infotecs

Безопасность ML-моделей при анализе сетевого трафика

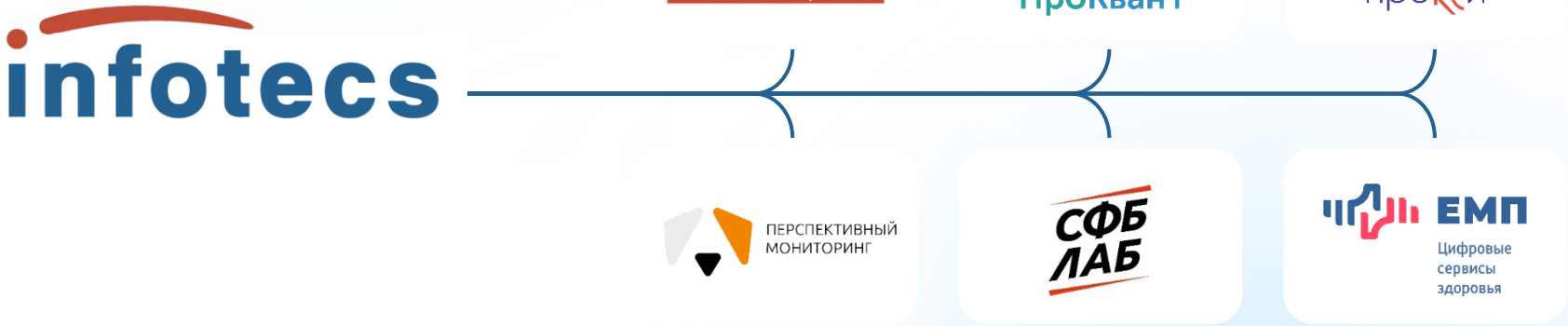
О спикере

Николай Кутьин

Архитектор системы обнаружения сетевых вторжений (IDS) и системы анализа сетевого трафика (NTA).

- В разработке с 2003 года.
- Основной язык программирования: C++.
- Занимаюсь встраиванием моделей машинного обучения в IDS и NTA.
- Разрабатываю решения DPI.

Группа компаний «ИнфоТеКС»



ИнфоТеКС в цифрах

Топ-10

Входит в пятёрку крупнейших ИТ-компаний в области разработки ПО

12 офисов

по всей стране

>10 млн

рабочих станций, защищённых продуктами ViPNet

>60 продуктов

для защиты информации

>30 лет

Работы на рынке ИБ

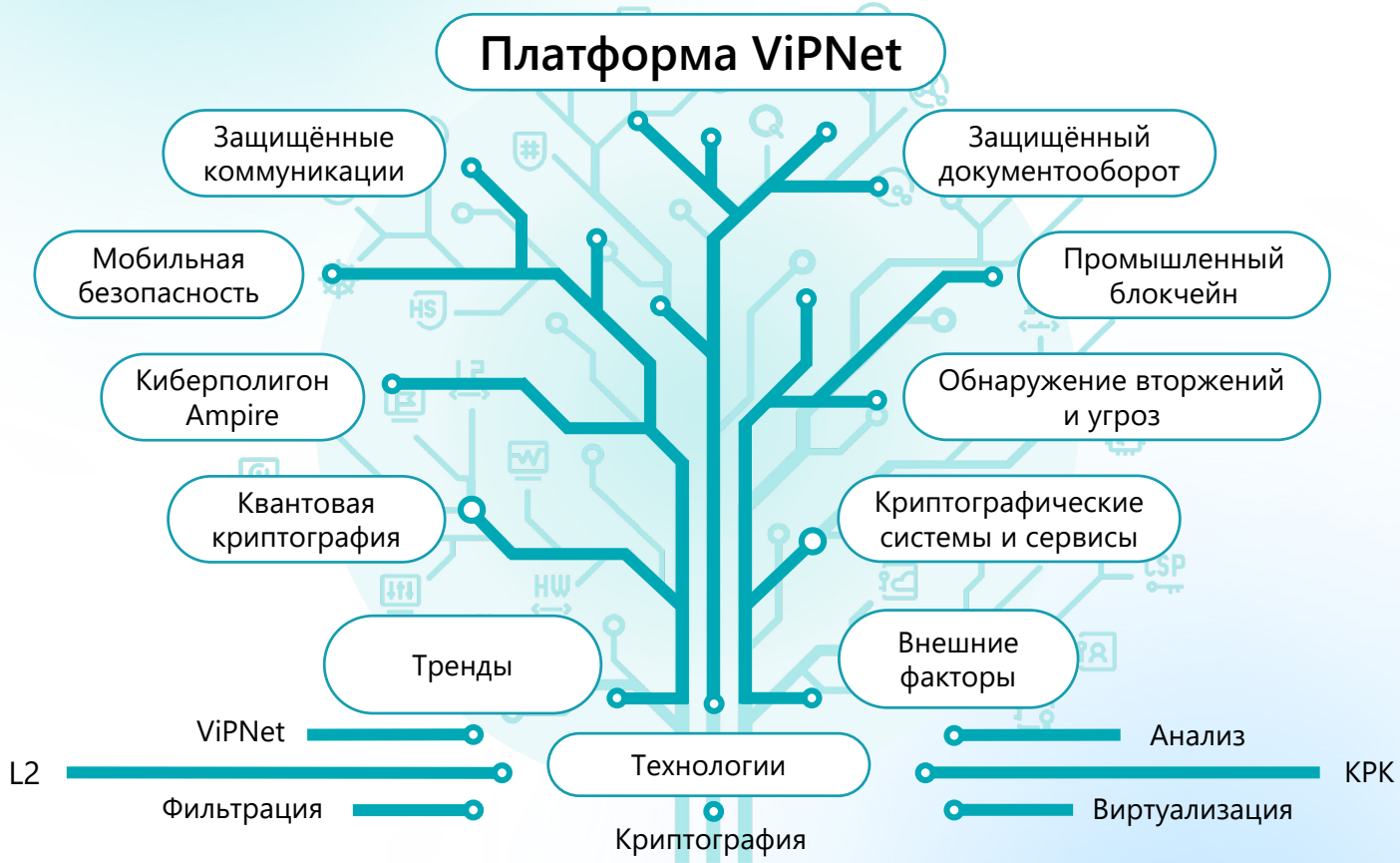
>1900

сотрудников

Топ-5

Входит в пятёрку компаний по количеству патентов в области цифровых технологий

Продукты и решения ИнфоТеКС



Поговорим про:

Противодействие угрозам информационной безопасности при внедрении моделей машинного обучения в анализ сетевого трафика.

Вопросы выбора сред исполнения моделей и методов подготовки данных для них. Рассмотрим реальный кейс на одном из продуктов компании.

Классы продуктов компании:

IDS (Intrusion Detection System) – система обнаружения вторжений (пассивная).

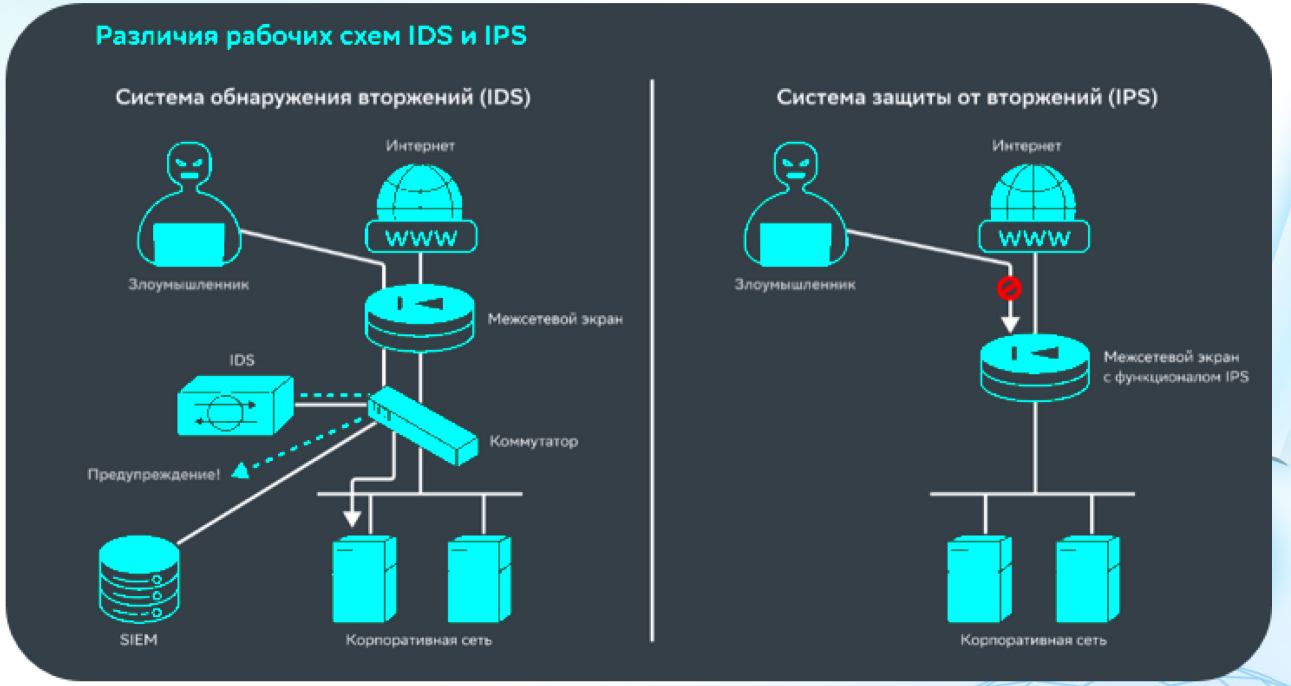
IPS (Intrusion Prevention System) – система предотвращения вторжений (активная).

NTA (Network Traffic Analysis) – категория продуктов безопасности, использующая сетевые коммуникации в качестве основного источника данных для обнаружения и расследования угроз.

infotecs

Актуальность защиты моделей ML при анализе сетевого трафика

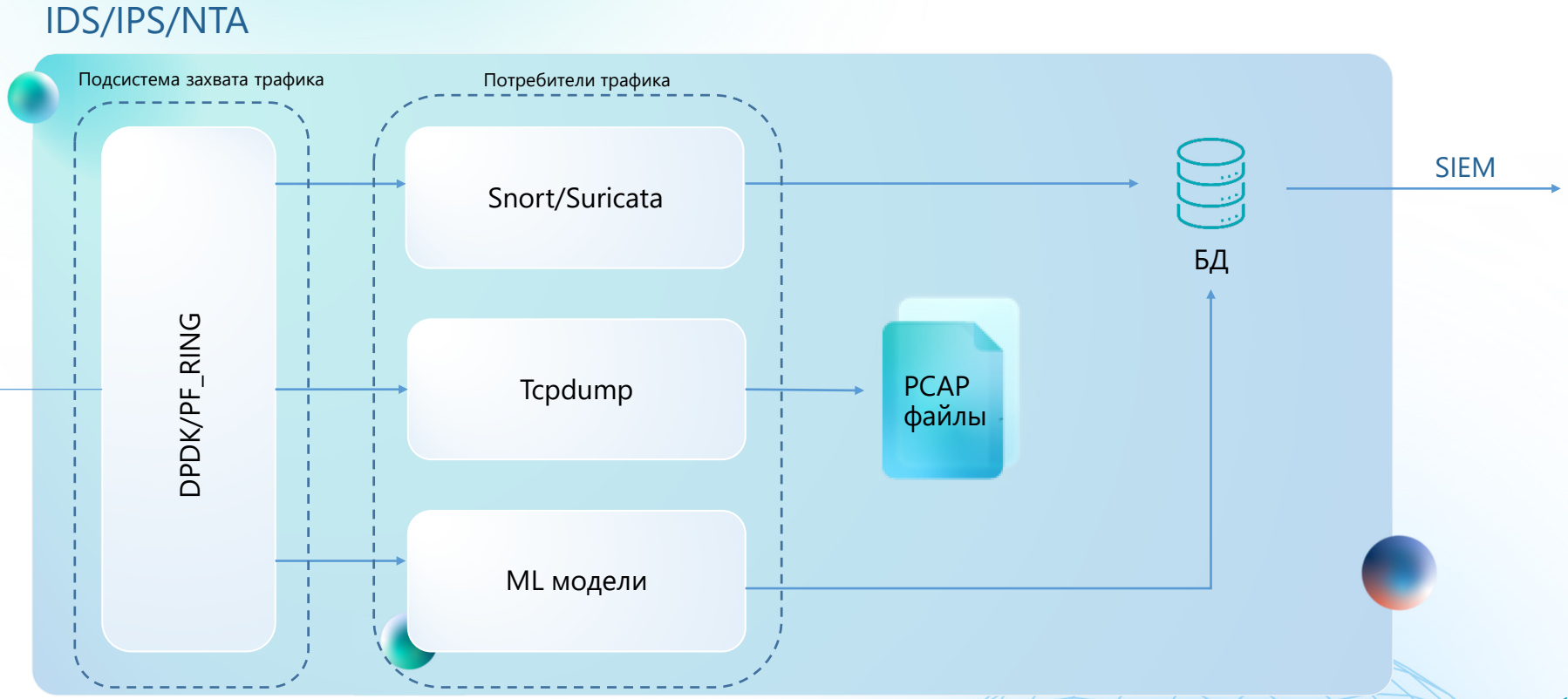
Угрозы продуктам безопасности



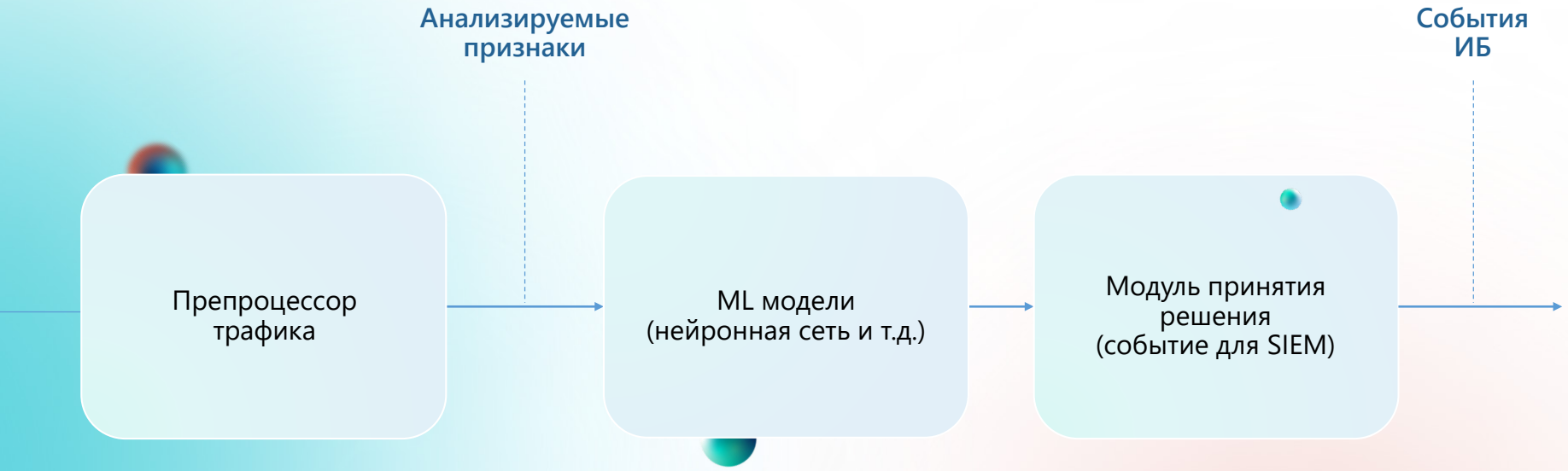
Поверхность атаки

Поверхность атаки – перечень потенциальных областей воздействия на продукт (функциональных подсистем, модулей ПО и их интерфейсов), для которых наличие уязвимостей приводит к нарушению свойств ИБ и которые могут быть использованы для проведения атаки.

Общая структура IDS/IPS/NTA

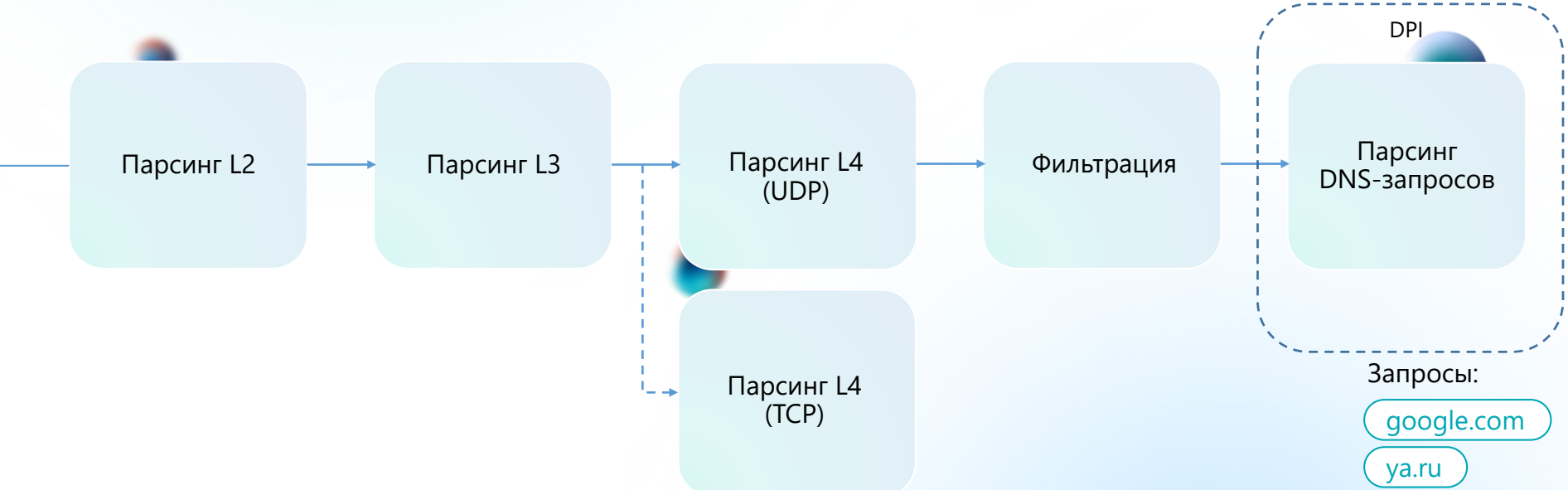


Pipeline обработки данных ML модели



- Модели ML не могут работать с «сырым» трафиком, для их работы необходимо применять предобработку трафика для выделения значащих признаков.

Pipeline предобработки трафика для модели DGA (обнаружения сгенерированных доменных имён)



Виды атак на модели ML

Через интерфейс захвата трафика:

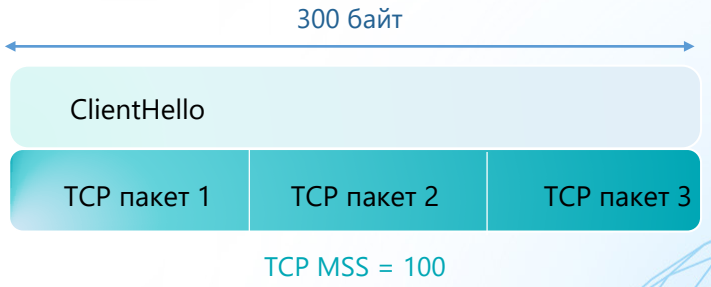
- Эксплуатация уязвимостей парсера трафика,
- Использование несовершенства алгоритмов машинного обучения.

Через управляющий интерфейс:

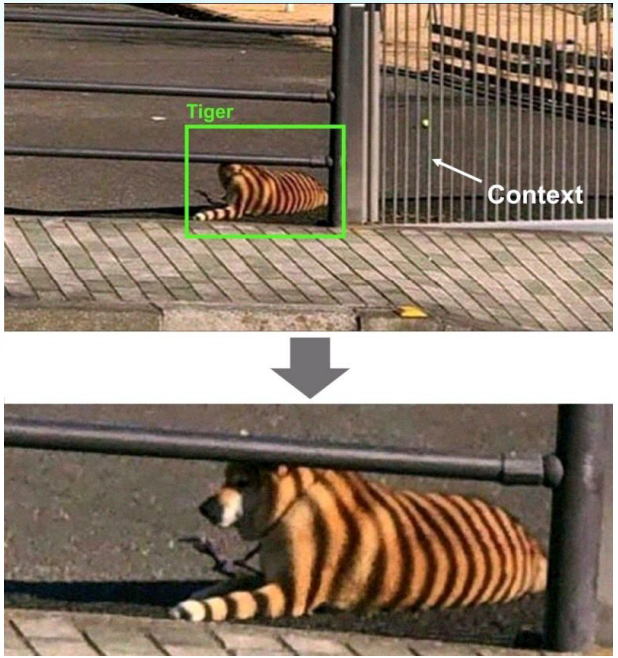
- Подмена данных,
- Подмена элементов среды исполнения моделей.

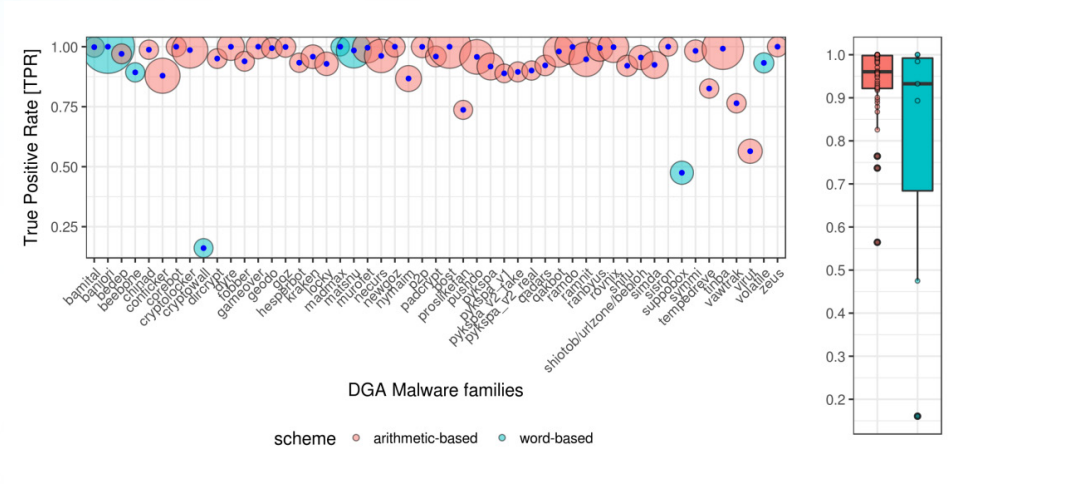
Пример эксплуатации уязвимостей парсера трафика протокола TLS для модели CnC TLS

- Данные выделяются на основе стартового сообщения ClientHello,
- Средняя длинна ClientHello – 200-300 байт,
- Стандартный MSS TCP – 1460 байт,
- В общем случае стартовое сообщение помещается в один пакет,
- С учётом большого объёма трафика затруднён statefull анализ (пересборка TCP), везде где возможно применяется stateless обработка (один пакет вместо потока).



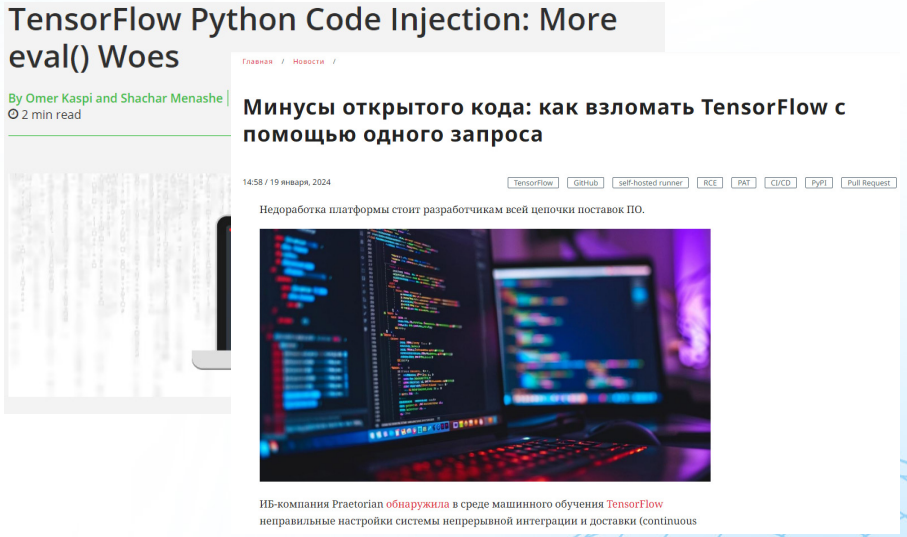
Использование несовершенства алгоритмов машинного обучения





Атаки через управляющий интерфейс

- Классические атаки с получением доступа к ОС IDS/IPS/NTA,
- Атакующий не всегда имеет полный доступ (root),
- Возможна подмена части библиотек, используемых в том числе и в ML (атака на среду выполнения моделей).



TensorFlow Python Code Injection: More eval() Woes

By Omer Kaspi and Shachar Menashe |
2 min read

Главная / Новости /

Минусы открытого кода: как взломать TensorFlow с помощью одного запроса

14:58 / 19 января, 2024

TensorFlow | GitHub | self-hosted runner | RCE | PAT | CI/CD | PyPI | Pull Request

Недоработка платформы стоит разработчикам всей цепочки поставок ПО.

ИБ-компания Praetorian обнаружила в среде машинного обучения TensorFlow неправильные настройки системы непрерывной интеграции и доставки (continuous

Среды выполнения моделей ML

- Python библиотеки (Tensorflow и т.д.),
- Удалённый ML-сервер (MLOps),
- Библиотеки для компилируемых языков (C++ OpenNN).

Способы защиты

01.

Ограничение доступа.

02.

Виртуализация.

03.

Отказ от динамически подгружаемых библиотек.

04.

Постобработка выходных данных модели.

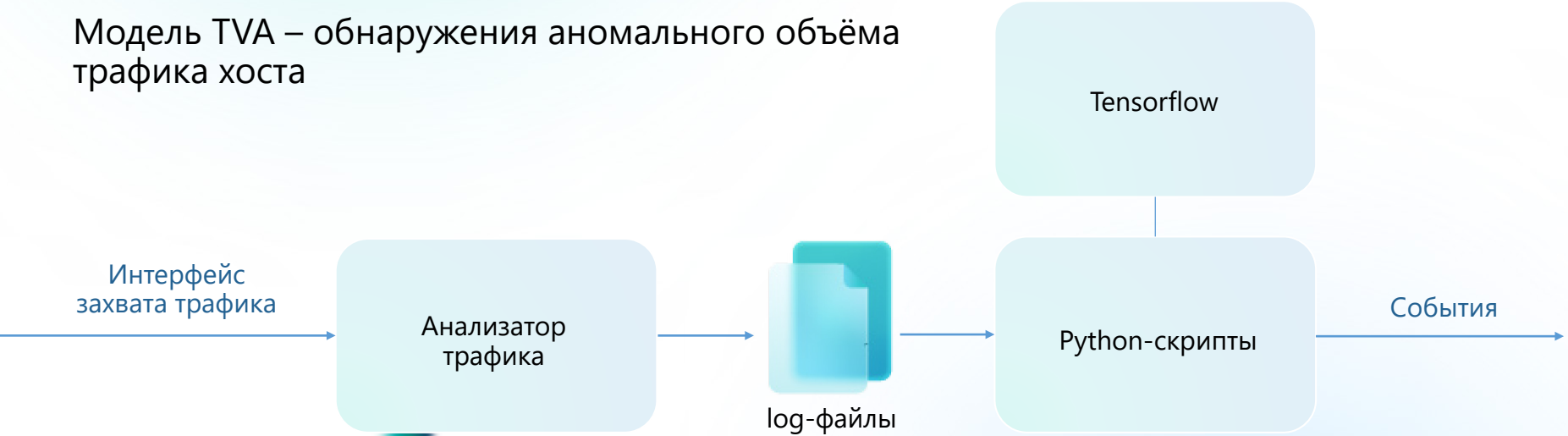
Постобработка выходных данных модели

Модель ML по возможности не принимает итоговое решение, а предобрабатывает данные для его принятия.

Итоговое решение принимает детерминированный алгоритм.

Выбор компонентов для реализации ML-модели на основе кейса компании

Модель TVA – обнаружения аномального объёма трафика хоста



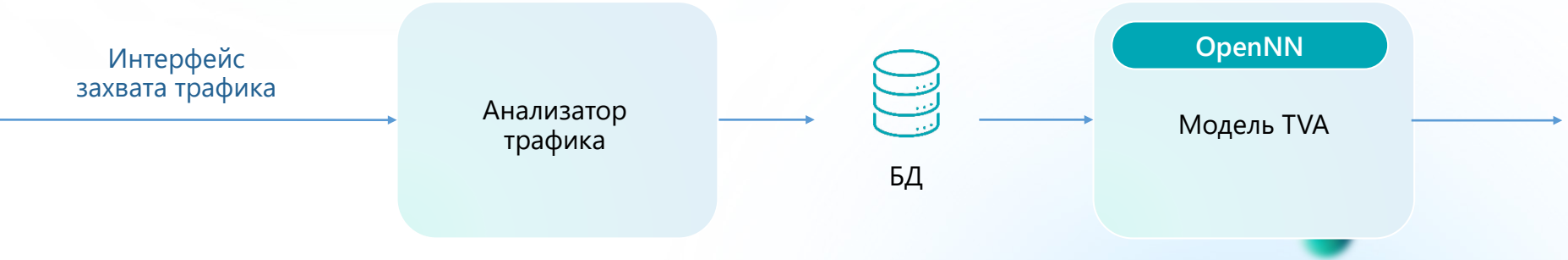
Проблемы реализации

- Log-файлы могут быть изменены,
- Библиотека Tensorflow имеет множество зависимостей, в общем случае используемые не только для ML, есть риск подмены компонентов.

- absl-py-0.1.13.tar.gz
- astor-0.6.2.tar.gz
- bazel-0.11.1-1.el7.centos.x86_64.rpm
- bleach-2.1.3.tar.gz
- chardet-3.0.4.tar.gz
- cudnn-8.0-linux-x64-v7.1.tar.gz
- dev-0.4.0.tar.gz
- elephas-0.3.tar.gz
- gast-0.2.0.tar.gz
- grpcio-1.10.0.tar.gz
- html5lib-1.0.1.tar.gz
- Keras-2.1.5.tar.gz
- numpy-1.14.2.zip
- opencv_python-3.4.0.12-cp36-cp36m-manylinux1_x86_64.whl
- Pillow-5.1.0.tar.gz
- protobuf-3.5.2.tar.gz
- PyYAML-3.12.tar.gz
- scipy-1.0.1.tar.gz
- setuptools-39.0.1.zip
- six-1.11.0.tar.gz
- tensorboard-1.7.0-py3-none-any.whl
- tensorflow-1.7.0-cp36-cp36m-linux_x86_64.whl
- tensorflowonspark-1.2.1-py2.py3-none-any.whl
- termcolor-1.1.0.tar.gz
- tensorflow v1.7.0.tar.gz
- webencodings-0.5.1.tar.gz
- wheel-0.30.0.tar.gz
- h5py-2.7.1-cp36-cp36m-manylinux1_x86_64.whl

Итоговая реализация

- Переход с Python на C++,
- Статическая линковка библиотек,
- Взаимодействие через БД с ограничением доступа.



Итог:

- Тестирование моделей ML и проверка гипотез может осуществляться с использованием скриптовых языков (Python), но внедрение в продукты желательно с применением компилируемых языков и статической линковкой библиотек,
- Есть смысл использовать более сложную обработку трафика для устранения уязвимостей,
- Желательно применение собственных решений для обработки трафика вместо свободного ПО.

Ответы на вопросы

infotecs

Спасибо за внимание!

infotecs.ru

Карьера

infotecs.team