BGP ANOMALIES
• Border Gateway Protocol (BGP) plays an essential role in routing data between Autonomous Systems (ASes).
• BGP anomalies affect Internet servers and hosts and are manifested by anomalous traffic behavior.
• They may be detected by analyzing collected traffic data and by generating various classification models.
• Machine learning techniques are the most common approaches for classifying BGP anomalies.
• We employ two supervised machine learning algorithms:
  - Support Vector Machine (SVM)
  - Long Short-Term Memory (LSTM).

FEATURE SELECTION
• Datasets are collected from the Route Views project, the Réseaux IP Européens (RIPE) Network Coordination Centre (NCC), and from BCNET.
• We extracted 37 features from BGP update messages originating from AS 513.
• Three cases of well-known anomalies are considered: Slammer, Nimda, and Code Red.
• 10 features were selected using minimum Redundancy Maximum Relevance (mRMR) algorithms.
• Mutual Information Deference (MID), Mutual Information Quotient (MIQ), and Mutual Information Base (MIBASE).

SUPPORT VECTOR MACHINE
• Support Vector Machine (SVM) is a supervised learning model for classification and regression tasks.
• SVM algorithm learns a classification hyperplane (decision boundary) by maximizing the minimum distance between data points belonging to various classes.
• We use soft-margin SVMs that allow certain data points to be misclassified.
• The hyperplane is acquired by solving a loss function with constraints:
  \[
  \sum_{i=1}^{N} \alpha_i y_i (x_i^T \omega + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, N
  \]
  \[
  C > 0 \text{ controls the trade-off between the margin and the penalty term } p \frac{1}{2} \|w\|.
  \]
• \(\alpha_i\) is the slack variable.

CLASSIFICATION ENVIRONMENT
• Machine learning tools:
  - SVM: a library developed in C language.
  - Tune the cost factor and the trade-off parameter that controls the training error and the margin.
• LSTM: PyBrain, a modular Machine Learning library for Python.
• Use to generate LSTM models with 37-dimensional inputs, 1 hidden layer, and 1-dimensional outputs.

PERFORMANCE EVALUATION
• SVM and LSTM training and testing datasets:
  - SVM_1 and LSTM_1: Slammer and Nimda.
  - SVM_2 and LSTM_2: Slammer and Code Red I.
  - SVM_3 and LSTM_3: Nimda and Code Red I.

EXPERIMENTAL PROCEDURE
• SVM and LSTM models are generated using both unbalanced and balanced datasets.
  - Unbalanced data: number of regular data is larger than number of anomalies.
  - Balanced data: contain all anomalies and randomly selected equivalent number of regular data.
• Step 1: Train and test the SVM and LSTM models using 37 features.
• Step 2: Select 10 most relevant features. Train and test SVM models. Skip this Step for LSTM models.
• Step 3: Evaluate performance of SVM and LSTM models.
• Step 4: Tune SVM and LSTM model parameters to achieve the best performance.

LONG SHORT-TERM MEMORY (LSTM) NEURAL NETWORK
• The LSTM approach employs a special form of the Recurrent Neural Networks.
• LSTM is capable of connecting time intervals to form a continuous memory.
• LSTM cell, also called the ‘memory block’:
  - Forget gate \(f_t\): discards the useless memories according to the cell state
  - Input gate \(i_t\): controls the information that will be updated in the LSTM cell
  - Output gate \(o_t\): controls the output.

NOTICE
• The accuracy and F-score using the SVM_2 models for unbalanced and balanced datasets.

REFERENCES