Application of Anomaly Detection Methods in the Housing and Utility Infrastructure Data

Ivan Shanin, Sergey Stupnikov, Viktor Zakharov
ivan.shanin@gmail.com

FRCCSC RAS

IVMEM 2019, Veliky Novgorod
13.09.2019

Outline

1. Internet of Things and predictive maintenance
2. Experimental Setup
   - Types of equipment
   - Proposed system architecture
   - Experimental setup
   - Data collection protocol
3. Fault detection models
   - Student's t-test and its applications
   - The two-sampled Kolmogorov-Smirnov test
   - Hidden Markov Model in anomaly detection
4. Results and discusstion

Machine condition monitoring

- The **Internet of Things** concept development in recent years:
  - Hardware items become equipped with means of interaction with each other or with the environment
  - Growing popularity of home sensor technology, which is becoming cheaper
  - Modern tools for data collection, storage and analysis
- **Predictive maintenance** system goals:
  - Support timely and effective maintenance of in-service equipment
  - Register and identify equipment problems in real time scenario

This work is carried out as part of a project that is aimed at development of software solutions for a housing and utility condition monitoring system.

Purpose of this study:

- To introduce the experimental dataset on housing and utility infrastructure equipment faults
- To propose a fault detection system based on statistical and probabilistic approaches

Housing and Utility Infrastructure Elements

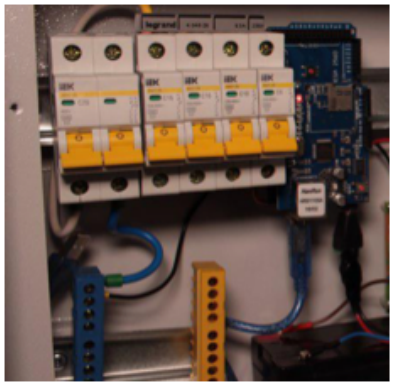
power transformer

pump

heater

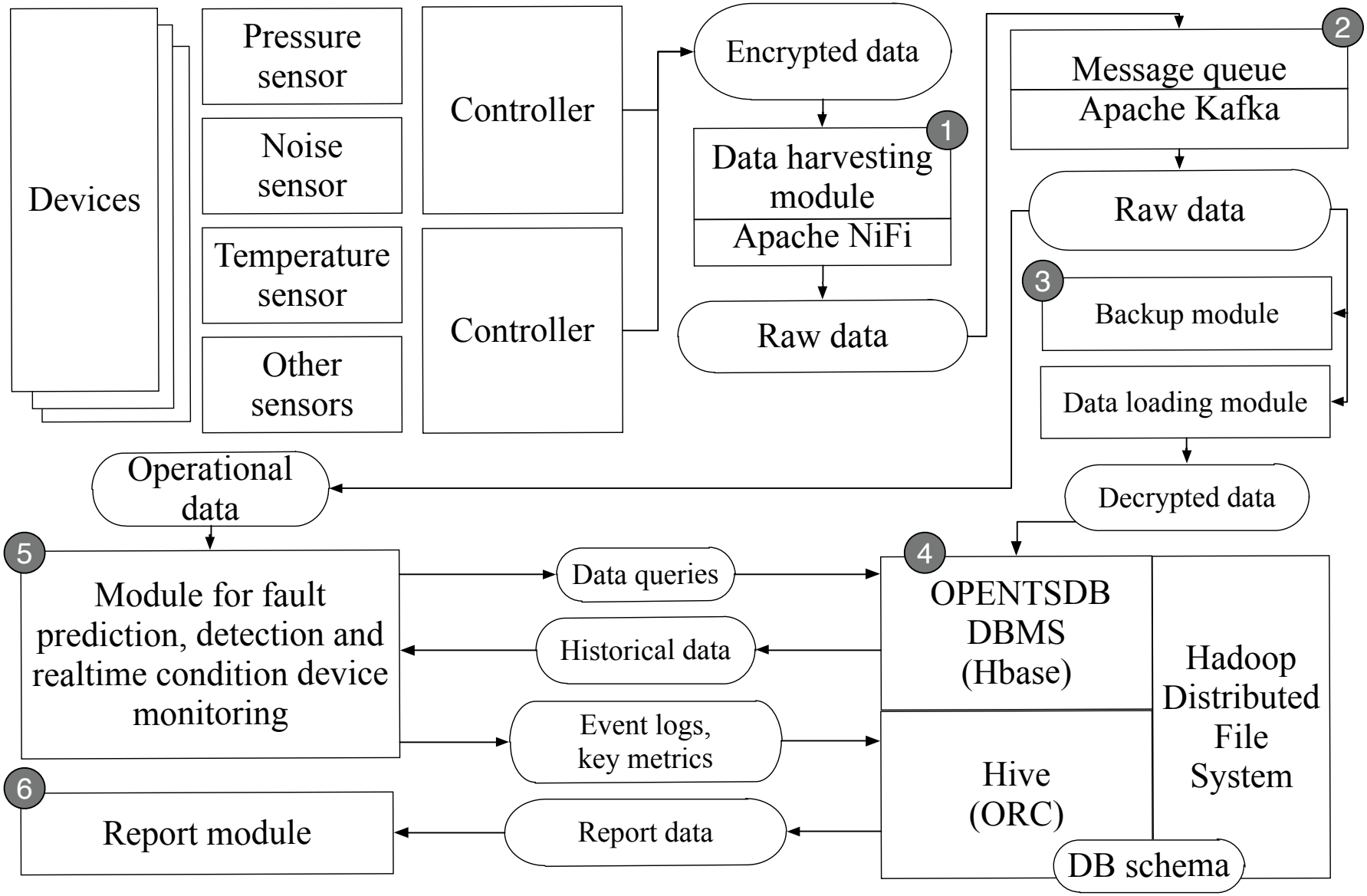
gas boiler

HVAC



controller installation

System Architecture



Experimental Setup

Hardware sensors were installed on the following elements of housing infrastructure:

- **electric heater** with current and temperature sensors
- **air conditioner** with current, temperature and acoustic noise sensors
- **ventilation blower fan** with anemometer, acoustic noise and current sensors
- **power transformer** with input/output current, voltage and temperature sensors
- **gas boiler system** with current, temperature, acoustic noise, leakage and flow sensors
- **water pump** with current and flow sensors
- **lamp** with light and current sensors

Experimental Dataset

The dataset used in this study was collected from abovementioned eperimental setup:

- Sensors were connected to hardware and software controllers
- Readings were obtained every 4 minutes, preprocessed and stored in time series DB
- Measurements were collected during one month
- Equipment went through a series of controlled simulations of various faults and breakdowns
- Every fault type was carefully analyzed, the measurements were used to fit anomaly detection models.

The preprocessing step includes:

- Missing value interpolation (linear, moving average)
- Flag and consistency check

Fault detection models: Student's t-test

In some fault types, a simple t-test provides acceptable results, i.e. in the ventilation clogging.

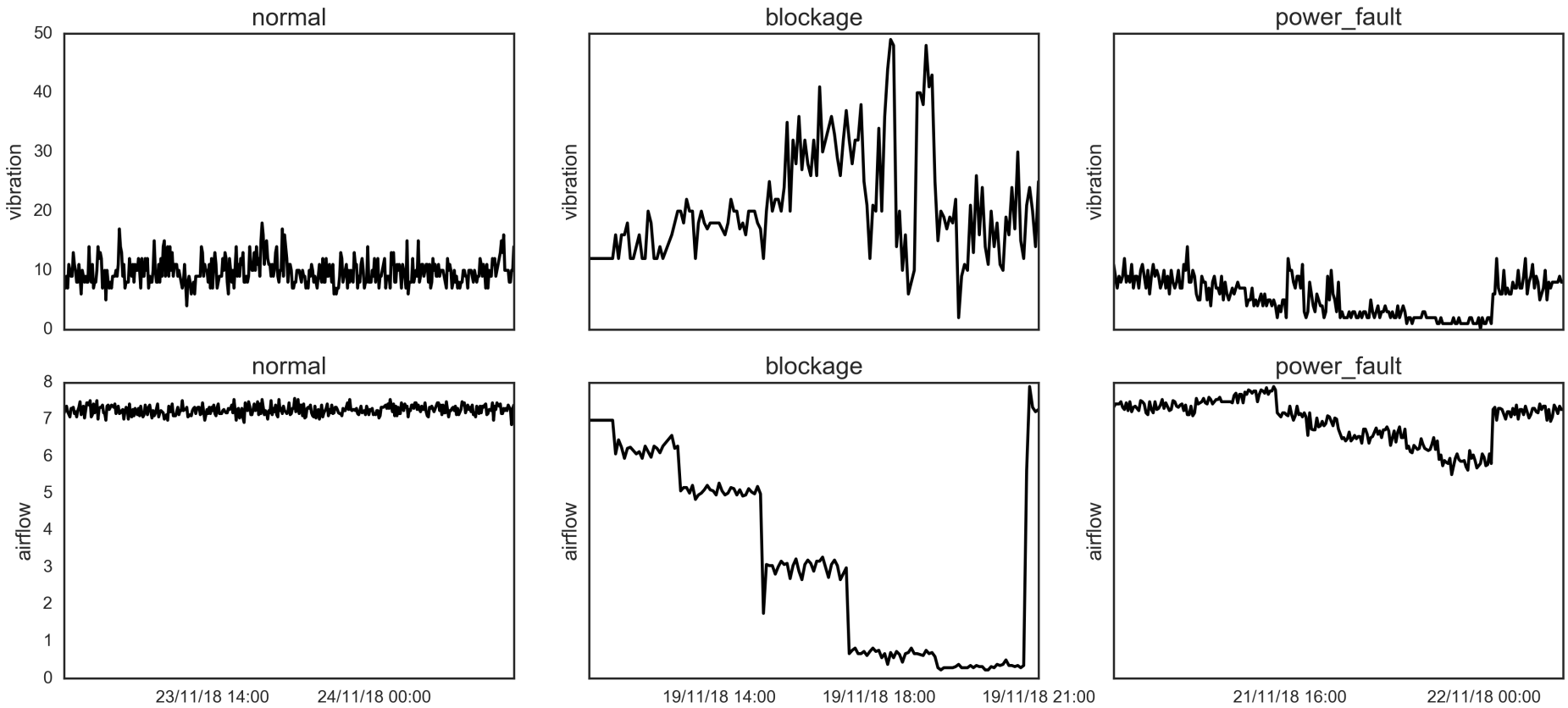


Figure 1: Airflow and vibration readings of a ventilation blower fan

Applications of Student's t-test

Student's t-test was used in three different scenarios:

- Point-wise anomaly detection (to test if the data point comes from a Student's t-distribution with known parameters)
- Segment-wise anomaly detection (compare means of values of two fragments of time series)
- To test abnormal slope of a linear model (for consumption sensor data)

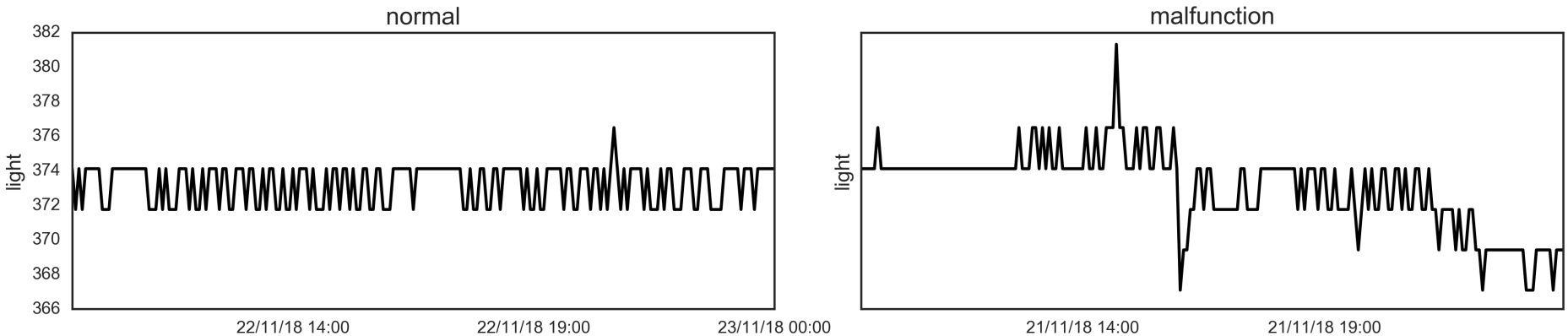


Figure 2: Light sensor readings of a lamp

Fault detection models: KS-test

- If the nature of the data does not imply a normal distribution, the non-parametrical **two-sample Kolmogorov-Smirnov** test can be used.
- An example of that behavior would be air conditioner temperature and current measurements.

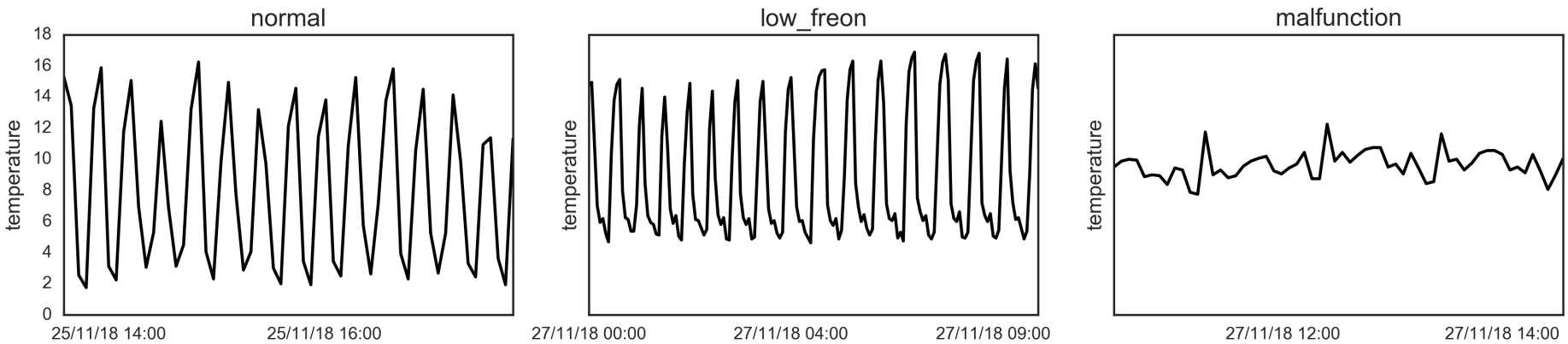


Figure 3: Temperature readings of an air conditioner

Hidden Markov Model

- **Hidden Markov model** can be used to detect a change in hidden states (operating modes) of the analyzed system.
- In condition monitoring, hidden states refer to operating modes, including both correct and faulty modes.
- The forward-backward algorithm is used to calculate likelihood of a test signal, which is then used as an anomaly score.

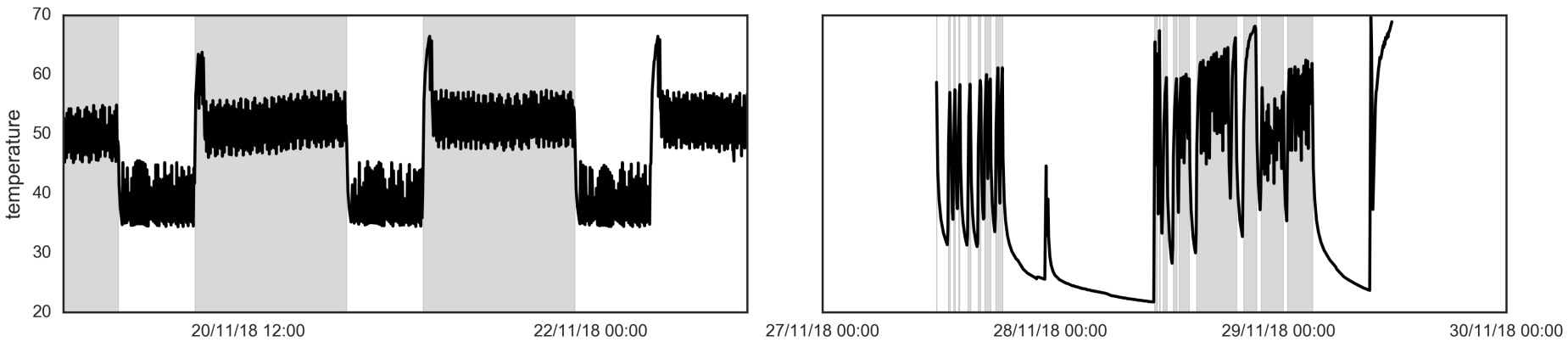


Figure 4: HMM-based signal segmentation of a gas boiler temperature readings

Results

- Proposed approach shows significant difference between correct and faulty operating modes on the training dataset.
- Obtained models are intended to be used as a theoretical basis for the design of industrial housing and utility condition monitoring systems.

Element type	Fault type	Fault symptoms	Model	Sensor	Criteria
Heater	thermostat failure	overheating	t-test	temperature	p-value <0.05
Power transformer	overload	overheating	t-test	current, temperature	p-value <0.05
Gas boiler	thermal overheating thermostat breakdown	overheating	t-test on HMM-generated likelihood values	temperature	p-value <0.05
Ventilation blower fan	clogging overload	increased consumption unstable airflow	t-test t-test	airflow, vibration airflow	p-value <0.05 p-value <0.05
Air conditioner	freon leakage filter contamination	lack of cooling frosting and icing	KS-test	temperature	p-value <0.05 p-value <0.05
Water pump	sedimentation	decrease in pumping efficiency	linear regression	flow	slope p-value <0.05
Lamp	flickering	lamp burnout	point-wise t-test	light	% of rejections >20

Table 1: Fault types and detection methods

Thank you!

Statistical Fault Detection

- **Student's t-test** is used to calculate a deviation of the test data from that normal distribution. That allows us to set a 95% confidence interval to detect point anomalies in real-time situations. T-test is used in two scenarios: to detect point anomalies and to compare means of values of two fragments of time series, thereby to test a whole time series fragment for a fault occurrence.
- the **two-sampled Kolmogorov-Smirnov test** is a statistical nonparametric approach to determine if two samples have significantly different probability distributions. It is used to detect and classify faulty modes of operation of an air conditioner.
- **linear regression model** is a statistical approach to model the dependency of two variables with a linear function. It is applied in water pump consumption monitoring. During a fault detection phase, if a linear model of current operating mode has a significantly different slope from the trained model, the mode is considered faulty. A regression task is performed to calculate slopes and their standard errors, a z-test is performed to test the equality of slopes.