STATISTICAL MODELS FOR LONG TERM NETWORK TRAFFIC IN ENTERPRISE NETWORKS

A.W.C.K. Atugoda

(148451H)

Degree of Master of Science

Department of Electronic and Telecommunication Engineering

University of Moratuwa

Sri Lanka

April 2019

STATISTICAL MODELS FOR LONG TERM NETWORK TRAFFIC IN ENTERPRISE NETWORKS

Atugoda Walawwe Chathurangi Kumari Atugoda

(148451H)

Dissertation submitted in partial fulfillment of the requirements of the degree Master of Science in Electronics and Automation Engineering

Department of Electronic and Telecommunication Engineering

University of Moratuwa

Sri Lanka

April 2019

Declaration

"I declare that this my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other university or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also I hereby grant to University of Moratuwa the non-exclusive right to produce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as article or books).

Name of student: A.W.C.K.Atugoda

Signature: Date:

The above candidate has carried out research for the Masters dissertation under my supervision.

Name of the supervisor: Dr.Upeka Premarathne

Signature: Date:

Abstract

With the rapid development of the internet it has converted the world into a global village and now a day we cannot even think of a micro second down time. For an instance, user demand has caused the internet to successfully combined with other networks. This expanded development has caused for huge internet traffic loads and network congestion.

For solving this key issue of the networks it is important to predict the traffic peaks in the network. These traffic peak is caused by a large amount of data being requested like in a download. If these traffic peaks are predictable then non critical traffic from another network can be scheduled to avoid peak to reduce the congestion and maximize utilization.

This dissertation introduces a method to solve that key issue. Curve fitting technique in Matlab and distributing fittings are used to build statistical models of predicting traffic. Once that identifying some drawbacks through curve fitting methodology it has been rejected and statistical models for long term network traffic in Enterprise network is used as the proposed technique. Pareto distribution, Beta-Prime distribution and Exponential distribution are derived as the statistical models to predict the traffic peak in Enterprise network. The analysis is conducted by looking at the predictability of a peak in terms of level crossing of a given level.

According to the available literature there was no such technique for predicting traffic peaks. As per the results curve fitting methodology error is significantly high. Beta-Prime and Exponential distribution are not good statistical models of predicting traffics due to huge error occurred when compared to the actual behavior of the network. But Pareto distribution is the best model of prediction on traffics in the network as it has vey less error when compared to the actual behavior of the network.

According to the results Pareto distribution is the best statistical model of predicting traffic peak. Once predicting the traffic peak can be scheduled the large data from other network for maximum utilization and to avoid the traffic congestion.

Key Words: traffic peak, level crossing, statistical model

Acknowledgement

I would like to acknowledge and express my heartfelt gratitude to my supervisor Dr.Upeka Premarathne, Senior Lecturer, Department of Electrionic and Telecommunication Engineering, University of Moratuwa for being with me in each and every step which took this research into a success while giving vital encouragement.

Further, I am heartly grateful to Dr. Roshan Regal, Director, Lanka Education and Research Network(LEARN), Dinesh Gunawardena, Secretory,LEARN and Senevi Herath, Network Engineer,LEARN for providing the data.

I am also heartily thankful to the coordinators of the MSc in Electronics and Automation Engineering pragramme and, all the staff members of Department of Electronics and Telecommunication Engineering, University of Moratuwa.

I am grateful to my parents and husband for encouraging me toward this work.

TABLE OF CONTENTS

Declaration…… ii

Abstract……. iii

Acknowledgment…….. iv

Table of contents…… v

List of figures……. vii

List of tables…….. ix

List of abbreviations……. xi

1. INTRODUCTION……. 1

1.1 Background and motivation……...1

1.2 Problem statement……..3

1.3 Aim and objectives…….3

1.4 Contribution……4

1.5 Report outline…… 5

2. LITREATURE SURVEY……..6

2.1 Related works…….6

2.1.1 Study based on packet length and inter arrival ……..6

2.1.2 Study based on self-similarity……... 10

3. METHODOLOGY…… 12

3.1 Data collection……... ……. 12

3.1.1 Institutional data collection ……... ……..…… 12

3.1.2 What is RRD tool …… 13

3.1.3 Using RRD tool for capturing data…… 13

3.1.4 Capture the graph through RRD tool …… 13

3.1.5 Individual data collection …… 15

3.1.6 Wireshark application …… 15

3.1.7 Using Wireshark tool for capturing data …… 15

3.2 Background analysis. ……... 16

3.2.1 Rearranging data in feasible way for analysis……... 16

3.2.2 Developing variables…… 20
3.2.3 Chi-square test for independence……. 21
3.2.4 Appling chi-square test for data set……... 22
3.2.5 Estimating probability density function…… 24
3.2.5.1verify the graphs obtain through the easy fit ……30
software using Matlab distributing fitting tool
3.2.6 Joint probability formula and its relevancy in this research. 31
3.2.7 Rice's formula……32
3.2.8 Predictability of level crossings…… 33
3.2.9 Actual level crossing rate……. 34
3.2.10 Institutional data analysis……. 35
4. RESULTS……. 36
4.1 Background analysis……. 36
4.1.1 Method A- using Matlab curve fitting tool…… ……... 36
4.1.2 Actual crossing rate……37
4.1.3 Comparison of results method A……... 37
4.1.4 Method B- using Easy-Fit software……... ……... 39
4.1.5 Predicted level crossing rate……. 39
4.1.6 Actual level crossing rate……. 42
4.1.7 Error percentage…….. 43
4.1.8 Average error percentage……. 45
4.2 Analysis for Institutional Data…….. 46
4.3 Mean error histograms…….. 46
5. CONCLUSION……. 52
References……. 53