Paper Summary
Deep Packet: A Novel Approach For Encrypted Traffic Classification Using Deep Learning

Sarthak | 2020CS10379

May 2024

Problem Statement

The paper addresses the challenge of accurately classifying network traffic and identifying application to ensure Quality of Service (QoS) using DeepLearning techniques.

Motivation

- Increased amount of encrypted traffic, turning data into pseudo-random making it difficult to identify discriminative patterns
- Previous methods require patterns or features to be extracted by experts, being prone to error, time-consuming and costly
- Need for blocking P2P applications(high bandwidth & copyright issues) using protocol embedding techniques to bypass traffic control systems

Key Idea

Using Deep Learning methods (Stacked Autoencoders and Convolutional Neural Networks) to classify network traffic, eliminating the need for manual feature extraction. Assumes all applications encrypt their packets with different pseudo-random generators.

Framework

Preprocessing

- Removal of the Ethernet Header
- Discarded irrelevant packets

- Truncation/Zero-padding done to have uniform packet size
- Packet bytes divided by 255 to have values between 0 and 1
- Batch Normalization done to speed up learning (keep distribution in all batches constant)
- IP masking done to prevent the NN from learning classification form IP addresses

Architecture

Stacked-AutoEncoders

- Five Fully Connected layers, stacked on top of each other
- Each Autoencoder tries to learn a compressed representation of the dataset, i.e., approximately learns the identity function
- Used as an unsupervised technique for automatic feature extraction
- Output of the encoder part considered as a high-level set of discriminative features for the classification task
- First trained in a greedy layer-wise fashion for 200 epochs with RMS as loss, i.e. train each layer while freezing all other layers
- Then fine-tuned for 200 epochs, with categorical cross entropy as loss

Convolutional Neural Network

- Two consecutive 1-D convolutional layers, followed by a pooling layer followed by a 3 layered fully connected layer with softmax classifier
- Captures the spatial codependency of data by employing pooling & convolutional layer
- Trained for 300 epochs with categorical cross entropy as loss function

Misc

- Used Dropout regularization with probability = 0.05
- Split Dataset into 64,16,20 for Train, Validation and Testing
- Early stopping to prevent over-fitting
- Used Adam Optimizer for Backpropagation
- ReLu used as activation function in both NNs
- Grid search done on the subspace of hyper-parameters space to select the ones which result in the best performance

Contributions

- Elimination of Manual Feature Extraction, Combines feature extraction and classification into one system
- High Accuracy in Traffic Classification, performs better than state-of-art methods
- Capable of identifying encrypted traffic and distinguishing between VPN and non-VPN traffic

Strengths

- No need to manually extract features
- Can identify traffic at both granular levels high accuracy
- Can accurately classify P2P applications
- Can handle more complex tasks like multi-channel (intra-app) classification
- Learns to categorize similar apps together (e.g. YouTube and Vimeo)

Weaknesses

- Difficult to integrate to High Speed Networks due to computational demands of DL methods
- Unable to classify Tor's traffic , having the same encryption scheme for all applications
- Rely heavily on the quality and diversity of training data