**Introduction**

Hello Professor, hello everyone. We are group 4 and our project is anomalous traffic prediction. My name is Tong Wu, and my teammates are Zhang Kaiyue and Zheng Wei. In this project, our group focus on the research and analysis of anomalous traffic detection methods. During our research, we found two mainstream approaches nowadays to detect the traffic: the clustering and LSTM model. Based on this, our group proposes a thought of fusing two models and tries to get a more effective approach.

**Outline**

We will do a full introduction for our project, including innovation points, approach, models we used, related works, questions we raised, optimise the objectives, data result and conclusion.

**Innovation Point**

When we were thinking about innovation, we looked up related work on abnormal network traffic monitoring. We found that in fact, the accuracy of network traffic anomaly detection of a single model is already very high. If we want to make a breakthrough in accuracy, we must innovate in the method.

**Approach**

Our approach is to use two different models, clustering and LSTM. We try to make a fusion to let these two different models can forecast the data together to get a higher accuracy rate.

**Clustering model**

So, let me introduce the two models we used first. The clustering method is classified as unsupervised learning to mining the structure of the dataset. The image in the left below shows how we apply the clustering method in our project code. We choose K-means algorithm and select the k value for 5, then we do training and prediction.

**LSTM model**

The LSTM model is seen as an improved version of the RNN model. RNN cannot cope with two attributes that are far apart because of its single register, and this problem is called long-term dependence. LSTM, as its name suggests, has long and short term memory capacity and can solve this problem. The image on the left below shows how we use LSTM to creates the training and testing dataset, construct the model, setup the optimiser and training the model.

**Related work**

By reading a lot of papers, we found three papers as our reference for this project.

The first paper focuses on the K-means clustering algorithm. Their team classifies time interval and destination port number, which results in two-class clusters of normal and abnormal traffic. Their work has been a reference for our project; they found that the data preprocessing is necessary and beneficial, so we did this to change the data type and select 11 features and labels to process clustering. Their result shows that there are 15% of traffic data don't direct to the unusual destination port. This also gives us a possibility to use other models to optimise it.

The second paper also employs the K-means clustering algorithm for anomalous traffic detection. Their team found that the clustering result may be unstable; this also allows us to optimise the clustering process to make sure it is convergent. And as a result, the false alarm rate remains at a relatively high level.

The third paper applies the LSTM algorithm to predict the traffic. Their team implemented the LSTM model while finding it difficult to go for a balanced dataset for training, which is also one of the issues we raised. Their result provides the guideline of the model, and by applying the LSTM model, the accuracy and recall rate can reach a high level.

**Question**

After reading the related work, we raised up three questions below:

How to make the detection method more precise and have a lower false alarm rate?

How to make the detection method can detect a large amount of data in real-time?

How to provide a balanced dataset to avoid naïve behaviour?

For the following part, my teammate will present. Thank you.