

LDA Classifier For Packet Status Prediction In A Wireless Star Network Topology

Mariana Ávalos Arce

Faculty of Engineering
Universidad Panamericana

0197495@up.edu.mx

May 22, 2022



Content

1 Data Collection

- Hardware
- Experimentation
- Building The Dataset

2 Data Analysis

- First Date
- Linear Discriminant Analysis
- LDA: Results

3 Performance

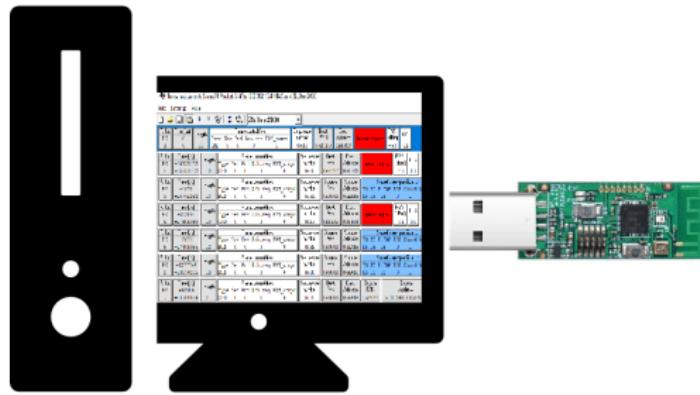
- Confusion Matrix, ROC-AUC Curve

4 Future Work

Data Collection: Hardware

The first step in the study was to determine **how to collect data on packets across a wireless network**, in order to analyze what factors influence and how to predict if a packet would be arrive OK or with Error:

- Sensor Node: Texas Instruments CC13x
- Collector Node: Texas Instruments CC26x
- Sniffer: CC2531 USB Dongle
- **Wireless Protocol: IEEE 802.15.4**



Data Collection: Experimentation

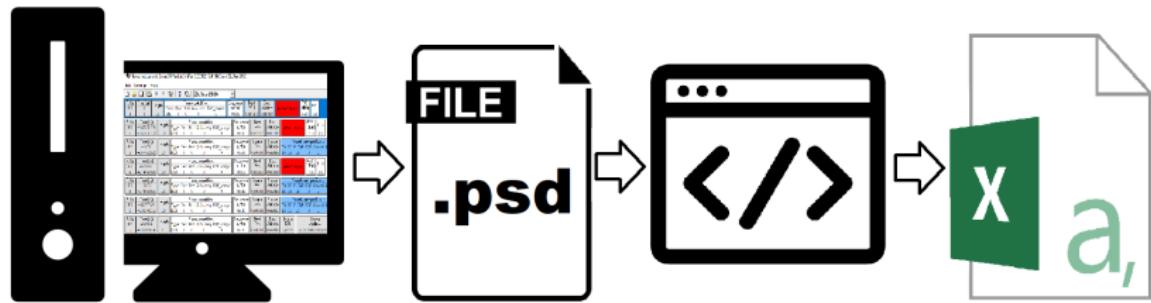
Over the course of three days we performed **7 Controlled-Environment Tests** that each exposed the installed network to different conditions over **1 hour of duration**:

- ① Control Test 1: No interference whatsoever.
- ② Control Test 2: No interference whatsoever.
- ③ Double Network Test: An additional star network apart from tested network.
- ④ Radiofrequencies Test: Microwave, Air Fryer, Blender, Toaster and a Sandwich Maker.
- ⑤ No restriction Test: People in the house living normally around the network.
- ⑥ Wind Test: Two Fans blowing air across the space.
- ⑦ Wireless Test: A SmartTV, 3 cellphones, a game console (PlayStation 4), 2 laptops, all downloading content from wifi network.



Data Collection: Building The Dataset

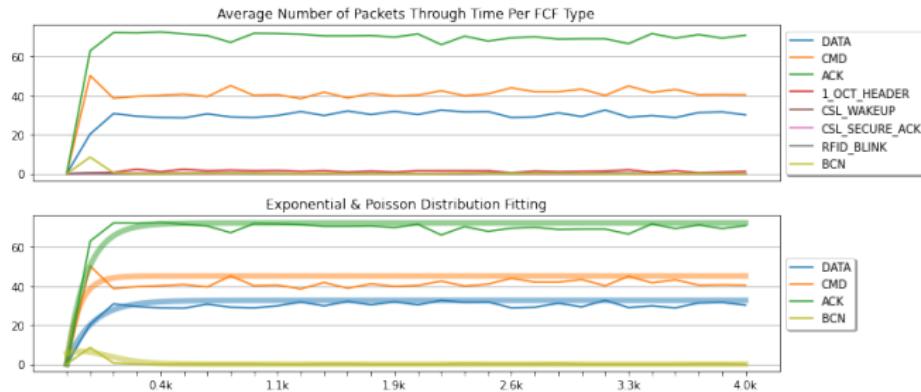
After each experiment, we stored using the Sniffer Software a **.psd file**, which is Texas Instruments' file format for Packet Sniffer Data.



We coded a program that reads the **.psd** files **byte-wise** and converts those 0's and 1's into the **overall packet info dataset with 33k rows and 24 columns**.

Data Analysis: Our First Date

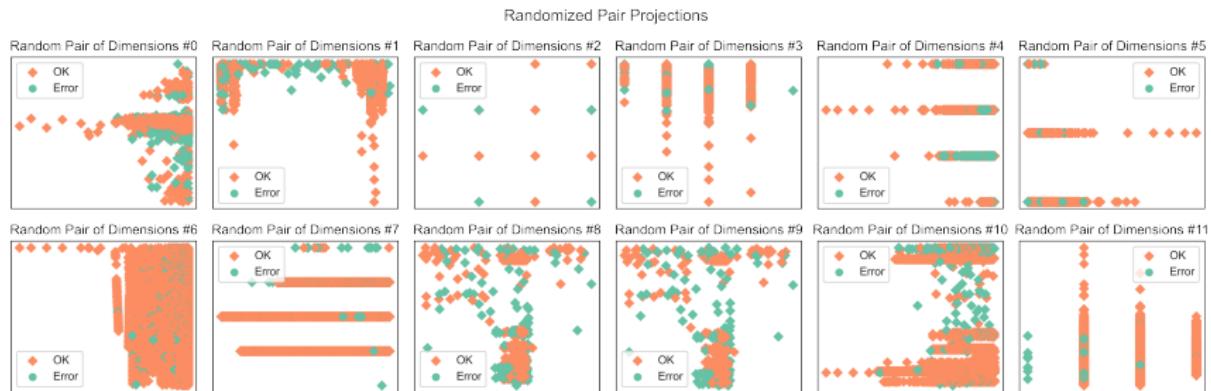
The next step was to get to know our dataset: for this, we performed **parametric distribution fitting**, since the packet arrival behaviour showed **Poisson and Exponential Distributions**.



with $f(x) = 1 - e^{-\frac{1}{120}x}$, $f(x) = 1 - e^{-\frac{1}{75}x}$, $f(x) = 1 - e^{-\frac{1}{150}x}$ and $f(x) = \frac{1.15^x}{x!} e^{-1.15}$

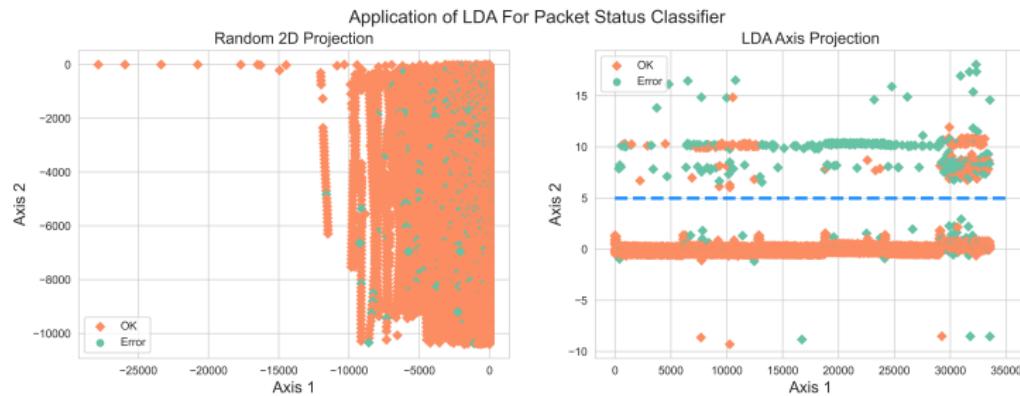
LDA: Projections

LDA seemed as the best option since the dataset had multiple columns, and the pairplot of each of the combinations of features did not show a separable data relationship with the Packet Status column (class):



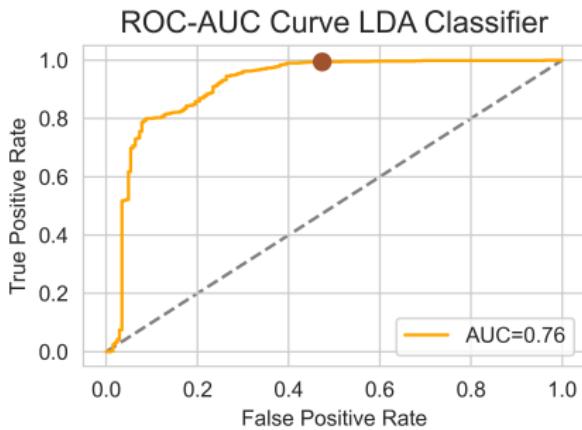
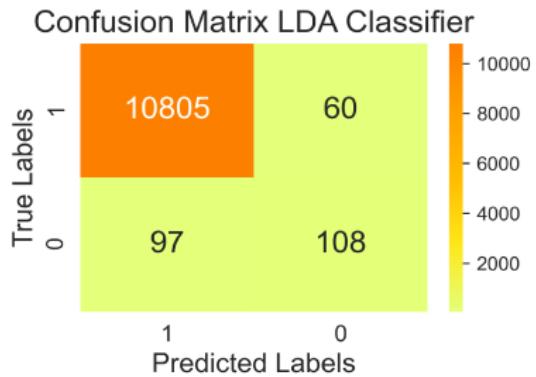
LDA: Results

After performing LDA for dimension reduction into **one eigenvalue**, since we had two classes, the model found an axis from where the classification could be possible:



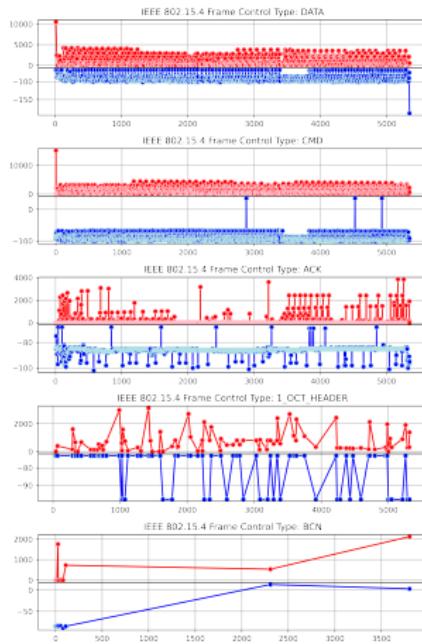
Performance: Confusion Matrix, ROC-AUC Curve

To evaluate the classifier that was created using Linear Discriminant Analysis, we computed the Confusion Matrix. **The amount of Packets with class 1 (OK) and 0 (Err) was equal**, and thus the matrix shows a **high TPR** and **almost mid-valued FPR**, which became the ROC curve with an **AUC = 0.76**, suggesting the classifier is moderately accurate.



Future Work

Important information aside from the useful for a binary classifier was found due to the extensive amount of data the .psd file stores, which leads to prominent future work related to **Time Taken, RSSI and Packet Loss analysis.**



Thank you!

