Методы обработки трафика для улучшения качества обнаружения сетевых атак при помощи нейронных сетей

Николай Змитрович
Исследователь ЦНИПР

Методы обнаружения компьютерных атак

Обнаружение признаков атак

- Сигнатуры
- Эмпирические правила
- Экспертные системы

Обнаружение аномалий

- Статистические методы
- Методы ML

Обнаружение признаков атак

Основаны на экспертных знаниях о **нелегитимном** поведении

Преимущества

- Высокая точность
- Производительность

Слабые стороны

- Возможность уклонения от известных методов обнаружения
- Использование легитимных инструментов
- Уязвимость к атакам нулевого дня

Обнаружение аномалий

Основаны на изучении **нормального** поведения

Преимущества

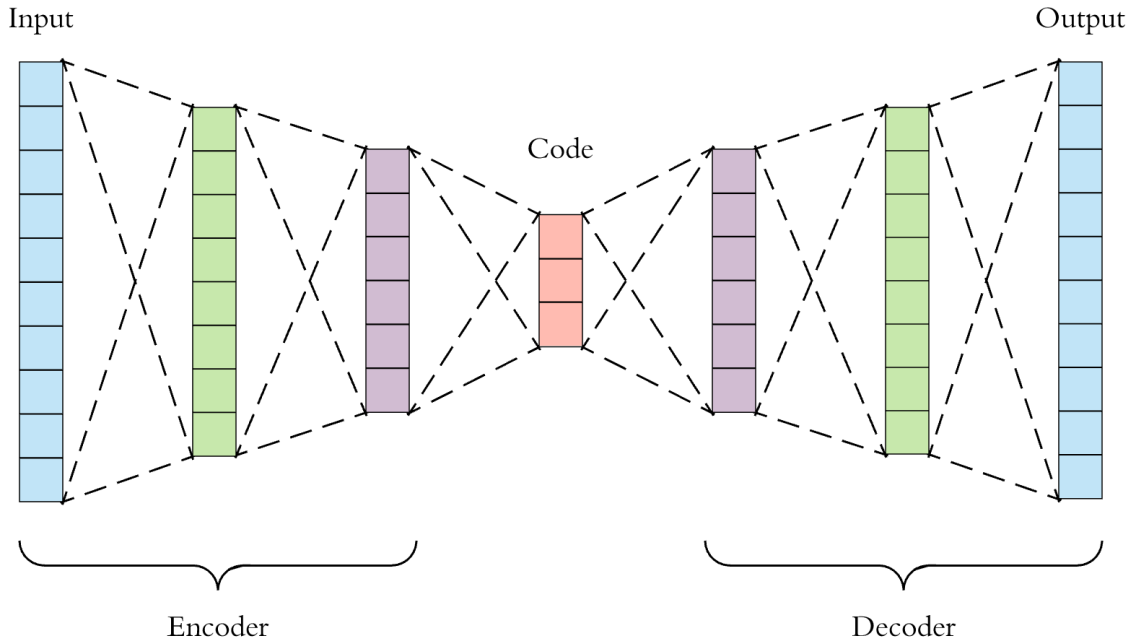
- Универсальность
- Затруднение обхода защиты
- Не требуют частого ручного обновления

Слабые стороны

- Требовательность к ресурсам для обучения модели
- Определение типа атаки

Архитектура автокодировщика

Архитектура нейронных сетей, основанная на восстановлении входного объекта из его сжатого (закодированного) представления

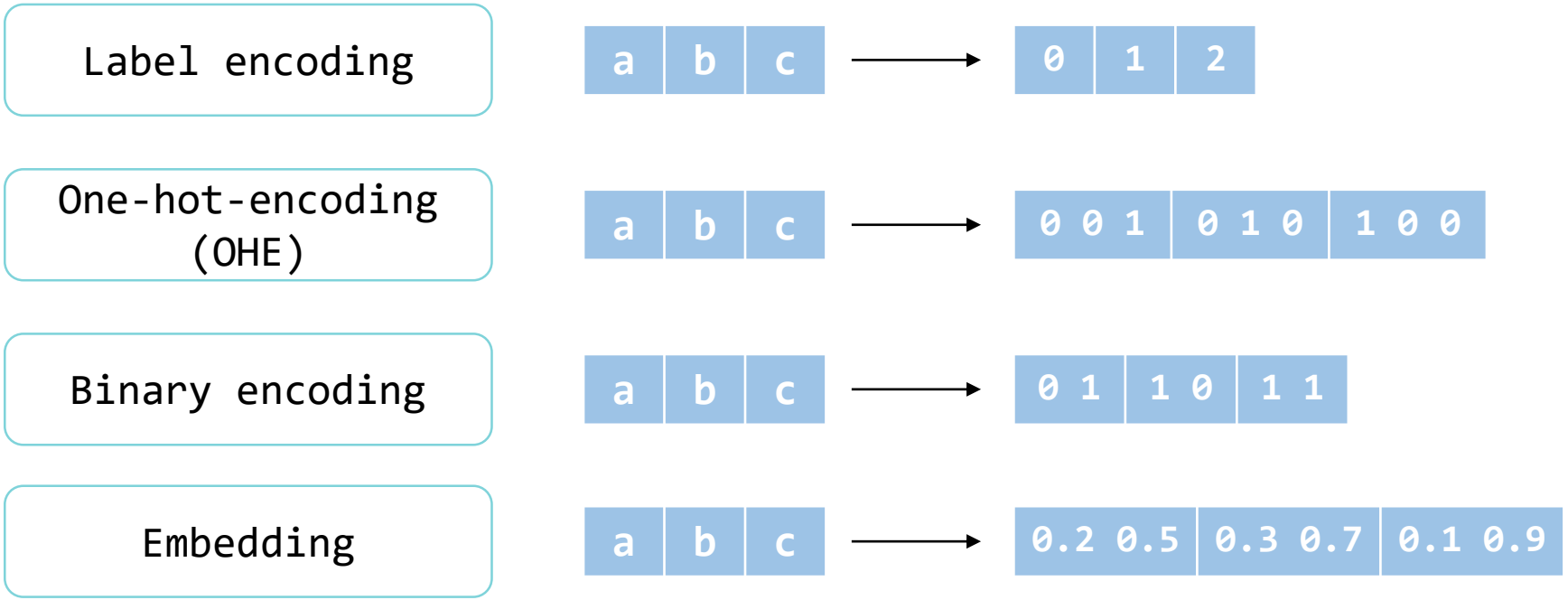


Модель данных

Объект анализа - сетевая сессия

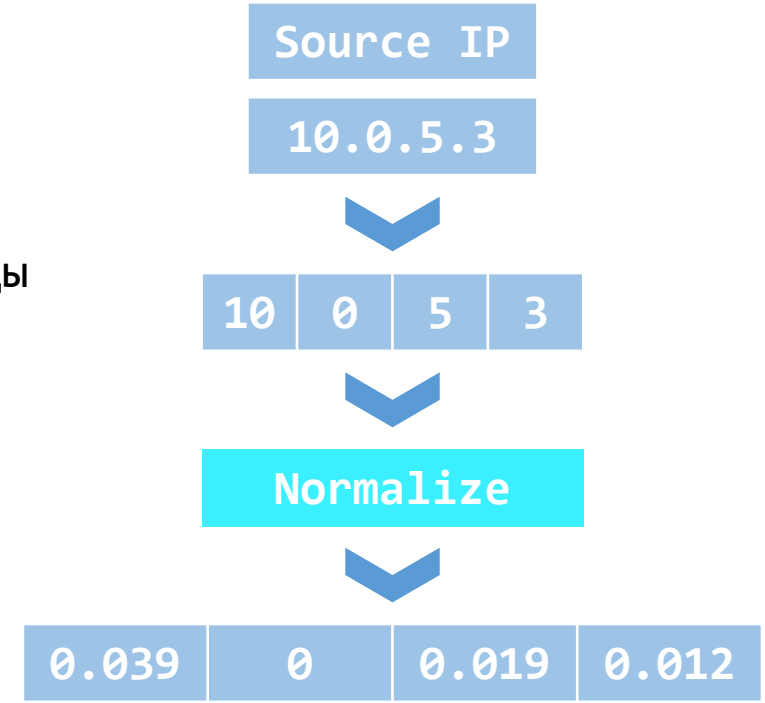
Параметры сессии	Вид параметра
Идентификаторы хостов (IP, MAC адреса)	Категориальный
Номер порта	Категориальный
Протокол	Категориальный
Длительность сессии	Численный
Количество пакетов и байт	Численный
Агрегирующие функции параметров сессий	Численный
Информация о флагах	Численный
…	…

Оцифровка категориальных параметров

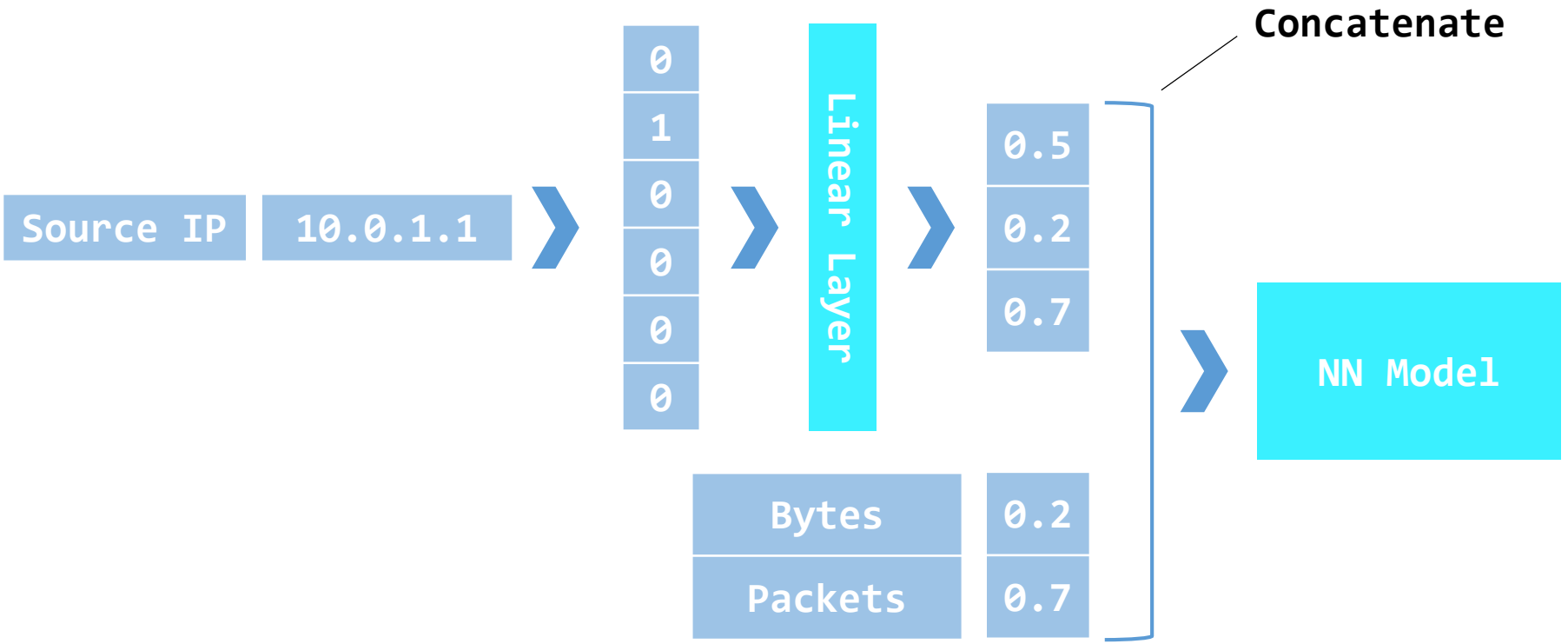


Метод оцифровки IP-адреса

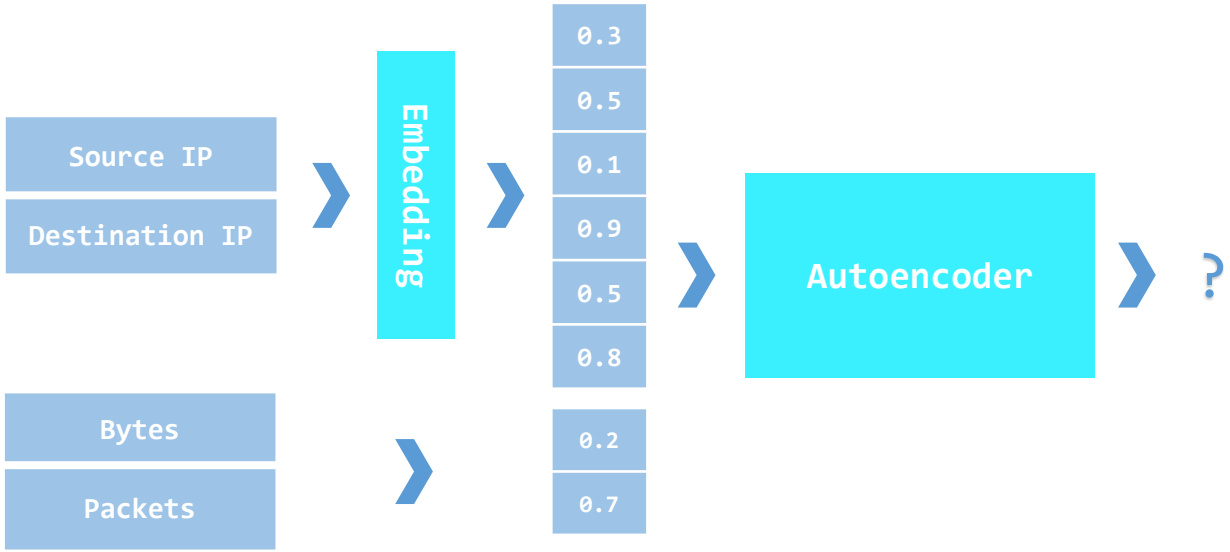
- Используется разбиение адреса на байты
- После нормализации столбцы конкатенируются с другими параметрами потока



Слой Embedding



Проблемы Embedding слоев в архитектуре автокодировщика

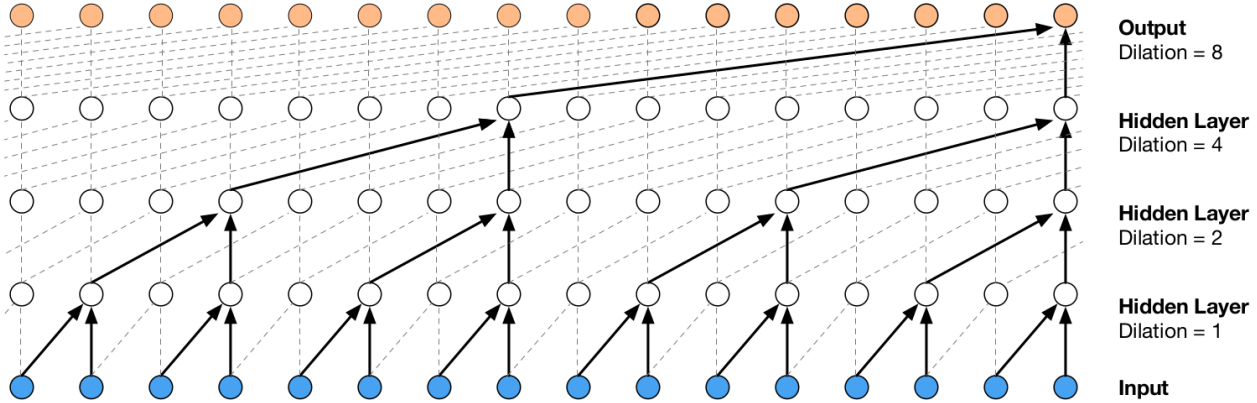


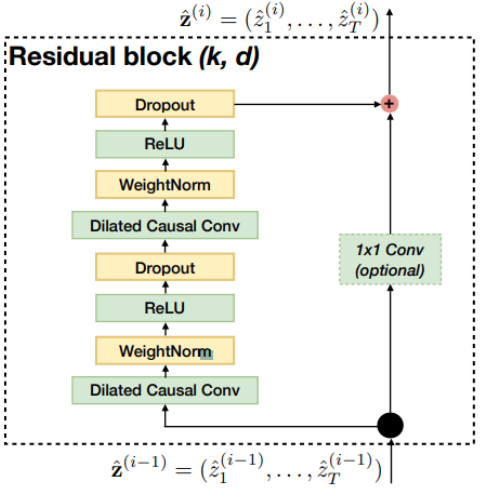
Вариации архитектур

1. Восстановление значений Emdedding-a
2. Восстановление на выходе целочисленных меток
3. **Восстановление OHE категориальных значений**

Temporal Convolutional Networks

Архитектура одномерной сверточной нейронной сети, использующая временну́ю свертку

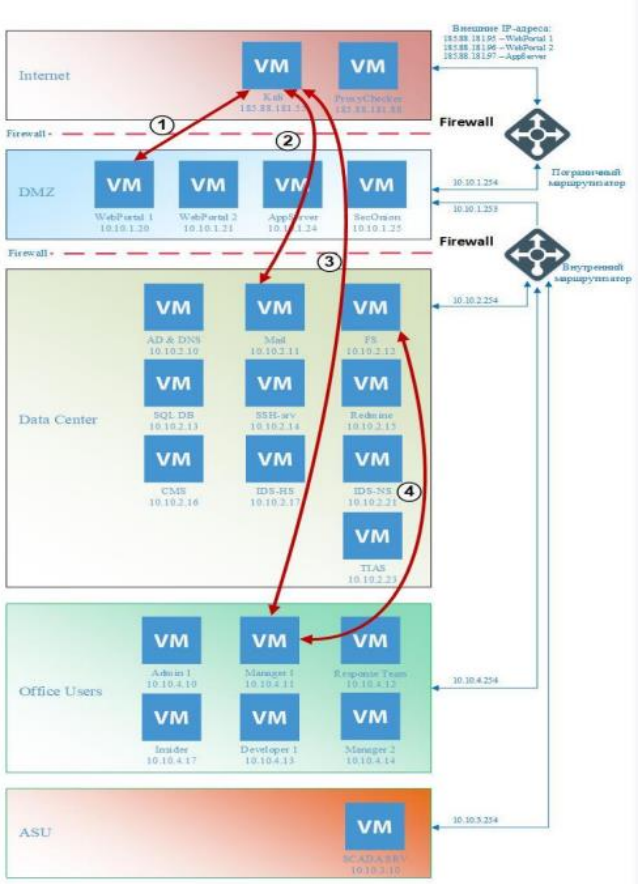




Набор данных

Набор данных основан на нормальном трафике сети ИнфоТеКС и APT-атаке, сгенерированной при помощи инструмента Ampire

Сессий нормального трафика: 325 968
Сессий атаки: 2834



Сценарий атаки

1. Сканирование активных хостов на предмет открытых HTTP/HTTPS портов.
2. Определение логина обнаруженной на этапе сканирования OWA с полученными аутентификационными данными менеджера, определение адреса RDP из электронного письма в OWA.
3. Подключение и получение шелла от хоста менеджера с помощью бота.
4. Сканирование доступных сегментов сети на предмет поиска файлового сервера.
5. Сканирование обнаруженного файлового сервера на предмет обнаружения уязвимости MS17-010.
6. Эксплуатация уязвимости MS17-010, получение доступа к файловому серверу.

Модель Temporal Convolutional Autoencoder (TCAE)

Методы оцифровки категориальных параметров:

Server Port – Label Encoding

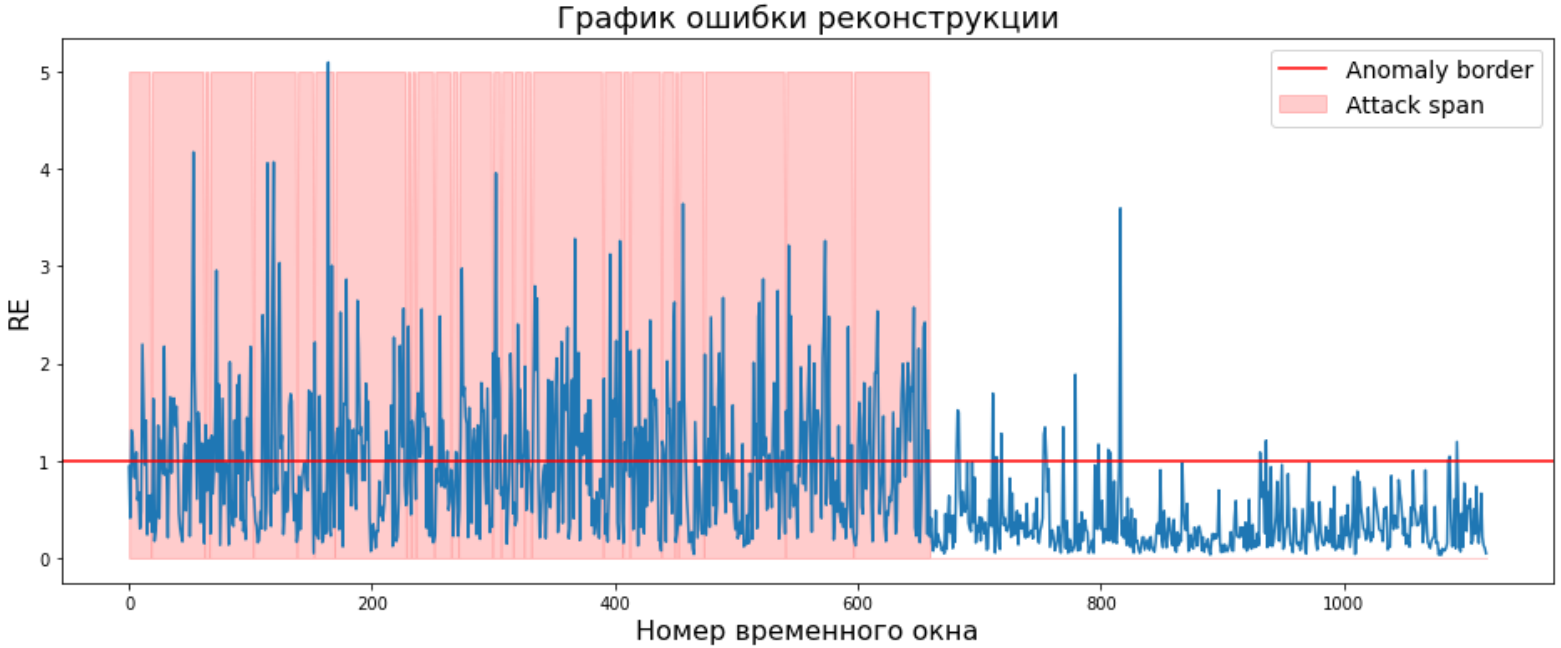
Protocol – Label Encoding

Source IP – разделение на байты

Destination IP – разделение на байты

Функция ошибки – **MSE loss** для всех параметров

Модель TCAE – график ошибки реконструкции



Модель TCAE Double Loss Function (DLF)

Методы оцифровки категориальных параметров:

Server Port – Binary Encoding

Protocol – Binary Encoding

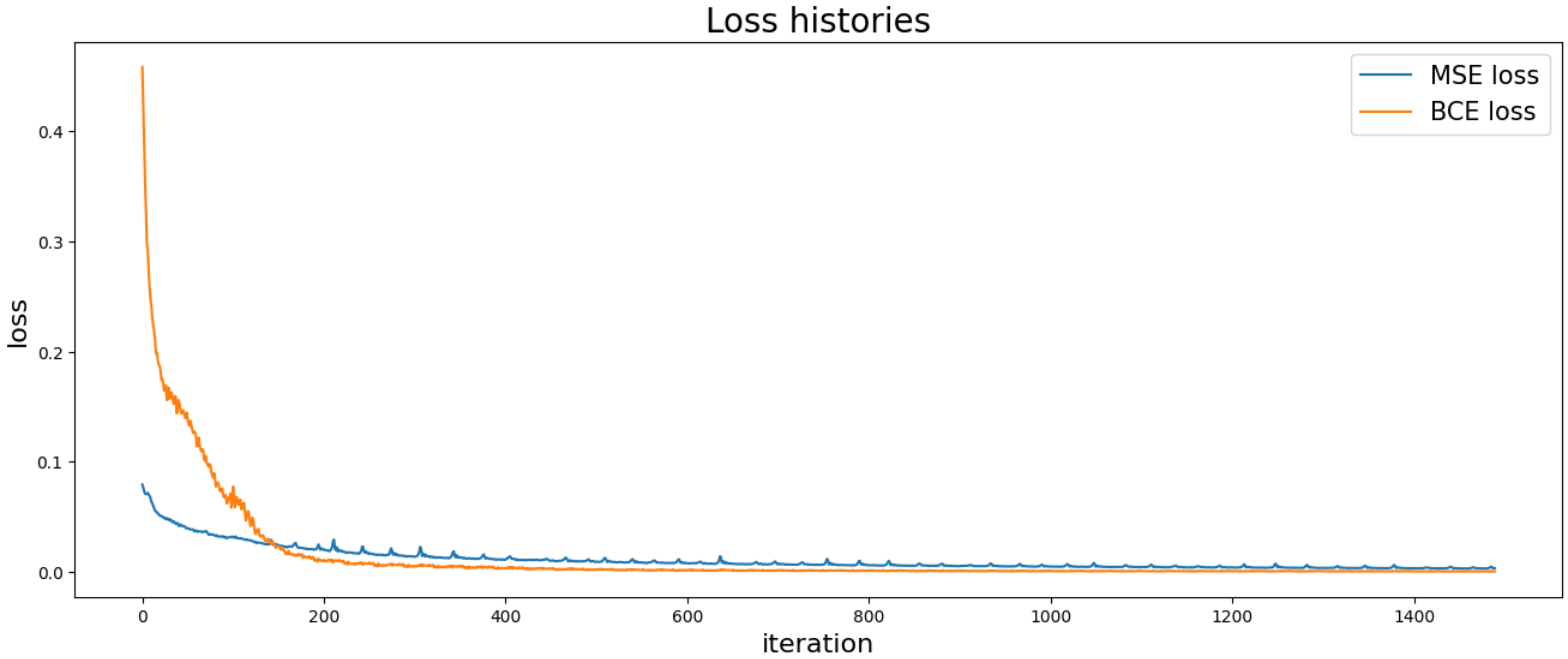
Source IP – разделение на байты

Destination IP – разделение на байты

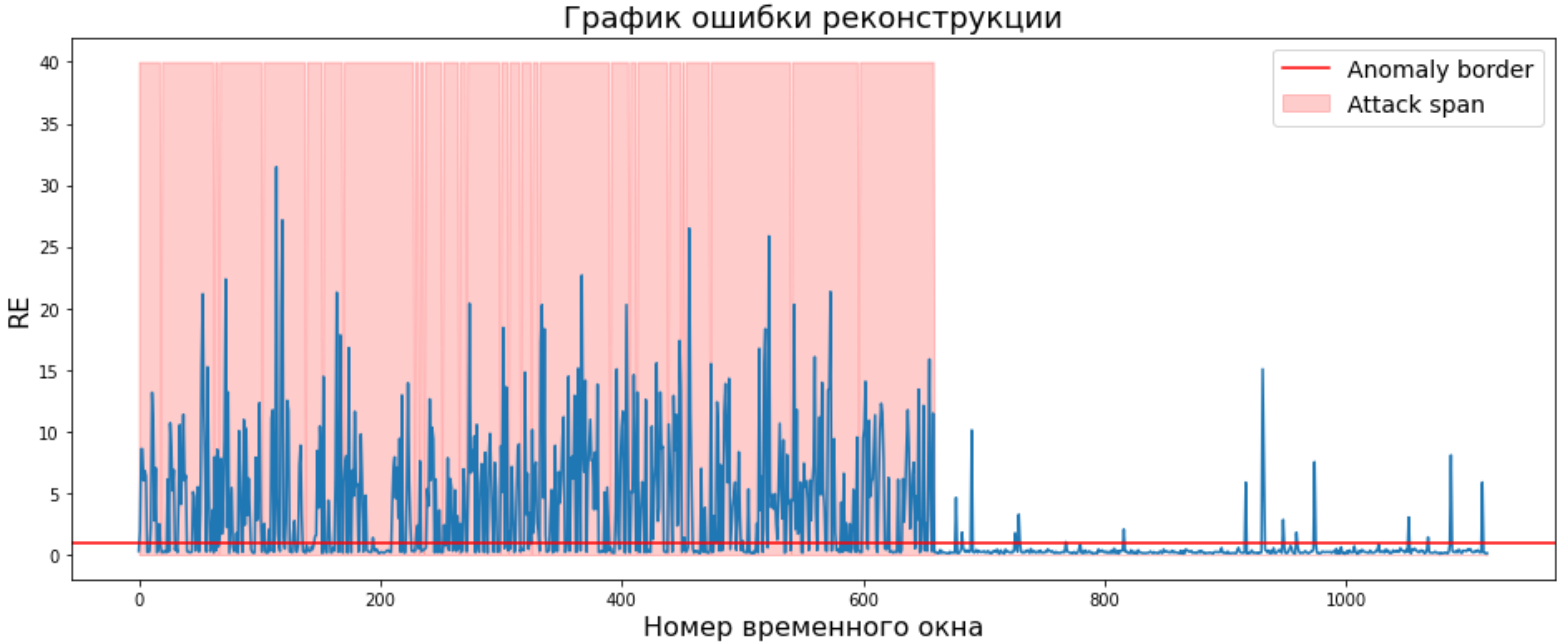
Функция ошибки – **MSE loss** для всех параметров кроме Server Port и Protocol

Функция ошибки – **BCE loss** для параметров Server Port и Protocol

Модель TCAE DLF – график ошибок при обучении модели



Модель TCAE DLF – график ошибки реконструкции



Модель TCAE Embedding (EMB)

Методы оцифровки категориальных параметров:

Server Port – Embedding

Protocol – Embedding

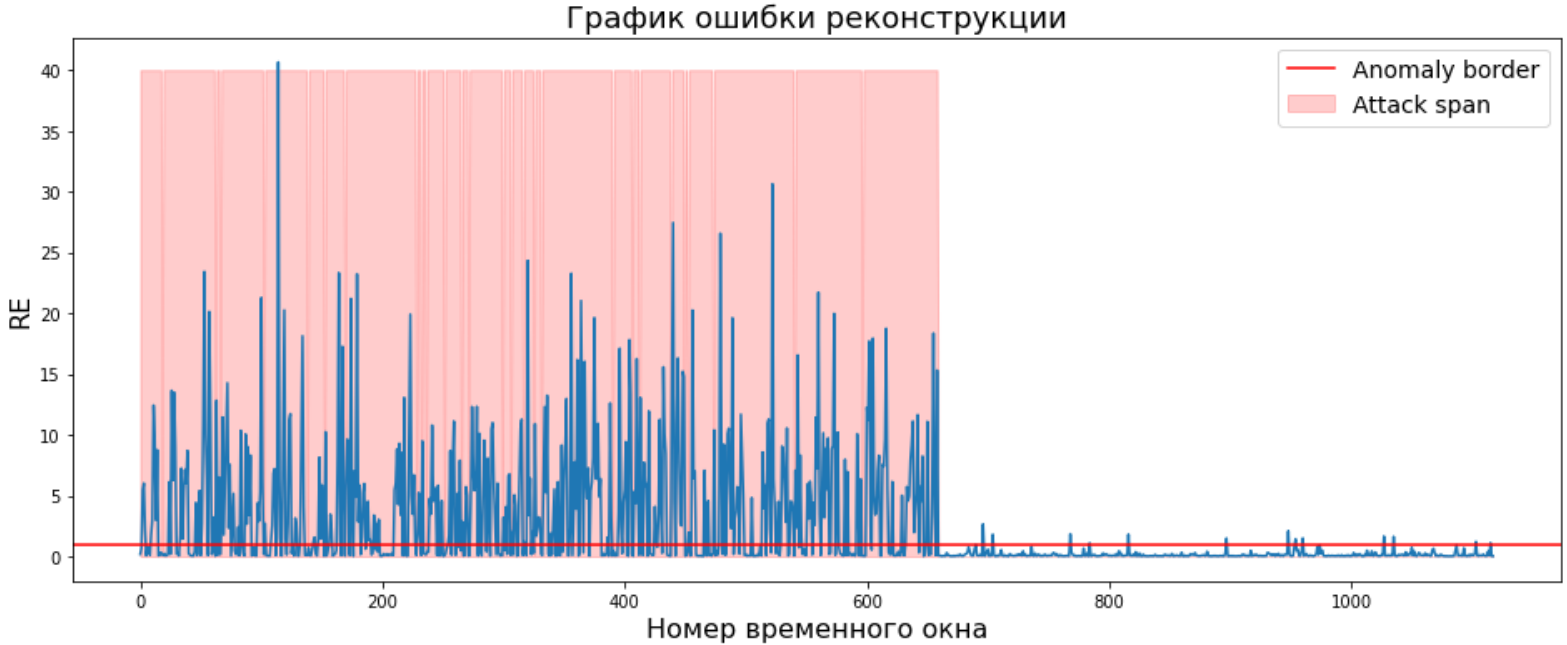
Source IP – Embedding

Destination IP – Embedding

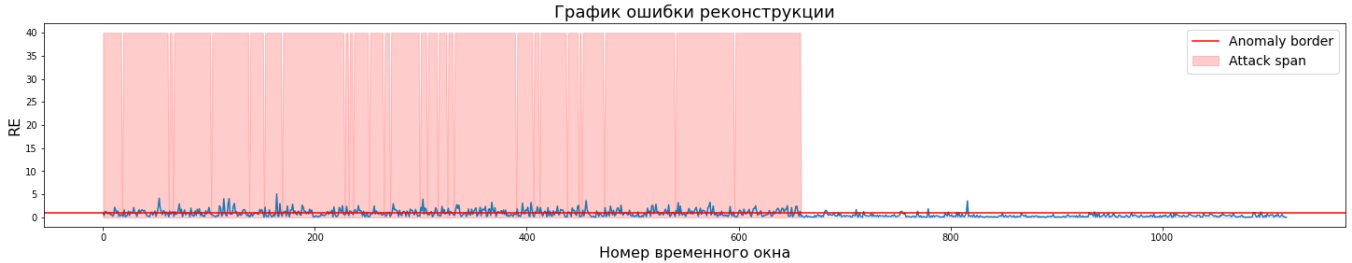
Функция ошибки – **MSE loss** для всех параметров кроме Server Port, Protocol и IP адресов

Функция ошибки – **Cross Entropy loss** для параметров Server Port, Protocol и IP адресов

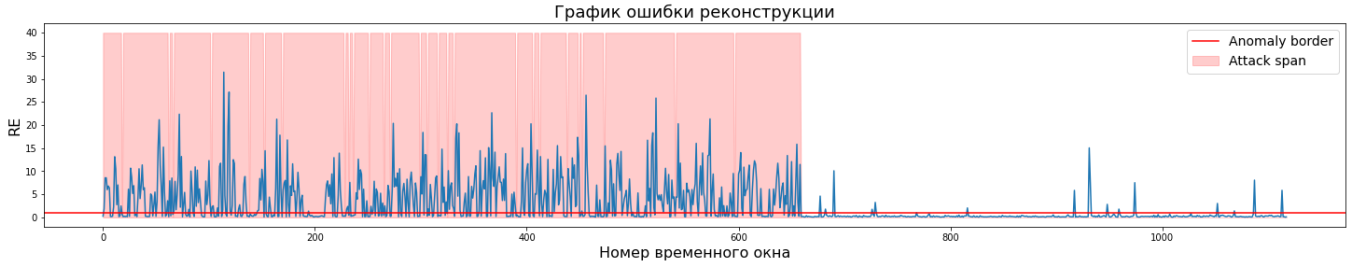
Модель TCAE DLF+EMB – график ошибки реконструкции



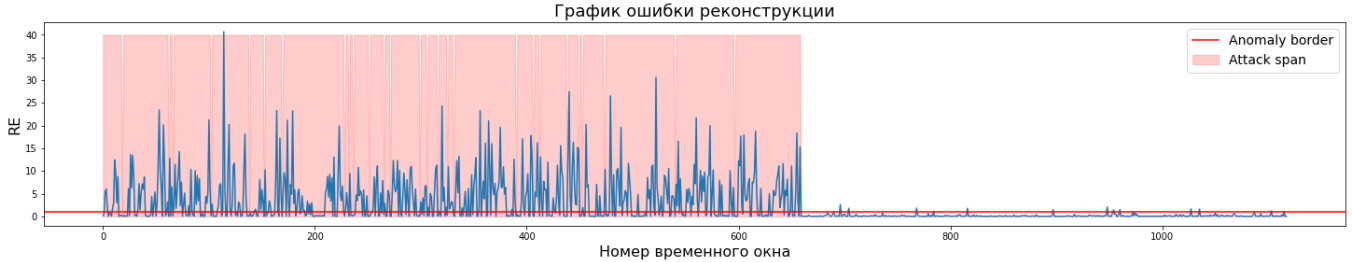
Сравнение результатов - графики



Model TCAE



Model TCAE+DLF



Model TCAE+DLF+EMB

Сравнение результатов - метрики

Model	BA	Precision	Recall	F1
TCAE	0.62	0.89	0.28	0.42
TCAE DLF	0.77	**0.95**	0.57	0.71
TCAE DLF+EMB	**0.91**	0.91	**0.93**	**0.92**

Balanced Accuracy – показывает процент верных предсказаний с учетом дисбаланса классов

Precision – показывает отношение верно предсказанных аномалий ко всем предсказанным аномалиями (чем выше, тем меньше FP)

Recall – показывает отношение верно предсказанных аномалий ко всем аномалиям в наборе данных (чем выше, тем меньше FN)

Выводы

- Способ оцифровки данных существенно влияет на качество обнаружения атак.
- Совместное использование двух различных функций ошибки дает значительный прирост в качестве.
- Использование дополнительных Embedding слоев позволяет улучшить качество обнаружения атак, но при этом требует больше ресурсов для обучения модели.

ТЕХНО 2023 infotecs ФЕСТ

Спасибо за внимание!

Подписывайтесь на наши соцсети

vk.com/infotecs_news

https://t.me/infotecs_official

rutube.ru/channel/24686363