2/14/202

crossfold validation

1st

recurrent neural network

2nd

hidden markov model

bayesian - takes into account previous

3rd

whatever output

2/28/2020

Sensors up

Calendar weber.edu\_amkd3eqithr5cuipkhveu1vj24@group.calendar.google.com

Account for Raspberry PI

zigbee-sniffer.users.weber.edu

u:nick

p:marietti

Setup the zigbee sniffer on the raspberry pie in Dr. Fuez office at EH.

Sniffer is running in screen and outputting data to /tmp/cc2531\_sniffer

TBD the need to backup the data to prevent overfilling the SD card.

Data packets show that there are different types that come in that the sniffer can parse the type.

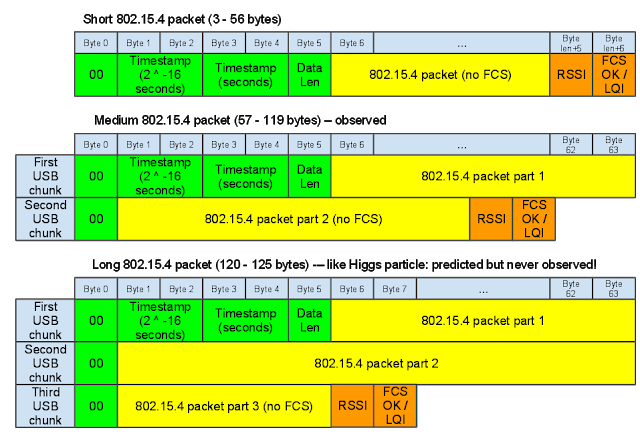
Currently only 3 packets have been captured

1. Type 3 MAC Command
2. Type 0 Beacon
3. Type 6 Unknown
   1. Type 6 Unknown is most likely candidate for ML algorithm. The other data types seem to be specific to the zigbee protocol and not related to sending data related to the motion sensors.

I need the write a parser for the raw data packets to preprocess it before being fed into the ML algorithm. I will focus my efforts on filtering out all known data types and focus only on Type 6. I will create multiple parser outputs to feed into algorithms that will be developed later.

Parser Methods to attempt:

1. Parse only time stamp and frame
2. Parse all data
3. Parse time stamp, addresses, and raw data
4. Parse time stamp, addresses, and raw data length
5. Parse data length according to the 802.15.4 and only parse out data packet according to the length based on picture as well as timestamp



3/11/2020

Observations

Most of the data packets are coming in on a limited number of channels. A histography might be a useful tool for analysis. It might be best for data capture to focus only on channels that have received data to get the data I am interested in to capture more relevant data to the project rather than random data packets that are not meaningful.

RSSI can be a useful tool to map out the sensor network.

Channel could be useful if I tune the sniffer to only focus on channels that it has seen data from while in a wide scanning mode where it sweeps a range at a certain interval. Perhaps a way to bias the readings towards channels that it has seen the most frequent traffic on.

Parser to parse out the raw data is complete. Next steps are to make the parser output only data that I am interested in running through a neural network.

Research focus should be on algorithms that are best suited to analyze raw data packets for patterns.

It might be useful to filter out non data packets. Commands could indicate that the sensor is communicating a time window to transmit or some other network related functionality that isn’t useful for analysis.

Channels 20, 24, 26, 16, 13, 14, 12

3/14/2020

Meet with Dr. Fuez and discussed setting up the sniffer to only sniff on channel 24 to gather more data packets. Also discussed finding out what the status is of presenting with COVID19 challenges.

3/19/2020

Logged into the zigbee sniffer at WSU and pulled the latest data capture. Ran it against the raw parser to generate csv output to be fed into a completed machine learning algorithm. Started working on filling out the Bayesian approach for a machine learning algorithm.

3/26/2020

Logged into the zigbee sniffer at WSU and pulled the latest data capture. Ran it against the raw parser to generate csv output to be fed into a completed machine learning algorithm. Continued to work on the Bayesian approach for a machine learning algorithm.

3/27/2020

Meeting today @10am with Dr. Fuez. Discussed possibility of moving sensors if needed and status of presentations. Plan is to continue to work without data on the web interface/algorithms and push tuning the algorithms to a later date when data is available.

Worked through the parsing issues related the Bayesian model. Looking at the data fit into the model there does not appear to be an obvious output to use that will tell me when something is occupied. I am considering using src address with destination address but even then, the model appears to indicate it will only predict where the data is being sent to based upon the other data fields. I might consider somehow identifying the base station and doing some sort of count or biasing towards two states based upon this data but for now this leads me to believe that I should be utilizing a Hidden Markov Model that is capable of having such states.

Reading about Hidden Markov Models I stumbled upon Recurrent Neural Networks and a subset Long Short-Term Memory Units. I will add those to my research list.

Reading more about Hidden Markov Models it is clear I am going to need to develop a probability distribution to calculate the probabilities that go into developing the hidden layer.

HMM:

<https://towardsdatascience.com/introduction-to-hidden-markov-models-cd2c93e6b781>

<https://towardsdatascience.com/hidden-markov-models-for-time-series-classification-basic-overview-a59b74e5e65b>

<https://www.aclweb.org/anthology/W16-5907.pdf>

<http://www.blackarbs.com/blog/introduction-hidden-markov-models-python-networkx-sklearn/2/9/2017>

<https://www.math.univ-toulouse.fr/~agarivie/Telecom/code/index.php>

LSTM:

<https://pathmind.com/wiki/lstm>