SDN Traffic Classification using Deep Learning

Shrinidhi Nayak(ENG20CS0348), Sriraksha P(ENG20CS0358), Swati Priya(ENG20CS0375)
Department of Computer Science and engineering
Dayananda sagar university

Abstract

Accurate traffic classification is the basis of various network activities such as network management and network security checks. Port-based approaches, deep packet inspections, and machine learning are widely used techniques for classifying and analyzing network traffic flows. However, in recent years, due to the rapid increase in the number of Internet users, the growth of Internet traffic has increased explosively. As a result, both the port-based approach and the deep packet inspection approach are inefficient due to the exponential growth of Internet applications that generate high computational costs. The new software-defined networking paradigm has restructured the network architecture by separating the control plane from the data plane and creating a centralized network controller that maintains a global view of the entire network.

Introduction

Traffic classification serves as the cornerstone for a multitude of network operations, encompassing tasks such as network management and security enforcement. Traditionally, techniques such as port-based methodologies, deep packet inspections, and machine learning have been prevalent for analyzing and categorizing network traffic flows. However, the burgeoning population of internet users has precipitated an unprecedented surge in internet traffic. Consequently, both port-based and deep packet inspection methods have proven inefficient, burdened by the soaring computational demands induced by the proliferation of diverse internet applications.

In response to these challenges, the emergence of the software-defined networking paradigm has instigated a significant restructuring of network architecture. This paradigm delineates the control plane from the data plane, establishing a centralized network controller tasked with maintaining a comprehensive global perspective of the entire network. Through this architectural overhaul, software-defined networking introduces a flexible and scalable framework, poised to address the escalating complexities inherent in modern network environments.

Challenges of Traditional Methods

The exponential growth of internet users and applications has led to a surge in network traffic. Traditional traffic classification methods like port-based approaches and deep packet inspection (DPI) are becoming inefficient due to:

- Scalability Issues: These methods struggle to handle the vast amount of data generated by modern applications.
- Computational Cost: DPI, in particular, requires significant processing power to inspect individual packets.

SDN and Deep Learning Approach

SDN offers a centralized network control plane that provides a global view of the network. This project leverages this architecture to implement a deep learning model for traffic classification.

Implementation

This project aimed to implement a Deep Learning model for SDN that can identify normal and abnormal traffic, aiding network administrators in detecting potential attacks. Using python RYU and Miniet, we were able to generate a random SDN traffic dataset containing more than 100,000 rows in both normal packets(1 labels) and abnormal packets (0 labels) We were able to reach a 98.38% accuracy on the test set with a total of 92,769 trainable parameters.

Dataset Generation: A random SDN traffic dataset containing over 100,000 samples (normal and abnormal packets) was generated using RYU and Mininet.

Data Preprocessing: The data was preprocessed using Python libraries like Pandas and scikit-learn to handle missing values, perform feature engineering, and normalize the data.

- Missing values were handled using df.dropna().
- Irrelevant features like source, destination IP, switch, and port number were removed using df.drop().
- Categorical features were converted to one-hot encoded using pd.get_dummies().
- Data was normalized using MinMaxScaler().

Model Development: A deep learning model with a sequential architecture was built using Keras. Batch Normalization and Dropout were used to prevent overfitting.The model was compiled with binary crossentropy loss, Adam optimizer, and accuracy metric.

Model Training and Evaluation: The model was trained on a portion of the dataset with validation set monitoring to prevent overfitting.

- Training involved fitting the model with the prepared data (X_train and y_train) for 30 epochs with a batch size of 250 and early stopping for loss minimization.
- Training and validation loss/accuracy were visualized using Matplotlib.

The final model achieved an accuracy of 98.38% on the test set with 92,769 trainable parameters.

1. Model training

```
model.fit(X_train,
        y_train,
        epochs=30,
        batch_size=250,
        verbose=1,
        validation_split=0.2,
        callbacks=[es])
```

2. Evaluating the test set accuracy

```
results = model.evaluate(X_test, y_test,
batch_size=250)
print("Test set accuracy = {} %".format(
results[1]*100))
```

Results

The Deep Learning model achieved a high accuracy of 98.38% on the test set, demonstrating its effectiveness in classifying normal and abnormal traffic in SDN environments.

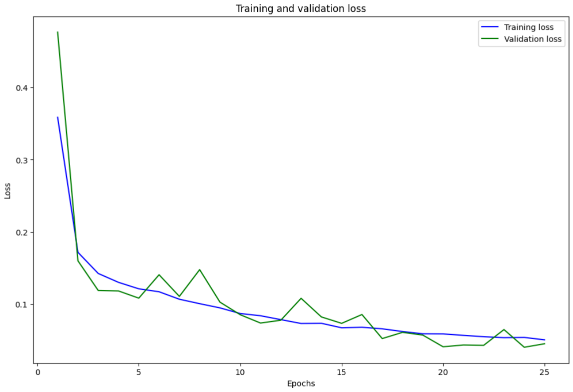

Fig. 1: Training and validation loss

Training loss indicates how well the model performs on the training data.

Validation loss indicates how well the model performs on a separate set of data (validation data).

Ideally, the training loss and validation loss should both decrease as the model is trained.

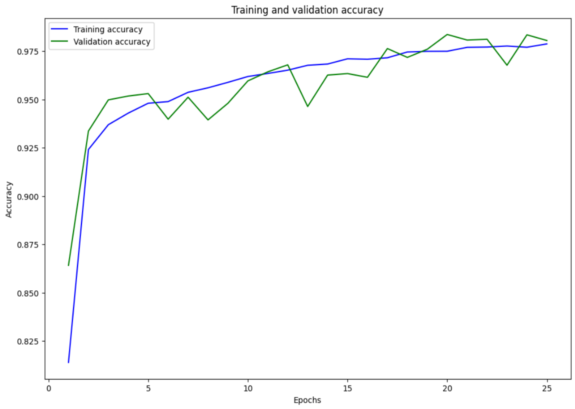

Fig. 2: Training and validation accuracy

Training accuracy indicates how well the model performs on the training data.

Validation accuracy indicates how well the model performs on a separate set of data (validation data).

The training accuracy and validation accuracy should both increase as the model is trained.

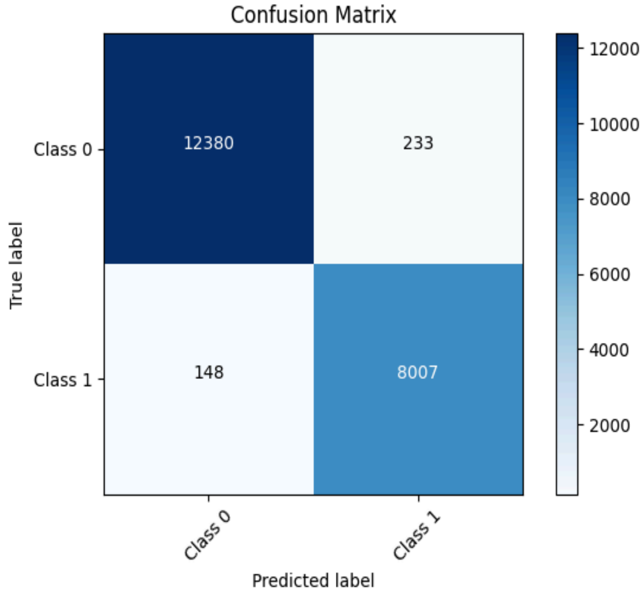

Fig 3: Confusion matrix

Conclusion

This project successfully implemented a deep learning model for SDN traffic classification, achieving high accuracy with a relatively simple architecture. This approach offers a promising alternative to traditional methods for managing and securing traffic in modern networks.

Future Work

This project demonstrates the potential of deep learning for SDN traffic classification. Further exploration could involve:

- Classifying traffic beyond normal and abnormal to identify specific applications (e.g., video streaming, email).
- Explore the use of different Deep Learning architectures like Convolutional Neural Networks (CNNs) for potentially improved performance.
- Investigate techniques for handling encrypted traffic data.
- Integrate the model into a real-world SDN controller for online traffic classification.

References

Ali Malik, Ruairi de Frein, Mohammed al-zeyadi, Javier Andreu-Perez, "Intelligent SDN Traffic Classification using Deep Learning" EEE ICCCI conference At: Nagoya, Japan.