

# Multivariate Industrial Time Series with Cyber-Attack Simulation: Fault Detection Using an LSTM

Артем Воронцов, Андрей Лаврентьев, Павел Филонов

Technology Research, Future Technologies

Kaspersky Lab

# Содержание

- Защита объектов критической инфраструктуры
- Источники данных
- Задача обнаружения атак на индустриальный объект
  - Описание данных
  - Формализация задачи
  - Архитектура прогнозной сети
  - Сравнение с известными методами
  - Оценки метрик качества
- Заключение

# Кибератаки на индустриальные объекты

- Иран в 2010 году [1]
  - вывод из строя множества центрифуг по обогащению урана
- Германия в 2014 году [2]
  - вывод из строя доменной печи на сталелитейном заводе
- Украина в 2014 году (BackEnergy)
  - на время было отключено электроснабжение нескольких районов

# Уровни защиты

- Уровень оконечных узлов
- Сетевое взаимодействие
- Контроль целостности технологических процессов
  - Методы основанные на правилах
  - Методы основанные на данных

# Сравнение подходов

## Rules Based Systems

- Прозрачные
- Нужен эксперт
- Ловят только то, что знают
- Сложны в реализации
- Долгое внедрение
- Ложные срабатывания

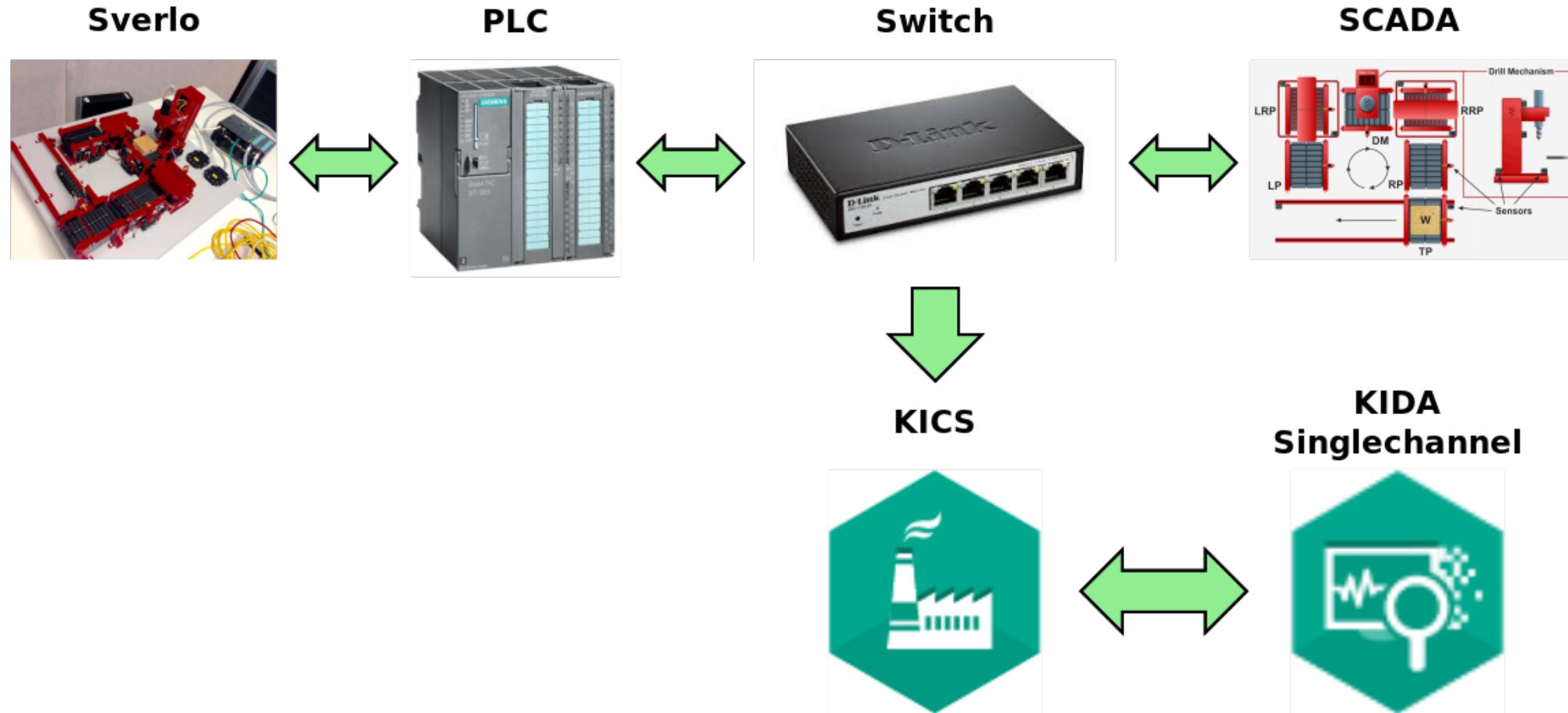
## Machine Learning

- Непрозрачные
- Нужны данные
- Могут поймать новые атаки
- Просты в реализации
- Быстрое внедрение\*
- Ложные срабатывания\*\*

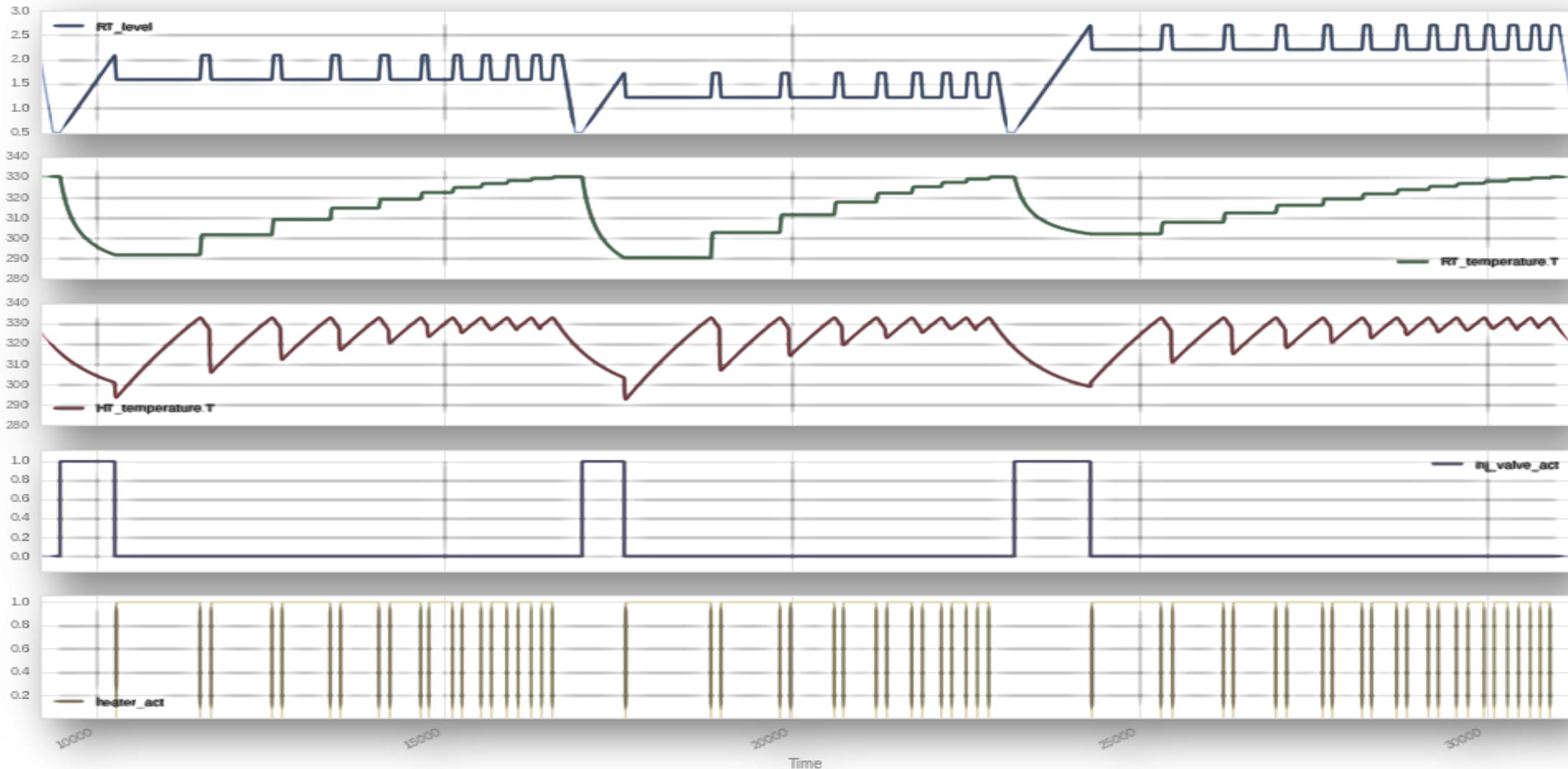
\* Нужно собрать данные

\*\* Можно легко управлять

# Приципиальная схема потоков данных



# Многомерный временной ряд



# Модель данных и характеристики

- Схема

```
{
```

```
  "id": 42,  
  "ts": 1023213221,  
  "value": 3.14
```

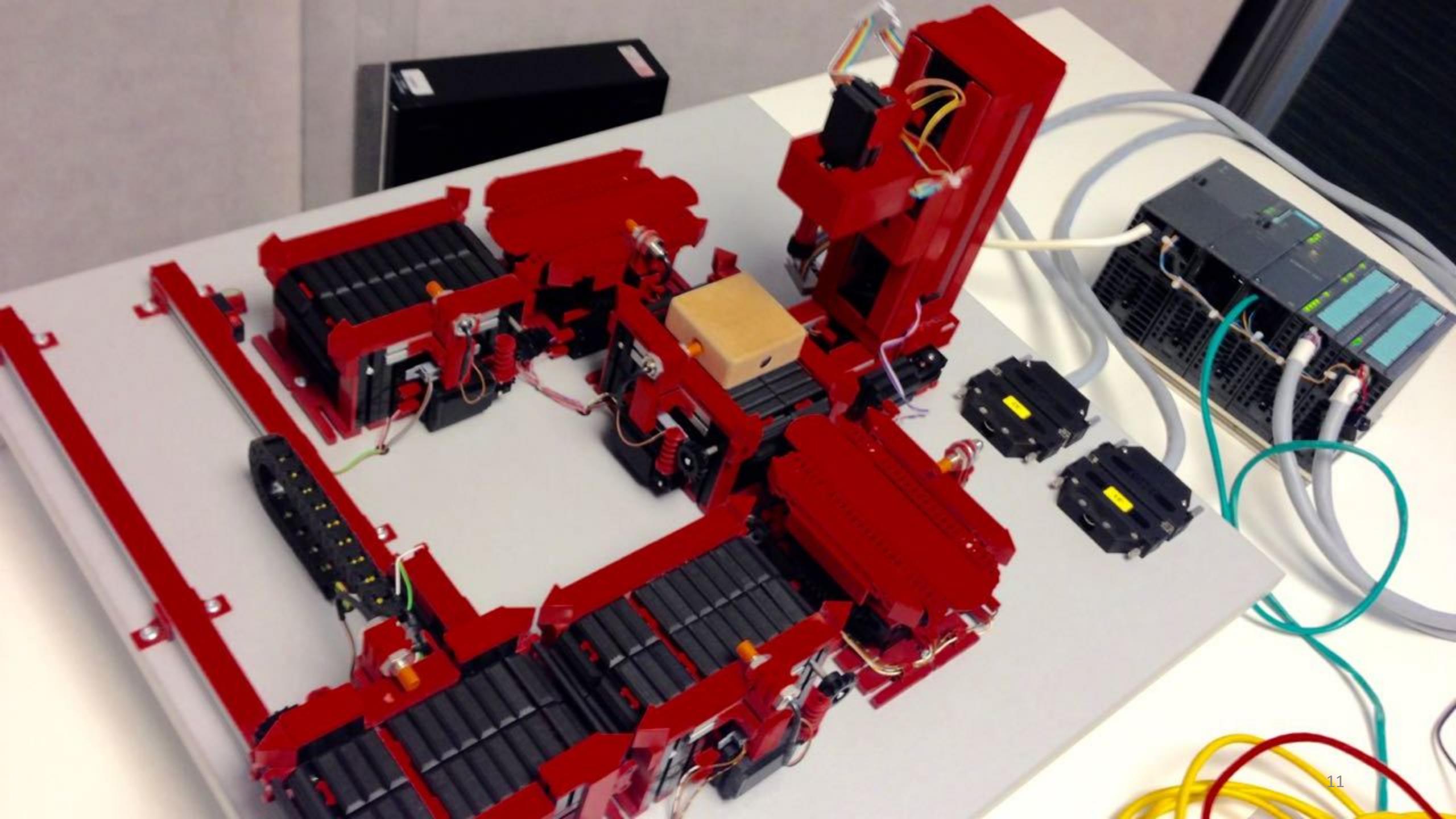
```
}
```

- Число сенсоров  $\sim 10^4$
- Частота опроса сенсоров  $\sim 10$  раз в секунду

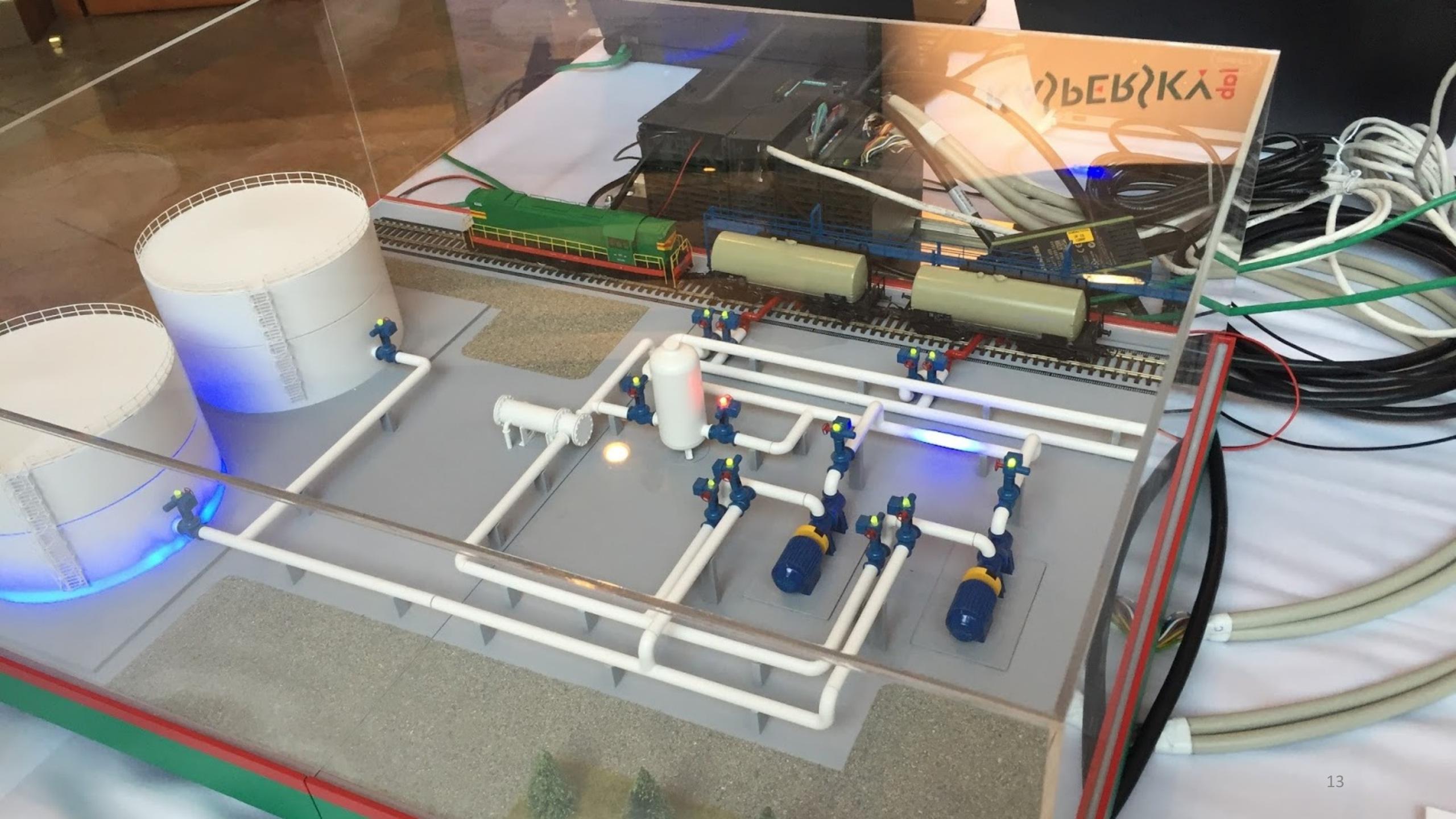
# Источники данных

- Реальные объекты
- Натурные модели
- Компьютерные модели



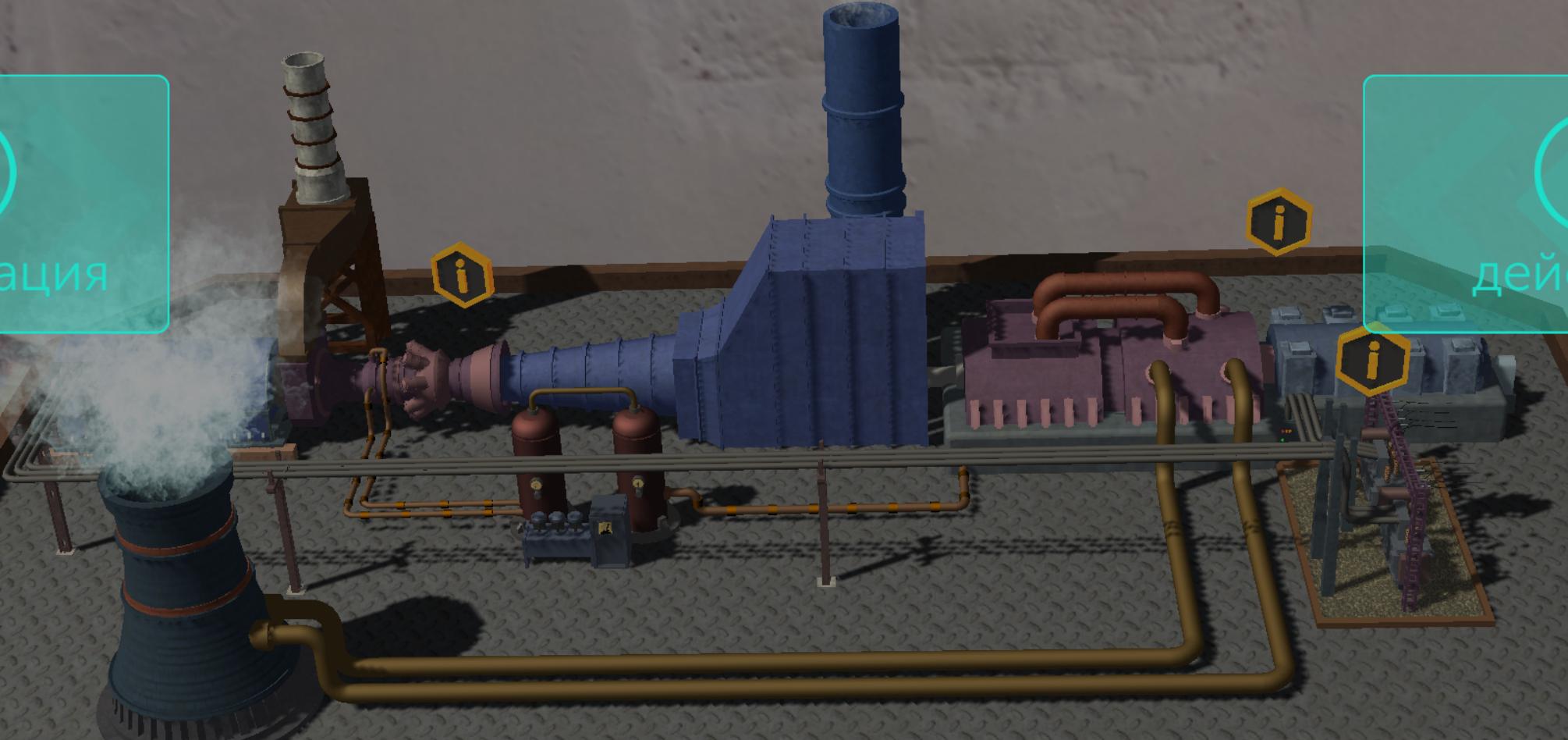








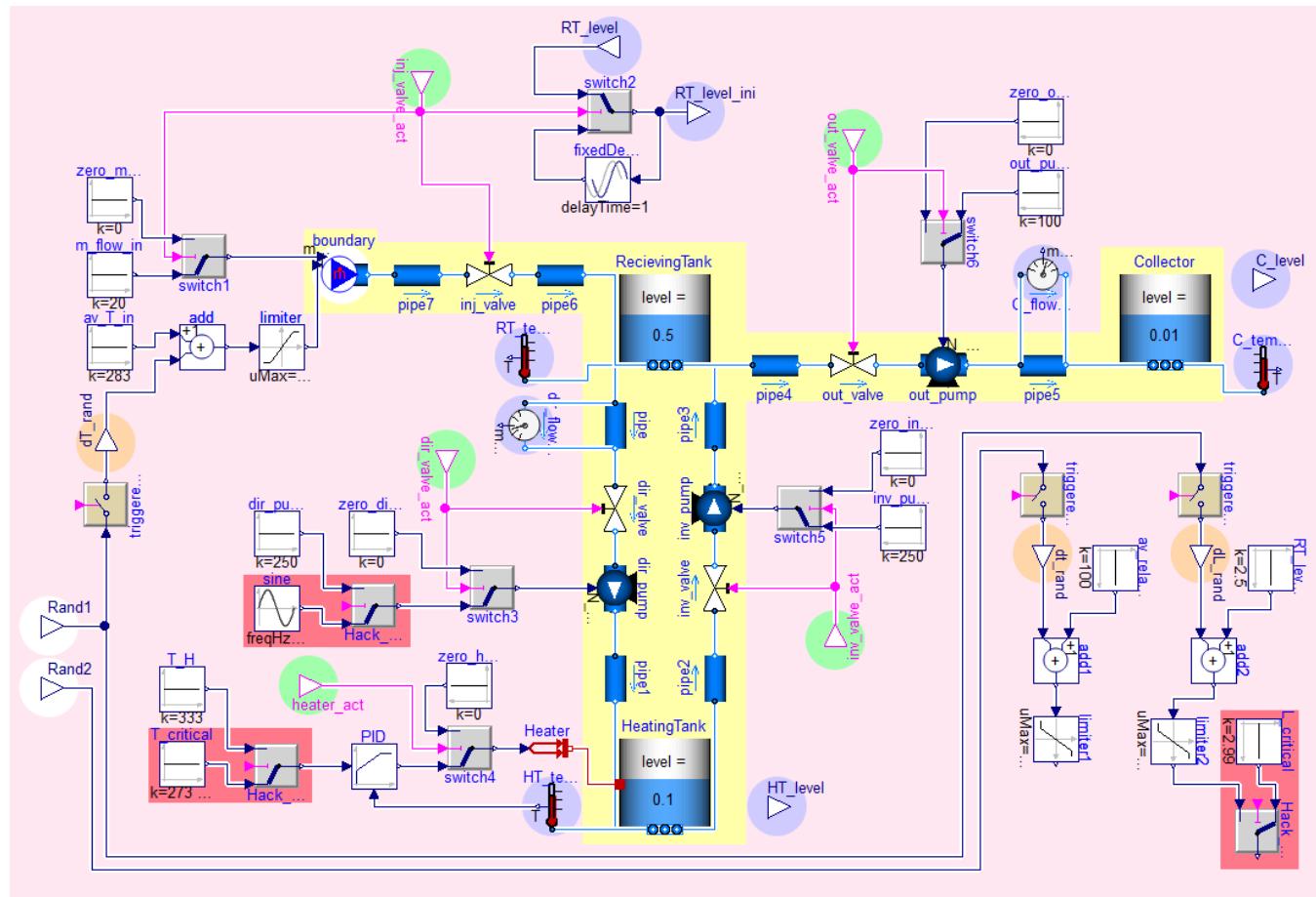
информация



действия

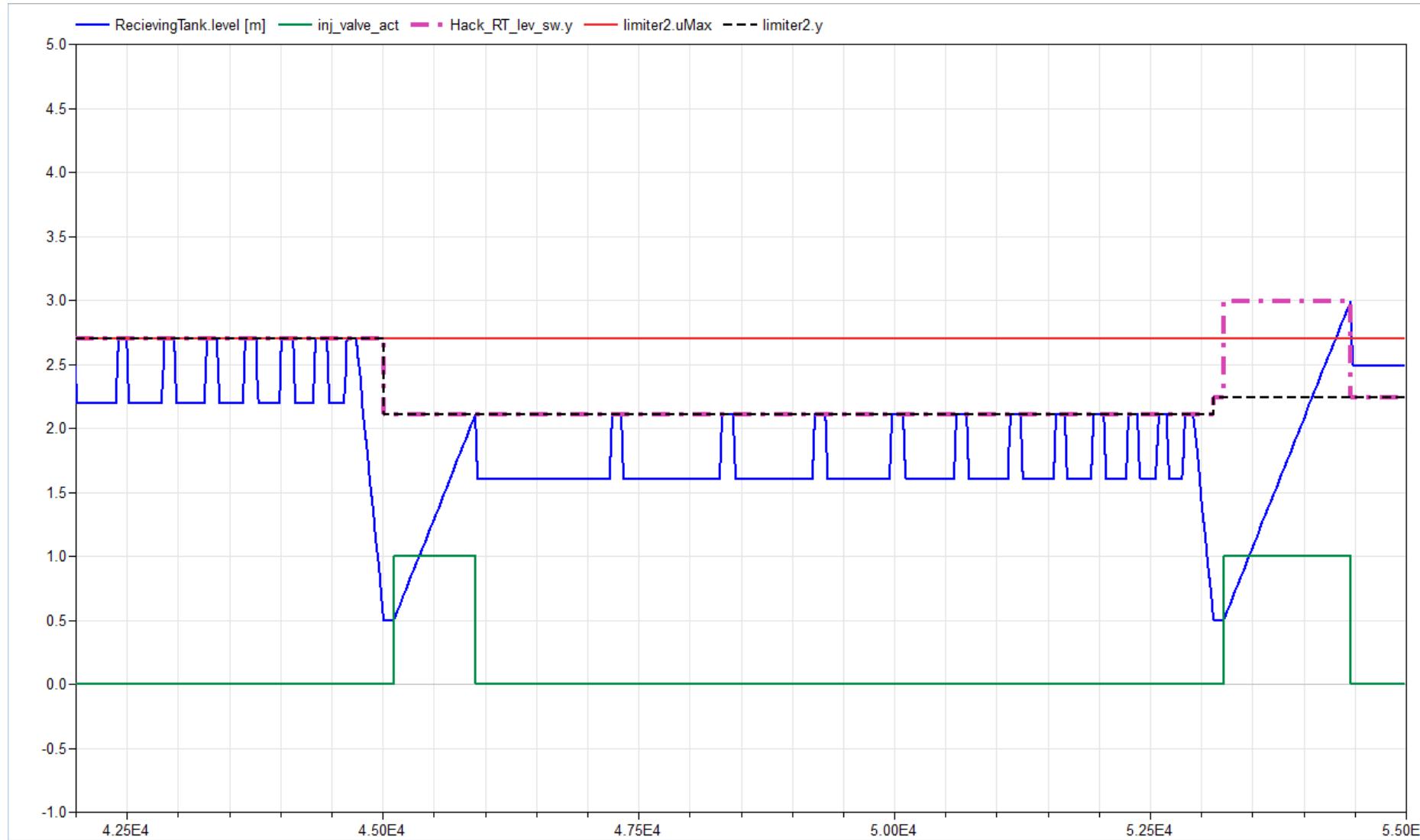
	<b>Физика</b>	<b>ПЛК</b>	<b>Полнота</b>	<b>Данные</b>
Реальный объект	Реальная	Реальные	Только «хорошие» примеры	Трудно добывать данные
Натурная модель	Упрощенная	Реальные	Мало «плохих» примеров	Долго добывать данные
Компьютерная модель	Модельная	Модельные	Много различных примеров	Легко добывать данные

# Gasoil Heating Loop



■ Technology loop  
■ Malicious logic for cyber-attack simulation  
● Control inputs  
● Disturbances  
● Sensors

# Пример атаки на уставку



# Сведение к задаче прогноза

- Представим задачу обнаружения кибер-атаки в виде следующих подзадач
  1. На основе предыдущих значений временного ряда получить прогноз будущих значений
  2. Сравнить посчитанный прогноз с полученными данными со сенсоров. Если ошибка прогноза превышает заданный порог, то считаем, что обнаружено отклонение в технологическом процессе (потенциальная атака)

# Постановка задачи прогноза

- Пусть  $\{x(t_i)\}_{i=1}^m$  – многомерный временной ряд  $x(t_i) \in R^n$
- Необходимо построить отображение (функцию прогноза)

$$F(\{x(t_i)\}_{i=1}^k) = \tilde{x}(t_{k+w}),$$

такую что - средне квадратичная ошибка будет минимальна

$$\|\{x(t_j) - \tilde{x}(t_j)\}\|_{l_2} \rightarrow \min$$

# Постановка задачи обнаружения

- Пусть для проверочной выборки задан прогноз  $\{\tilde{x}(t_j)\}_{j=1}^h$  и реальные значения  $\{x(t_j)\}_{j=1}^h$
- Рассмотрим квадрат ошибки прогноза  $e(t_j) = (\tilde{x}(t_j) - x(t_j))^2$
- Обозначим за порог принятия решения  $H$  0.999 квантиль СВ  $\{e(t_j)\}$
- Для тестовой выборки введем следующее решающее правило для тестовой выборки
  - Если значение  $e(t_j) < \alpha H$ , то система работает в стабильном режиме
  - Если  $e(t_j) \geq \alpha H$ , то наблюдается отклонение в ТП (потенциальная атака)

# Использование RNN

- Формально АСУ ТП может быть представлена как многомерная динамическая система, в которой осуществляется свертка по времени сигналов и управления [refs].
  - Можно показать, что любая динамическая система может быть описана с произвольной точностью с помощью RNN [8, 9]
- 
- Ведется исследование по задаче восстановления вида динамической системы на основе коэффициентов «эквивалентной» RNN

# Использование LSTM

- Коэффициенты нейронной сети Элмана сходятся очень медленно. Результат сходимости сильно зависит от начального приближения
- Характерной корреляционный радиус заранее неизвестен, поэтому хочется использовать работы с небольшими окнами и долгой памятью, сохраняемой между ними.
- Хорошо зарекомендовали себя в близких работах [4, 5]

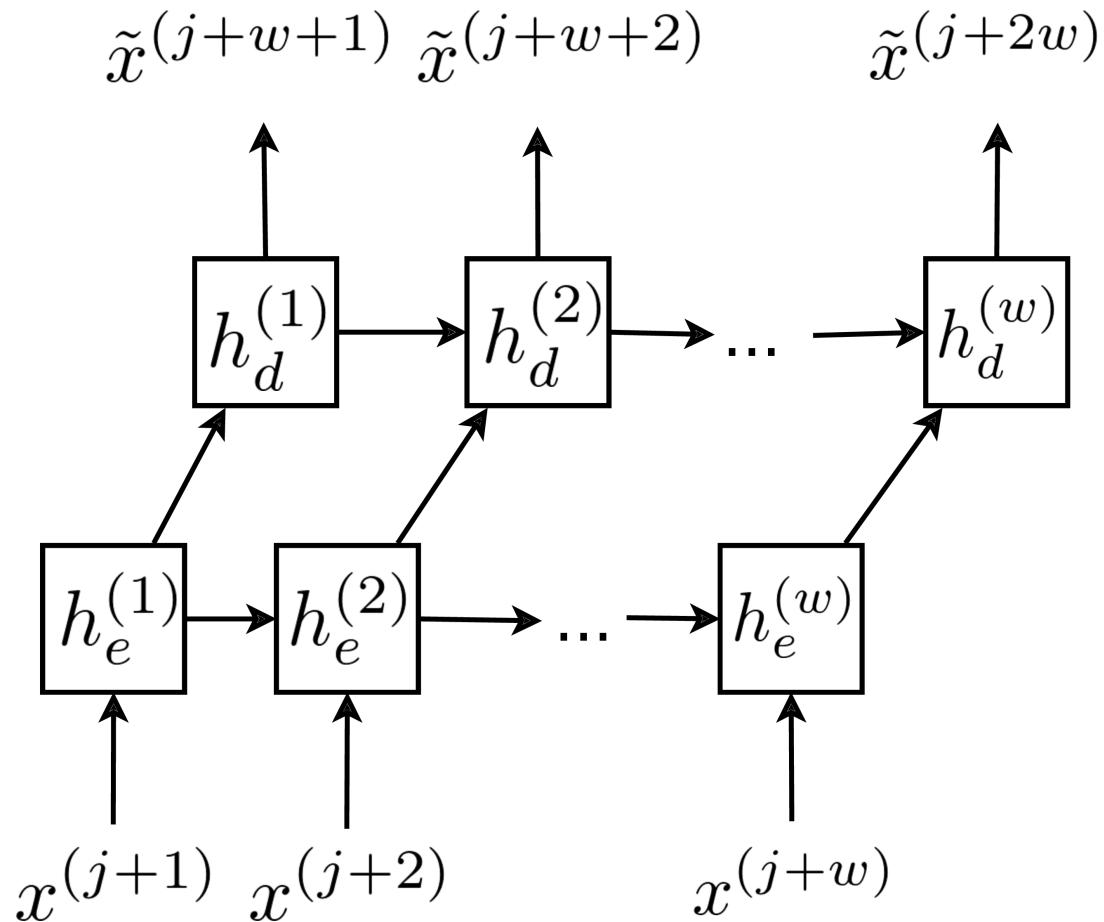
# Обучающая выборка

- АСУ ТП работает в штатном режиме. Атак нет.
- Размерность временного ряда  $n = 17$
- Число точек  $m = 1.5 \cdot 10^6$
- Доля проверочной выборки – 0.1

# Предобработка

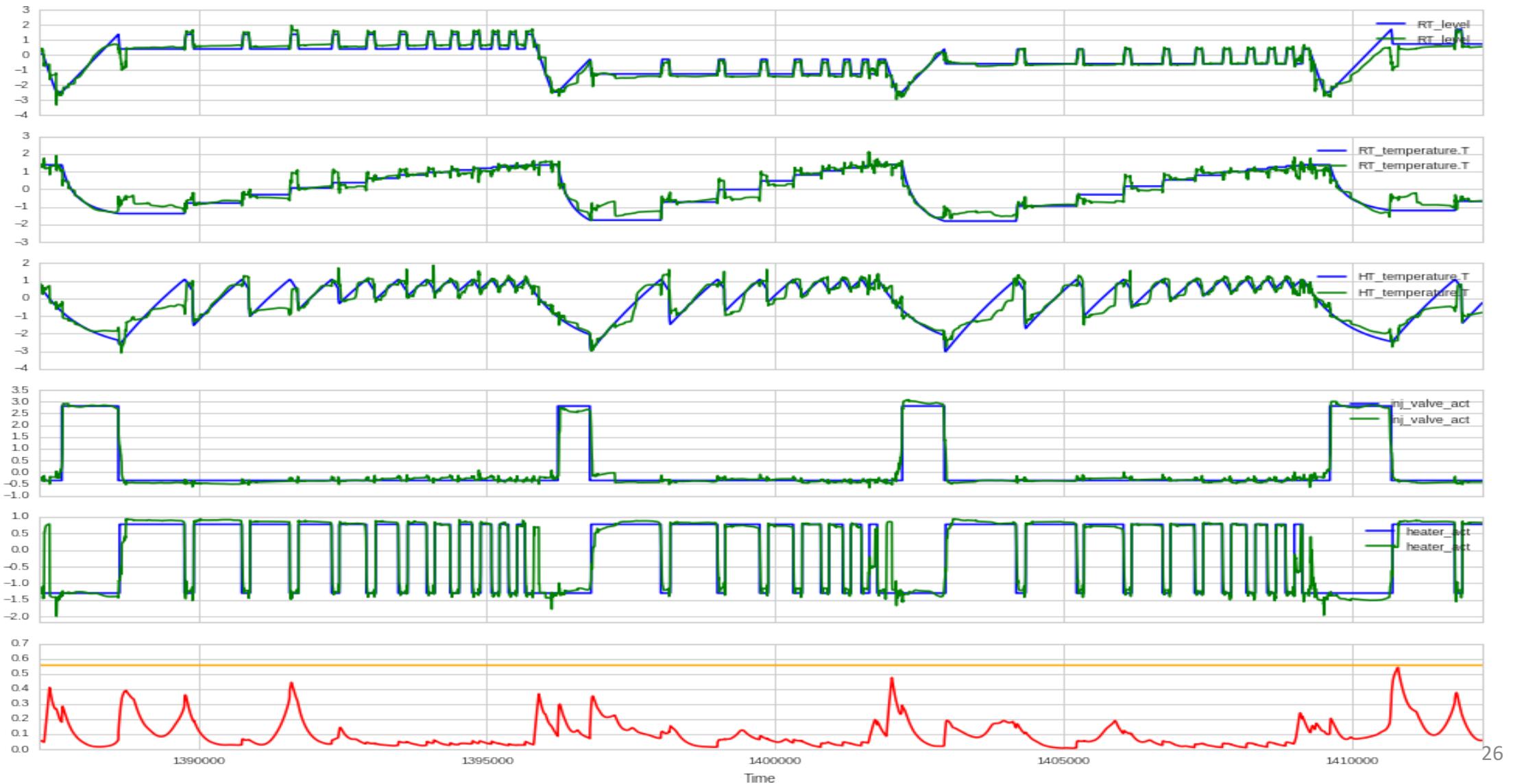
- Нормализация каждого измерения по отдельности
- Число точек усечено до числа кратного размеру горизонта прогноза

# Архитектура сети



- 2 слоя LSTM (synced many-to-many)
- Функция активации
  - Скрытый слой – ReLU
  - Выходной слой – линейный
- Размер входного окна соответствует горизонту прогноза ( $w$ )
- В качестве регуляризации Dropout между слоями
- Алгоритм обучения – RMSProp
- Функция потерь - MSE

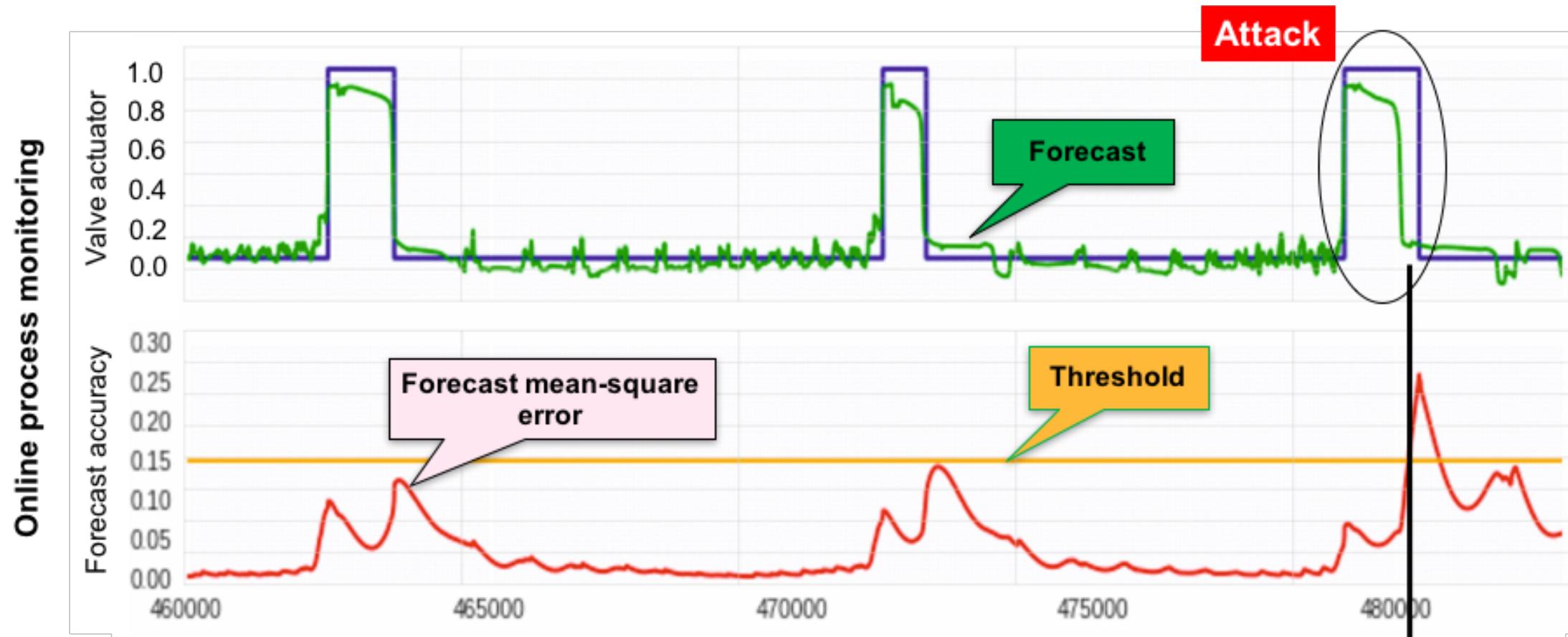
# Результаты экспериментов



# Тестовая выборка

- 49 временных рядов
- Протяженность каждого ряда – 200 000 точек
- В каждом ряду один раз в случайный момент времени возникает атака одного из 3-х типов
  - Атака на уровень RT
  - Атака на температуру HT
  - Атака на насос
- В тестовой выборке присутствуют дополнительные размерности
  - ATTACK – время начала атаки (может не наблюдаться в данных)
  - DANGER – системы вышла за технологические пределы
  - FAULT – точка невозврата

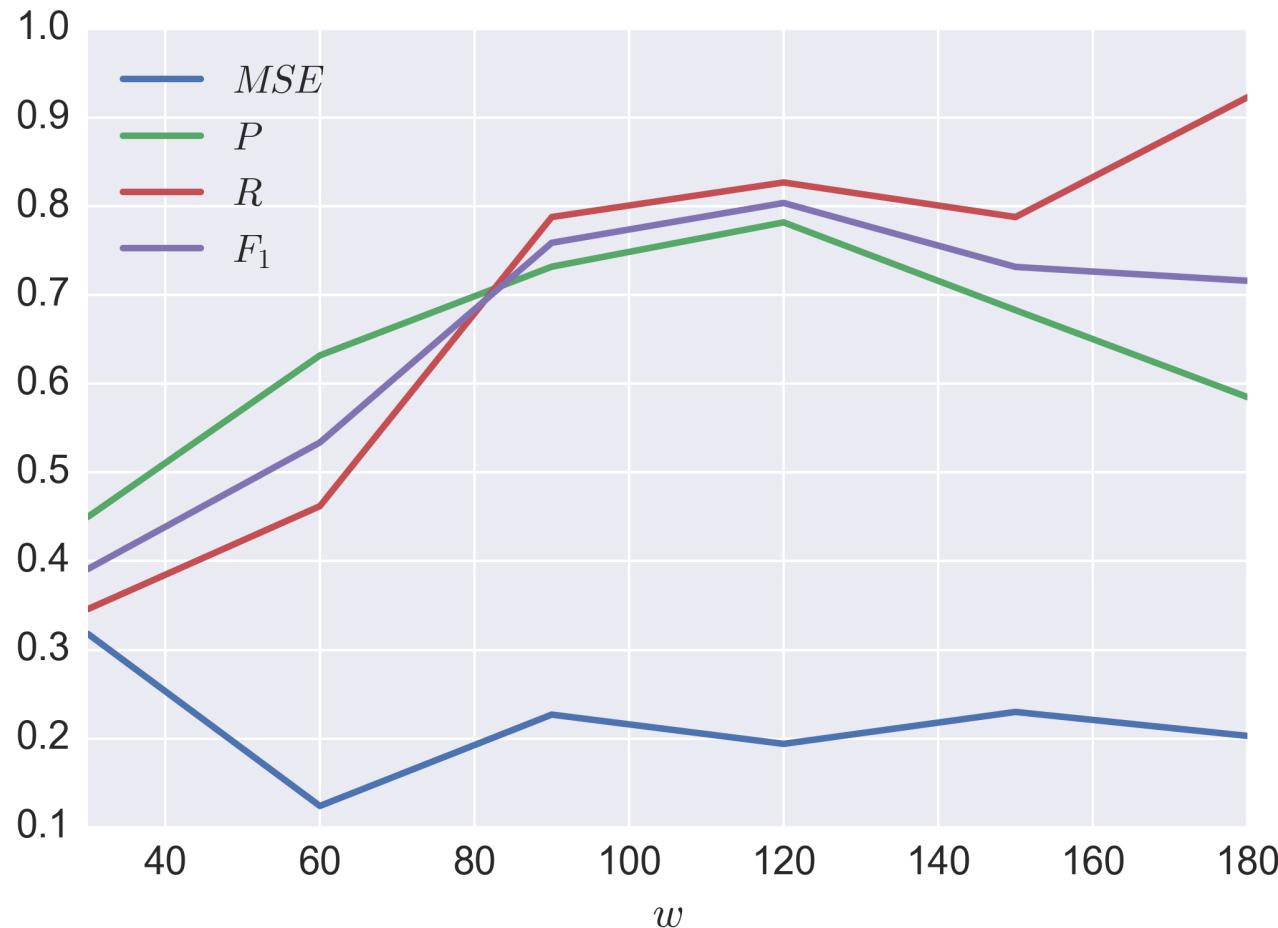
# Пример обнаружения атаки



# Метрики качества

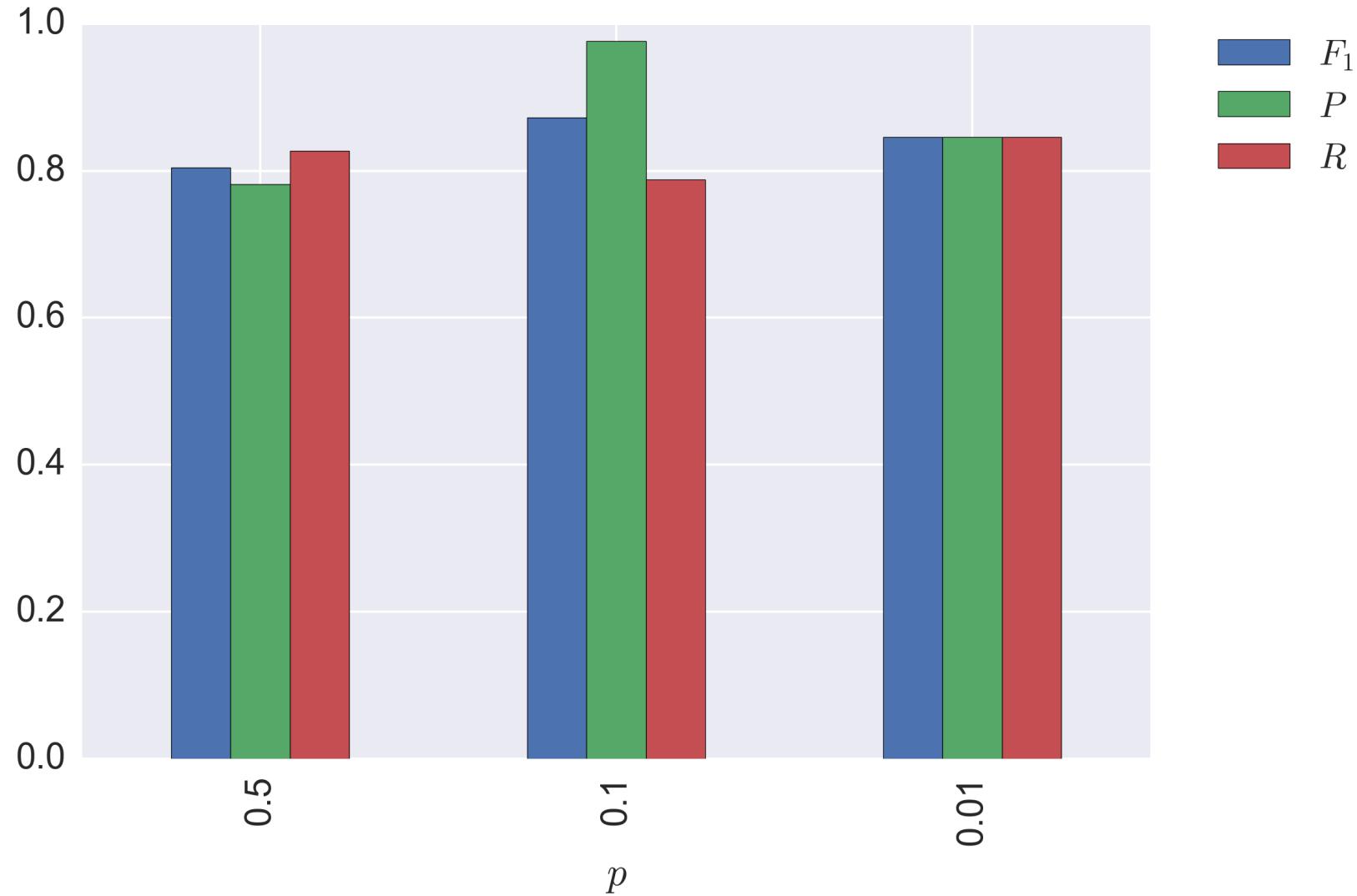
- Разобьем каждый временной ряд в тестовой выборки на интервалы по 10 000 точек. Для каждого интервала сформулируем задачу бинарной классификации – {есть атака, нет атаки}
- Оценим по всем рядам точность  $P$  и полноту  $R$
- В качестве скалярного критерия качества будем рассматривать  $F_1$

# Подбор горизонта прогноза

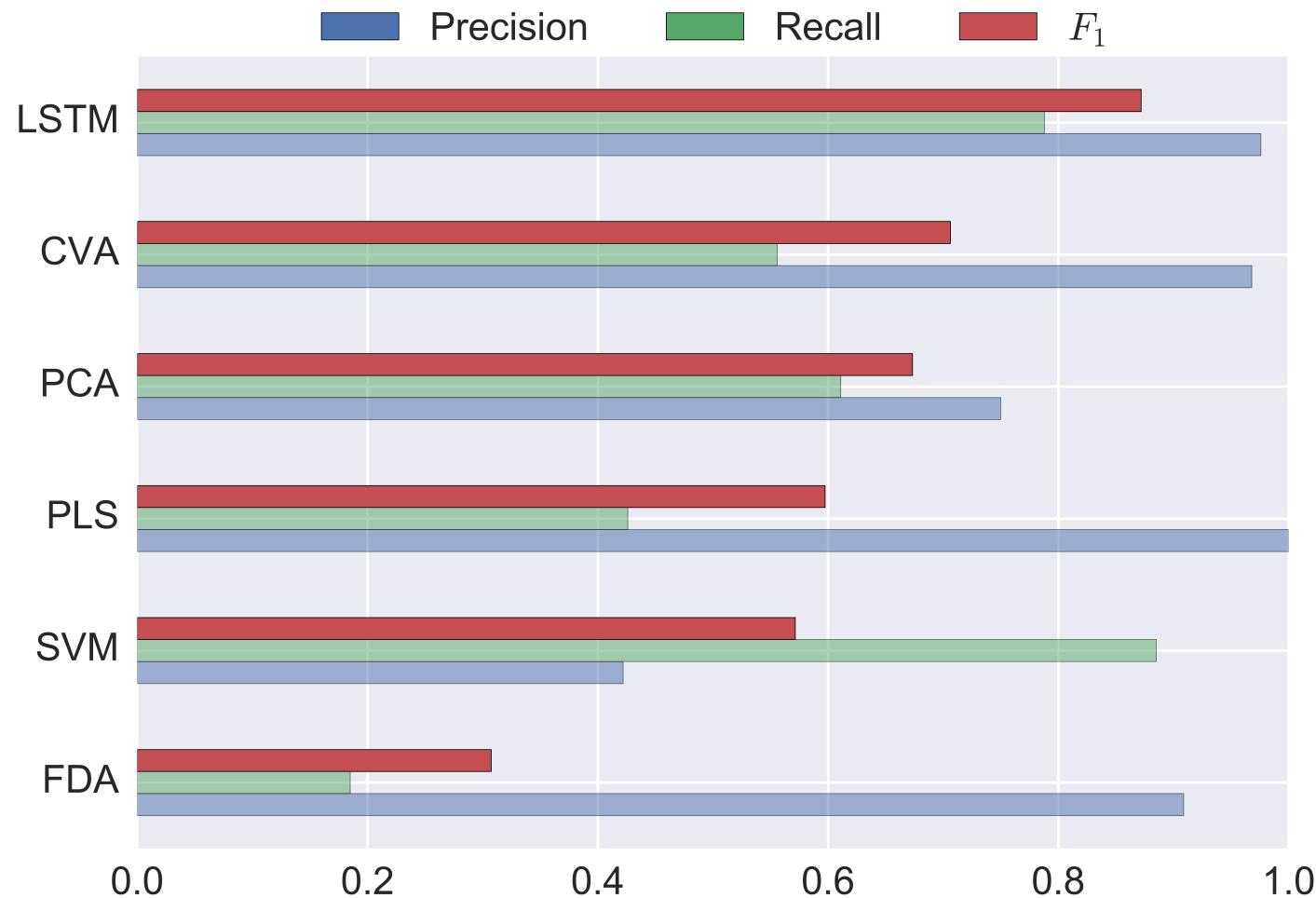


- $w$  – горизонт прогноза
- $MSE$  – среднеквадратичная ошибка прогноза
- $P$  – точность
- $R$  – полнота
- $F_1$  –  $F$ -мера

# Подбор параметров dropout

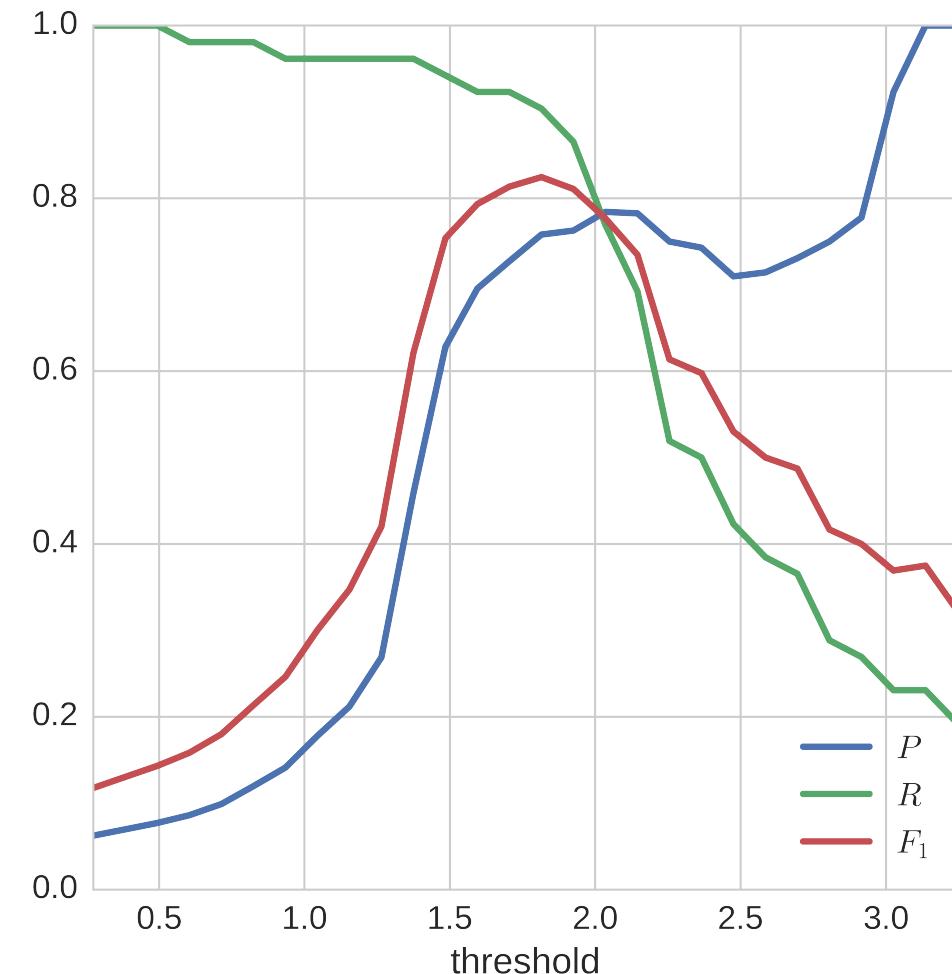


# Сравнение с известными методами [3]



# Баланс между точностью и полнотой

- $\alpha$ -варируемый параметр (коэффициент при пороге  $H$ )
- При крайних значениях  $\alpha$  можно достигать значительных показателей по точности ценой уменьшения полноты



# ИСТОЧНИКИ

1. Stuxnet: первые жертвы – [blogpost](#)
2. Будни немецких сталеваров – [blogpost](#)
3. L.H. Chiang, E.L. Russel, and R.D. Bratz. Fault Detection and Diagnosis in Industrial Systems – [eBook](#)
4. P. Malhotra, A. Ramakrishnan, G. Anand, and L. Vig. LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection – [paper](#)
5. P. Malhotra, L. Vig, G. Shroff, P. Agarwal. Long Short Term Memory Networks for Anomaly Detection in Time Series – [paper](#)
6. A. Nanduri, L. Sherry. Anomaly detection in aircraft data using recurrent neural networks (RNN) – [paper](#)
7. J. Brownlee. Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras – [blogpost](#)
8. Xiao-Dong Li, John K. L. Ho, and Tommy W. S. Chow. Approximation of Dynamical Time-Variant Systems by Continuous-Time Recurrent Neural Networks – [paper](#)
9. Coryn A.L. Bailer-Jones , David J.C. MacKay. A Recurrent Neural Network for Modelling Dynamical Systems [paper](#)
10. SIMATIC PCS 7 APC-Portfolio – [paper](#)
11. Fluidized Bed Dryer - Design of Model Predictive Control with Economical Steady State Optimization - [paper](#)

# Спасибо за внимание!

**Контакты:** [Pavel.Filonov@Kaspersky.com](mailto:Pavel.Filonov@Kaspersky.com)

[Andrey.Lavrentyev@Kaspersky.com](mailto:Andrey.Lavrentyev@Kaspersky.com)

[Artem.Vorontsov@Kaspersky.com](mailto:Artem.Vorontsov@Kaspersky.com)