Department of Electrical and Electronic Engineering

Mechanisms for Root Cause Analysis for Packet Loss in LTE networks implemented over on the Sea

Menglin Yao

Supervisor: Prof. Robert J. Piechocki

Introduction

With the utilization of excellent wind energy resources in the North Sea, large-scale offshore wind power projects have been built over the past decades. The maintenance vessel of these wind farms requires wireless network to provide information support. An offshore private LTE network over the Moray East wind farm is contracted by Villicom UK Ltd. for such a purpose.

- ❖ Due to the environments on the sea and land are different, the evaluation and diagnosis of network performance need to be specially treated. This project focuses on this theme to perform RCA (root cause analysis) on packet loss issue in this network.
- ❖ By analysing patterns of the network KPI (key performance indicators) data through machine learning, a great deal of valuable information can be extracted, and in some cases the RC (root cause) of the problem can even be directly identified. Further, an automated root cause analysis system is expected to be proposed.

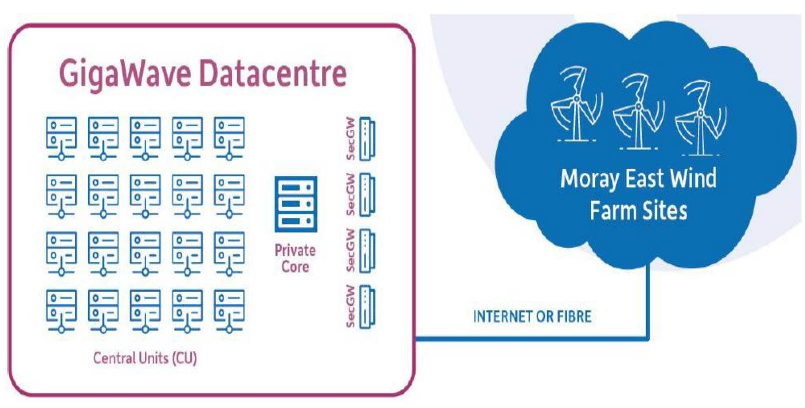

Fig.1: Network Connectivity for the Sea Vessels [1]

Aims

1. Perform anomaly detection on the target *i.e.* the PLR (packet loss rate), so as to achieve the purpose of labelling data.
2. Using the machine learning algorithm to analyse other KPIs during abnormal periods to achieve fault diagnosis.
3. Build an automated root cause analysis system to simplify the diagnosis process.
4. Automatically intercept abnormal fragments in real-time data and feed them into the above algorithm to realize real-time automatic root cause analysis.

Packet Loss Rate Labelling

- The UL (uplink) and DL (downlink) PLR data are obtained from cell stations' database on 5 minutes average.
- They are classified and labelled as normal and abnormal behaviours.
- Two labelling techniques are applied:
  1. Threshold setting for manual labelling
  2. Isolation forest anomaly detection for Auto-labelling
- Uplink and downlink PLRs are labelled separately in cell CPU level RCA.

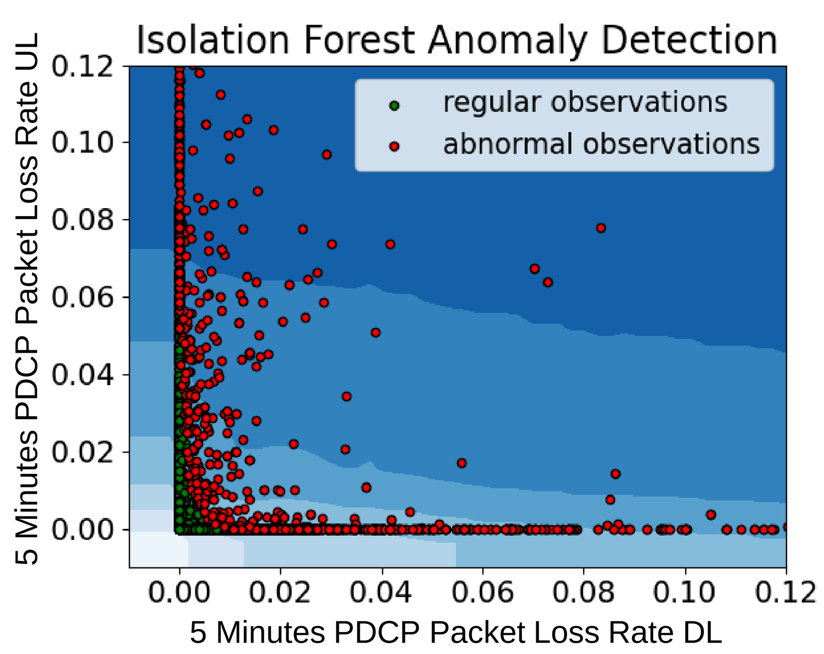

Fig.2: Result of Anomaly Detection on UL and DL Packet Loss Rate Based on Isolation Forest

Algorithm 1: CPE-End RCA

Background: The data provided by the CPE (customer premises equipment) includes physical environment indicators on the UE (user equipment) side that may affect network performance, such as the geographic location of the UE and the radio environment KPIs. From these metrics, we have a good chance of finding the RC of the abnormal PLR.

Problem: The sampling frequency of the CPE database is every 10 seconds, which is inconsistent with that of the PLR data. A higher level of uncertainty would be faced if the five-minute average data were distributed over all CPE data within that five minutes.

Step By Step Solution:

1. We can be more confident that within the five minutes with a normal average PLR, CPE data with corresponding timestamps can be classified into the normal category (a whitelist).
2. A suitable semi-supervised novelty detection model is assigned to each feature, models include e.g., One-Class SVM and Local Outlier Factor.
3. We use all existential normal data to train the novelty detection machines and determine whether the unseen data is abnormal.
4. A final detection report is then generated, which contains the data rows where each feature is labelled as abnormal under the PLR criteria. These abnormal cases are screened out as possible RCs.

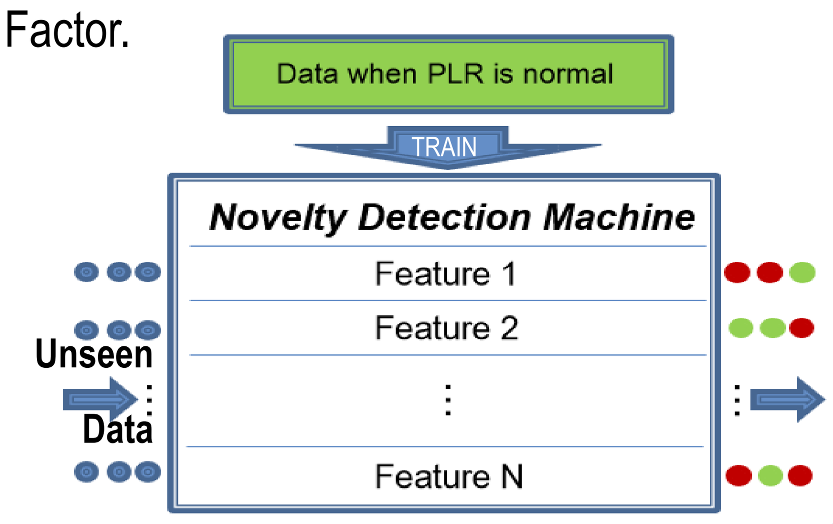

Fig.3: Mechanism of CPE-End RCA

Algorithm 2: Cell Level Automatic RCA

Background: Another set of KPIs are obtained from the cell CPU, they are recorded in the cell database along with the PLR. They can explain data flow performance, such as data volume, delay and throughput. These metrics may also have a critical impact on the packet loss rate.

Problem: The KPI data from the cell CPU does not have the problem of sampling frequency inconsistency, each labelled target (PLR) corresponds to a set of KPI data. In that case, why can't we pursue a more efficient algorithm?

Solution: Figure 4 shows a decision tree based automatic RCA system I designed. The decision tree model can play multiple roles in this algorithm:

- ❑ At 1, data patterns with known RC can be stored in the model, and saved to the library.
- ❑ At 2, the target of the test dataset (PLR performance) will be predicted by every model in the library, the most fit model can then be selected through its prediction accuracy.
- ❑ At 3, When the RC model is selected, from its built-in feature importance scoring system based on features' information gain contributions, the most important features can be found as RCs.
- ❑ At 4, whether the RC can be found from this piece of data is determined from the quality of the model (number of layers) trained by this dataset.

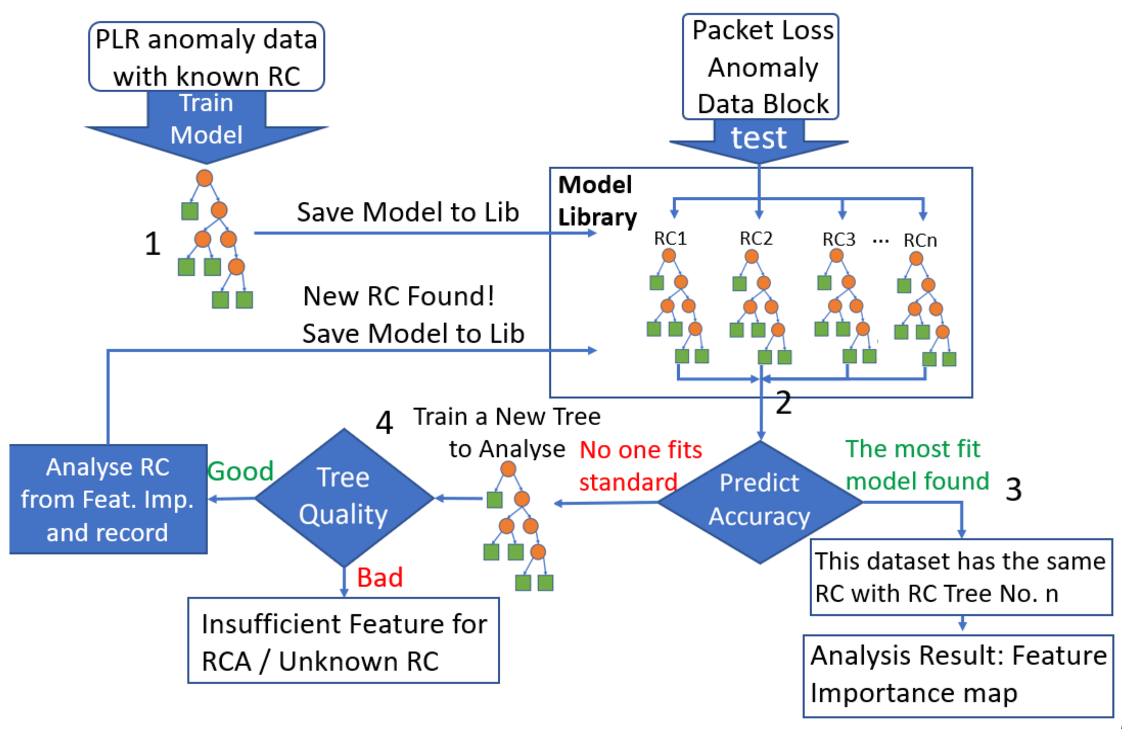

Fig.4: Mechanism of Cell-Level Automatic RCA

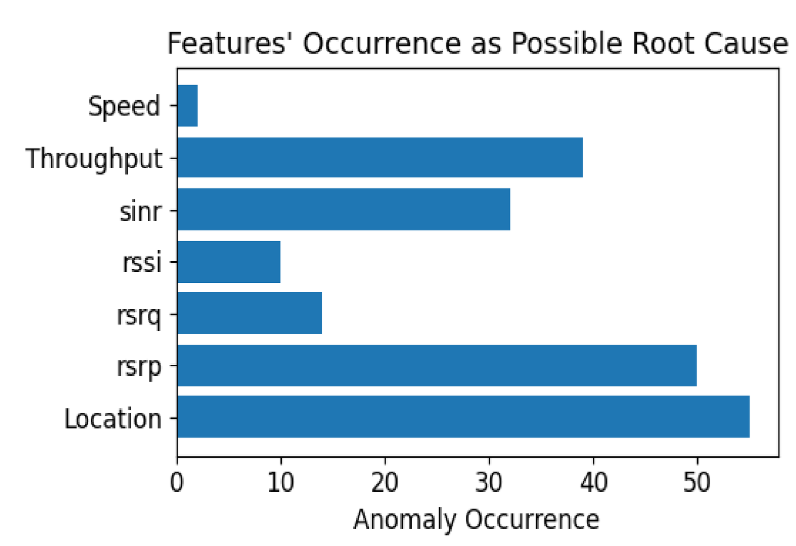

Fig.5: RCA Result Statistics on the CPE-End – Algorithm 1

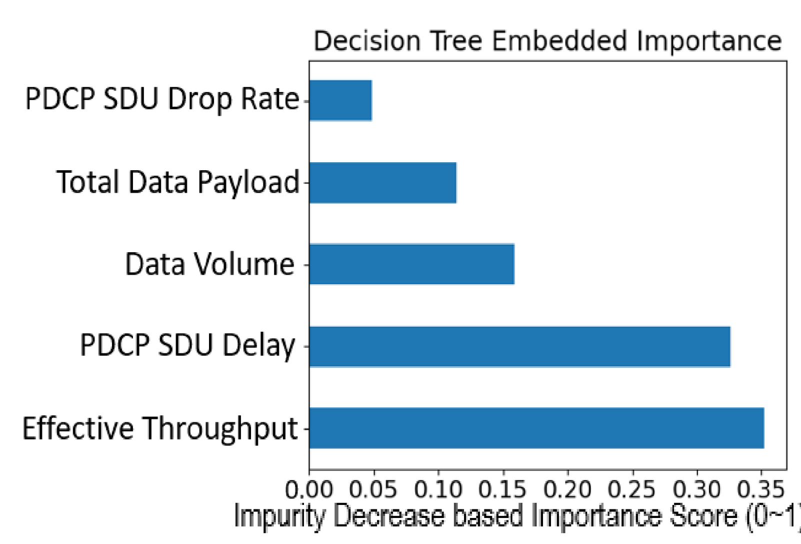

Fig.6: Downlink Packet Loss RCA Result on Cell CPU Level - Algorithm 2

Results

- ➢ Figure 5 shows the results of the CPE-end test data, in this period of data, the anomalies that most often lead to packet loss were found to be the ship's location (poor coverage), low received signal strength and overloaded throughput. When cross-validating this model with other whitelisted data, it achieved over 94% accuracy.
- ➢ In the result shown in Figure 6, the most likely causes of downlink data packet loss in this test dataset are peaked cell throughput and PDCP layer SDU (Service Data Unit) delay. The accuracy of this model is still in the process of verification.
- ➢ Note that due to different mechanisms, for a single data segment, the result shown in Figure 5 is the occurrence of each feature as RC, while the result in Figure 6 is the quantized impact of each feature on the packet loss (likelihood of being RC).

Conclusions, Ongoing and Future Work

In the progress of this project so far, the target of the RCA algorithm, which is the PLR, has been classified either manually or automatically, through which the RCA can be performed in the following situations:

- ✓ A semi-supervised RCA algorithm based on novelty detection is deployed at the CPE side.
- ✓ A highly automated RCA algorithm based on the supervised decision trees is established at the Cell level.

In general, these two algorithms solve the RCA problem under the scenario of inconsistent and consistent target-features sampling frequencies, respectively. These algorithms will still function even if the participating features are different or expanded from the current datasets.

The current work in progress includes building RC models and test data for the validation of Algorithm 2, then verifying its accuracy with them.

Both algorithms belong to empirical research, in future runs, as more training samples are accumulated and fed into the algorithm, their accuracy will be greatly improved.

References

[1]. Vilicom (2021) *Moray Wind Farm| Private LTE Use Case Study* [Online]. Available: https://www.vilicom.com/case-studies/private-lte-network-for-moray-east-offshore/

University of BRISTOL