Predicting Future Network Status and Potential Network Errors with Deep Learning

ECGR 4106 – Real-Time Machine Learning
Final Project Proposal
Adam Shigapov, Hunter Brock, Patrick Ballou

Project Summary

This project is the development of a deep learning model that will analyze network traffic to predict the future network state and predict potential network performance bottlenecks/latency. The overall impact of this project is reducing network infrastructure loads, increasing the reliability, security, and cost efficiency of modern networks. All while aiding network administrators in automating network management and solving problems before they occur.

Datasets

- Cybersecurity Datasets - Kaggle
- NTA w/ Wireshark - Kaggle
- Measurement Lab - Google, Inc
- Internet Traffic Dataset - Kaggle
- Networking Datasets - GitHub
- Encrypted Traffic Feature Dataset for ML
- Synthetic Network Traffic - Kaggle
- Wireless Network Traffic Time Series of an Enterprise Network - IEEE Dataport
- 5G Traffic Dataset - IEEE Dataport (Warning: ~430 GB compressed)

Training Plan

1. Preprocessing
    a. Choosing which datasets provide useful data for training
    b. Get rid of unnecessary data
    c. Combining and formatting data
2. Tokenize network information
    a. Test what length of tokens would work best for the specified data
3. Define Sequence Length & Projection Window
    a. Decide on and test what length of sequence data would be best (1 Hour, 24 Hours, Week)
4. Experiment with DL models (BiDir LSTM, seq2seq, Transformers, etc.)
    a. Adjust models
        i. Adjust Hidden layers
        ii. LSTM, GRU, RNN
    b. Experiment w/ Attention
    c. Experiment w/ Regularization & Dropout

5. Model Tuning
    a. Find which models work best with different parameters
    b. Optimize for best results

Evaluation Plan

- Create a Validation Set from the datasets, separate from the training set
- Gather model performance metrics
- Create a Virtual network to run the model on

Project Outcomes

This project aims to create a Deep Learning Model that can predict network metrics that provide the user information to enhance the management of the network. Our metrics can be broken down into the following:

- Regression Outcomes
    - Latency Prediction
    - Bandwidth Usage
    - Throughput Predictions
    - Resource Usage Predictions
    - Anomaly Detection (optional/extra)
        - DDoS
        - Network Intrusion (Unauthorized Access)
- Classification Outcomes
    - Bottleneck Identification (Type vs True/False)
    - Traffic Type (Ex: VoIP, Video Streaming, Web browsing)
    - Quality of Service (High/Low Priority Packets)
    - Network Resource Performance Status
        - Healthy
        - Underperforming
        - Overburdened

Aspects that we have choose to remove from this project in regards to the project timeline, available resources, scope that can be added onto this project include:

- Network Security Threat Identification and Classification
- Maintenance Prediction
- Resource Allocation for Bottleneck Removal Suggestions (Topology & Routing)
- IoT Integration

Individual responsibilities

Adam (Project Lead):

Will be responsible for making sure the project stays on task, and manage team-wide communication. Adam will design and execute the evaluation of the models that are created and validate that model tuning is meeting expectations. Adam will aid Patrick in the tokenization of network information and the testing of the sequence lengths and projection window experiments.

Hunter:

Will be responsible for organizing the data as well as preprocessing what data will be used and putting it into a usable dataset. Will also be responsible for the validation of the data, and gather metrics from said data. Will assist Adam with the evaluation of the models and tuning.

Patrick:

Will be responsible for tokenization of network data and selecting appropriate sequence and projection window lengths. Also, Patrick will experiment with various DL models, decide on their architectures, and fine tune their parameters to optimize evaluation results. Lastly, Patrick will help Hunter with the data preprocessing.