# MS-LSTM: a Multi-Scale LSTM Model for BGP Anomaly Detection

LyuJiuyang, Dec 21st, 2021.

## Introduction

### Info

**Authors:** Min Cheng, Qian Xu, Jianming Lv, Wenyin Liu, Qing Li, Jianping Wang. ICNP 2016

**Code:** https://github.com/jayvischeng/MSLSTM (Tensorflow).

**Link:** MS-LSTM

### Target

**Detecting anomalous Border Gateway Protocol traffic.**

BGP: Border Gateway Protocol (边界网关协议), which is designed to exchange routing and reachability information among autonomous systems (AS) on the Internet. TCP layer. Zhihu-Link

### Motivation

Existing solutions apply classic classifiers to make real-time decision based on **the traffic features of present moment**. However, due to the frequently happening burst and noise in dynamic Internet traffic, the decision based on short-term features is not reliable.

### Contribution

- propose to adopt MS-LSTM, a multi-scale LSTM model, for BGP anomaly detection.
- show that applying optimal and time scale to the existing classification model in BGP anomaly detection can improve their performance by 10%.

## Related Work

- Based on statistics pattern and signal processing techniques, where the anomalies are identified as correlated abrupt changes occurring in the underlying distribution.
- Rule-based method, which is applying Internet Routing Forensics (IRF) to classify anomalies.
- Machine-learning methods have been employed to build traffic classification models and predict anomaly.

## Time Series Analysis

- Integrating the historical information into the classifier can make the decision more cautious and more accurate.
- In a larger time scale, the global trend of the time sequence is easier to be captured, but it becomes harder to sense a local change.

## Methodology

**Input:** previous traffic features $x_{t_1}$ , $x_{t_2}$ , ..., $x_{t_n}$ ,

**Output:** the current state of traffic $x_{t_{n+1}}$

## Preprocessing

$e$, the size of window; $p$, the time scale.

$$S_n = x_{t_{n-e+1}}, x_{t_{n-e+2}}, \ldots, x_{t_n}$$
$$S_n = \left(d_1, d_2, \ldots, d_{e/p}\right)$$
$$d_1 = 1/p\left(x_{t_{n-e+1}} + x_{t_{n-e+2}} + \ldots x_{t_{n-e+p}}\right)$$

## Model

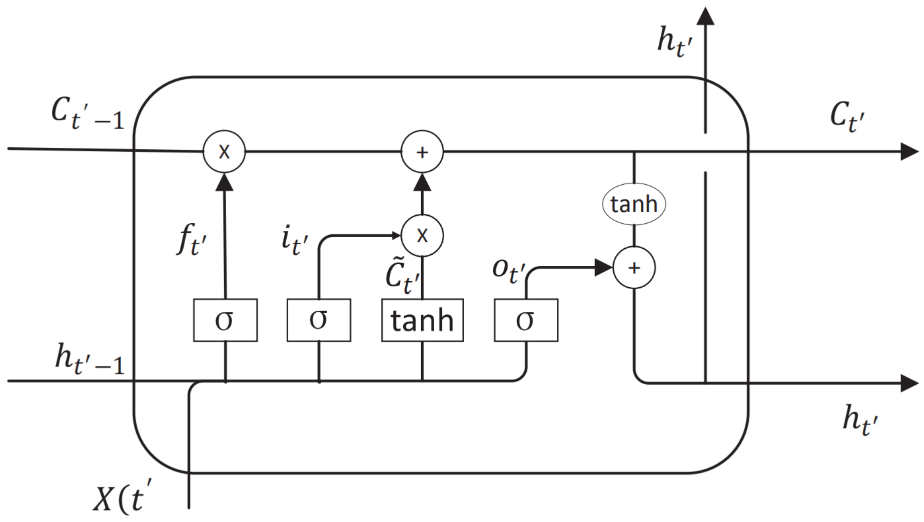


**Figure 3: Structure of LSTM memory cell.**

**Step1:** Throw away redundant old information.

$$f_t = \sigma\left(W_f \cdot \left[h_{t'-1}, X_t'\right] + b_f\right)$$

**Step2:** Store new useful information.

$$i_t = \sigma\left(W_i \cdot \left[h_{t'-1}, X_t'\right] + b_i\right)$$
$$\widetilde{C_t'} = \tanh\left(\left(W_c \cdot \left[h_{t'-1}, X_t'\right] + b_c\right)\right)$$

**Step3:** Update the cell state.

$$C_t = f_t * C_t' - 1 + i_t * \widetilde{C_t'}$$

**Step4:** Output for next memory cell.

$$o_t = \sigma\left(W_o \cdot \left[h_{t'-1}, X_t'\right] + b_0\right), h_t = o_t * \tanh\left(C_t'\right)$$

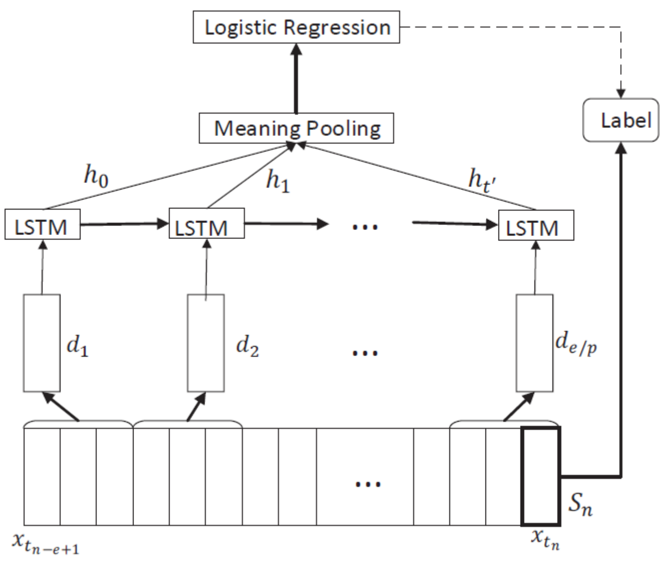


Figure 4: MS-LSTM classification model.

# Experiment

Dataset: RIPE

Optimal window size: 40.

Optimal time scale: 8.

## Other

- The highlight of this article is using LSTM to detect anomalous BGP and time scale.