

**A NEURAL NETWORK MODAL DECOMPOSITION
MECHANISM IN PREDICTING NETWORK TRAFFIC**



SHI JINMEI

UMS
UNIVERSITI MALAYSIA SABAH

**FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH
2023**

**A NEURAL NETWORK MODAL DECOMPOSITION
MECHANISM IN PREDICTING NETWORK TRAFFIC**

SHI JINMEI



**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

**FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH
2023**

UNIVERSITI MALAYSIA SABAH
BORANG PENGESAHAN STATUS TESIS

JUDUL : **A NEURAL NETWORK MODAL DECOMPOSITION MECHANISM
IN PREDICTING NETWORK TRAFFIC**

IJAZAH : **DOKTOR FALSAFAH SAINS KOMPUTER**

BIDANG : **SAINS KOMPUTER**

Saya **SHI JINMEI**, Sesi **2019-2023**, mengaku membenarkan tesis Doktoral ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:

1. Tesis ini adalah hak milik Universiti Malaysia Sabah
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/):

SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)

TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

TIDAK TERHAD

SHI JINMEI

SHI JINMEI
DI1821011A

Disahkan Oleh,



ANITA BINTI ARSAD
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH

(Tandatangan Pustakawan)



DR. LEAU YU BENG
Faculty of Computing and Informatics
Universiti Malaysia Sabah
84300 BIMP
leuybeng@ums.edu.my

Tarikh : 25 July 2023

(Dr. Leau Yu Beng)

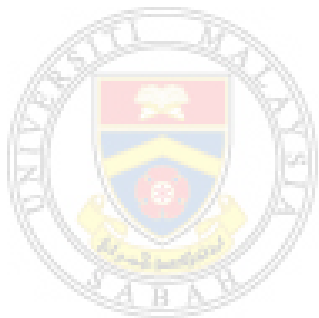
DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries, and references, which have been duly acknowledged.

11 April 2023

SHI JINMEI

Shi Jinmei
DI1821011A



UMMS
UNIVERSITI MALAYSIA SABAH

CERTIFICATION

NAME : **SHI JINMEI**

MATRIC NUM. : **DI1821011A**

TITLE : **A NEURAL NETWORK MODAL DECOMPOSITION
MECHANISM IN PREDICTING NETWORK TRAFFIC**

DEGREE : **DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE**

FIELD : **COMPUTER SCIENCE**

VIVA DATE : **11 APRIL 2023**



CERTIFIED BY;
UNIVERSITI MALAYSIA SABAH

Signature

1. MAIN SUPERVISOR
Dr. Leau Yu Beng


DR. LEAU YU BENG
Faculty of Computing and Informatics
Universiti Malaysia Sabah
Jalan UMS, 88400 Bdr. Lusa, Sabah
lybeng@ums.edu.my

2. CO-SUPERVISOR
Prof. Dr. Li Kun



ACKNOWLEDGEMENT

I would like to take this opportunity to express my deepest gratitude to my supervisors Dr. Leau Yu Beng and Associate Prof. Dr. Li Kun for their invaluable guidance, patience, and encouragement until the end of my academic journey. In particular, Dr. Leau's guidance to me made me feel warm as a Chinese student and also motivated me to continue my Ph.D. career.

I would like to thank the researchers in the field of network traffic prediction for providing us with academic achievements on time series prediction models, which can be used for reference to help us complete this research.

At the same time, I would like to thank the President of my university Dr. Yang Xiuying, who gave me full tuition support and gave me the opportunity to pursue my Ph.D. in the faculty of computing and informatics Universiti Sabah Malaysia.

Finally, I would like to thank my dear husband for his love, understanding, support, and encouragement until the completion of my academic journey.

Shi Jinmei

11 April 2023

ABSTRACT

Network traffic prediction is essential for effective network management as it can provide an early warning to the administrator before an incident occurs. This study designs a novel network traffic prediction model namely SAVE-AS. It embeds a new proposed Scalable Artificial Bee Colony (SABC) algorithm, Phase Space Reconstruction, Variational Mode Decomposition (VMD) and an integrated Extreme Learning Machine (ELM). The proposed mechanism starts by using SABC to update the model with a new solution and fine-tune the disturbances in each iteration to deal with the interference in order to find the best values that are also synchronously optimal. The SAVE-AS then constructs an adaptive selection operator. It adaptively selects the number of datasets after VMD optimization decomposition to precisely set the number of hidden layer nodes in an ELM to improve prediction accuracy. Meanwhile, the ELM model is trained using a variety of sub-data sequences that meet the requirements for minimizing computational complexity in modeling. Furthermore, the mechanism eliminates the poor sub-sequence caused by the volatility of the results to accelerate the convergence rate stability. The effectiveness of the model is evaluated using three datasets, i.e. Mackey-Glass, Lorenz chaotic time series of recognized benchmarks and a WIDE backbone of actual network traffic datasets. By comparing six existing model algorithms in all datasets, the results show that SAVE-AS can achieve faster convergence and high predictive accuracy while maintaining stability. Specifically, the predictive accuracy indexes such as Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) can reach a lowest optimum value of 1.1410, 0.1758 and 0.2263, and the average training time is reduced by 25.25%, 23.87% and 41.36%, respectively. The findings demonstrate that the proposed mechanism can predict network traffic more stably, accurately and rapidly in a short time regardless of time intervals or data sequence behavior. Consequently, it can provide

effective security warning guidance for network management as well as further improve network service quality.



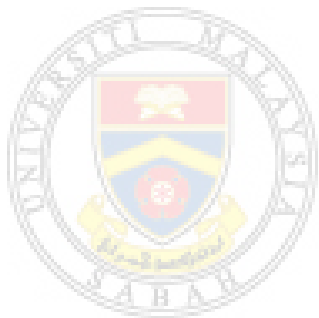
UMS
UNIVERSITI MALAYSIA SABAH

ABSTRAK

MEKANISME PENGURAIAN MODAL RANGKAIAN NEURAL DALAM MERAMAL TRAFFIC RANGKAIAN

Ramalan trafik rangkaian adalah penting untuk pengurusan rangkaian yang berkesan kerana ia boleh memberi amaran awal kepada pentadbir sebelum insiden berlaku. Kajian ini merekabentuk satu model ramalan trafik rangkaian yang baru iaitu SAVE-AS. Ia menyertakan algoritma Scalable Artificial Bee Colony (SABC) yang baru dicadangkan, Pembinaan Ruang Fasa, Penguraian Mod Variasi (VMD) dan Mesin Pembelajaran Ekstrim Bersepadu (ELM). Mekanisma yang dicadangkan ini bermula dengan menggunakan SABC untuk mengemas kini model dengan penyelesaian baru dan menala gangguan dalam setiap iterasi untuk menangani gangguan agar dapat mencari nilai-nilai terbaik yang juga secara serentak optimum. SAVE-AS kemudian membina operator pemilihan penyesuaian. Ia secara adaptif memilih jumlah set data selepas penguraian pengoptimuman VMD untuk menetapkan dengan tepat jumlah nod lapisan tersembunyi dalam ELM untuk meningkatkan ketepatan ramalan. Sementara itu, model ELM dilatih menggunakan pelbagai jujukan data yang memenuhi keperluan untuk meminimumkan kompleksiti komputasi dalam pemodelan. Selain itu, mekanisma ini menghilangkan jujukan yang lemah disebabkan oleh kerentanan keputusan untuk mempercepatkan kadar kestabilan penumpuan. Keberkesanan model ini dinilai menggunakan tiga set data, iaitu Mackey-Glass, Lorenz Siri masa huru-hara Lorenz sebagai piawaian yang diiktiraf dan set data trafik rangkaian sebenar, WIDE. Dengan membandingkan enam algoritma model yang sedia ada dalam semua set data, hasil kajian menunjukkan bahawa SAVE-AS dapat mencapai penumpuan yang lebih cepat dan ketepatan ramalan yang tinggi sambil mengekalkan kestabilan. Secara khusus, indeks ketepatan ramalan seperti Ralat Peratus Mutlak Rata-Rata (MAPE), Ralat Mutlak Rata-Rata (MAE) dan Ralat Purata Kuasa Dua (RMSE) boleh mencapai nilai optimum terendah masing-masing sebanyak 1.1410, 0.1758 dan 0.2263, dan masa

latihan purata dikurangkan sebanyak 25.25%, 23.87% dan 41.36% masing-masing. Dapatan kajian ini menunjukkan bahawa mekanisma yang dicadangkan dapat meramalkan trafik rangkaian dengan lebih stabil, tepat dan cepat dalam masa yang singkat tanpa mengira selang masa atau kelakuan rangkaian data. Oleh itu, ia dapat memberikan panduan amaran keselamatan yang berkesan untuk pengurusan rangkaian serta meningkatkan lagi kualiti perkhidmatan rangkaian.



UMS
UNIVERSITI MALAYSIA SABAH

LIST OF CONTENTS

	Page
TITLE	i
DECLARATION	ii
CERTIFICATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
<i>ABSTRAK</i>	vii
LIST OF CONTENTS	ix
LIST OF TABLES	xiii
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xviii
CHAPTER 1: INTRODUCTION	1
1.1 Overview and Motivation	1
1.2 Research Background	6
1.3 Problem Statement	7
1.4 Research Aim and Hypothesis	9
1.5 Research Questions	9
1.6 Research Objectives	10
1.7 Research Scope	10
1.8 Research Methods	11
1.9 Research Contribution	16
1.10 Organization of the Thesis	18
1.11 Summary	20

CHAPTER 2: LITERATURE REVIEW	22
2.1 Introduction	22
2.2 Linear traffic prediction model	23
2.2.1 Traditional model	23
2.2.2 Linear model	24
2.3 Nonlinear traffic prediction model	27
2.3.1 Chaos Theory and Phase Space Reconstruction	28
2.3.2 Statistical Theory and Support Vector Machine	29
2.3.3 Neural Network Models	31
2.4 Decomposition technique	38
2.4.1 Wavelet Analysis	39
2.4.2 Empirical Mode Decomposition	40
2.4.3 Ensemble Empirical Mode Decomposition	41
2.4.4 Variational Mode Decomposition	42
2.5 Optimization algorithm	44
2.5.1 Genetic Algorithm	45
2.5.2 Particle Swarm Optimization	46
2.5.3 Quantum Genetic Algorithm	47
2.5.4 Fruit Fly Optimization Algorithm	47
2.5.5 Artificial Bee Colony	48
2.5.6 Other recent related algorithms	49
2.6 Integrated traffic prediction model	52
2.6.1 Overview of the integrated model	52
2.6.2 Existing integrated models and applications	53
2.7 Summary	82

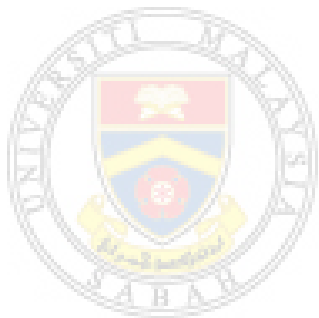
CHAPTER 3: METHODOLOGY	84
3.1 Introduction	84
3.2 The Novel SAVE-AS Network Traffic Prediction Model	85
3.2.1 Construction Method of Prediction Model	85
3.2.2 Design structure of prediction model	87
3.2.3 Proposed SAVE-AS prediction model for network traffic	90
3.3 Experimental Data Set	96
3.3.1 Lorenz chaotic time series	96
3.3.2 Mackey-glass chaotic time series	98
3.3.3 WIDE backbone actual network traffic	101
3.3.4 Predictability of Experimental Data	103
3.4 Summary	106
CHAPTER 4: A SCALABLE ARTIFICIAL BEE COLONY ALGORITHM IN DECOMPOSITION AND OPTIMIZATION OF NETWORK TRAFFIC PREDICTION MODEL	107
4.1 Introduction	107
4.2 Main methods	108
4.2.1 Variational Mode Decomposition	108
4.2.2 Extreme Learning Machine	112
4.2.3 Phase Space Reconstruction	115
4.3 Optimization Method	117
4.3.1 Classical Artificial Bee Colony Algorithm	118
4.3.2 Problem in Artificial Bee Colony Algorithm	121
4.3.3 New Scalable Artificial Bee Colony Algorithm	121
4.4 Proposed SABC-VMD-PSR-ELM Prediction Model	124
4.5 The Simulation Verification	129

4.5.1	Lorenz chaotic time series	132
4.5.2	Mackey-Glass chaotic time series	136
4.5.3	WIDE backbone network traffic data	141
4.6	Summary	146
CHAPTER 5: ACCELERATING CONVERGENCE IN NETWORK TRAFFIC PREDICTION THROUGH ADAPTIVE SELECTION OPTIMIZED VARIATIONAL MODE DECOMPOSITION AND AN INTEGRATED EXTREME LEARNING MACHINE		148
5.1	Introduction	148
5.2	New Thought for SABC-VMD-PSR-ELM	149
5.2.1	Limitations of SABC-VMD-PSR-ELM	149
5.2.2	Proposed Solution	151
5.3	Proposed SABC-VMD-PSR-ELM with Adaptive Selection (SAVE-AS) Model	156
5.4	Performance Verification	159
5.4.1	Mackey-glass chaotic time series	160
5.4.2	Lorenz chaotic time series	168
5.4.3	WIDE Backbone Network traffic data	175
5.5	Summary	184
CHAPTER 6: CONCLUSION AND FUTURE WORK		186
6.1	Introduction	186
6.2	Achievements of the Research	188
6.3	Suggestions for Future Work	192
BIBLIOGRAPHY		194

LIST OF TABLES

		Page
Table 2.1	: Neural network model for network traffic prediction	36
Table 2.2	: Strengths and Limitations of Decomposition Techniques	43
Table 2.3	: Strengths and Limitations of Optimization Algorithms	51
Table 2.4	: Analysis of application status of integrated network traffic models	63
Table 3.1	: Statistical information of Hour network traffic series	101
Table 3.2	: m and τ of Hour network traffic series	105
Table 4.1	: Parameters Setting	130
Table 4.2	: Setup of experimental running environment	132
Table 4.3	: Comparison of Predictive Performance Metrics of the Models (Lorenz)	135
Table 4.4	: Comparison of Predictive Performance Metrics of the Models (Mackey-Glass)	139
Table 4.5	: Comparison of Predictive Performance Metrics of the Models ("Hour")	145
Table 5.1	: Configure the experiment running environment	160
Table 5.2	: Training time reduction percentage of Mackey-Glass chaotic time series	163
Table 5.3	: Comparison of predictive performance indicators belonging to the three models (Mackey-Glass)	164
Table 5.4	: Parameter Settings in different models	165
Table 5.5	: Comparison of the predictive performance indicators belonging to several models (Mackey-Glass)	167
Table 5.6	: Training time reduction percentage of Lorenz chaotic time series	171
Table 5.7	: Comparison of prediction performance index values of the three models (Lorenz)	172
Table 5.8	: Comparison of prediction performance indicators of several comparison models (Lorenz)	174
Table 5.9	: Training time reduction percentage of "Hour" chaotic time series	178

Table 5.10	: Comparison of predicted performance index values of the three models (" Hour ")	179
Table 5.11	: Comparison of predicted performance indicators of several comparison models (" Hour ")	181
Table 5.12	: Computational stability of SAVE-AS model in three different datasets	184



UMS
UNIVERSITI MALAYSIA SABAH

LIST OF FIGURES

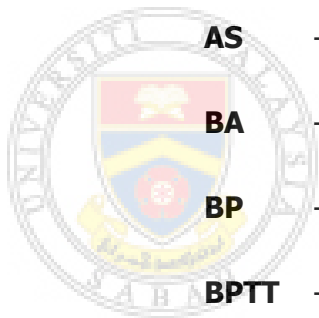
	Page
Figure 1.1 : The number of Global Internet Users	1
Figure 1.2 : Schematic diagram of research process	13
Figure 2.1 : Decomposition techniques in network traffic prediction models	39
Figure 2.2 : Related optimization algorithm in network traffic prediction model	45
Figure 3.1 : Calculation flow chart of the designed integrated prediction model	90
Figure 3.2 : Research framework of network traffic prediction model	91
Figure 3.3 : A novel SAVE-AS prediction model for network traffic time series	95
Figure 3.4 : Three-dimensional evolution trajectory of Lorenz model with different initial values	96
Figure 3.5 : Channel evolution diagram of $x(t)$ and $y(t)$ under different initial $z(t)$ values	97
Figure 3.6 : Mackey-Glass chaotic sequence at 300s	98
Figure 3.7 : Data series and phase difference diagram for $x(0) = 1.2$ and $\eta = 13$	99
Figure 3.8 : Data series and phase difference diagram for $x(0) = 1.2$ and $\eta = 30$	100
Figure 3.9 : Hour network traffic data series	102
Figure 3.10 : Hour network traffic data sequence after logarithm	103
Figure 4.1 : SABC algorithm model implementation of pseudo code	127
Figure 4.2 : Flow chart of parameter optimization in prediction method	128

Figure 4.3	: Lorenz chaotic time series	132
Figure 4.4	: Mean fitness curve of different models in the training stage (Lorenz)	133
Figure 4.5	: Average predicted output curve of different models in the test phase (Lorenz)	134
Figure 4.6	: Training and testing time of different models (Lorenz)	136
Figure 4.7	: Mackey-Glass chaotic time series	137
Figure 4.8	: Average fitness curve of different models in the training stage (Mackey-Glass)	138
Figure 4.9	: Average predicted output curves of different models in the test phase (Mackey-Glass)	139
Figure 4.10	: Training and testing time of different models (Mackey-Glass)	141
Figure 4.11	: Network traffic data set of "Hour"	142
Figure 4.12	: Mean fitness curves of different models at the training stage ("Hour")	143
Figure 4.13	: Average predicted output curves of different models in the test phase ("hour")	144
Figure 4.14	: Training and testing time of different models ("Hour")	146
Figure 5.1	: Flow chart of SABC-VMD-PSR-ELM Model	150
Figure 5.2	: Flow chart of SAVE-AS Model	159
Figure 5.3	: Mackey-Glass chaotic time series	161
Figure 5.4	: Average fitness curve of VMD-ELM models optimized by three different ABC-based algorithms in the training stage (Mackey-Glass)	162

Figure 5.5	: Training and testing time of VMD-ELM models optimized by three different ABC-based algorithms (Mackey-Glass)	163
Figure 5.6	: Training and testing time of different comparison models (Mackey-Glass)	167
Figure 5.7	: Lorenz chaotic time series	169
Figure 5.8	: Average fitness curve of VMD-ELM models optimized by three different ABC-based algorithms in the training stage (Lorenz)	170
Figure 5.9	: Training and testing time of VMD-ELM models optimized by three different ABC-based algorithms (Lorenz)	171
Figure 5.10	: Training and testing time of different comparison models (Lorenz)	174
Figure 5.11	: "Hour" network traffic data set	176
Figure 5.12	: Average fitness curve of VMD-ELM models optimized by three different ABC-based algorithms in the training stage (" Hour ")	177
Figure 5.13	: Training and testing time of VMD-ELM models optimized by three different ABC-based algorithms (" Hour ")	178
Figure 5.14	: Training and testing time of different comparison models ("Hour")	181

LIST OF ABBREVIATIONS

- ABC** - Artificial Bee Colony algorithm
- ANN** - Artificial Neural Network
- APSO** - Adaptive Particle Swarm Optimization
- AR** - Auto Regressive
- ARMA** - Auto-Regressive Moving Average
- ARIMA** - Auto Regressive Integrated Moving Average
- AS** - Adaptive Selection
- BA** - Bat Algorithm
- BP** - Back Propagation
- BPTT** - Back Propagation Trough Time
- CAIDA** - Cooperative Association for Internet Data Analysis
- CEEMD** - Complementary Ensemble Empirical Mode Decomposition
- CT** - Chaos Theory
- CVNN** - Complex-valued Neural Network
- CNN** - Convolutional Neural Networks
- DBN** - Deep Belief Network
- DE** - Differential Evolution
- DR** - Dynamic Reservoir
- DWT** - Discrete Wavelet Transform



UMS
UNIVERSITI MALAYSIA SABAH

- EEMD** - Ensemble Empirical Mode Decomposition
- EELM** - Efficient Extreme Learning Machine
- ELM** - Extreme Learning Machines
- EM** - Error Minimized
- EMD** - Empirical Mode Decomposition
- ESN** - Echo State Network
- EWT** - Empirical Wavelet Transform
- FA** - Firefly Algorithm
- FARIMA** - Fractional Auto Regressive Integrated Moving Average
- FOA** - Fruit Fly Optimization Algorithm
- FP** - Fitness Prediction
- FS** - Free Search
- GA** - Genetic Algorithm
- GM** - Gaussian model
- GMDH** - Group Method of Data Handling
- GPR** - Gaussian Process Regression
- GSA** - Gravity search algorithm
- GST** - Grey System Theory
- HHT** - Hilbert-Huang Transform
- IABC** - Improved Artificial Bee Colony
- ICA** - Imperialist Competitive Algorithm



IEPM - Internet End-to-End Performance Monitoring

IFOA - Improved Fruit Optimization Algorithm

IFS - Improved Free Search

IGSA - Improved Gravity search algorithm

IHS - Improved Harmony Search

IMF - Intrinsic Mode function

IPPM - IP Performance Metrics

IPSO - Improved Particle Swarm Optimization

ITU - International Telecommunication Union

LMD - Local Mean Decomposition

LSSVM - Least Squares Support Vector Machines

LSTM - Long short-term Memory

LVCMFOA - Levy flight function and cloud generator Fruit Fly Optimization Algorithm

MA - Moving Average

MAE - Mean Absolute Error

MAD - Mean Absolute Deviation

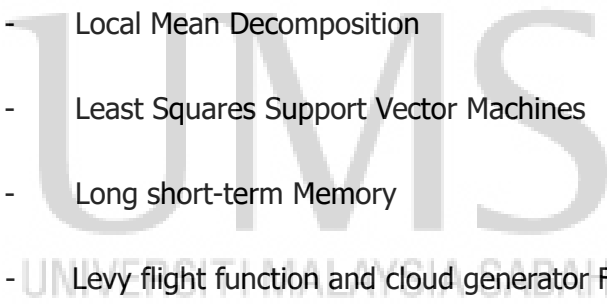
MAPE - Mean Absolute Percentage Error

MAWI - Measurement and Analysis on the WIDE Internet

NNT - Neural Network Theory

OBL - Opposition-Based Learning

PSO - Particle Swarm Optimization



PSR	-	Phase Space Reconstruction
QFOA	-	Quantum Fruit Optimization Algorithm
QGA	-	Quantum Genetic Algorithm
QNN	-	Quantum neural network
QPSO	-	Quantum Particle Swarm Optimization
RBF	-	Radial Basis Function
RMSE	-	Root Mean Squared Error
RNN	-	Recurrent Neural Network
SABC	-	Scalable Artificial Bee Colony
SAVE-AS	-	SABC-VMD-AS-PSR-ELM
SCS	-	Spatiotemporal Compressive Sensing
SLAC	-	Stanford Linear Accelerator Center
SLFN	-	Single Hidden Layer Feedforward Network
SLT	-	Seasonal Loess Trend
ST	-	Statistical Theory
ST	-	Seasonal Transform
SVM	-	Support Vector Machines
TSD	-	Time Series Decomposition
VMD	-	Variational Mode Decomposition
WA	-	Wavelet Analysis
WD	-	Wavelet Decomposition
WT	-	Wavelet Transform



UMS
UNIVERSITI MALAYSIA SABAH

CHAPTER 1

INTRODUCTION

1.1 Overview and Motivation

With the rapid expansion of information in the 21st century coupled with increasing internet penetration rate worldwide, the Internet has consequently become a part and parcel of our everyday life. According to the International Telecommunication Union's (ITU) 2022 Annual Internet Survey report, the global Internet user-base continues to increase rapidly from 2013 to 2022. As of 29 January 2022, the number of global Internet users reached 4.95 billion, accounting for 62.5% of the world's population.

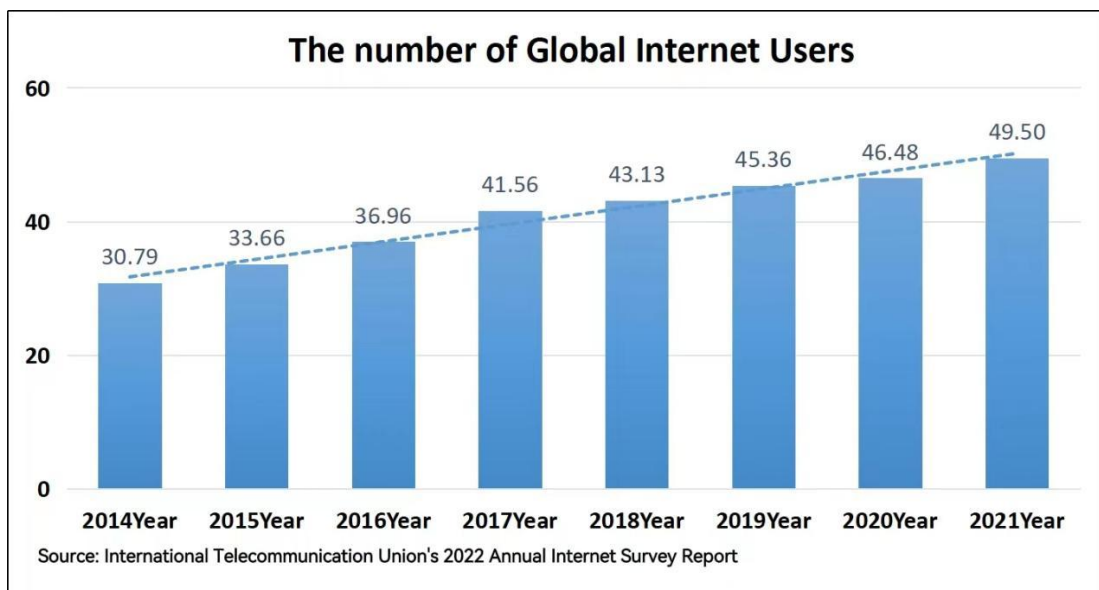


Figure 1.1 : The number of Global Internet Users