

DEEP LEARNING BASED NETWORK TRAFFIC CLASSIFICATION OF P2P AND VOIP APPLICATIONS

A PROJECT REPORT

Submitted by

V.SIVA CHANDIRAN (810015104315)

S. SYED IBRAHIM (810015104718)

in partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING



**UNIVERSITY COLLEGE OF ENGINEERING, BIT CAMPUS,
TIRUCHIRAPPALLI**

ANNA UNIVERSITY: CHENNAI 600 025

APRIL 2019

**UNIVERSITY COLLEGE OF ENGINEERING,
BIT CAMPUS,**

TIRUCHIRAPPALLI-620 024

BONAFIDE CERTIFICATE

Certified that this project report “**DEEP LEARNING BASED NETWORK TRAFFIC CLASSIFICATION OF P2P AND VOIP APPLICATIONS**” is the bonafide work of “**V. SIVA CHANDIRAN (810015104315) and S. SYED IBRAHIM (810015104718)**” who carried out the project work under my supervision.

SIGNATURE

Dr. D. VENKATESAN

Head of the department

Assistant Professor

Department of CSE

University College of Engineering

Anna University, BIT Campus

Tiruchirappalli - 620 024

Submitted for the ANNA UNIVERSITY Viva-voce Examination held on

.....

SIGNATURE

Mrs. A. JENEFA

Supervisor

Teaching Fellow

Department of CSE

University College of Engineering

Anna University, BIT Campus

Tiruchirappalli - 620 024

Internal Examiner

External Examiner

DECLARATION

We hereby declare that the work entitled “**DEEP LEARNING BASED NETWORK TRAFFIC CLASSIFICATION OF P2P AND VOIP APPLICATIONS**” is submitted in partial fulfilment of the requirements for the award of the degree in B.E.-Computer Science and Engineering, University College of Engineering, BIT Campus, Tiruchirappalli, is a record of our own work carried out by us during the academic year 2018-2019 under the guidance of **Mrs.A.JENEFA**, Teaching Fellow, Department of Computer Science and Engineering, University College of Engineering, BIT Campus, Tiruchirappalli. The extent and source of information are derived from the existing literature and have been indicated through the dissertation at the appropriate places. The matter embodied in this work is original and has not been submitted for the award of any other Degree, either in this or any other University.

V. SIVACHANDIRAN (810015104315)

S. SYED IBRAHIM (810015104718)

I certify that the declaration made above by the candidate is true.

Signature of the Project Guide

Mrs. A. JENEFA

Teaching Fellow

Department of Computer Science and Engineering

University College of Engineering (BIT Campus)

Tiruchirappalli – 620 024

ACKNOWLEDGEMENT

First and foremost, of all, we would like to thank our beloved Parents and God Almighty for giving us the strength, knowledge, ability and opportunity to undertake this project study and to persevere and complete it satisfactorily.

We would like to convey our thanks to our honourable Dean **Dr.T.SENTHILKUMAR**, Associate Professor, for having provided us with all required facilities not only to complete our project without hurdles but also for the entire course of study.

We extremely indebted to our Head of the Department **Dr. D. VENKATESAN**, Assistant Professor, Department of Computer Science and Engineering, for the devoted attention shown upon us and making the project a grant success.

We also express our sincere thanks to our Project Coordinators **Mr.C.SANKARRAM**, Assistant Professor, **Mr. P. KARTHIKEYAN**, Assistant professor, **Mr. C. SURESH KUMAR**, Teaching Fellow, Department of Computer Science and Engineering, for their constant inspiration and support.

We whole heartedly express our gratitude to our Internal Project Guide **Mrs.A.JENEFA**, Teaching Fellow, Department of Computer Science and Engineering for her valuable guidance and help for the successful completion of the project.

We extend our thanks to all other Teaching and Non-Teaching Staff members for their constant encouragement and moral support.

And we would like to thank all our well-wishers who have encouraged us to do this project.

ABSTRACT

Network Traffic Classification is the crucial phase of network monitoring. The network traffic once categorized per application it can be imposed with appropriate security policies to improve the performance of the network. The port and payload based traffic classification techniques used in the past decade relapsed owing to new techniques of encryption and tunnelling emerging day-by-day. Recently, the statistical classification employing data mining techniques and analysing the attributes for characterization of network traffic proved high efficiency. This proposed work characterized the network traffic based on Deep learning approach. It initially involves learning of spatial features of network traffic by RNN and comparing the obtained results with other deep learning algorithms like Stack auto Encoder, DNN to check the efficiency of proposed system. This system is designed to classify the network traffic accurately that resulted in high F-Score value of 0.93 when it is compared with other Deep learning algorithms

TABLE OF CONTENTS

Chapter No	Title	Page No
	ABSTRACT	v
	LIST OF TABLES	ix
	LIST OF FIGURES	x
1.	INTRODUCTION	01
	1.1 Port based Classification	02
	1.2 Payload based Classification	02
	1.3 Deep Packet Inspection	03
	1.3.1 Pattern analysis	03
	1.3.2 Numerical analysis	04
	1.3.3 Behavioural and Heuristic analysis	04
	1.4 Statistical Analysis	05
	1.5 Machine Learning Approaches	05
	1.5.1 Supervised Machine Learning	06
	1.5.2 unsupervised machine learning	07
	1.5.3 Semi supervised machine learning	07
	1.6 Application of ML approaches in internet traffic classification:	
	1.6.1 Bayes Net	08
	1.6.2 Feed forward Neural Network	08
	1.6.3 C4.5 Decision tree	09
	1.6.4 DBSCAN	09
	1.6.5 K-Means	10
	1.6.6 Expectation maximisation	10
	1.6.7 C5.0 Decision Tree	11

TABLE OF CONTENTS

Chapter No	Title	Page No
	1.6.8 Stacked Auto Encoder	11
	1.7 Introduction to Deep Learning based network traffic classification	12
2.	LITERATURE SURVEY	13
3.	METHODOLOGY	20
	3.1 Proposed System	20
	3.2 Modules of the system	20
	3.2.1 Data Extraction	20
	3.2.2 Data pre-processing	20
	3.2.3 Training stacked Auto Encoder	22
	3.2.4 Evaluation of classifier	22
	3.3 Stacked Auto Encoder	23
4.	SYSTEM REQUIREMENTS	25
	4.1 Hardware Requirements	25
	4.2 Software Requirements	25
5.	RESULTS AND DISCUSSION	26
	5.1 Experimental setup	26
	5.2 Implementation	26
	5.2.1 Data Extraction	26
	5.2.2 Pre-processing	27
	5.2.3 Training of SAE	29
	5.2.4 Evaluation of classifier	29
	5.3 Performance Evaluation	31

TABLE OF CONTENTS

Chapter No	Title	Page No
6.	CONCLUSION AND FUTURE ENHANCEMENT	32
7.	APPENDIX	33
8.	REFERENCES	43

LIST OF TABLE

SNo	Tb No	Description	Pg No
1.	Tb 1.1	Patterns of strings commonly found in P2P applications	02
2.	Tb 5.1	Precision and Recall values of Applications	30
3.	Tb 5.7	Confusion Matrix	31

LIST OF FIGURES

S.No	Fig.No	Description	Pg.No
1.	Fig 3.1	System Architecture	21
2.	Fig 3.2	Structure of SAE	22
3.	Fig 3.3	SAE with two hidden layers	23
4.	Fig 3.4	Training of SAE in Fig3.3	24
5.	Fig5.1(a)	Facebook .pcap file analysed in wireshark	26
6.	Fig5.1(b)	Skype .pcap file analysed in wireshark	27
7.	Fig 5.2	Extracting statistical parameters from .pcap file	28
8.	Fig5.3(a)	Pre-processed Skype .csv file	28
9.	Fig5.3(b)	Pre-processed Facebook .csv file	29
10.	Fig 5.4	H2O cluster initialization for SAE	30
11.	Fig 5.6	Performance of Skype and eDonkey application	31
12.	Fig 5.7	Confusion matrix	31