**Introduction**

In this presentation, our group focuses on the research and analysis of anomalous traffic detection methods. During our research, we found two mainstream approaches nowadays to detect anomalous traffic: the clustering model algorithm and LSTM neural network model to analyse and predict the network traffic. Based on this, our group proposes a bold hypothesis of fusing the clustering algorithm and the LSTM model and tries to propose a more effective approach. Next, I will briefly introduce the two different models.

**Clustering model**

The clustering method is classified as unsupervised learning to mining the structure of the dataset. The basic structure of the model is shown in the image on the right below. When implementing this algorithm, Euclidean distance has been used to measure the similarity between points and clusters. The Euclidean distance equation has been shown in left below.

**LSTM model**

The Long Short Term Memory model is seen as an improved version of the RNN model, which has a storage function for the previous output. The diagram on the left below shows the structure of an RNN. RNN cannot cope with two attributes that are particularly far apart because of its single register, this problem is called long-term dependence. LSTM, as its name suggests, has long and short term memory capacity and can solve long-term dependency problems. The diagram in the right below shows the chain structure of the LSTM. The data from the previous unit is selected, added and output through the forget gate, input gate and update gate, respectively. This gives the LSTM the ability to long-term memory.

**Related work**

We found three papers as our reference for this project. The first paper focuses on the K-means clustering algorithm. Their team classifies the time interval and destination port number of network traffic, which results in two class clusters of normal traffic and anomalous traffic. Their result shows that there are 15% of traffic data don’t direct to unusual destination port.

The second paper also employs the K-means clustering algorithm for anomalous traffic detection. Their team uses a k-value of 5 for classification detection, and as a result, the false alarm rate is remaining at a relatively high level.

The third paper applies on LSTM algorithm. 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 is that, by the application of 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 large amount of data in real time?

How to provide a balance dataset to avoid naïve behaviour?

That is all my part, for the innovation points and provements, my teammate will present. Thank you.

**Approach**

我们已经编写了一个程序对两种模型进行融合。在此我将贴上代码图片进行演示。其中第一部分是对依赖和数据库的导入。在数据库方面，我们选用了加拿大网络安全研究所2017年发布的数据集，CISCIDS2017。我们通过pandas对数据集进行预处理，清洗和标准化。其中我们对object类型转换为离散数值，同时把数据类型转换为float。

第二部分中，我们对数据库进行特征选取，一共选取了11个特征，并对数据进行聚类，K值取5.并将聚类结果进行标准化以继续进行LSTM的计算。

第三部分为设置LSTM模型并对