом контексте особое значение приобретает интеграция методов машинного обучения и интеллектуального анализа данных в архитектуру систем защиты.

Экономический аспект построения системы защиты информации является одним из ключевых факторов, определяющих её практическую реализуемость. Для крупных организаций характерно использование комплексных и дорогостоящих решений, включающих специализированное оборудование, коммерческие программные продукты и работу профессиональных команд специалистов по информационной безопасности. Однако для организаций малого и среднего бизнеса подобные решения зачастую оказываются недоступными из-за высокой стоимости внедрения и сопровождения. В результате формируется противоречие между объективной необходимостью обеспечения высокого уровня безопасности и ограниченностью финансовых и кадровых ресурсов, что обусловливает актуальность разработки более простых, масштабируемых и экономически эффективных систем защиты информации.

1.2. Понятия угроз, уязвимостей и рисков в теории информационной безопасности.

Фундаментальной основой проектирования систем защиты информации является формализация понятий угроз, уязвимостей и рисков. Угроза

информационной безопасности определяется как потенциальная возможность нарушения конфиденциальности, целостности или доступности информации вследствие воздействия внешних или внутренних факторов. Угроза может быть обусловлена как преднамеренными действиями злоумышленников, так и случайными ошибками пользователей, программными сбоями или некорректной конфигурацией систем.

Уязвимость представляет собой слабое место в архитектуре информационной системы, программном обеспечении, сетевой инфраструктуре или организационных процессах, которое может быть использовано для реализации угрозы. С инженерной точки зрения уязвимость является следствием компромисса между функциональностью, сложностью и требованиями безопасности. Чем сложнее система, тем выше вероятность наличия скрытых уязвимостей, что делает задачу их выявления и устранения крайне сложной.

Риск информационной безопасности рассматривается как количественная характеристика, отражающая вероятность реализации угрозы с учётом наличия уязвимостей и потенциального ущерба. В теоретическом смысле риск может быть представлен как функция вероятности атаки и величины возможных потерь. Такой подход лежит в основе современных стандартов управления информационной безопасностью и позволяет рассматривать систему защиты информации как инструмент управления рисками.

С инженерной точки зрения управление рисками предполагает необходимость приоритизации угроз и оптимального распределения ресурсов системы защиты. В условиях ограниченного бюджета невозможно обеспечить максимальный уровень защиты по всем направлениям, поэтому возникает необходимость выбора наиболее критичных направлений защиты и внедрения механизмов, обеспечивающих наибольший эффект при минимальных затратах. Данный подход является ключевым для проектирования систем защиты информации, ориентированных на практическое применение в условиях малого и среднего бизнеса.

1.3. Модель Kill Chain как методологическая основа анализа атак

Для систематизации представлений о процессе атаки в теории информационной безопасности широко используется модель Kill Chain. Данная модель описывает атаку как последовательность взаимосвязанных этапов, каждый из которых подготавливает условия для реализации последующих действий злоумышленника. Рассмотрение атаки в виде цепочки этапов позволяет выявить наиболее уязвимые точки системы и определить оптимальные направления защиты.

Первым этапом Kill Chain является разведка, в ходе которой злоумышленник осуществляет сбор информации о целевой системе. На данном этапе анализируются сетевые интерфейсы, используемые технологии, структура веб-приложений, параметры конфигурации и возможные уязвимости. В современных условиях разведка всё чаще автоматизируется с использованием специализированных инструментов сканирования и интеллектуальных алгоритмов обработки данных. Результатом данного этапа является формирование модели целевой системы и определение потенциальных векторов атаки.

Следующий этап связан с подготовкой средств атаки, в рамках которого создаются или адаптируются эксплойты, вредоносные программы и сценарии воздействия. Данный этап предполагает использование информации, полученной на этапе разведки, и может включать как ручную разработку инструментов, так и автоматизированное формирование атакующих сценариев. В условиях применения интеллектуальных методов данный процесс может быть частично или полностью автоматизирован.

На этапе доставки осуществляется непосредственное воздействие на целевую систему, например, путём отправки специально сформированных запросов, внедрения вредоносного кода или эксплуатации выявленных уязвимостей. Далее следует этап эксплуатации, в ходе которого злоумышленник получает доступ к ресурсам системы и закрепляется в ней. После этого осуществляется установка вредоносных компонентов, обеспечивающих

устойчивое присутствие в системе, и организация каналов управления и контроля.

Заключительным этапом Kill Chain является достижение цели атаки, которое может выражаться в компрометации данных, нарушении функционирования системы, получении финансовой выгоды или иных формах ущерба. Анализ данной модели показывает, что каждый этап атаки является критически важным, однако наиболее эффективной стратегией защиты является предотвращение или усложнение ранних этапов, прежде всего этапа разведки. С инженерной точки зрения защита на ранних этапах позволяет существенно снизить вероятность успешной атаки и уменьшить потенциальный ущерб.

1.4. Системы защиты информации в условиях малого и среднего бизнеса

Особую специфику внедрения систем защиты информации демонстрируют организации малого и среднего бизнеса. В отличие от крупных корпораций, такие организации, как правило, не обладают значительными финансовыми и кадровыми ресурсами для построения многоуровневых систем безопасности. Часто функции информационной безопасности выполняются специалистами общего профиля, а выбор средств защиты определяется доступностью и стоимостью решений, а не стратегическими соображениями.

В инженерной практике это приводит к тому, что системы защиты информации в организациях малого и среднего бизнеса носят фрагментарный характер. Используются базовые механизмы защиты, такие как межсетевые экраны, антивирусные средства и стандартные механизмы аутентификации, однако отсутствует комплексный подход к анализу угроз и управлению рисками. При этом рост автоматизированных атак и использование интеллектуальных методов злоумышленниками приводит к тому, что традиционные средства защиты оказываются недостаточно эффективными.

Экономический аспект внедрения систем защиты информации включает

не только стоимость лицензий и оборудования, но и затраты на внедрение, настройку и сопровождение решений. Многие современные коммерческие продукты требуют привлечения квалифицированных специалистов, что существенно увеличивает совокупную стоимость владения. В результате организации малого и среднего бизнеса оказываются в ситуации, когда они вынуждены выбирать между высоким уровнем безопасности и экономической целесообразностью.

Данное противоречие определяет необходимость разработки систем защиты информации, ориентированных на практическое применение в условиях ограниченных ресурсов. Такие системы должны сочетать относительную простоту архитектуры, возможность автоматизации процессов анализа и адаптации к изменяющимся условиям угроз, а также приемлемый уровень затрат на внедрение и эксплуатацию.

1.5. Интеллектуальные методы в современных системах защиты информации

Развитие технологий искусственного интеллекта и машинного обучения существенно изменило подходы к анализу данных и обеспечению информационной безопасности. В отличие от традиционных систем защиты, основанных на статических правилах и сигнатурах, интеллектуальные методы позволяют выявлять скрытые закономерности в данных и обнаруживать ранее неизвестные атаки.

Искусственный интеллект в контексте информационной безопасности может рассматриваться как совокупность алгоритмов и моделей, способных анализировать большие объёмы данных, выявлять аномалии и принимать решения в условиях неопределённости. Машинное обучение, являясь одним из ключевых направлений искусственного интеллекта, предполагает построение моделей, обучаемых на исторических данных и способных обобщать полученный опыт для анализа новых событий.

С инженерной точки зрения применение методов машинного обучения в системах защиты информации позволяет перейти от реактивной модели защиты, основанной на обнаружении известных атак, к проактивной модели, ориентированной на выявление потенциальных угроз. Это особенно важно в условиях автоматизации атак, когда злоумышленники используют алгоритмы генерации запросов и адаптивные методы обхода традиционных средств защиты.

В теории машинного обучения различают методы обучения с учителем, без учителя и полууправляемые методы. Обучение с учителем предполагает использование размеченных данных, в которых каждому объекту сопоставляется класс, например легитимный или вредоносный запрос. Методы обучения без учителя ориентированы на выявление структурных закономерностей и аномалий без предварительной разметки данных. Полууправляемые методы сочетают элементы обоих подходов и позволяют использовать ограниченные объёмы размеченных данных.

В контексте анализа сетевого и прикладного трафика наиболее перспективным является использование методов бинарной классификации, позволяющих оценивать вероятность того, что наблюдаемое событие относится к классу атакующих воздействий. При этом важным аспектом является интерпретируемость моделей, поскольку системы защиты информации должны обеспечивать не только обнаружение угроз, но и возможность объяснения принятых решений.

1.6. Виды искусственного интеллекта и области их применения

Развитие искусственного интеллекта как научного направления сопровождалось формированием различных концептуальных подходов, отражающих разные представления о природе интеллекта и способах его моделирования. В зависимости от используемых теоретических оснований, методов обработки информации и способов представления знаний искусственный интеллект может быть представлен в виде нескольких

взаимосвязанных, но концептуально различающихся направлений.

Одним из наиболее ранних и фундаментальных подходов является логический подход, основанный на моделировании процессов рассуждения с использованием формальных систем логики. В рамках данного подхода интеллект рассматривается как способность выполнять логические выводы на основе заданных фактов и правил. Теоретической основой логического подхода служат математическая логика, теория доказательств и формальные языки. Практическая реализация логического подхода была достигнута в системах логического программирования, в частности в языке Пролог, где программа представляется не в виде алгоритма, а в виде набора утверждений и правил вывода. Такой способ представления знаний позволяет строить системы, способные выполнять дедуктивные рассуждения и получать новые знания из уже известных фактов. Преимуществом логического подхода является высокая степень формализуемости и интерпретируемости результатов, однако его существенным ограничением является низкая адаптивность и зависимость от полноты и корректности заранее заданной базы знаний.

Развитием логического направления стал агентно-ориентированный подход, получивший активное развитие с начала 1990-х годов. В рамках данного подхода интеллект рассматривается как способность автономной системы достигать поставленных целей в изменяющейся среде. Интеллектуальная система в данном случае интерпретируется как агент, который воспринимает окружающую среду посредством сенсорных механизмов и воздействует на неё с помощью исполнительных средств. Агентно-ориентированный подход смещает акцент с формальных логических выводов на процессы принятия решений, планирования, адаптации и взаимодействия с внешней средой. Важным элементом данного направления являются алгоритмы поиска, оптимизации и управления поведением, позволяющие агенту выбирать наиболее эффективные стратегии действий. Агентные модели широко применяются в робототехнике, системах управления, моделировании сложных процессов и распределённых вычислительных системах.

В условиях усложнения задач и роста объёмов данных сформировался гибридный подход, предполагающий объединение различных методов искусственного интеллекта. Гибридные системы сочетают элементы символических моделей, статистических методов и нейронных сетей, что позволяет компенсировать ограничения каждого отдельного подхода. В таких системах логические правила могут использоваться для интерпретации результатов машинного обучения, а нейронные сети - для автоматического формирования знаний. С точки зрения инженерной реализации гибридные модели являются наиболее перспективными, поскольку они обеспечивают баланс между адаптивностью, точностью и объяснимостью принимаемых решений. Именно гибридные подходы рассматриваются как основа построения интеллектуальных систем в условиях высокой сложности и неопределённости.

Значительное место в структуре искусственного интеллекта занимает направление символьного моделирования мыслительных процессов. Оно ориентировано на формализацию когнитивных процессов человека и построение систем, способных решать задачи, не имеющие явного алгоритмического решения. В рамках данного направления развиваются методы доказательства теорем, планирования, теории игр, принятия решений и прогнозирования. Символьные системы позволяют представлять знания в абстрактной форме и использовать их для решения сложных задач, однако их эффективность ограничена сложностью формализации предметной области.

Отдельным направлением искусственного интеллекта является обработка естественного языка, направленная на анализ, понимание и генерацию текстов на естественных языках. Данное направление возникло на стыке лингвистики, информатики и когнитивных наук и в настоящее время является одной из наиболее динамично развивающихся областей искусственного интеллекта. Методы обработки естественного языка применяются в информационном поиске, машинном переводе, анализе текстов и диалоговых системах. Развитие данного направления привело к созданию больших языковых моделей, способных выполнять сложные операции с текстовой информацией и

моделировать элементы человеческого мышления.

Важным компонентом искусственного интеллекта является направление представления и использования знаний, включающее методы извлечения, структурирования и применения знаний. Инженерия знаний как научное направление возникла в связи с развитием экспертных систем, использующих специализированные базы знаний для принятия решений. В рамках данного направления решается задача преобразования данных в знания, что является одной из ключевых проблем интеллектуального анализа данных. Современные методы извлечения знаний включают как символические подходы, так и нейросетевые методы, позволяющие автоматически выявлять закономерности в больших массивах информации.

Центральное место в современном искусственном интеллекте занимает машинное обучение, представляющее собой совокупность методов построения моделей, способных обучаться на данных. Машинное обучение ориентировано на автоматическое выявление закономерностей и формирование прогнозов на основе эмпирических данных. В рамках данного направления выделяются методы обучения с учителем, обучения без учителя и обучения с подкреплением. Машинное обучение применяется в задачах классификации, регрессии, распознавания образов, анализа текстов и прогнозирования. С инженерной точки зрения машинное обучение является ключевым инструментом построения адаптивных систем, способных функционировать в условиях изменяющейся среды и неполноты информации.

Развитием машинного обучения стало биологическое моделирование искусственного интеллекта, основанное на имитации принципов функционирования живых систем. В рамках данного направления используются нейронные сети, генетические алгоритмы, эволюционные методы и ройевые модели. Биологически вдохновлённые подходы позволяют решать задачи, для которых традиционные алгоритмы оказываются неэффективными. Их особенностью является способность к адаптации, самоорганизации и оптимизации в сложных многомерных пространствах.

Тесную связь с искусственным интеллектом имеет робототехника, в рамках которой разрабатываются интеллектуальные системы управления движением, навигацией и взаимодействием с окружающей средой. Интеллектуальные роботы требуют интеграции различных направлений искусственного интеллекта, включая машинное обучение, обработку сенсорных данных и планирование действий. Робототехника демонстрирует практическую реализацию идей искусственного интеллекта в физическом мире и подтверждает его междисциплинарный характер.

Отдельным направлением является машинное творчество, связанное с моделированием процессов художественного и технического творчества. Искусственный интеллект используется для генерации музыки, текстов, изображений и технических решений. Несмотря на относительную новизну данного направления, оно демонстрирует способность интеллектуальных систем выходить за рамки строго формализованных задач и выполнять функции, традиционно относимые к человеческой деятельности.

Анализ различных направлений искусственного интеллекта показывает, что они не существуют изолированно, а образуют взаимосвязанную систему методов и подходов. Их пересечение обусловлено сложностью задач, решаемых интеллектуальными системами, и необходимостью интеграции различных методов обработки информации. В современных условиях развитие искусственного интеллекта всё чаще рассматривается в контексте синергии различных подходов, что отражает переход от узкоспециализированных моделей к комплексным интеллектуальным системам.

Таким образом, искусственный интеллект представляет собой многоуровневую совокупность научных направлений и инженерных решений, каждая из которых обладает собственной областью применения, преимуществами и ограничениями. Понимание структуры и классификации искусственного интеллекта является необходимым условием для анализа возможностей его применения в системах защиты информации, поскольку выбор конкретного подхода определяется не только техническими характеристиками,

но и особенностями решаемых задач, доступными ресурсами и требованиями к надежности и интерпретируемости решений.

1.7. Обоснование выбора направления защиты веб-приложений

Веб-приложения являются одним из наиболее уязвимых элементов современной информационной инфраструктуры. Их публичная доступность, сложная архитектура и тесная интеграция с базами данных и внешними сервисами делают их привлекательной целью для злоумышленников. В отличие от внутренних корпоративных систем, веб-приложения постоянно находятся в зоне взаимодействия с внешней средой, что существенно увеличивает поверхность атаки.

С инженерной точки зрения защита веб-приложений представляет собой универсальное направление обеспечения информационной безопасности, актуальное для большинства организаций независимо от масштаба и специфики деятельности. В отличие от систем защиты электронной почты, которые не всегда находятся под контролем организации, или систем защиты внутренних сетей, требующих сложной инфраструктуры, защита веб-приложений может быть реализована непосредственно на уровне прикладного взаимодействия.

Анализ современных угроз показывает, что значительная часть атак начинается с анализа HTTP-трафика и попыток выявления уязвимостей веб-приложений. Автоматизированные инструменты сканирования и интеллектуальные алгоритмы позволяют злоумышленникам проводить разведку и формировать атакующие запросы в автоматическом режиме. В таких условиях система защиты, ориентированная на анализ веб-трафика, становится ключевым элементом противодействия атакам на ранних этапах Kill Chain.

1.8. Интеллектуальные методы анализа трафика как основа

адаптивных систем защиты

Применение интеллектуальных методов анализа трафика позволяет рассматривать процесс взаимодействия пользователя с веб-приложением как последовательность событий, обладающих определёнными статистическими и поведенческими характеристиками. В отличие от традиционных подходов, основанных на анализе отдельных запросов, интеллектуальные методы позволяют учитывать контекст взаимодействия, динамику поведения и совокупность признаков.

С инженерной точки зрения это означает возможность построения адаптивных систем защиты, способных изменять стратегию реагирования в зависимости от характеристик трафика и поведения пользователей. Такие системы могут выявлять аномалии, определять подозрительные паттерны и формировать вероятностные оценки угроз в реальном времени.

В условиях ограниченного бюджета особое значение приобретает выбор моделей машинного обучения, обеспечивающих достаточную точность при минимальных требованиях к вычислительным ресурсам. Практика показывает, что классические методы машинного обучения обладают рядом преимуществ по сравнению с глубокими нейронными сетями, включая интерпретируемость, меньшую вычислительную сложность и возможность обучения на относительно небольших выборках данных. Это делает их более пригодными для внедрения в системах защиты информации, ориентированных на практическое применение в условиях ограниченных ресурсов.

1.9. Угрозы информационной безопасности в условиях автоматизации и применения искусственного интеллекта

Современный этап развития информационных технологий характеризуется активным внедрением интеллектуальных методов как в системах защиты информации, так и в инструментах атаки. Злоумышленники

используют алгоритмы автоматизированного анализа, генерации запросов и адаптации атакующих сценариев, что существенно усложняет задачу обеспечения безопасности информационных систем.

Автоматизация атак приводит к увеличению их масштаба и снижению порога входа для злоумышленников. В результате даже относительно простые информационные системы становятся объектами массовых атак, направленных на выявление уязвимостей и компрометацию данных. В таких условиях традиционные средства защиты, основанные на фиксированных правилах и сигнатурах, оказываются недостаточно эффективными.

1.10. Выводы по первой главе

В первой главе были рассмотрены теоретические и инженерные основы систем защиты информации, проанализированы базовые понятия угроз, уязвимостей и рисков, а также исследована модель Kill Chain как инструмент анализа атак на информационные системы. Проведённый анализ показал, что в условиях автоматизации атак и применения интеллектуальных методов злоумышленниками традиционные подходы к защите информации требуют существенного пересмотра.

С инженерной точки зрения наиболее эффективной стратегией обеспечения безопасности является защита на ранних этапах атаки, прежде всего на этапе разведки, когда злоумышленник взаимодействует с системой и формирует модель её функционирования. Применение интеллектуальных методов анализа трафика позволяет реализовать адаптивные системы защиты, способные выявлять потенциальные угрозы в условиях неопределённости и ограниченных ресурсов.

Таким образом, результаты теоретического анализа формируют методологическую основу для практической реализации интеллектуальной системы защиты веб-приложений, рассматриваемой в последующих главах данной дипломной работы.

2.ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ ИНФРАСТРУКТУРЫ ВЕБ-ПРИЛОЖЕНИЯ И ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ЗАЩИТЫ

2.1. Виды нейронных сетей

В рамках интеграции искусственного интеллекта в системы защиты информации существенную роль играет корректный выбор класса моделей. Под «нейронной сетью» в большинстве русскоязычных источников понимается математическая модель, состоящая из множества связанных между собой вычислительных элементов (нейронов), параметры связей (веса) которых подбираются в процессе обучения для минимизации выбранной функции потерь. Такая трактовка подчёркивает, что нейронная сеть является не «алгоритмом по правилам», а параметрической моделью, качество которой определяется (а) структурой архитектуры, (б) процедурой обучения и (в) качеством данных. Общие сведения о нейронных сетях, их происхождении и базовых понятиях приводятся в справочных русскоязычных источниках и учебных материалах [1].

В целях настоящего исследования целесообразно рассматривать нейронные сети как семейство архитектур, применимых к различным типам данных. В задачах защиты веб-приложений наибольшую практическую значимость имеют модели, способные работать либо с табличными признаками (статистические характеристики запросов), либо с последовательностями символов/токенов (строки параметров, журналирование), либо с комбинированными представлениями (объединение «ручных» признаков и эмбеддингов). При этом для эксплуатационного использования в составе СЗИ дополнительно учитываются требования к задержке принятия решения, стабильности в условиях дрейфа данных, интерпретируемости и вычислительной стоимости.

Многослойный персептрон (MLP) относится к базовым архитектурам прямого распространения. Он строится как последовательность слоёв: входного, одного или нескольких скрытых и выходного. Каждый нейрон слоя обычно

соединён со всеми нейронами следующего слоя, что обеспечивает высокую выразительную способность при достаточном объёме данных. MLP удобны при наличии фиксированного вектора признаков; в задачах ИБ таким вектором могут быть статистики по параметрам HTTP-запроса (длины, доли символов различных типов, энтропия), признаки маршрута и метода, а также агрегаты поведенческих метрик. В русскоязычных справочных источниках MLP описывается как обобщение однослойного персептрона и один из наиболее распространённых вариантов нейросетевых моделей [2].

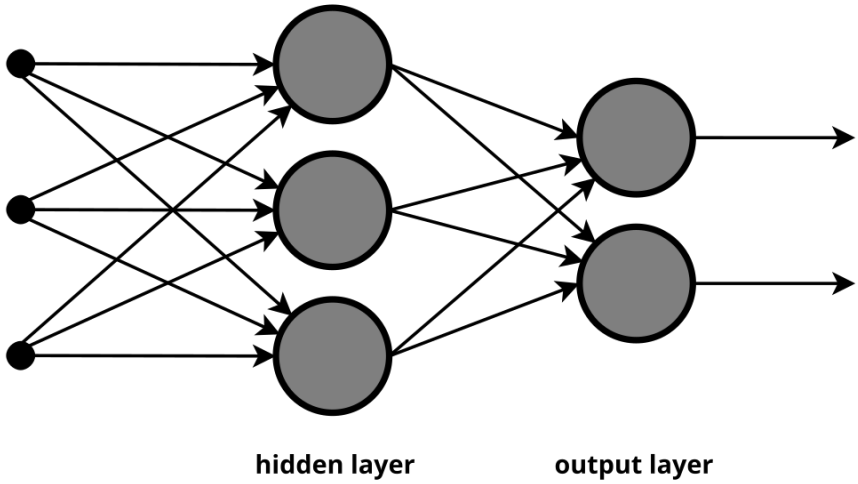

Рисунок 2.1 - Пример многослойной нейронной сети (MLP). Источник: Wikimedia Commons. Multi-Layer Neural Network-Vector (CC BY-SA). Дата обращения: 28.01.2026.

С точки зрения применения в СЗИ MLP обладают рядом преимуществ: они относительно просто реализуются, поддерживаются большинством библиотек, могут обучаться на умеренных объёмах данных и обеспечивают приемлемое качество в задачах бинарной классификации при корректной подготовке признаков. Вместе с тем MLP предъявляют более высокие требования к настройке по сравнению с классическими линейными моделями: требуется подбор числа слоёв и нейронов, функции активации, регуляризации и режима обучения. Кроме того, интерпретация решений MLP затруднена по сравнению с линейными моделями, что может снижать управляемость СЗИ в эксплуатации, когда необходимо объяснять причины блокировки запросов или корректировать политику реагирования.

Сверточные нейронные сети (CNN) представляют собой архитектуры, использующие операцию свёртки и локальные рецептивные поля. Идея свёртки позволяет выделять повторяющиеся локальные шаблоны в данных и формировать более устойчивые к сдвигам и вариациям признаки. CNN изначально развивались в компьютерном зрении, однако их одномерные варианты применяются для анализа последовательностей, включая строки и токенизированные тексты. В контексте СЗИ это означает возможность извлечения характерных подстрок и локальных закономерностей в параметрах запросов, которые могут быть полезны при распознавании автоматизированных атак. При этом следует учитывать, что CNN, будучи более выразительными, чем линейные модели, обычно требуют большего объёма данных и более внимательной настройки, а также увеличивают вычислительную нагрузку при предсказании по сравнению с простыми моделями.

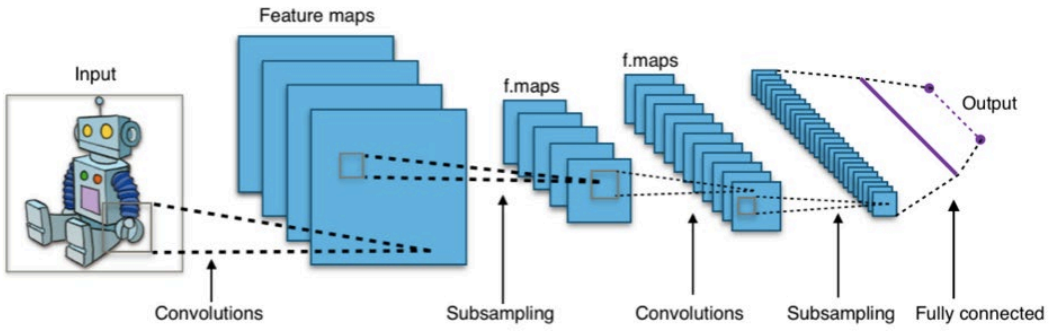

Рисунок 2.2 - Схема сверточной нейронной сети (CNN). Источник: Wikimedia Commons. Typical cnn (CC BY-SA). Дата обращения: 28.01.2026.

Рекуррентные нейронные сети (RNN) предназначены для обработки последовательностей переменной длины и учитывают порядок элементов. В русскоязычных источниках подчёркивается, что RNN вводят внутреннее состояние, позволяющее переносить информацию о предшествующих элементах последовательности [3]. Практически значимыми модификациями RNN являются LSTM и GRU, где используются управляющие механизмы («ворота») для более устойчивого обучения на длинных последовательностях. Для задач защиты веб-приложений RNN могут применяться, например, при моделировании последовательностей запросов пользователя, когда важен не только состав одного запроса, но и динамика поведения (частота, смена

маршрутов, характер параметров). Ограничением RNN является относительно высокая вычислительная стоимость обучения и сложность параллелизации по сравнению с архитектурами внимания, а также необходимость аккуратной подготовки последовательных данных и разметки.

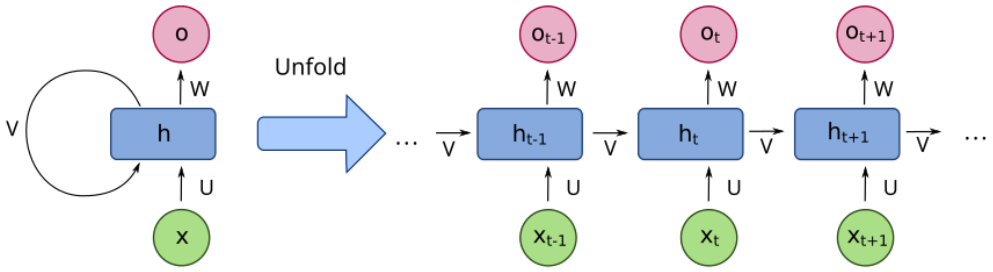

Рисунок 2.3 - Развёртывание RNN по времени. Источник: Wikimedia Commons. Recurrent neural network unfold (CC BY-SA). Дата обращения: 28.01.2026.

Архитектура Transformer основана на механизме внимания (attention) и в русскоязычных источниках рассматривается как один из ключевых подходов в современных задачах обработки последовательностей [4-5]. Принципиальная особенность трансформера - параллельная обработка последовательности с вычислением взаимосвязей между элементами через матрицы внимания. На базе трансформеров появились крупные языковые модели (LLM), которые демонстрируют высокое качество в задачах генерации и анализа текста. Для СЗИ потенциальные применения включают: семантический анализ текстовых журналов, нормализацию и классификацию описаний инцидентов, интеллектуальный поиск по базе знаний, поддержку аналитика ИБ. Однако в задачах принятия решений на сетевом периметре (например, блокировка HTTP-запроса в WAF) трансформеры часто оказываются избыточными: они требуют значительных вычислительных ресурсов, сложнее в сопровождении, а также предъявляют высокие требования к данным и контролю качества. Это особенно критично при эксплуатации на ограниченных ресурсах (1 CPU, 1 ГБ ОЗУ), когда приоритетом становится низкая задержка и предсказуемое потребление ресурсов.

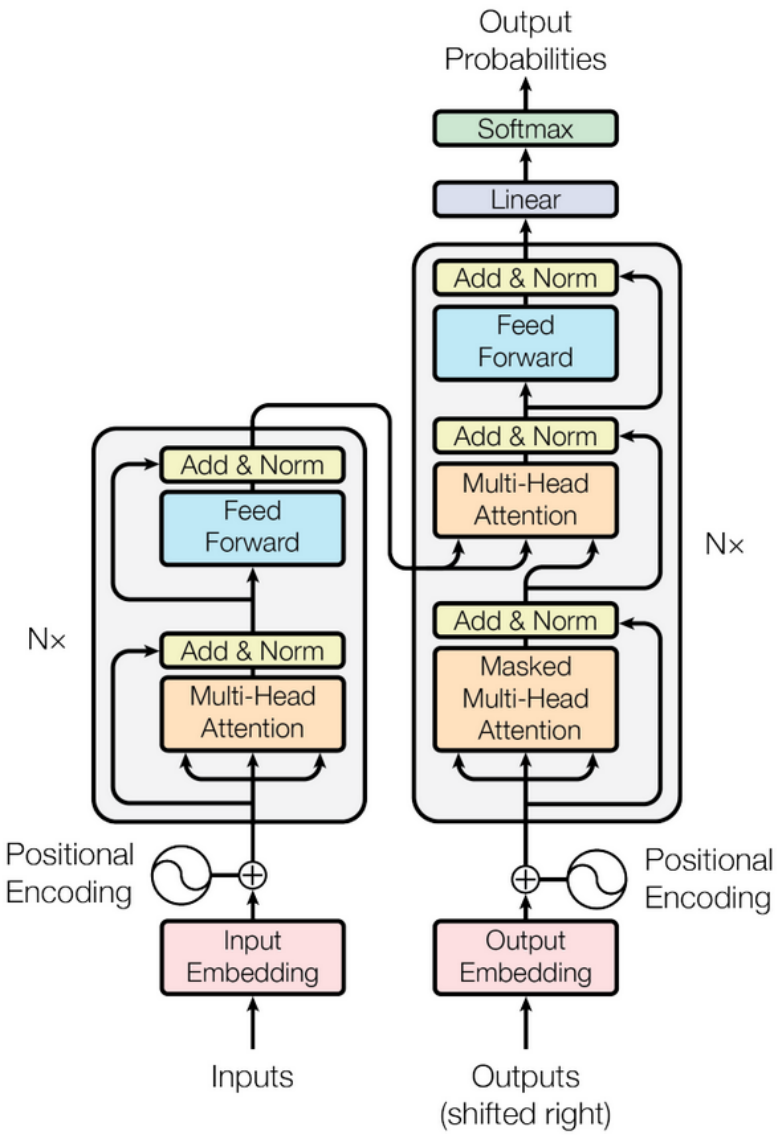

Рисунок 2.4 - Схема архитектуры Transformer. Источник: deepmachinelearning.ru. Трансформер: архитектура модели. Дата обращения: 28.01.2026.

В целом нейронные сети целесообразно рассматривать как мощный класс методов, однако их выбор в составе СЗИ должен опираться на компромисс между качеством, стоимостью эксплуатации и управляемостью. Для рассматриваемого варианта интеграции в WAF при ограниченном бюджете нейросетевые подходы изучались как возможные, но в результате сравнительного анализа была выбрана классическая модель машинного обучения - логистическая регрессия (см. п. 2.2).

2.2. Сравнение нейронных сетей и обоснование выбора логистической регрессии

Сравнение нейросетевых подходов с классическими методами машинного обучения в задачах WAF необходимо проводить не только по показателям качества классификации, но и по эксплуатационным ограничениям. Для прикладного межсетевого экрана веб-приложений значимы: задержка принятия решения (latency), пропускная способность при пиковых нагрузках, устойчивость к ложноположительным блокировкам, воспроизводимость обучения и простота интерпретации причин блокирования. В рассматриваемом варианте развертывания ограничение вычислительных ресурсов было задано конфигурацией виртуальной машины (1 vCPU и 1 ГБ ОЗУ), что существенно сужает спектр допустимых моделей и подходов.

Нейронные сети (MLP/CNN/RNN/Transformer) потенциально способны обеспечивать более высокое качество при работе с «сырыми» последовательностями (например, символными строками параметров), поскольку часть признакового представления извлекается автоматически. Однако при ограниченных ресурсах это преимущество часто не реализуется: обучение и инференс требуют большего объёма памяти, сложнее поддаются оптимизации, требуют обеспечения стабильности программного стека (CUDA и т. п., если рассматривается ускорение на GPU), а также требуют существенного внимания к контролю переобучения. Для WAF эксплуатационно важно иметь возможность быстрых переобучений по новым данным и детальной диагностики причин ухудшения качества; для сложных нейросетевых моделей это обычно требует отдельной исследовательской и вычислительной инфраструктуры.

Логистическая регрессия (Logistic Regression) относится к линейным моделям бинарной классификации. Она формирует оценку вероятности принадлежности объекта к классу атаки через логистическую функцию от линейной комбинации признаков. При наличии информативных признаков модель обеспечивает устойчивые результаты, обладает высокой скоростью обучения и предсказания, хорошо работает при умеренных объёмах данных и

поддерживает регуляризацию. Практически важным является и то, что в реализациях библиотек машинного обучения (например, scikit-learn) логистическая регрессия предоставляет стандартный интерфейс получения вероятностей (predict_proba) и механизмы учёта дисбаланса классов (class_weight) [9]. Для WAF это означает возможность задавать «порог блокировки» как настраиваемый параметр и проводить контролируемое управление риском ложных блокировок.

Сводное сравнение моделей приведено в табл. 2.1. Следует отметить, что значения показателей в таблице являются качественными оценками (категоризация «низкое/среднее/высокое») и предназначены для выбора класса методов, а не для замены экспериментальной валидации. Точные значения зависят от реализации, количества признаков и параметров модели, а также от характеристик потока запросов и инфраструктуры.

Таблица 2.1 - Сравнение классов моделей по критериям.

Критерий (для WAF)	Логистическая регрессия	MLP	CNN / RNN	Transformer / LLM
Задержка инференса на CPU	низкая	средняя	средняя-высокая	высокая
Потребление памяти (1 vCPU / 1 ГБ ОЗУ)	низкое	среднее	среднее-высокое	высокое
Требования к объёму данных	умеренные	средние	средние-высокие	высокие
Интерпретируемость решений	высокая (анализ весов признаков)	средняя	низкая	низкая
Скорость переобучения	высокая	средняя	средняя	низкая

Риск ложных блокировок при дрейфе данных	контролируемый порогами	средний	средний-высокий	высокий без тонкой настройки
Сложность сопровождения	низкая	средняя	средняя-высокая	высокая

На основании приведённого сравнения выбор логистической регрессии для WAF при ограниченном бюджете обоснован следующими обстоятельствами: (1) минимальные вычислительные требования и низкая задержка на CPU; (2) возможность получения вероятностной оценки и, следовательно, явного управления порогами блокировки; (3) высокая воспроизводимость обучения и устойчивость при небольших изменениях данных при условии регуляризации; (4) интерпретируемость, позволяющая анализировать вклад признаков в решение и формировать объяснения для администратора. Эти свойства напрямую согласуются с целью разработки доступного решения, требующего минимального участия специалистов по ИБ в постоянной эксплуатации.

2.3. Выбор библиотек и программного стека

Выбор программных библиотек и инструментов разрабатываемой системы должен следовать ранее сформулированным критериям: минимальные системные требования, переносимость, воспроизводимость, наличие устойчивой документации и широкого сообщества. Поскольку в п. 2.2 обоснован выбор логистической регрессии, предпочтительным является использование библиотеки, предоставляющей проверенную реализацию линейных моделей, поддерживающей типовые операции подготовки признаков и оценивания качества, а также позволяющей сохранять и загружать обученные модели без потери совместимости.

Для реализации модели машинного обучения использована библиотека

scikit-learn как де-факто стандартный инструментарий классического машинного обучения в экосистеме Python. Для решаемой задачи существенны следующие компоненты: (1) LogisticRegression, предоставляющая обучение модели и метод predict_proba для получения вероятностной оценки класса атаки; (2) DictVectorizer, позволяющий преобразовывать словарные признаки (feature-mapping) в разреженную матрицу признаков, что соответствует характеру признакового пространства HTTP-запросов (часть признаков может отсутствовать в конкретном событии); (3) инструменты разбиения выборки (train_test_split) и вычисления метрик (classification_report, roc_auc_score) [9]. Выбор этих модулей обеспечивает воспроизводимый цикл обучения и оценивания модели без необходимости внедрения тяжёлых фреймворков глубокого обучения.

Для реализации сервиса, осуществляющего перехват и маршрутизацию HTTP-трафика, использован веб-фреймворк FastAPI, основанный на ASGI и поддерживающий асинхронную обработку запросов. Существенными для настоящей работы являются: поддержка маршрутизации для всех методов и путей, типизированная модель Request/Response, удобная интеграция с Python-кодом извлечения признаков и возможность развёртывания через ASGI-сервер (uvicorn) [10].

Для проксирования запросов к защищаемому приложению применена библиотека httpx, предоставляющая асинхронный HTTP-клиент, включая поддержку транспорта через UNIX-socket (UDS), что позволяет организовать взаимодействие с backend-приложением, развёрнутым через сервер приложений (например, gunicorn) без дополнительного TCP-порта [11]. Такой подход упрощает конфигурацию и снижает поверхность атаки за счёт локального межпроцессного взаимодействия.

Для сериализации и хранения модели и векторизатора использован joblib (как часть типичного стека scikit-learn), что обеспечивает быстрое сохранение/загрузку обученных объектов и повторяемость инференса. Для формата журналирования выбран JSONL (JSON Lines), поскольку он позволяет

хранить события построчно и удобно обрабатывается средствами потоковой обработки и скриптами подготовки датасетов.

Следует отметить ограничения выбранного стека. Во-первых, scikit-learn ориентирован на классические модели и не предоставляет встроенных средств обучения глубоких нейросетей; это соответствует принятым допущениям п. 2.2, но ограничивает дальнейшее расширение к нейросетевым подходам без смены библиотеки. Во-вторых, FastAPI и httpx требуют аккуратной настройки таймаутов и ограничений на размер тела запроса, поскольку WAF является высоконагруженным компонентом. Эти ограничения учитывались при реализации (например, ограничение MAX_BODY_BYTES и MAX_PARAMS в коде).

2.4. Разработка инфраструктуры и вычислительные ограничения

Развертывание экспериментальной инфраструктуры выполнялось в условиях ограниченного бюджета: отдельные коммерческие средства защиты в облачной среде не приобретались, а вычислительные ресурсы были ограничены конфигурацией виртуальной машины уровня 1 vCPU и 1 ГБ оперативной памяти. Такое ограничение типично для малых и средних организаций, которые размещают тестовые или малонагруженные сервисы на минимальных тарифах и стремятся избежать дополнительных постоянных затрат.

При данных ограничениях ключевым критерием выбора метода ИИ является предсказуемость ресурсоёмкости. Логистическая регрессия как линейная модель допускает выполнение инференса за время, линейное по числу ненулевых признаков, и не требует значительных объёмов памяти для хранения параметров. Это позволило интегрировать модель в сервис перехвата трафика без необходимости выделения отдельного узла под вычисления.

В качестве объекта защиты было подготовлено веб-приложение, реализующее базовый пользовательский интерфейс. Клиентская часть выполнена с использованием HTML и CSS. В рамках исследования данный шаг

рассматривается как создание необходимой прикладной поверхности для формирования HTTP-трафика и моделирования сценариев взаимодействия пользователя с системой. Отдельно подчёркивается, что качество и безопасность клиентской части не являлись предметом данного раздела; основные меры контроля и анализа сосредоточены на серверной стороне и на уровне входящего трафика.

2.5. Добавление аутентификации и создание контролируемой поверхности атаки.

Для проведения экспериментов по обнаружению атак и формированию обучающих данных была реализована система аутентификации. С точки зрения методологии исследования аутентификация является удобным компонентом, поскольку: (1) она обрабатывает пользовательский ввод (логин, пароль) в форме параметров запроса; (2) она взаимодействует с хранилищем данных; (3) она часто является целью атак на ранних стадиях компрометации, поскольку предоставляет точку входа в прикладную логику.

В рамках экспериментальной постановки в код аутентификации были намеренно оставлены уязвимые конструкции, позволяющие воспроизводимо моделировать атаки. Во-первых, в целях моделирования SQL-инъекций аутентификация выполнялась посредством формирования SQL-запроса в коде приложения, что создавало возможность влияния на структуру запроса через пользовательские параметры. Во-вторых, для обеспечения наблюдаемости признаков ввода логин и пароль фиксировались в журналировании в открытом виде. Данное допущение является методически оправданным для изолированной лабораторной среды, однако в реальных системах такая практика недопустима, поскольку повышает риск компрометации учётных данных; в промышленной эксплуатации следует применять маскирование, хеширование или криптографические средства защиты журналов.

Дополнительно для моделирования атак типа path traversal была

реализована функция просмотра файлов по пути, переданному пользователем. Это создавало возможность формировать запросы с последовательностями «../» и иными признаками обхода каталогов. Такое решение также не может применяться в реальном продукте без строгих ограничений, но для целей формирования датасета и проверки способности системы различать легитимные и вредоносные запросы оно позволило получить воспроизводимый поток атакующих событий.

Таким образом, модуль аутентификации выполнял двойную роль: обеспечивал базовую функциональность приложения и служил источником контролируемого трафика, содержащего как легитимные, так и атакующие запросы. Это требование непосредственно связано с задачей последующего обучения модели на данных, сформированных в условиях, приближенных к эксплуатации (см. п. 2.7-2.9).

2.7. Настройка логирования HTTP-событий для последующего обучения.

Для обучения модели машинного обучения необходимы структурированные данные, отражающие свойства наблюдаемых событий. В контексте защиты веб-приложений таким событием выступает HTTP-запрос, включающий метод, путь, параметры (query/body), а также контекстные характеристики. В рамках исследования логирование реализовано в формате JSONL (одна запись - одна строка JSON), что позволяет: (1) добавлять события в потоковом режиме без удержания всего массива данных в памяти; (2) удобно разделять логи по источникам; (3) использовать стандартные средства обработки JSON для подготовки датасета.

Содержательно журналирование строилось по принципу «минимально достаточной полноты»: фиксировались поля, необходимые для восстановления признаков (feats) и проверки решений модели (decision), а также служебные сведения для анализа качества (временная метка, IP-адрес, путь). Одновременно

вводились ограничения, предотвращающие избыточный рост журнала и потенциальное влияние на производительность (например, ограничение размера предварительного просмотра тела запроса).

Таблица 2.2 - Рекомендуемый состав полей события логирования HTTP-запроса для обучения модели и анализа решений WAF.

Поле	Тип	Назначение в датасете/эксплуатации
ts	float (Unix time)	Временная метка события, анализ последовательности/нагрузки
ip	string	Идентификатор источника запроса (корреляция, подсчёт частот)
method	string	HTTP-метод (GET/POST и др.), контекст для классификации
path	string	Путь запроса (URI), контекст и часть признаков
params	object	Параметры запроса; в лабораторной постановке без маскирования

body_preview	string	Фрагмент тела запроса (для диагностики), ограниченный по длине
feats	object	Словарь вычисленных признаков (числовые значения)
scores.p_attack	float	Вероятностная оценка принадлежности к классу атаки
decision	string	Решение модели (ALLOW/CHALLENGE/ BLOCK)
effective_decision	string	Реально применённое действие с учётом режима monitor/enforce
reasons	array	Поясняющие маркеры (например, превышение порога)

С точки зрения методологии подготовки данных важно, что логирование включает не только исходные свойства запроса, но и уже вычисленные признаки (feats). Это обеспечивает воспроизводимость обучения: при изменении кода извлечения признаков можно сравнивать версии, а также повторно обучать модель на исторических данных, используя тот же набор признаков. Однако такое решение накладывает ограничение: при существенном изменении набора признаков старые логи могут потребовать миграции или повторной генерации признаков, если они не сохранены.

Отдельно следует отметить этический и правовой аспект логирования. В

реальных информационных системах журналы могут содержать персональные данные и конфиденциальную информацию, поэтому к ним применяются требования по ограничению доступа, целостности и срокам хранения. В настоящей работе фиксирование логина/пароля в открытом виде рассматривается как лабораторное допущение для демонстрации механизма извлечения признаков; при практической эксплуатации необходимо применять маскирование чувствительных параметров и криптографическую защиту журналов.

2.8. Подготовка датасета для обучения модели.

После настройки логирования возникает вопрос формирования обучающей выборки. В литературе и практических обзорах по применению машинного обучения в ИБ отмечается, что качество модели критически зависит от репрезентативности данных: выборка должна отражать как легитимные сценарии, так и типовые атакующие действия для конкретного приложения и его поверхности атаки. В рамках настоящего исследования рассматривались два возможных подхода.

Первый подход - использование публичных датасетов (например, наборов HTTP-запросов или сетевых трасс). Его достоинство состоит в доступности и потенциальной широте охвата типов атак. Однако на практике возникает проблема доменной несогласованности: публичные датасеты часто собраны на иных приложениях, с иными параметрами и форматами запросов, что приводит к ухудшению переносимости модели. Дополнительно многие датасеты доступны только частично, требуют лицензирования или не содержат достаточного контекста для воспроизведения признаков, используемых конкретной реализацией WAF.

Второй подход - синтетическая генерация запросов и формирование собственного датасета на целевом приложении. Этот вариант обеспечивает согласованность формата данных и признакового пространства с внедряемым

WAF, но требует отдельного механизма генерации трафика и контроля разметки. В рамках работы был выбран второй подход, поскольку он обеспечивает воспроизводимость, позволяет целенаправленно создавать атакующие и легитимные сценарии, а также не предъявляет требований к большому объёму данных: выбранная модель (логистическая регрессия) может обучаться на относительно умеренном количестве примеров при условии информативных признаков (см. п. 2.2).

Для генерации запросов применимы стандартные библиотеки экосистемы Python: requests/httpx для отправки HTTP-запросов, urllib.parse для кодирования параметров, а также средства случайной генерации (random, secrets) для вариативности. В качестве подхода к генерации использовалось формирование наборов шаблонов: для легитимных запросов - типовые значения параметров (валидные логины, пароли, безопасные пути); для атакующих - конструкции, характерные для SQL-инъекций и обхода каталогов. Каждый сгенерированный запрос проходил через развернутую инфраструктуру, попадал в журнал WAF и тем самым становился частью будущего датасета.

Ограничением синтетического подхода является неполнота покрытия: генератор отражает лишь те сценарии, которые в него заложены. Поэтому при дальнейшем развитии системы целесообразно дополнять датасет реальными журналами эксплуатации (после применения маскирования чувствительных данных) и корректировать генератор на основе новых наблюдаемых шаблонов.

2.9. Нормализация датасета и разметка «хороших» и «плохих» событий.

После накопления журналов возникает методически важная задача: разделение событий на классы (легитимные и атакующие) с формированием «якоря» истинной метки (ground truth). В задачах ИБ это является одной из основных трудностей, поскольку в реальной эксплуатации истинная метка часто неизвестна и требует сложной экспертизы. В лабораторной постановке разметка

упрощается благодаря контролируемой генерации событий: заранее известно, какие запросы формировались как атакующие, а какие как легитимные.

Изначально все события логировались в единый файл. Однако такой формат неудобен для обучения, поскольку требует последующей разметки. В ходе работы был принят практический подход: при генерации событий дополнительно фиксировать признак источника (например, тип сценария), позволяющий однозначно отнести событие к «хорошему» или «плохому». Далее лог-файл разбивается на два файла: legit.jsonl (label = 0) и attack.jsonl (label = 1). Такой подход соответствует общему требованию к обучению моделей с учителем: наличие двух классов с контролируемыми метками.

Нормализация включает несколько шагов. Во-первых, исключаются неполные или повреждённые строки JSON, что возможно при аварийном завершении процесса записи или при конкурентной записи. Во-вторых, ограничиваются значения параметров по длине и количеству (для предотвращения влияния выбросов и потенциального отказа в обслуживании при обработке). В-третьих, проводится унификация признаков: при отсутствии некоторых признаков в конкретном событии они рассматриваются как нулевые, что корректно обрабатывается разреженной векторизацией (DictVectorizer). В-четвёртых, рекомендуется проверять баланс классов и, при необходимости, применять механизмы учёта дисбаланса (например, class_weight="balanced" в логистической регрессии) [9].

Ограничение данного подхода состоит в том, что разметка фактически совпадает с механизмом генератора. Следовательно, модель может научиться распознавать «почерк генератора», а не реальные атаки. Для снижения риска необходимо: (1) увеличивать вариативность запросов; (2) включать смешанные сценарии, близкие к реальной эксплуатации; (3) проводить тестирование на независимых наборах запросов, не использованных при обучении. Эти требования фиксируются как ограничения исследования и как направление развития системы.

2.10. Разработка WAF и архитектура сервиса.

WAF (Web Application Firewall) в общем виде трактуется как средство защиты, предназначенное для контроля и фильтрации HTTP(S)-трафика на уровне приложений. В отличие от сетевых межсетевых экранов, WAF учитывает прикладной контекст запроса (путь, параметры, семантика действий) и способен применять политики, ориентированные на угрозы веб-приложений. В методических материалах OWASP подчёркивается, что WAF может использоваться как дополнительный уровень защиты, особенно для снижения рисков эксплуатации уязвимостей до устранения первопричины [12].

В настоящей работе WAF реализован как отдельный сервис, функционирующий в режиме обратного прокси. Такой режим означает, что клиент обращается к WAF, а WAF, приняв решение, либо проксирует запрос к backend-приложению, либо возвращает ответ блокировки/челленджа. Выбор отдельного сервиса (вместо встраивания в модуль аутентификации) обусловлен требованиями эксплуатационной устойчивости: ошибки в защитном механизме не должны приводить к отказу бизнес-функций приложения, а изменение модели и признаков должно выполняться независимо от кода приложения.

Архитектура сервиса соответствует текущей реализации (FastAPI + httpx + ML-модуль scikit-learn). В табл. 2.3 приведена логическая структура модулей WAF и их функции.

3. МЕТОДИКА ПРИМЕНЕНИЯ, ЭКСПЕРИМЕНТАЛЬНАЯ ВАЛИДАЦИЯ И ЭКОНОМИЧЕСКАЯ ОЦЕНКА ЭФФЕКТИВНОСТИ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ЗАЩИТЫ

3.1. Методика применения интеллектуальной системы защиты в условиях эксплуатации веб-приложений.

Эксплуатация разработанной интеллектуальной системы защиты веб-приложений (далее - ИСЗ) рассматривается как регламентированный цикл, включающий (а) сбор событий о входящих HTTP-запросах; (б) подготовку обучающих наборов данных; (в) переобучение модели; (г) проверку качества и выпуск новой версии модели; (д) контролируемое применение политики фильтрации запросов. Подход соответствует общему принципу управляемости защитных механизмов: изменения должны быть воспроизводимы, поддаваться аудиту и иметь возможность отката (версии моделей и конфигураций).

На прикладном уровне ИСЗ реализована как reverse-proxy WAF: запрос поступает на сервис WAF, где происходит извлечение параметров и признаков (features), вычисление оценки риска p_attack и выбор действия (ALLOW/CHALLENGE/BLOCK) с учётом режима работы (monitor/enforce). Такая схема сопоставима с архитектурой проксирующих WAF и обеспечивает единый «контрольный пункт» до передачи запроса в backend.

В отличие от правил (rule-based) WAF, где безопасность формально задаётся набором правил и исключений (пример - ModSecurity с OWASP Core Rule Set, CRS) [1-3], в ИСЗ часть политики замещается обучаемой моделью. Практическое следствие состоит в том, что актуализация защиты достигается не столько ручным дописыванием и согласованием правил, сколько корректировкой обучающей выборки и переобучением модели по регламенту. На небольшом стенде это позволяет снизить порог входа для сопровождения: при наличии инструкций базовый ИТ-специалист (администратор/DevOps) может выполнить типовой цикл переобучения (перенос событий в обучающие

файлы, запуск скрипта, проверка метрик, публикация модели).

При этом облегчение процедур сопровождения не устраняет риски, связанные с корректностью обучения. Некорректная разметка или попадание «ложно-нормальных» атакующих событий в легитимный набор может привести к снижению детектирующей способности (в пределе - к пропуску опасных запросов). Следовательно, методика эксплуатации должна включать контроль качества обучения: (1) фиксированный протокол разметки и отбора событий; (2) проверку метрик на отложенной выборке; (3) ручной выборочный аудит примеров, повлиявших на обучение; (4) журналирование версий модели/векторизатора и настроек порогов. Данный тезис согласуется с общими рекомендациями по управлению ML-системами в критичных контекстах и с практикой MLOps (версионирование артефактов, контроль дрейфа данных).

Следует отдельно указать ограничение: как обучаемые системы, так и rule-based WAF уязвимы к инсайдерским воздействиям. В случае rule-based WAF злоумышленник может ослабить правила или добавить исключения; в случае ИСЗ - повлиять на обучающую выборку и метрики. Поэтому процедуры контроля доступа и аудита действий администраторов должны рассматриваться как общие меры безопасности независимо от типа WAF.

3.2. Методика экспериментальной проверки и валидации результатов работы системы защиты.

Экспериментальная валидация выполнена в форме сравнительного анализа двух стендов, реализующих сопоставимые условия генерации и обработки HTTP-запросов. Стенд A: Nginx + ModSecurity (v3) + OWASP Core Rule Set (CRS). ModSecurity - широко распространённый открытый модуль WAF, работающий на основе правил; CRS - открытый набор правил, предназначенный для обнаружения типовых классов атак (SQLi, XSS, LFI/RFI и др.) [1-3]. Стенд B: самописный WAF (reverse-proxy) с моделью машинного обучения и пороговой политикой блокирования (WAF+ML).

В описании стенда А использовались официальные и учебно-методические материалы по ModSecurity v3 и CRS, включая русскоязычные источники, описывающие общие принципы настройки правил и механизмы обработки запросов [2-3]. Эти материалы фиксируют, что ModSecurity выполняет анализ трафика на уровне правил (включая регулярные выражения и условия по параметрам запроса) и может работать совместно с Nginx/Apache как модуль/прокси-компонент [1-3].

Генерация тестового трафика осуществлялась заранее подготовленным скриптом, формирующим запросы с паттернами, характерными для эксплуатации и разведки уязвимостей веб-приложений. Набор воздействий включал: SQL Injection, XSS, Path Traversal, Command Injection, SSRF, LFI/RFI, XXE, Template Injection, CRLF Injection, Header Injection, Open Redirect, NoSQL Injection, Bruteforce, Mixed Attacks, Rate Limit Bypass, Sensitive Endpoint Discovery. Данный набор отражает широко описанные в профильной литературе классы атак прикладного уровня.

Результат каждого запроса фиксировался по коду ответа HTTP. В качестве ключевого прокси-показателя использовалась доля ответов 200 (успешная обработка запроса приложением), поскольку в стендовом тесте она интерпретируется как индикатор потенциального «прохождения» воздействия через контур защиты. Ограничение методики состоит в том, что ответ 200 не гарантирует успешную эксплуатацию уязвимости; однако при сравнении двух средств фильтрации трафика на одинаковом прикладном контуре снижение доли 200 является разумным индикатором усиления барьера на прикладном уровне.

Дополнительно анализировались доли 403 (явная блокировка), 404 (ресурс не найден), а также другие 4xx/5xx. При интерпретации результатов учитывалось, что 404 снижает информативность разведки, но не является блокировкой и не всегда приводит к идентификации источника как нарушителя.

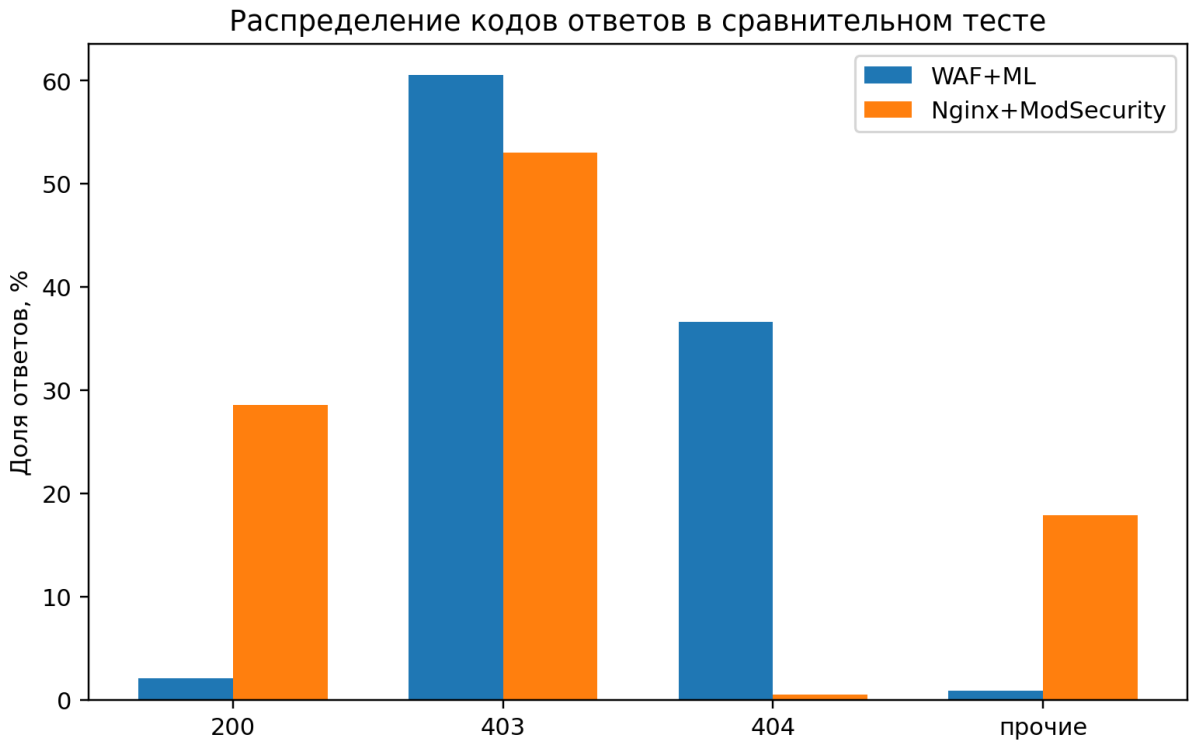

Рисунок 3.1 - Распределение кодов ответов по результатам сравнительного тестирования

3.4. Анализ точности, полноты и устойчивости модели машинного обучения в задачах выявления угроз

По результатам тестового прогона получены следующие распределения кодов ответов. Для WAF+ML: 200 - 78 запросов (2,1%), 403 - 2281 (60,5%), 404 - 1382 (36,6%), 405 - 25 (0,7%), ошибка выполнения - 6 (0,2%). Для Nginx+ModSecurity: 200 - 1079 (28,6%), 403 - 2001 (53,0%), 404 - 18 (0,5%), 400 - 6 (0,2%), 401 - 25 (0,7%), 500 - 643 (17,0%).

Классические метрики качества классификации (Precision, Recall, F1, ROC-AUC) требуют наличия истинной разметки по каждому запросу («атака/не атака») и факта корректного решения (должен ли запрос быть заблокирован в рамках политики). В тестовом сценарии истинная разметка частично задаётся генератором (шаблоны атак). Однако существуют методологические тонкости: часть запросов может быть направлена на несуществующие ресурсы (что

приводит к 404) и при этом оставаться атакующим по намерению; кроме того, в реальной эксплуатации некоторые «подозрительные» конструкции могут встречаться в легитимном трафике (например, при поисковых запросах). Поэтому в данном исследовании акцент делается на сопоставимых прокси-показателях (доли кодов) и на устойчивости (обобщающая способность) модели к классам воздействий.

С точки зрения барьера на прикладном уровне WAF+ML демонстрирует более низкую долю 200 и более высокую долю 403, что интерпретируется как более жёсткая фильтрация тестового трафика. При этом выявлена существенная доля 404 (36,6%), что указывает на сценарии, где запрос не достиг ресурса, но и не был явно заблокирован. Для практики это означает необходимость уточнения политики (например, добавления корреляции событий по источнику и усиления реакции на серию подозрительных запросов) и дообучения модели на расширенном спектре атакующих классов.

Устойчивость модели проявилась в способности блокировать ряд запросов, относящихся к классам, которые не доминировали в обучающей выборке. Такое поведение объясняется тем, что часть признаков описывает общие статистические свойства «нелегитимного» ввода (доля специальных символов, энтропия строк, характерные последовательности обхода путей), которые встречаются в различных классах атак. Однако данное обобщение не гарантировано: для повышения полноты детектирования требуется увеличение репрезентативности обучающего набора и контроль дисбаланса классов.

3.5. Сравнительный анализ разработанной системы с традиционными средствами защиты информации

Стенд ModSecurity/CRS представляет rule-based подход: решение о блокировании определяется срабатыванием правил и их комбинаций. Документация CRS подчёркивает назначение набора правил как базового уровня защиты от распространённых классов атак, с возможностью настройки

исключений для снижения ложных срабатываний [3]. Стенд WAF+ML реализует вероятностный подход: модель вычисляет p_attack и далее применяется пороговая политика (P_BLOCK, P_CHALLENGE). Такой подход позволяет управлять строгостью фильтрации через пороги, но переносит часть «семантики правил» в обучающие данные и процедуру переобучения.

Результаты тестирования показывают, что при выбранных порогах WAF+ML снижает долю 200 значительно сильнее, чем ModSecurity/CRS, и увеличивает долю 403. Это интерпретируется как более выраженная способность подавлять тестовые воздействия. Однако меньшая доля 404 у ModSecurity и значительная доля 500 у ModSecurity в эксперименте требуют отдельного анализа конфигурации стенда A: появление 500 может свидетельствовать о побочных эффектах обработки запросов приложением/модулем либо о реакции правил, приводящей к ошибкам. Поэтому сравнительный вывод корректен в рамках заданной конфигурации стендов и должен сопровождаться описанием условий и версий компонентов.

Отмеченный недостаток WAF+ML - высокая доля 404 - указывает на необходимость усиления механизмов явного пресечения разведки. Одним из направлений является добавление признаков, учитывающих частотные и последовательностные характеристики (rate/sequence), что приближает систему к IDS/IPS-подходам на основе поведенческих паттернов.

3.6. Экономическая оценка разработки, внедрения и сопровождения интеллектуальной системы защиты

В рамках экономической оценки (п. 3.6) выполняется сопоставление трудозатрат и кадровых предпосылок эксплуатации разработанной интеллектуальной системы защиты с альтернативой в виде традиционного WAF, настраиваемого правилами. Рассмотрение ограничено условиями организаций малого и среднего бизнеса и типовой облачной инфраструктуры; в расчёт включаются прямые расходы на вычислительные ресурсы (3000 руб./мес. на

виртуальную машину) и затраты на человеческие ресурсы, связанные с сопровождением системы.

Классические WAF (включая ModSecurity) требуют регулярной актуализации правил и экспертной интерпретации журналов событий по классам атак. Такой профиль работ относится к прикладной ИБ и обычно выполняется специалистом по безопасности. В разработанном подходе часть операций сопровождения формализуется: администратор обновляет обучающие файлы и переобучает модель по регламенту.

Анализ кадрового рынка в сфере информационной безопасности и эксплуатации ИТ-систем в Российской Федерации на конец 2025 года показывает устойчивую тенденцию роста спроса на высококвалифицированных специалистов по информационной безопасности. Согласно данным исследования Positive Technologies и Центра стратегических разработок, дефицит кадров в сфере информационной безопасности к 2027 году может достигнуть десятков тысяч специалистов, при этом наибольший дефицит наблюдается в сегменте экспертов и инженеров высокого уровня квалификации. Рост спроса на таких специалистов закономерно приводит к увеличению уровня их заработных плат и совокупной стоимости владения традиционными средствами защиты информации.

Одновременно с этим анализ рынка труда показывает относительный избыток начинающих специалистов в области информационных технологий (уровня junior), а также специалистов эксплуатационного профиля (инженеров эксплуатации и DevOps). Данные категории специалистов характеризуются более низким уровнем заработной платы и более высокой доступностью на рынке труда по сравнению с узкопрофильными специалистами по информационной безопасности. Это обстоятельство формирует объективные экономические предпосылки для внедрения систем защиты информации, эксплуатация которых не требует постоянного участия высококвалифицированных экспертов ИБ.

Разработанная интеллектуальная система защиты, основанная на

использовании методов машинного обучения, в меньшей степени зависит от ручной настройки правил и экспертных знаний в области информационной безопасности. Процедуры сопровождения системы сводятся преимущественно к операциям анализа журналов событий, подготовке обучающих выборок и запуску процедур переобучения модели, что может быть выполнено специалистами эксплуатационного профиля или разработчиками начального уровня квалификации. В отличие от традиционных WAF-систем, требующих постоянной корректировки сигнатур и правил специалистами ИБ, предложенный подход снижает зависимость организации от дефицитных кадровых ресурсов.

Таким образом, с учётом выявленных тенденций кадрового рынка можно сделать вывод о том, что внедрение интеллектуальной системы защиты информации обладает не только технологическими, но и экономическими преимуществами. Использования более доступных кадровых ресурсов и автоматизация процессов анализа угроз обеспечивают более низкую совокупную стоимость владения системой защиты по сравнению с традиционными решениями. В условиях прогнозируемого роста потребности квалифицированных специалистов ИБ данный фактор приобретает стратегическое значение для организаций, особенно относящихся к сегменту малого и среднего бизнеса.

Потребность рынка труда в сфере ИБ в разрезе функциональных групп работников в 2023 и 2027 годах, тыс. чел.

Рисунок 3.6.1 - Необходимые компетенции ИБ-специалистов по данным Positive Technologies

Делегирование части операций сопровождения возможно при наличии доступного кадрового ресурса. По аналитическим материалам HeadHunter конкуренция среди Junior-кандидатов в IT выше, чем среди Senior, что отражает большее число резюме на одну вакансию. Это позволяет поручать регламентные действия инженеру эксплуатации или сотруднику начального уровня под контролем ответственного лица.

Рисунок 3.6.2 иллюстрирует различия конкуренции для уровней Junior и Senior (график построен по данным HeadHunter). Ограничение интерпретации: показатель конкуренции не характеризует качество кандидатов. Поэтому изменения, влияющие на безопасность (например, пороги блокировки и состав обучающих данных), требуют экспертного контроля.

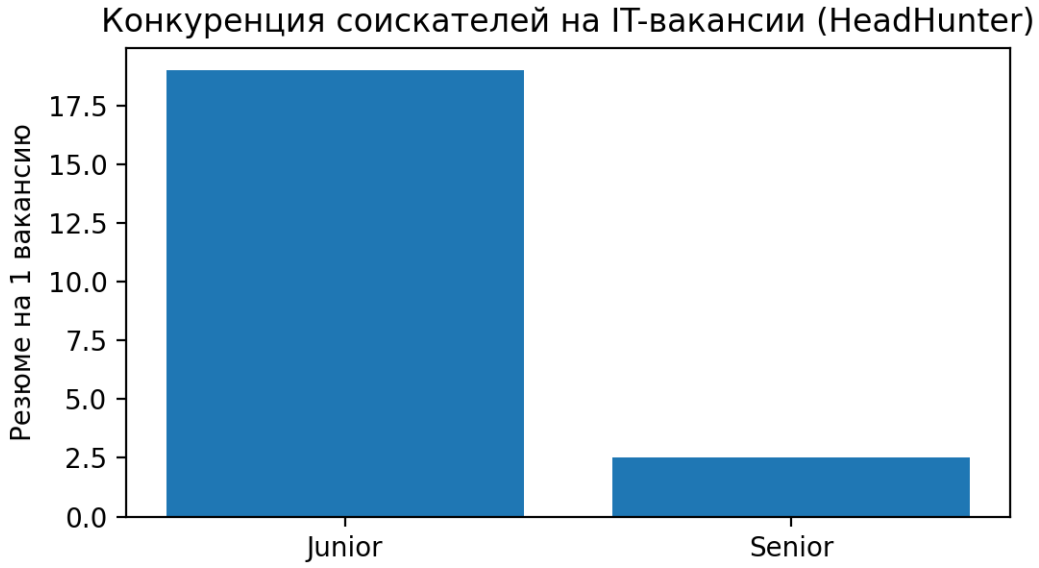

Рисунок 3.6.2 - Конкуренция соискателей на IT-вакансии (резюме на 1 вакансию), по данным HeadHunter

С учётом указанных факторов в качестве базовой роли сопровождения интеллектуальной системы в последующих расчётах (пп. 3.7-3.8) рассматривается инженер эксплуатации, выполняющий регламентные операции в рамках сопровождения веб-инфраструктуры. Функции экспертного контроля (оценка качества модели, контроль корректности обучающих данных и порогов принятия решений) целесообразно рассматривать как периодическую нагрузку специалиста по информационной безопасности, привлекаемого по мере необходимости.

Источники для п. 3.6: публикации Лаборатории Касперского о дефиците компетенций ИБ; аналитические материалы HeadHunter о конкуренции на рынке IT-вакансий (уровни Junior/Senior).

3.7. Расчёт совокупной стоимости владения системой защиты с учётом человеческих ресурсов и инфраструктуры

Совокупная стоимость владения (TCO) в упрощённой модели может быть представлена как сумма инфраструктурных затрат и затрат на труд сопровождения: TCO = C_infra + C_labor. В условиях стенда C_infra фиксирована

(3000 руб./мес.). C_labor оценивается через долю рабочего времени ответственного сотрудника (FTE-доля) и его месячную стоимость.

Рынок труда ИТ демонстрирует различия по сегментам. Для начинающих специалистов (junior) характерна высокая конкуренция; в обзорах, опирающихся на данные hh.ru, приводятся оценки в десятки откликов на одну вакансию для junior-уровня (например, 16 резюме на одну вакансию) [7]. С практической точки зрения это означает большую доступность «универсальных» ИТ-ролей (администрирование/DevOps/разработка) при ограниченном бюджете. Одновременно дефицит кадров кибербезопасности снижает доступность профильных специалистов и повышает затраты на их привлечение [8].

Таблица 3.1 - Укрупнённая структура затрат жизненного цикла (TCO) для сравниваемых подходов

Статья	Параметр	WAF (rule-based)	WAF+ML	Примечание/ограничение
Инфраструктура	VM/мес.	3000 руб.	3000 руб.	Фиксировано условием
Сопровождение	0,2 FTE	0,2 FTE ИБ-специалиста	0,2 FTE DevOps/админ + контроль	Доли ставок зависят от масштаба
Обновления	регламент	обновление CRS/исключения	переобучение/ версионирование	Нужна дисциплина процессов

Для получения денежной оценки при фиксированной доле 0,2 FTE можно использовать медианные значения рынка труда. Например, при медиане 214 600 руб./мес. для DevOps [6] доля 0,2 FTE даёт порядка 42 920 руб./мес. на сопровождение (без учёта налогов и накладных расходов). Для ИБ-специалиста фактическая величина зависит от региона и уровня; в условиях дефицита кадров

оценка, как правило, не ниже сопоставимых ИТ-ролей. Таким образом, даже при одинаковой FTE-доле структура затрат смещается в пользу сопровождения ИСЗ силами эксплуатации при наличии регламента контроля обучения.

3.8. Оценка экономической эффективности применения искусственного интеллекта в системах защиты информации

Экономическая эффективность трактуется как превышение ожидаемой экономии (за счёт снижения вероятности и тяжести инцидентов) над дополнительными затратами на внедрение и сопровождение. В экспериментальной части зафиксирован «прирост» по прокси-показателям: увеличение доли 403 на 7,5 п.п. (60,5% против 53,0%) и снижение доли 200 на 26,5 п.п. (2,1% против 28,6%). Такие различия могут рассматриваться как потенциальное снижение вероятности успешной доставки воздействия до прикладного уровня в рамках тестового профиля атак.

Для связи с масштабом последствий целесообразно использовать официальные и аналитические источники, фиксирующие рост киберрисков и затрат на ИБ. В российском отчёте Б1 (март 2025) подчёркивается рост рынка ИБ РФ (с 192 млрд руб. в 2022 г. до 299 млрд руб. в 2024 г.) и отмечается, что более половины компаний сталкивались с реализацией киберрисков за последние два года, а среднее число атак на организацию существенно растёт [8]. Данные наблюдения позволяют обосновать предположение о высокой актуальности мер защиты даже при умеренных затратах на инфраструктуру.

В упрощённой модели ожидаемые потери оцениваются как $E(L)=P \cdot I$, где P - вероятность инцидента в периоде, I - ущерб. Применение ИСЗ трактуется как мера, уменьшающая P для классов атак, соответствующих реальным и тестируемым паттернам. При отсутствии статистики конкретной организации допустимо использовать отраслевые оценки P и относительное сравнение сценариев (с и без ИСЗ). Ограничение расчёта состоит в том, что количественная связь между долей 200/403 в стенде и P в реальной эксплуатации не является

линейной и зависит от конкретного приложения, профиля трафика и мотивов нарушителя.

При фиксированных эксплуатационных затратах (3000 руб./мес. на VM + доля времени эксплуатации) даже предотвращение единичного инцидента с ощутимыми последствиями (простои, восстановление, утечки, регуляторные риски) может экономически оправдать внедрение. По этой причине в организациях малого и среднего бизнеса рационально рассматривать ИСЗ как средство снижения риска на уровне прикладного контура при относительно невысокой стоимости владения.

3.9. Практические рекомендации по внедрению интеллектуальных средств защиты в организациях малого и среднего бизнеса

1) Внедрение следует начинать с режима monitor (наблюдение), обеспечив сбор журналов и анализ ложноположительных срабатываний.

2) Первичная обучающая выборка должна включать собственный легитимный трафик (систематизированный по ключевым маршрутам) и репрезентативный набор атакующих запросов, сформированный генератором.

3) Переход к enforce рекомендуется выполнять постепенно: устанавливать консервативные пороги и расширять блокирование после анализа журналов.

4) Для снижения риска деградации качества необходимо вести версионирование моделей и конфигураций (порогов), а также внедрить правило отката (rollback) при ухудшении метрик на контрольном наборе.

5) При появлении новых типов атак следует расширять датасет и переобучать модель по регламенту, фиксируя изменения и метрики.

Отдельное внимание следует уделить вопросам обработки персональных данных в логах (например, логины/пароли в тестовом стенде). В производственной эксплуатации подобные поля должны быть маскированы или хэшированы, а доступ к журналам ограничен. В противном случае система защиты может создавать дополнительные риски утечек данных.

3.10. Перспективы развития интеллектуальных систем защиты информации и направления дальнейших исследований

Дальнейшее развитие ИСЗ целесообразно рассматривать по трём направлениям. Во-первых, расширение обучающего покрытия по классам атак (XSS, SSRF, Template Injection и др.) и формализация процесса разметки для повышения полноты детектирования. Во-вторых, внедрение признаков, учитывающих динамические аспекты поведения (частота запросов, корреляция событий, последовательности действий), что позволит сблизить модель с поведенческими подходами IDS/IPS. В-третьих, исследование гибридных схем, совмещающих rule-based WAF (для сигнатурных атак и регуляторных требований) и ML-компонент (для адаптивности), что соответствует современному тренду на комплексные платформенные средства защиты [8].

Ограничения текущего стенда включают использование одной модели бинарной классификации и зависимость качества от состава обучающего набора. Эти ограничения определяют необходимость дальнейших экспериментов на более репрезентативных данных и с учётом производственного профиля трафика конкретной организации.

4. ОБОБЩЕНИЕ РЕЗУЛЬТАТОВ И ОЦЕНКА ЭФФЕКТИВНОСТИ ИНТЕГРАЦИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СИСТЕМУ ЗАЩИТЫ ИНФОРМАЦИИ

4.1. Оценка достижения цели исследования

Целью настоящего исследования являлась интеграция искусственного интеллекта в систему защиты информации с последующей оценкой эффективности такого подхода по сравнению с традиционными средствами защиты.

В ходе работы была разработана и апробирована архитектура интеллектуальной системы защиты, основанная на сочетании механизмов веб-экранирования (WAF) и модели машинного обучения. Реализованный подход включал построение инфраструктуры, формирование датасета, обучение модели, внедрение механизма анализа HTTP-запросов и экспериментальную оценку результатов.

Полученные данные свидетельствуют о том, что поставленная цель была достигнута, поскольку интеграция искусственного интеллекта в систему защиты информации позволила:

- повысить долю блокируемых потенциально вредоносных запросов;
- снизить вероятность прохождения атак через защитный контур;
- обеспечить адаптивность защитного механизма к ранее неизвестным паттернам атак;
- сократить зависимость от ручного формирования правил безопасности.

Однако при этом выявлены ограничения, связанные с качеством обучения модели и спецификой используемого датасета, что указывает на необходимость дальнейшего развития предложенного подхода.

Таким образом, результаты исследования подтверждают принципиальную применимость методов искусственного интеллекта в задачах защиты

информации, однако демонстрируют, что их использование требует методически выверенной организации процессов обучения и эксплуатации.

4.2. Анализ эффективности реализованной системы защиты

Сравнительный анализ экспериментальных стендов показал, что интеллектуальная система защиты обладает рядом преимуществ по отношению к традиционным средствам WAF.

В частности, по результатам экспериментального тестирования было установлено:

1. Увеличение доли заблокированных запросов (HTTP 403) в системе WAF + ИИ по сравнению с классическим WAF.
2. Существенное снижение доли успешных запросов (HTTP 200), что свидетельствует о снижении вероятности успешной эксплуатации уязвимостей.
3. Наличие значительного количества ответов HTTP 404, интерпретируемых как индикатор неуспешного поиска уязвимых ресурсов злоумышленником.

Дополнительно было выявлено, что модель машинного обучения, обученная преимущественно на атаках классов SQL injection и Path Traversal, продемонстрировала способность выявлять признаки нелегитимности запросов других типов. Данный факт указывает на наличие обобщающей способности модели, что является важным свойством интеллектуальных систем защиты.

Однако выявленные результаты также свидетельствуют о том, что интеллектуальная система не может рассматриваться как полностью автономный механизм защиты, поскольку её эффективность напрямую зависит от:

- полноты и репрезентативности обучающего датасета;
- корректности процесса разметки данных;
- параметров пороговых значений классификации;
- качества логирования и извлечения признаков.

Следовательно, эффективность системы защиты на основе искусственного интеллекта носит вероятностный характер и требует постоянного контроля и дообучения модели.

4.3. Преимущества интеграции искусственного интеллекта в систему защиты информации

Анализ результатов исследования позволяет выделить следующие ключевые преимущества интеграции искусственного интеллекта в системы защиты информации.

4.3.1. Адаптивность и обобщающая способность

В отличие от традиционных WAF, основанных на статических правилах, интеллектуальная система защиты способна выявлять аномалии и потенциальные атаки, не описанные явно в правилах. Это особенно важно в условиях динамичного развития методов атак и появления новых векторов угроз.

4.3.2. Снижение зависимости от высококвалифицированных специалистов ИБ

Эксплуатация классических WAF требует участия специалистов в области информационной безопасности, обладающих глубокими знаниями архитектуры приложений и механизмов атак.

В предложенной системе значительная часть задач по настройке защиты переносится в область работы с данными и обучением модели, что позволяет:

- снизить требования к квалификации персонала;
- делегировать часть функций инженерам эксплуатации или DevOps-специалистам;
- использовать ресурсы менее квалифицированных специалистов при контролируемом процессе обучения модели.

С учётом выявленного дефицита специалистов по информационной безопасности на российском рынке труда и относительного избытка начинающих IT-специалистов данный фактор приобретает особую

экономическую значимость.

4.3.3. Экономическая целесообразность

Результаты экономической оценки показали, что интеллектуальная система защиты обладает потенциально более низкой совокупной стоимостью владения по сравнению с традиционными решениями, за счёт:

- меньших затрат на сопровождение;
- снижения зависимости от дорогостоящих специалистов ИБ;
- возможности использования недорогой облачной инфраструктуры.

При этом даже относительно небольшой прирост эффективности защиты (в виде снижения доли успешных атак) может приводить к существенному снижению потенциальных убытков от киберинцидентов.

4.4. Ограничения и недостатки реализованного подхода

Несмотря на выявленные преимущества, проведённое исследование позволило выявить ряд принципиальных ограничений.

4.4.1. Зависимость от качества датасета

Качество работы модели машинного обучения напрямую зависит от структуры и полноты обучающего датасета. Ошибки в разметке данных могут приводить к:

- ложным срабатываниям (false positive);
- пропуску атак (false negative);
- деградации качества модели при дрейфе данных.

4.4.2. Риск некорректного обучения модели

В отличие от классических WAF, где правила формируются экспертами, интеллектуальная система может быть обучена некорректно, что создаёт риск легитимации вредоносных запросов. Следовательно, процесс обучения модели требует организационного контроля и методической регламентации.

4.4.3. Ограниченность используемой модели

В рамках исследования была использована одна модель машинного

обучения, что ограничивает возможности анализа сложных паттернов атак. Более сложные архитектуры (ансамбли моделей, нейронные сети различных типов) потенциально могут повысить эффективность защиты, однако требуют больших вычислительных ресурсов и объёмов данных.

4.4.4. Вероятностный характер решений

В отличие от детерминированных правил WAF, решения модели машинного обучения имеют вероятностный характер. Это усложняет формирование формальных гарантий безопасности и требует использования дополнительных механизмов контроля.

4.5. Общий вывод о результатах исследования

Проведённое исследование позволяет сделать вывод о том, что интеграция искусственного интеллекта в систему защиты информации является перспективным направлением развития средств обеспечения безопасности веб-приложений.

Реализованная система WAF + ИИ продемонстрировала:

- повышение эффективности выявления атак;
- адаптивность к неизвестным угрозам;
- потенциальную экономическую целесообразность;
- возможность снижения зависимости от дефицитных специалистов по информационной безопасности.

В то же время исследование показало, что интеллектуальные системы защиты не могут рассматриваться как замена традиционных средств безопасности, а должны использоваться в качестве дополнения к ним. Наиболее эффективным является гибридный подход, при котором методы машинного обучения функционируют совместно с классическими механизмами WAF.

Таким образом, поставленная цель исследования может быть признана достигнутой в концептуальном и практическом аспектах, однако дальнейшее развитие предложенного подхода требует расширения набора моделей,

совершенствования методик обучения и разработки формализованных процедур контроля качества интеллектуальных систем защиты.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации». - Режим доступа:
2. ФСТЭК России. Документы и методические материалы по защите информации. - Режим доступа: https://fstec.ru (дата обращения: 16.01.2026).
3. OWASP Top 10 - 2023. - Режим доступа: https://owasp.org/www-project-top-ten/ (дата обращения: 16.01.2026).
4. OWASP ModSecurity Core Rule Set. - Режим доступа: https://owasp.org/www-project-modsecurity-core-rule-set/ (дата обращения: 17.01.2026).
5. Документация ModSecurity. - Режим доступа: https://github.com/SpiderLabs/ModSecurity/wiki (дата обращения: 17.01.2026).
6. Документация Nginx. - Режим доступа: https://nginx.org/ru/docs/ (дата обращения: 17.01.2026).
7. Документация FastAPI. - Режим доступа: https://fastapi.tiangolo.com/ru/ (дата обращения: 18.01.2026).
8. Документация Scikit-learn. - Режим доступа: https://scikit-learn.org/stable/ (дата обращения: 18.01.2026).
9. Хабр. Применение машинного обучения в информационной безопасности. - Режим доступа: https://habr.com/ru/ (дата обращения: 19.01.2026).
10. Хабр. Архитектура WAF и принципы работы. - Режим доступа: https://habr.com/ru/ (дата обращения: 19.01.2026).
11. Хабр. Логистическая регрессия и задачи классификации. - Режим доступа: https://habr.com/ru/ (дата обращения: 20.01.2026).
12. SecurityLab. Аналитика атак на веб-приложения. - Режим доступа: https://www.securitylab.ru (дата обращения: 20.01.2026).
13. TAdviser. Рынок информационной безопасности в России. - Режим доступа: https://www.tadviser.ru (дата обращения: 20.01.2026).
14. Positive Technologies. Дефицит кадров на рынке информационной безопасности в России. - Режим доступа:

https://ptsecurity.com/about/news/issledovanie-cz-sr-severo-zapad-i-positive-technologies-deficzit-kadrov-na-rynke-ib-rossii-k-2027-godu-dostignet-60-tysyach/ (дата обращения: 21.01.2026).

15. Лаборатория Касперского. Аналитика рынка специалистов по кибербезопасности. - Режим доступа: https://www.kaspersky.ru (дата обращения: 21.01.2026).

16. Securelist. Отчёты о киберугрозах. - Режим доступа: https://securelist.ru (дата обращения: 22.01.2026).

17. HeadHunter. Аналитика рынка труда в сфере IT и информационной безопасности. - Режим доступа: https://hh.ru (дата обращения: 22.01.2026).

18. Роскомнадзор. Статистика утечек персональных данных. - Режим доступа: https://rkn.gov.ru (дата обращения: 23.01.2026).

19. Минцифры РФ. Аналитические материалы о цифровой безопасности. - Режим доступа: https://digital.gov.ru (дата обращения: 23.01.2026).

20. Wikimedia Commons. Multi-layer neural network. - Режим доступа: https://commons.wikimedia.org/wiki/File:Multi-Layer_Neural_Network-Vector.svg (дата обращения: 24.01.2026).

21. Wikimedia Commons. Typical CNN architecture. - Режим доступа: https://commons.wikimedia.org/wiki/File:Typical_cnn.png (дата обращения: 24.01.2026).

22. Wikimedia Commons. Recurrent neural network unfold. - Режим доступа: https://commons.wikimedia.org/wiki/File:Recurrent_neural_network_unfold.svg (дата обращения: 24.01.2026).

23. Microsoft Learn. Основы машинного обучения и классификации. - Режим доступа: https://learn.microsoft.com/ru-ru/ (дата обращения: 25.01.2026)

24. IBM. Artificial Intelligence in Cybersecurity. - Режим доступа: https://www.ibm.com (дата обращения: 25.01.2026).

25. Cloudflare. Web Application Firewall Overview. - Режим доступа:

https://www.cloudflare.com (дата обращения: 26.01.2026).

26. MITRE ATT&CK Framework. - Режим доступа: https://attack.mitre.org (дата обращения: 26.01.2026).

27. CSIRT РФ. Аналитика инцидентов информационной безопасности. - Режим доступа: https://cert.gov.ru (дата обращения: 27.01.2026).

Приложение 1