WaveAccess
We Make IT Easy

Антифрод-система на основе алгоритмов машинного обучения

WaveAccess – это

Международная ИТ-компания, которая создает технически сложное, высоконагруженное и отказоустойчивое программное обеспечение для компаний по всему миру.

18
лет опыта разработки IT решений

300+
специалистов в штате

4
глобальных центра разработки

9
индустрий – от банкинга до здравоохранения

280+
успешно реализованных проектов

72%
постоянных клиентов

Санкт-Петербург
головной офис

Великобритания, США, Дания и Германия
офисы продаж

Academy Award-winning Mocha for Imagineer Systems

ISO 27001:2013 CERTIFIED

ISO 9001 CERTIFIED

Microsoft Partner Network WORLDWIDE DYNAMICS PARTNER OF THE YEAR 2009, 2010, 2011 & 2016

Silver Microsoft Partner

Microsoft Partner Network | 2011 PARTNER OF THE YEAR Microsoft Dynamics Professional Services CRM4Legal for Client Profiles Winner

Microsoft Partner | 2017 Partner of the Year Winner Business Analytics Award

Microsoft Partner | 2018 Partner of the Year Artificial Intelligence Award

Обзор проекта

Модуль на основе алгоритмов машинного обучения и нечеткой логики. Разработан для защиты данных крупной организации в сфере телекоммуникаций. Компания обслуживает web-портал, насчитывающий более миллиона пользователей.

Описание проблемы

Для повышения безопасности системы нашему клиенту потребовался модуль защиты, который выявляет факты мошенничества на миллионном трафике.

Но мошеннические действия очень сложно предсказать. Попытка доступа по поддельному IP, хищение номера кредитной карты, подбор пароля, а также несколько десятков других непрогнозируемых событий — мошенники изобретательны и непредсказуемы.

Чтобы распознавать подозрительные действия как угрозы, нашему заказчику требовался модуль на основе алгоритмов нечеткой логики и машинного обучения. Его требовалось не только разработать, но и обучить.

Чтобы распознавать подозрительные действия как угрозы, нашему заказчику требовался модуль на основе алгоритмов нечеткой логики и машинного обучения. Его требовалось не только разработать, но и обучить.

Почему WaveAccess?

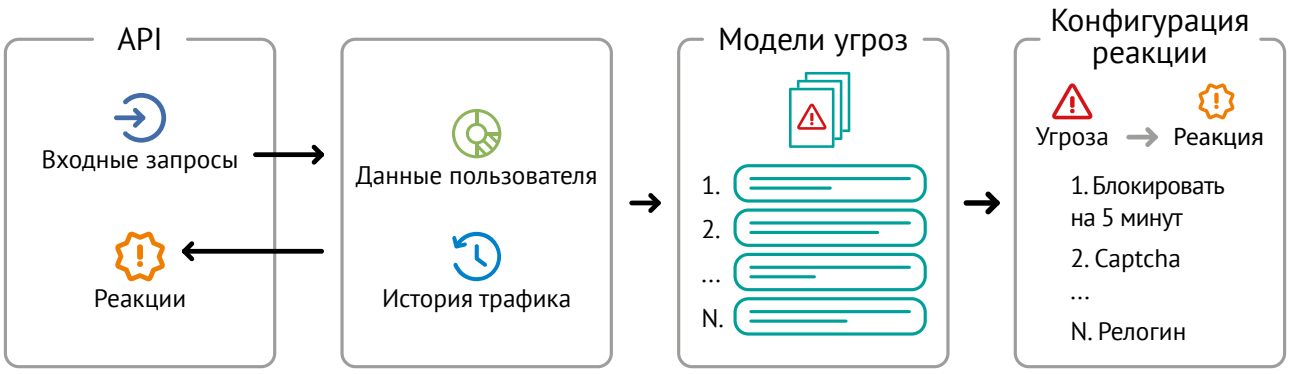
Системы и отдельные модули на основе алгоритмов машинного обучения— одна из наших специализаций. Решая задачи e-commerce, специалисты WaveAccess создают подобные модули для «умных» рекомендаций в online-магазинах; эти же алгоритмы наши клиенты используют для расчета вероятности наступления страхового случая, для предсказания вероятности покупки и других бизнес-задач. Эти решения с успехом можно применять, изучая и распознавая аномальное поведение пользователей.

Как разработчик полного цикла, команда WaveAccess обеспечивает работу над каждым этапом проекта: от анализа до составления документации. Все это убедило клиента стать нашим партнером.

Этапы работы

Подробнее о проекте

Мы изучили web-портал клиента, смоделировали и проанализировали всевозможные действия пользователей. На основе этого анализа были описаны и формализованы 16 типовых угроз. Для каждой угрозы разработан алгоритм определения и реакция на нее, для части предусмотрен запрос обратной связи от пользователя. В последнем случае система принимает решение о реакции на угрозу на основе этого ответа.



Второй этап работы заключался в разработке самого модуля защиты. Разработка модуля защиты началась с каркасов подмодулей, самые значимые из них — подмодуль аналитики, API и утилита проверки. Именно в нем размещены все алгоритмы и условия прохода по ним.

В систему заложены два принципа обработки:

Для реализации **нечеткой логики** был выбран готовый движок расчета входов и выходов на основе шаблонов. Мы усовершенствовали процесс подключения движка; кэширование данных; создали API для работы с набором шаблонов. Осуществили перегенерацию кэша «на горячую» при изменении граничных значений.

Машинное обучение позволяет настраивать модуль аналитики с помощью тренировочных данных и затем распознавать похожие ситуации с некоторым отклонением от эталонных данных. Хотя такой подход и требует затраты на поддержание тренировочных и проверочных сценариев, он позволяет покрывать больше смежных случаев.

Внешнее API для интеграции с модулем реализовали, следуя канонам REST архитектуры. Данные, приходящие через API, проходят базовую валидацию.

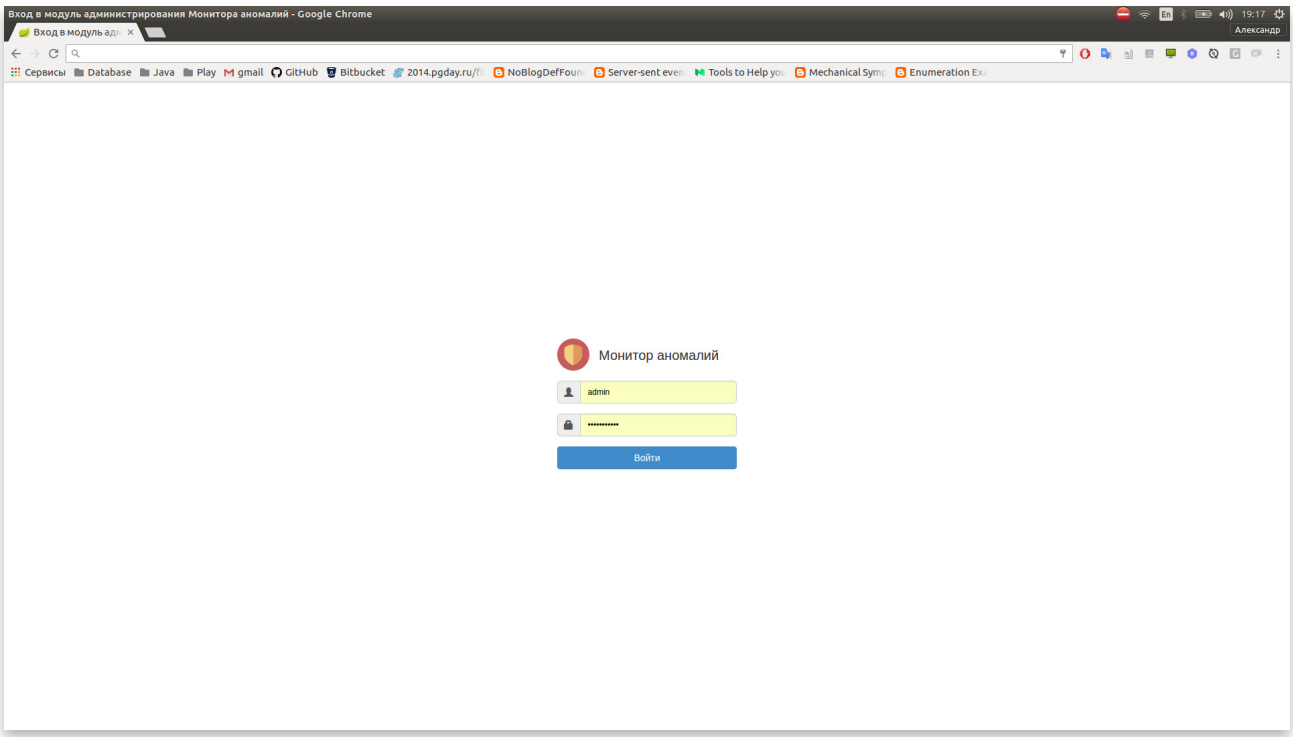
API-сервисы, как и вся система, сделаны согласно stateless архитектуре, благодаря этому разработанную систему легко масштабировать при необходимости. Размер расчётного кластера в промышленной эксплуатации — четыре сервера приложений с балансировкой и резервированием нагрузки.

UX

Несмотря на то, что наш клиент не требовал этого — для удобства мы разработали подмодуль администрирования с интерфейсом пользователя. Он помогает следить за ситуацией и заметно облегчит клиенту администрирование:

- Открывает администратору доступ к логам безопасности и событий.
- Реализует возможность получения и анализа данных.
- Позволяет настраивать параметры модулей.

Администраторов эта разработка избавляет от большого количества сложной ручной работы, что было встречено заказчиком с благодарностью.



В ходе проекта применялись различные виды тестирования, чтобы обеспечить его максимальное качество и полное соответствие требованиям заказчика.

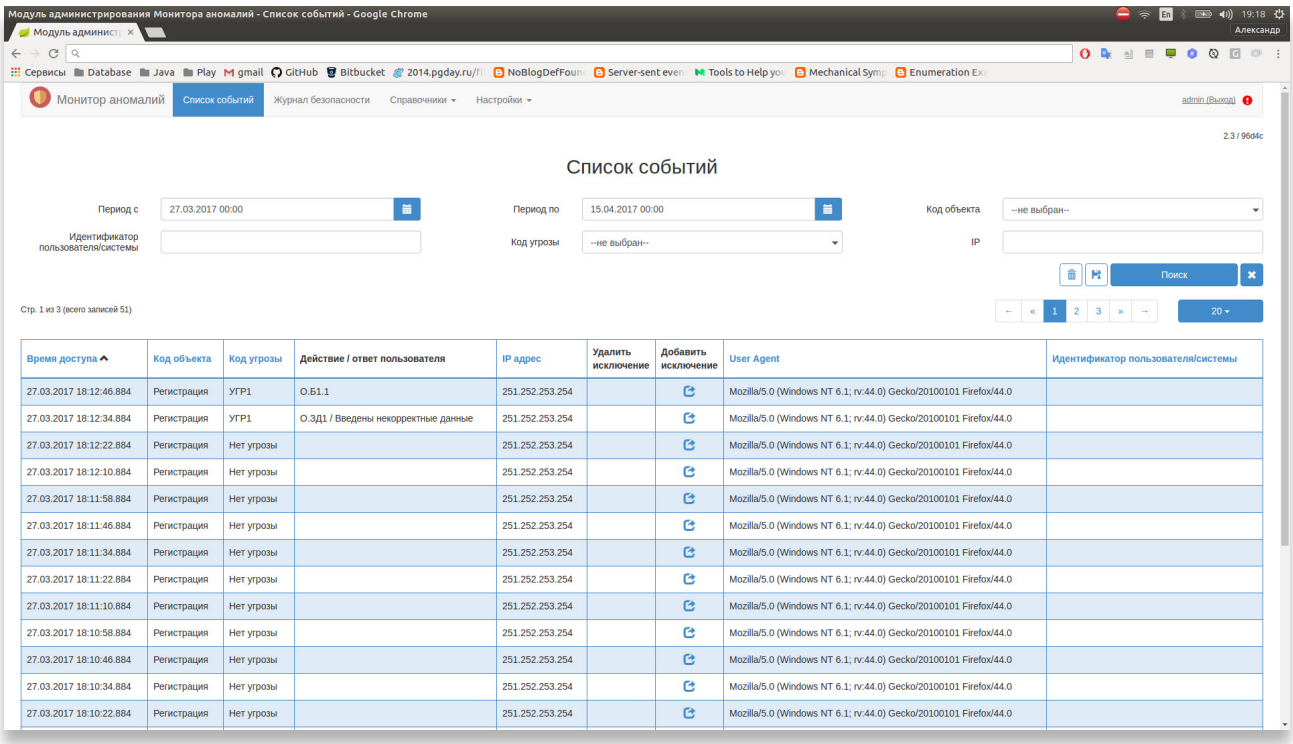
Тестирование

Для интеграционного тестирования был специально разработан проигрыватель сценариев — утилита проверки, которая сократила затраты на ручное тестирование и позволила включить проверку в Continuous Integration.

Антифрод-система предназначена для работы с большим количеством входящих данных, поэтому особенное внимание было уделено тестированию производительности и нагрузки.

Чтобы нагрузить систему и проверить различные варианты её работы, использовался тот же проигрыватель сценариев. Утилита способна не только определять правильность работы алгоритма по заданным параметрам, но и создавать необходимый поток данных.

Результаты тестов показали, что при постоянной работе утилиты идет быстрое накопление статистической информации, что вызывает разрастание размера базы и замедляет запросы. Чтобы избежать проблем, связанных с таким поведением, была оптимизирована хранимая модель данных, добавлена очистка ненужной статистической информации, а на особенно больших таблицах сделано автоматическое ежемесячное партиционирование.

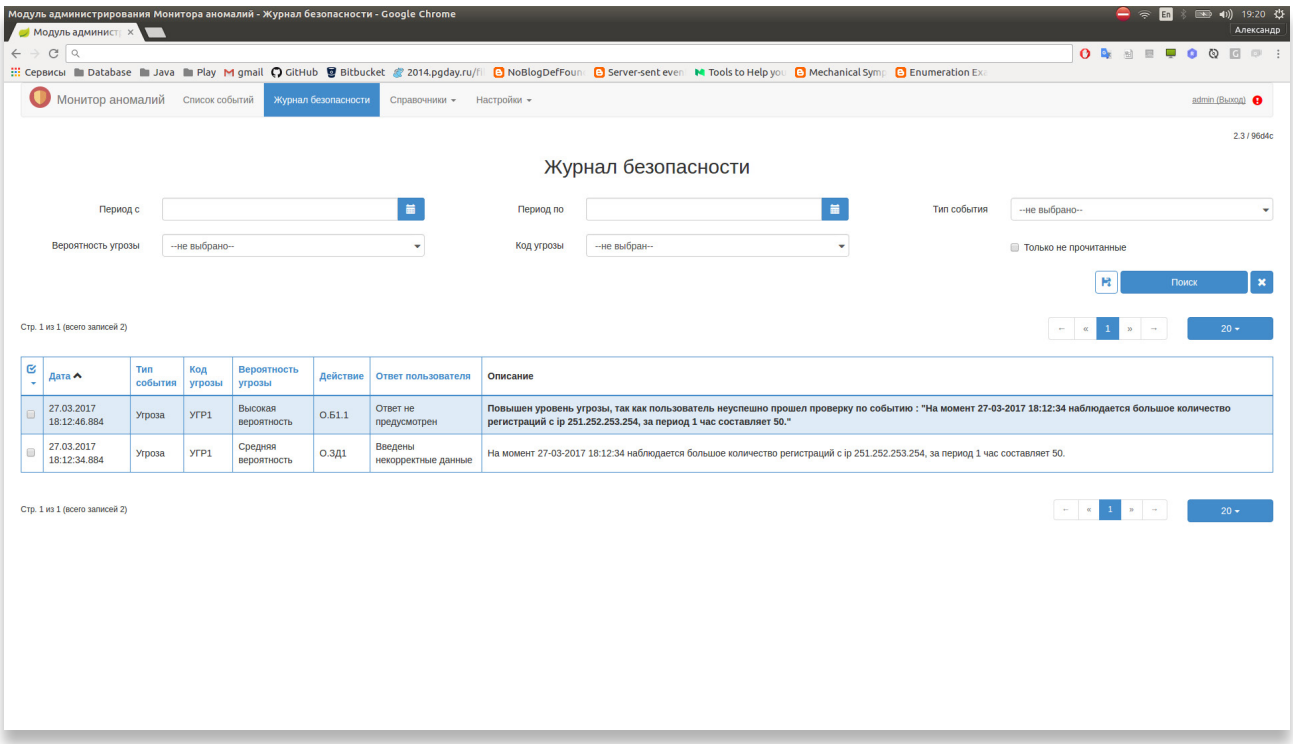


Для проверки основной логической части системы применялся инструмент SoapUI. Были написаны функциональные автотесты, которые отправляли rest-запросы определенное количество раз с определёнными параметрами.

Для проверки модуля администрирования, обладающего пользовательским интерфейсом, использовалось ручное функциональное тестирование по сценариям, разработанным и описанным на этапе анализа в программе и методике испытаний.

С помощью Selenium были написаны основные сценарии проверки подмодуля администрирования, которые запускались при каждой сборке проекта.

Перед передачей системы в опытную эксплуатацию заказчику проводился тест инсталляции на физически разных машинах с балансировщиком нагрузки и без. Результаты теста показали, что система ведет себя согласно требованиям.



Технологии, которые использованы в ходе проекта:

Основные:

База данных:

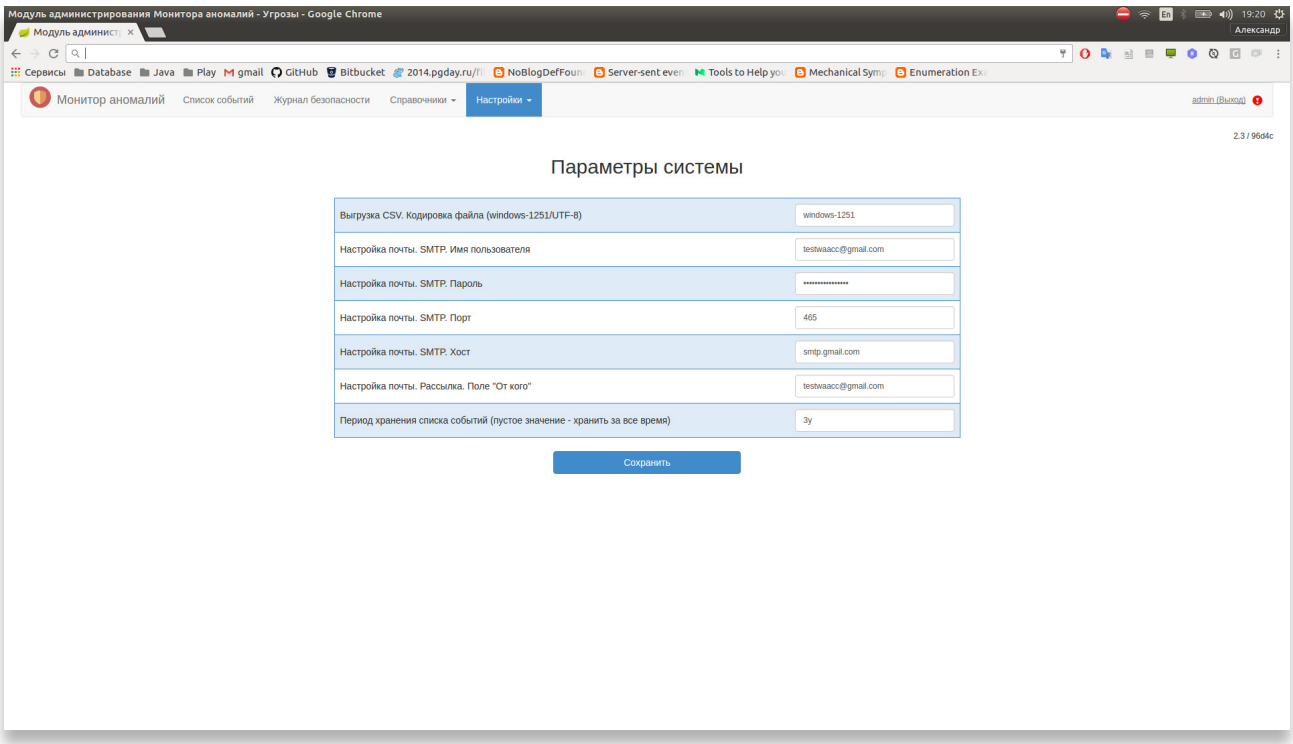
Клиентская часть:

Тестирование:

Результат

Для защиты веб-сервиса нашего клиента от мошеннических атак были разработаны:

- Основное решение: антифрод-система для мониторинга угроз веб-сервису.
- Утилита проверки для имитации нагрузки и интеграционного тестирования с помощью шаблонов.
- Модуль администратора для просмотра журналов событий и безопасности, настройки подсистем.

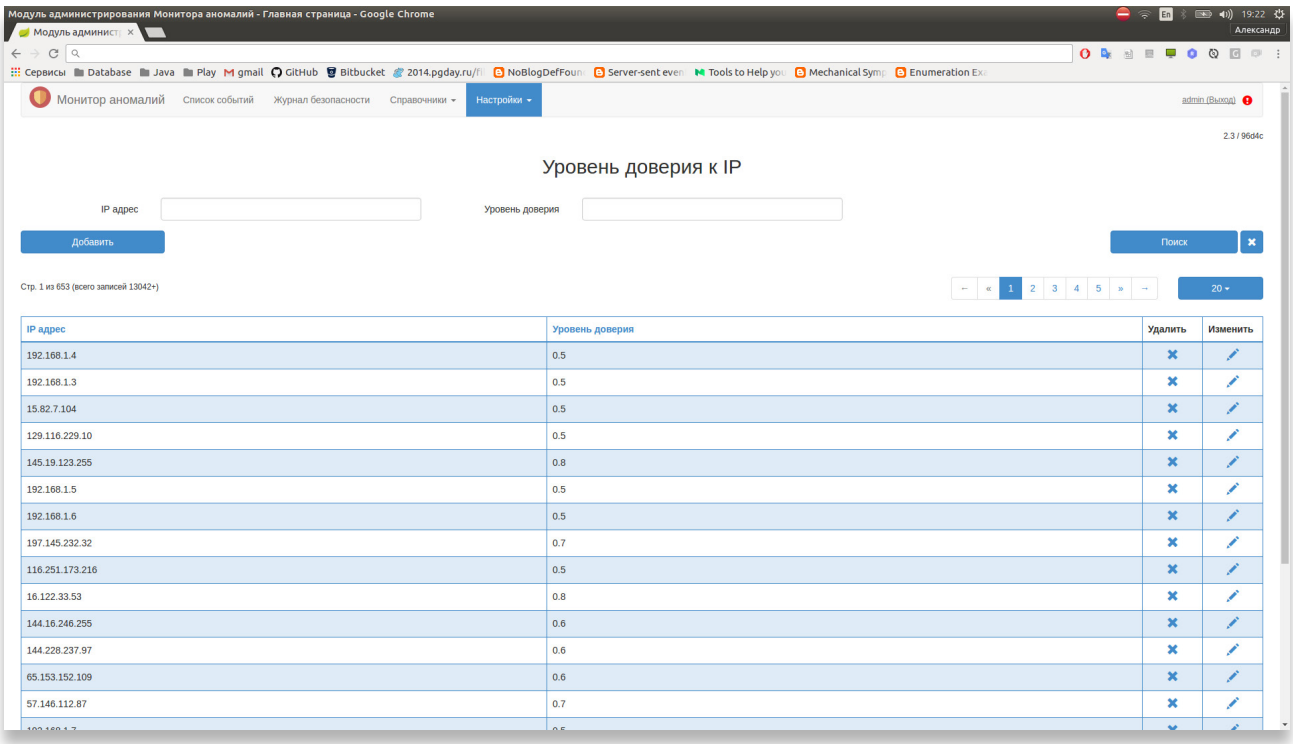


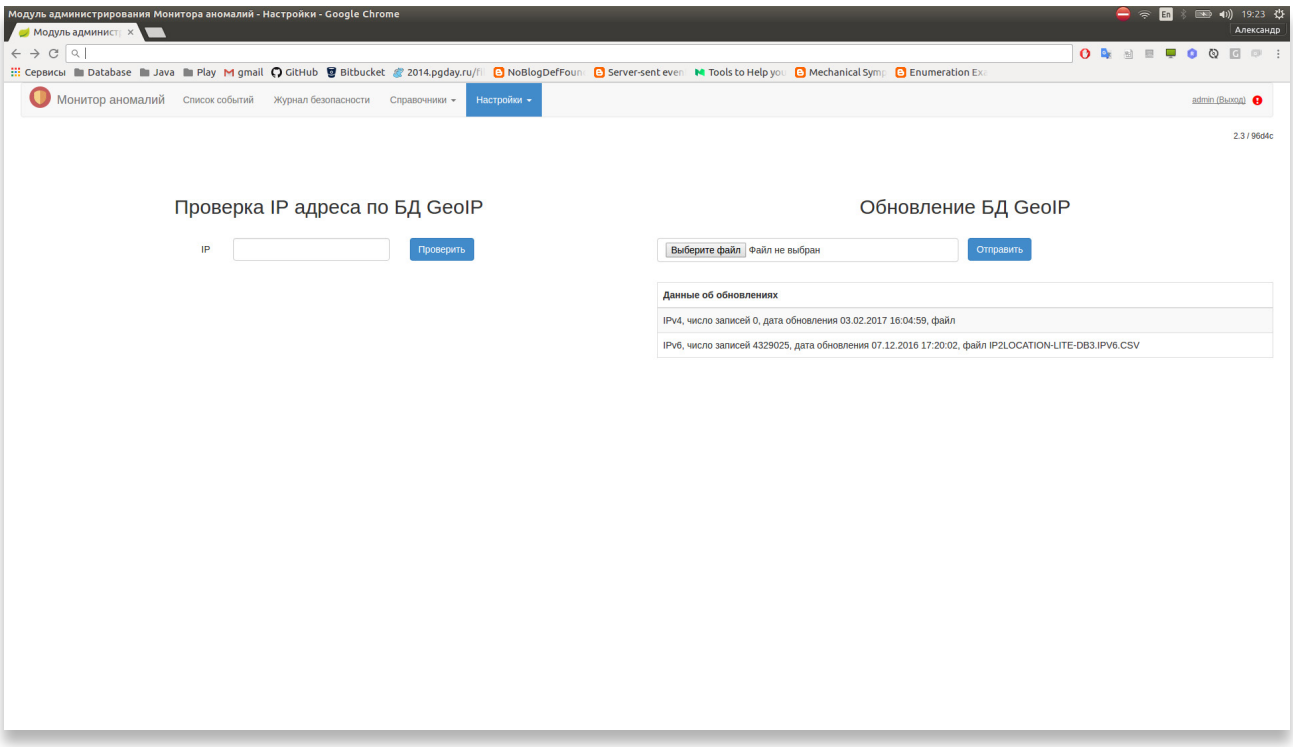
На основе известных системе 16-ти типов угроз, при помощи алгоритмов машинного обучения, стало возможным защитить веб-сервис от всевозможных мошеннических действий.

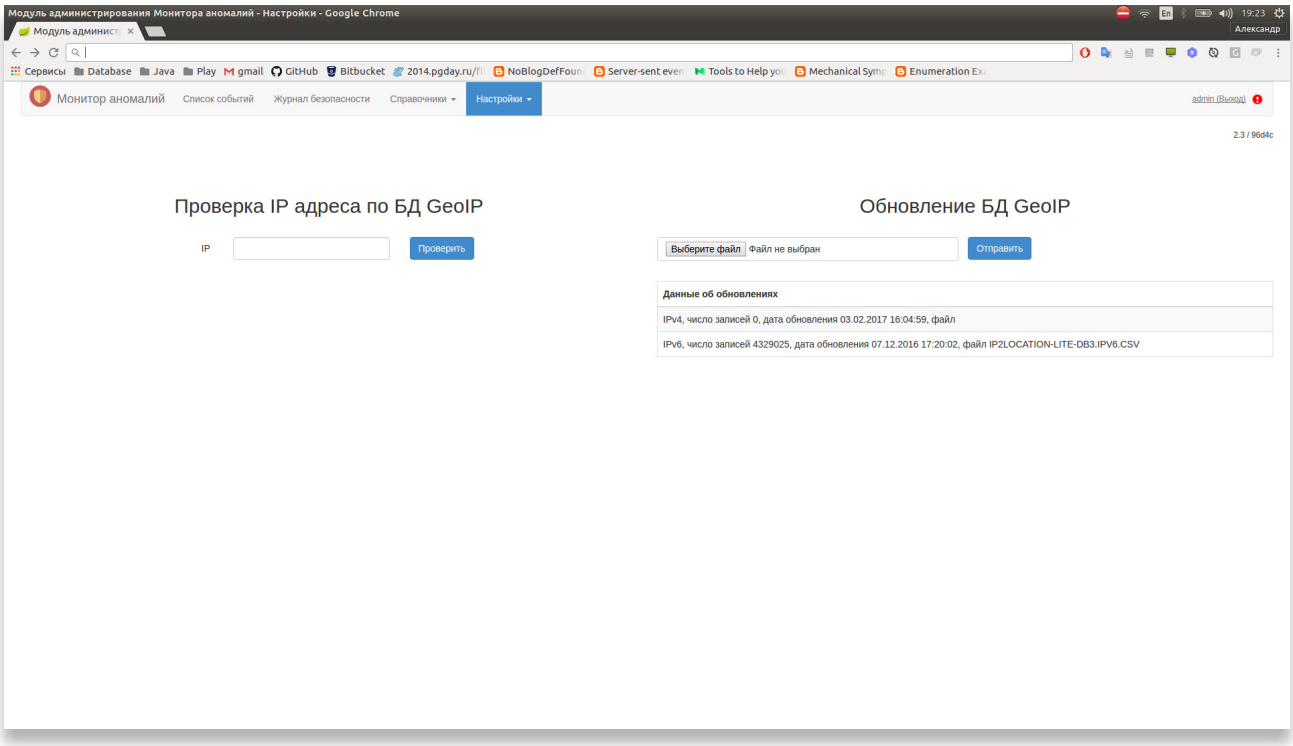
Производительность системы превысила ожидания заказчика. Система предоставляет возможность роста потока обрабатываемых данных и добавления новых функций.

Использованные технологии

1. **Основные:** Java 8, Spring 4, Spring Boot 1.4, JDBC Template, jFuzzyLogic 1.3, FlywayDB, ThymeLeaf, OpenCSV
2. **Клиентская часть:** Bootstrap, JQuery
3. **База Данных:** PostgreSQL 9.4+
4. **Тестирование:** JUnit, Selenium + PhantomJS driver, JXLS Reader, SoapUI.




Модуль администрирования Монитора аномалий - Настройки - Google Chrome
Монитор аномалий
Список событий
Журнал безопасности
Справочники
Настройки
admin (Выход)
2.3 / 96d4c
Проверка IP адреса по БД GeoIP
IP
Проверить
Обновление БД GeoIP
Выберите файл
Файл не выбран
Отправить
Данные об обновлениях
IPv4, число записей 0, дата обновления 03.02.2017 16:04:59, файл
IPv6, число записей 4329025, дата обновления 07.12.2016 17:20:02, файл IP2LOCATION-LITE-DB3.IPV6.CSV


Модуль администрирования Монитора аномалий - Настройки - Google Chrome
Монитор аномалий
Список событий
Журнал безопасности
Справочники
Настройки
admin (Выход)
2.3 / 96d4c
Проверка IP адреса по БД GeoIP
IP
Проверить
Обновление БД GeoIP
Выберите файл
Файл не выбран
Отправить
Данные об обновлениях
IPv4, число записей 0, дата обновления 03.02.2017 16:04:59, файл
IPv6, число записей 4329025, дата обновления 07.12.2016 17:20:02, файл IP2LOCATION-LITE-DB3.IPV6.CSV

WaveAccess
We Make IT Easy

Если вы хотите оценить возможности машинного обучения для вашего проекта, будем рады поделиться опытом:

Телефон: +7 812 407 2350

E-mail: hello@wave-access.com

Материал предоставлен компанией WaveAccess,
узнайте больше на
waveaccess.ru