



**UNIVERSITI PUTRA MALAYSIA**

***ADAPTIVE LINUX-BASED TCP CONGESTION CONTROL ALGORITHM  
FOR HIGH-SPEED NETWORKS***

**MOHAMED A. ALRSHAH**

**FSKTM 2017 31**



**ADAPTIVE LINUX-BASED TCP CONGESTION CONTROL ALGORITHM  
FOR HIGH-SPEED NETWORKS**

**By**

**MOHAMED A. ALRSHAH**

**Thesis Submitted to the School of Graduate Studies, Universiti Putra  
Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of  
Philosophy**

**February 2017**



© COPYRIGHT UPM

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial uses of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright ©Universiti Putra Malaysia



## DEDICATIONS

*I would like to dedicate this thesis to my beloved motherland  
"LIBYA".*

&

*To my family and all whom I love.*



©

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in  
fulfilment of the requirement for the degree of Doctor of Philosophy

## **ADAPTIVE LINUX-BASED TCP CONGESTION CONTROL ALGORITHM FOR HIGH-SPEED NETWORKS**

By

**MOHAMED A. ALRSHAH**

**February 2017**

**Chairman: Mohamed Othman, PhD**  
**Faculty: Computer Science and Information Technology**

Recently, high-speed networks are widely deployed and their necessity is rapidly increasing everyday. In general, high-speed networks are deployed to provide connectivity among computing elements, storage devices and/or data centers in order to provide fast and reliable services for end-users. High-speed networks can be classified as: (1) short-distance networks, such as local area networks and data center networks, and (2) long-distance networks, such as metropolitan and wide area networks, which occasionally employ the oceanic and/or transatlantic links to provide a fast connectivity among the scattered data centers located in different places around the world.

Indeed, the overall performance of such networks is significantly influenced by the Transmission Control Protocol (TCP). Although TCP is the predominant transmission protocol used in Internet, its Congestion Control Algorithm (CCA) is still unable to adapt to high-speed networks, which are not the typical environment for which most CCAs were designed. For this reason, the employment of TCP over high-speed networks causes an extreme performance degradation leads to a poor bandwidth utilization due to the unavoidable network characteristics such as small buffer, long RTT and non-congestion loss.

In order to reduce the sensitivity to packet loss and to improve the ability of TCP CCA on dealing with small buffer regimes as in short-distance and low-BDP networks, this work proposes a novel loss-based TCP CCA, namely AF-based, designed for high-speed and short-distance networks. Thereafter, extensive simulation experiments are carried out to evaluate the performance of the proposed AF-based CCA compared to C-TCP and Cubic-TCP, which are the default CCAs of the most commonly used operating systems. The results

show that AF-based CCA outperforms the compared CCAs in terms of average throughput, loss ratio and fairness, especially when a small buffer regime is applied. Moreover, the AF-based CCA shows lower sensitivity to the change of buffer size and packet error rate, which increases its efficiency.

Further, we propose a novel mathematical model to calculate the average throughput of the AF-based CCA. The main contributions of this model are: First, to validate the simulation results of AF-based CCA by comparing them to the numerical results of this model and to the results of NewReno as a benchmark. Second, to study the impact of  $\lambda_{max}$  parameter on the throughput and epoch time. Third, to formulate an equation to automate the configuration of  $\lambda_{max}$  parameter in order to increase the scalability of AF-based CCA. Fortunately, the results confirm the validity of the proposed algorithm.

Furthermore, we propose a new delay-based CCA to increase bandwidth utilization over long-distance networks, in which RTTs are very long, buffers are very large and packet loss is very common. This CCA contributes the novel Window-correlated Weighting Function (WWF), which correlates the value of the increase in cwnd to the magnitude of it. Thereafter, the gained increase is balanced using the weighting function according to the variation of RTT in order to maintain the fairness. Consequently, this behavior improves the ability of TCP to adapt to different long-distance network scenarios, which especially improves bandwidth utilization over high-BDP networks. Extensive simulation experiments show that WWF-based CCA achieves higher performance than the other CCAs while maintaining fairness. Moreover, it shows higher efficiency and stability than the compared CCAs, especially in the cases of big buffers which cause an additional delay.

Fundamentally, TCP-based applications naturally need to deal with links of any-distance without the need of human reconfiguration. For this reason, it becomes very necessary to design an adaptive CCA, which is able to serve simultaneously any-distance networks. Thus, we propose a novel adaptive TCP CCA, namely Agile-TCP, which combines both AF-based and WWF-based approaches. This combination reduces the sensitivity to packet loss, buffer size and RTT variation, which in turn, improves the total performance of TCP over any-distance networks. Beyond that, a Linux kernel CCA module is implemented as a real product of the Agile-TCP. For evaluation purpose, a real test-bed of single dumb-bell topology is carried out using the well-known Dummynet network emulator. Fortunately, the results show that Agile-TCP outperforms the compared CCAs in most scenarios, which is very promising for many application such as cloud computing and big data transfer.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia  
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

## ADAPTIF LINUX-BERASASKAN TCP KESESAKAN ALGORITMA KAWALAN BAGI RANGKAIAN KELAJUAN TINGGI

Oleh

**MOHAMED A. ALRSHAH**

Februari 2017

**Pengerusi: Mohamed Othman, PhD**  
**Fakulti: Sains Komputer dan Teknologi Maklumat**

Kebelakangan ini, rangkaian kelajuan tinggi telah digunakan secara meluas dan keperluan rangkaian tersebut semakin meningkat dengan pesat setiap hari. Secara umum, rangkaian kelajuan tinggi dikerahkan untuk menyediakan sambungan antara elemen pengkomputeran, peranti penyimpanan dan/atau pusat data dalam usaha untuk menyediakan perkhidmatan yang cepat dan boleh dipercayai untuk pengguna akhir. Rangkaian kelajuan tinggi boleh diklasifikasikan sebagai: (1) rangkaian jarak pendek, seperti rangkaian kawasan tempatan dan rangkaian pusat data, dan (2) rangkaian jarak jauh, seperti rangkaian kawasan metropolitan dan luas, yang kadang-kadang menggunakan lautan dan/atau pautan transatlantik untuk menyediakan sambungan yang cepat antara pusat data yang bertaburan di tempat-tempat yang berbeza di seluruh dunia.

Malah, prestasi keseluruhan rangkaian tersebut dipengaruhi dengan ketara oleh Protokol Kawalan Penghantaran (TCP). Walaupun TCP ialah protokol penghantaran yang utama yang digunakan dalam Internet, Algoritma Kawalan Kesyakan (CCA) masih tidak dapat menyesuaikan diri dengan rangkaian berkelajuan tinggi, dimana ia adalah bukan persekitaran yang biasa untuk rekaan CCA. Oleh itu, bekerja dengan TCP melalui rangkaian kelajuan tinggi menyebabkan kemerosotan prestasi yang melampau dan membawa kepada penggunaan jalur lebar yang lemah menyebabkan ciri-ciri rangkaian yang tidak dapat dielakkan seperti penimbal kecil, kepanjangan RTT dan kehilangan bukan kesesakan.

Dalam usaha untuk mengurangkan sensitiviti kepada kehilangan paket dan untuk meningkatkan keupayaan TCP CCA untuk berurusan dengan rejim



penimbal kecil seperti dalam jarak pendek dan rangkaian rendah BDP, kerja ini mencadangkan novel TCP CCA berdasarkan badan, yang berasaskan AF, ia direka untuk kelajuan tinggi dan rangkaian jarak pendek. Selepas itu, eksperimen simulasi menyeluruh dijalankan untuk menilai prestasi CCA berasaskan AF yang dicadangkan berbanding C-TCP dan Cubic-TCP, dimana ia adalah sistem operasi yang paling biasa digunakan dalam CCA. Keputusan menunjukkan bahawa CCA berasaskan AF melebihi prestasi CCA yang dibandingkan dari segi pemprosesan purata, nisbah kerugian dan keadilan, terutamanya apabila rejim penimbal kecil digunakan. Selain itu, CCA berasaskan AF menunjukkan sensitiviti yang lebih rendah kepada perubahan saiz penimbal dan kesilapan kadar paket, yang meningkatkan kecekapan.

Selanjutnya, kami mencadangkan satu novel model matematik untuk mengira daya pemprosesan purata CCA berasaskan AF. Sumbangan utama model ini ialah: Pertama, untuk mengesahkan keputusan simulasi CCA berasaskan AF dengan membandingkannya dengan keputusan berangka model ini dan keputusan NewReno sebagai penanda aras. Kedua, untuk mengkaji kesan parameter  $\bar{I}z_{max}$  pada pemprosesan dan zaman masa. Ketiga, merumuskan persamaan untuk mengautomatiskan konfigurasi parameter  $\bar{I}z_{max}$  bagi meningkatkan kebolehan untuk diskala oleh CCA berasaskan AF. Mujurlah, keputusan mengesahkan kesahihan algoritma yang dicadangkan.

Tambahan pula, kami mencadangkan CCA berdasarkan kelewatan yang baru - untuk meningkatkan penggunaan jalur lebar melalui rangkaian jarak jauh, di mana RTTS yang sangat panjang, penimbal adalah sangat besar dan kehilangan paket yang sangat biasa. CCA ini menyumbang satu novel fungsi pemberat tetingskap-rapat (WWF), yang ada hubung kait nilai peningkatan cwnd yang besar. Selepas itu, peningkatan yang diperolehi adalah seimbang menggunakan fungsi pemberat mengikut perubahan RTT untuk mengekalkan keadilan itu. Oleh itu, tingkah laku ini meningkatkan keupayaan TCP untuk menyesuaikan diri dengan senario rangkaian jarak jauh yang berbeza, terutama meningkatkan penggunaan jalur lebar melalui rangkaian tinggi BDP. Eksperimen simulasi meluas menunjukkan bahawa CCA berasaskan WWF mencapai prestasi yang lebih tinggi daripada CCA lain, di samping mengekalkan keadilan. Selain itu, ia menunjukkan kecekapan dan kestabilan yang lebih tinggi berbanding CCA yang, terutamanya dalam kes-kes penimbal besar yang menyebabkan kelewatan tambahan.

Pada dasarnya, aplikasi berasaskan TCP-secara semula jadi perlu menangani kedua-dua pautan pendek dan jarak jauh pada masa yang sama. Atas sebab ini, ia menjadi keperluan untuk mereka bentuk CCA penyesuaian, yang mampu untuk berkhidmat untuk rangkaian secara serentak pendek dan jarak jauh. Oleh itu, kami mencadangkan satu novel TCP penyesuaian CCA, iaitu Agile-TCP, yang menggabungkan kedua-dua pendekatan berasaskan AF dan berasaskan WWF. Gabungan ini mengurangkan sensitiviti kepada kehilangan paket, saiz penimbal dan perubahan RTT, yang seterusnya, meningkatkan jumlah prestasi

TCP melalui rangkaian pendek dan jarak jauh pada masa yang sama. Selain itu, modul Linux kernel CCA dilaksanakan sebagai produk sebenar Agile-TCP. Bagi tujuan penilaian, yang sebenar-ujian topologi dumbbell tunggal dijalankan menggunakan emulator rangkaian Dummynet yang terkenal. Mujurlah, keputusan menunjukkan bahawa Agile-TCP melebihi prestasi CCA berbanding di kebanyakan senario, yang sangat menjanjikan untuk banyak aplikasi seperti pengkomputeran awan dan pemindahan data yang besar. Kebelakangan ini, rangkaian kelajuan tinggi telah digunakan secara meluas dan keperluan rangkaian tersebut semakin meningkat dengan pesat setiap hari. Secara umum, rangkaian kelajuan tinggi dikerahkan untuk menyediakan sambungan antara elemen pengkomputeran, peranti penyimpanan dan/atau pusat data dalam usaha untuk menyediakan perkhidmatan yang cepat dan boleh dipercayai untuk pengguna akhir. Rangkaian kelajuan tinggi boleh diklasifikasikan sebagai: (1) rangkaian jarak pendek, seperti rangkaian kawasan tempatan dan rangkaian pusat data, dan (2) rangkaian jarak jauh, seperti rangkaian kawasan metropolitan dan luas, yang kadang-kadang menggunakan lautan dan/atau pautan transatlantik untuk menyediakan sambungan yang cepat antara pusat data yang bertaburan di tempat-tempat yang berbeza di seluruh dunia.

Malah, prestasi keseluruhan rangkaian tersebut dipengaruhi dengan ketara oleh Protokol Kawalan Penghantaran (TCP). Walaupun TCP ialah protokol penghantaran yang utama yang digunakan dalam Internet, Algoritma Kawalan Kesesakan (CCA) masih tidak dapat menyesuaikan diri dengan rangkaian berkelajuan tinggi, dimana ia adalah bukan persekitaran yang biasa untuk rekaan CCA. Oleh itu, bekerja dengan TCP melalui rangkaian kelajuan tinggi menyebabkan kemerosotan prestasi yang melampau dan membawa kepada penggunaan jalur lebar yang lemah menyebabkan ciri-ciri rangkaian yang tidak dapat dielakkan seperti penimbal kecil, kepanjangan RTT dan kehilangan bukan kesesakan.

Dalam usaha untuk mengurangkan sensitiviti kepada kehilangan paket dan untuk meningkatkan keupayaan TCP CCA untuk berurusan dengan rejim penimbal kecil seperti dalam jarak pendek dan rangkaian rendah BDP, kerja ini mencadangkan novel TCP CCA berdasarkan badan, yang berasaskan AF, ia direka untuk kelajuan tinggi dan rangkaian jarak pendek. Selepas itu, eksperimen simulasi menyeluruh dijalankan untuk menilai prestasi CCA berasaskan AF yang dicadangkan berbanding C-TCP dan Cubic-TCP, dimana ia adalah sistem operasi yang paling biasa digunakan dalam CCA. Keputusan menunjukkan bahawa CCA berasaskan AF melebihi prestasi CCA yang dibandingkan dari segi pemrosesan purata, nisbah kerugian dan keadilan, terutamanya apabila rejim penimbal kecil digunakan. Selain itu, CCA berasaskan AF menunjukkan sensitiviti yang lebih rendah kepada perubahan saiz penimbal dan kesilapan kadar paket, yang meningkatkan kecekapan.

Selanjutnya, kami mencadangkan satu novel model matematik untuk mengira daya pemrosesan purata CCA berasaskan AF. Sumbangan utama model ini

ialah: Pertama, untuk mengesahkan keputusan simulasi CCA berasaskan AF dengan membandingkannya dengan keputusan berangka model ini dan keputusan NewReno sebagai penanda aras. Kedua, untuk mengkaji kesan parameter  $\hat{I}zmax$  pada pemprosesan dan zaman masa. Ketiga, merumuskan persamaan untuk mengautomasikan konfigurasi parameter  $\hat{I}zmax$  bagi meningkatkan kebolehan untuk diskala oleh CCA berasaskan AF. Mujurlah, keputusan mengesahkan kesahihan algoritma yang dicadangkan.

Tambahan pula, kami mencadangkan CCA berdasarkan kelewatan yang baru - untuk meningkatkan penggunaan jalur lebar melalui rangkaian jarak jauh, di mana RTTS yang sangat panjang, penimbal adalah sangat besar dan kehilangan paket yang sangat biasa. CCA ini menyumbang satu novel fungsi pemberat tetingskap-rapat (WWF), yang ada hubung kait nilai peningkatan cwnd yang besar. Selepas itu, peningkatan yang diperolehi adalah seimbang menggunakan fungsi pemberat mengikut perubahan RTT untuk mengekalkan keadilan itu. Oleh itu, tingkah laku ini meningkatkan keupayaan TCP untuk menyesuaikan diri dengan senario rangkaian jarak jauh yang berbeza, terutama meningkatkan penggunaan jalur lebar melalui rangkaian tinggi BDP. Eksperimen simulasi meluas menunjukkan bahawa CCA berasaskan WWF mencapai prestasi yang lebih tinggi daripada CCA lain, di samping mengekalkan keadilan. Selain itu, ia menunjukkan kecekapan dan kestabilan yang lebih tinggi berbanding CCA yang, terutamanya dalam kes-kes penimbal besar yang menyebabkan kelewatan tambahan.

Pada dasarnya, aplikasi berasaskan TCP-secara semula jadi perlu menangani kedua-dua pautan pendek dan jarak jauh pada masa yang sama. Atas sebab ini, ia menjadi keperluan untuk mereka bentuk CCA penyesuaian, yang mampu untuk berkhidmat untuk rangkaian secara serentak pendek dan jarak jauh. Oleh itu, kami mencadangkan satu novel TCP penyesuaian CCA, iaitu Agile-TCP, yang menggabungkan kedua-dua pendekatan berasaskan AF dan berasaskan WWF. Gabungan ini mengurangkan sensitiviti kepada kehilangan paket, saiz penimbal dan perubahan RTT, yang seterusnya, meningkatkan jumlah prestasi TCP melalui rangkaian pendek dan jarak jauh pada masa yang sama. Selain itu, modul Linux kernel CCA dilaksanakan sebagai produk sebenar Agile-TCP. Bagi tujuan penilaian, yang sebenar-ujian topologi dumbbell tunggal dijalankan menggunakan emulator rangkaian Dummynet yang terkenal. Mujurlah, keputusan menunjukkan bahawa Agile-TCP melebihi prestasi CCA berbanding di kebanyakan senario, yang sangat menjanjikan untuk banyak aplikasi seperti pengkomputeran awan dan pemindahan data yang besar.

## ACKNOWLEDGEMENTS

First and foremost, all praise is for *Allah Subhanahu Wa Taala* for giving me the strength, guidance, and patience to complete this thesis. I thank Allah for His immense grace and blessing every stage of my entire life. May blessing and peace be upon Prophet Muhammad *Sallallahu Alaihi Wasallam*, who was sent for mercy to the world.

I would like to express my sincere gratitude to my supervisor Prof. Dr. Mohamed Othman for the continuous support of my study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. His encouragement and help made me feel confident to overcome every difficulty I encountered in all the stages of this research. What I really learned from him, however, is his attitude to work and life - always aiming for excellence.

I would like to extend my gratitude and thanks to the distinguished supervisory committee members, Professor Dr. Borhanuddin Mohd Ali, and Dr. Zurina Mohd Hanapi for their encouragement and insightful comments.

I am very grateful to the Faculty of Computer Science and Information Technology and the staff of Postgraduate office, School of Graduate Studies, Library and Universiti Putra Malaysia, for providing me excellent research environment. Thanks to every person who has supported me to pursue and finish my Ph.D.

Words fail to express my love and appreciation to my father and my mother, your love and prayers are really the reason of my success. I am very grateful to my lovely wife Aida whose dedication, love and persistent confidence in me have taken the load off my shoulder. I owe her for being unselfish, her intelligence, passions, and ambitions pushed me to success. Special thanks for my sons Jihad, Ahmad and Ziad, you are the joy and the hope. Thanks for giving me your valuable time through all this long process. I promise I will never let you alone anymore. I am very thankful to my brothers, and sisters for their unflagging love and support throughout my life. I have no suitable words that can fully describe my everlasting love to them except, I love you all.

Last but not least, it gives me immense pleasure to express my deepest gratitude to my friends, colleagues and lab mates, especially Al-jubari, Al-maqri, Ejmaa, and Abdulazim for their unlimited support and encouragement.

Finally, I would like to thank everybody who was important to the successful realization of this thesis, as well as I express my apology that I could not mention you all personally.

I certify that a Thesis Examination Committee has met on 28 February 2017 to conduct the final examination of Mohamed .A. Alrshah on his thesis entitled "Adaptive Linux-Based TCP Congestion Control Algorithm for High-Speed Networks" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

**Abdul Azim bin Abd Ghani, PhD**

Professor  
Faculty of Computer Science and Information Technology  
Universiti Putra Malaysia  
(Chairman)

**Shamala a/p K Subramaniam, PhD**

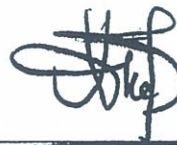
Professor  
Faculty of Computer Science and Information Technology  
Universiti Putra Malaysia  
(Internal Examiner)

**Abdul Hanan Abdullah, PhD**

Professor  
Universiti Teknologi Malaysia  
Malaysia  
(External Examiner)

**Carlo Caini, PhD**

Associate Professor  
University of Bologna  
Italy  
(External Examiner)



---

**NOR AINI AB. SHUKOR, PhD**  
Professor and Deputy Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date: 28 April 2017



This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy.

The members of the Supervisory Committee were as follows:

**Mohamed Othman, PhD**

Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Chairman)

**Borhanuddin Mohd Ali, PhD**

Professor

Faculty of Engineering

Universiti Putra Malaysia

(Member)

**Zurina Mohd Hanapi, PhD**

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Member)

---

**ROBIAH BINTI YUNUS, PhD**

Professor and Dean

School of Graduate Studies

Universiti Putra Malaysia

Date:

## Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

Name and Matric No.: \_\_\_\_\_ Mohamed A. Alrshah (GS35294) \_\_\_\_\_

## Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: \_\_\_\_\_  
Name of  
Chairman of  
Supervisory  
Committee: \_\_\_\_\_

Signature: \_\_\_\_\_  
Name of  
Member of  
Supervisory  
Committee: \_\_\_\_\_

Signature: \_\_\_\_\_  
Name of  
Member of  
Supervisory  
Committee: \_\_\_\_\_



## TABLE OF CONTENTS

	Page
ABSTRACT	i
ABSTRAK	iii
ACKNOWLEDGEMENTS	vii
APPROVAL	viii
DECLARATION	x
LIST OF TABLES	xv
LIST OF FIGURES	xvi
LIST OF ABBREVIATIONS	xviii
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 TCP Overview	1
1.1.1 Slow-Start Mechanism	1
1.1.2 Congestion Avoidance Mechanism	2
1.2 Approaches to Congestion Control	2
1.2.1 Loss-based ( <i>Reactive</i> ) Approach	3
1.2.2 Delay-based ( <i>Proactive</i> ) Approach	3
1.2.3 Loss-delay-based ( <i>Hybrid</i> ) Approach	4
1.3 Short-distance Networks	4
1.4 Long-distance Networks	5
1.5 Motivation	5
1.6 Problem Statement	6
1.7 Research Objectives	7
1.8 Research Scope	7
1.9 Research Significance	8
1.10 Thesis Organization	8
<b>2 LITERATURE REVIEW</b>	<b>9</b>
2.1 Introduction	9
2.2 TCP Variants in High-speed Networks	9
2.2.1 TCP Vegas	11
2.2.2 TCP NewReno	12
2.2.3 TCP Westwood	12
2.2.4 FAST TCP (FAST-TCP)	12
2.2.5 Scalable TCP (S-TCP)	13
2.2.6 High-speed TCP (HS-TCP)	13
2.2.7 Hamilton TCP (H-TCP)	15
2.2.8 TCP-Hybla	15

2.2.9	BIC-TCP	16
2.2.10	TCP-AFRICA	17
2.2.11	TCP-illinois	18
2.2.12	Compound TCP (C-TCP)	18
2.2.13	YeAH-TCP	19
2.2.14	TCP-Fusion	19
2.2.15	Cubic-TCP	20
2.3	Latest Enhancements on TCP	21
2.4	Summary	23
<b>3</b>	<b>RESEARCH METHODOLOGY</b>	<b>25</b>
3.1	Introduction	25
3.2	Notations and Definitions	25
3.3	Research Framework	25
3.3.1	Problem Formulation	26
3.3.2	Previous Algorithms Implementation	26
3.3.3	The Proposed Work	27
3.3.4	Implementation and Comparison for Evaluation	27
3.3.5	Performance Metrics Evaluation	30
3.4	Experiments Environment	30
3.4.1	Software and Hardware Tools	30
3.4.2	Network Topologies	31
3.4.3	Experimental Setups	32
3.5	Performance Metrics	33
3.5.1	Throughput	33
3.5.2	Packet Loss Ratio	34
3.5.3	Sharing Fairness Index	34
3.6	Summary	34
<b>4</b>	<b>AGILITY FACTOR BASED CCA FOR SHORT-DISTANCE NETWORKS</b>	<b>35</b>
4.1	Introduction	35
4.2	AF-based CCA: The Proposed Model	35
4.2.1	The Agility Factor Mechanism	35
4.2.2	The Decrement of $cwnd$	39
4.2.3	AF-based CCA Overall Behavior	40
4.3	Performance Evaluation of AF-based CCA	41
4.3.1	The Experiments Setup	42
4.3.2	Results and Discussion	42
4.4	Summary	54
<b>5</b>	<b>A MATHEMATICAL MODEL FOR STEADY STATE THROUGHPUT OF AF-BASED CCA</b>	<b>55</b>
5.1	Introduction	55
5.2	System Model for AF-Based CCA	55
5.2.1	Congestion Control of AF-based CCA	56

5.2.2	Congestion Loss and Random Packet Loss	59
5.2.3	Markov Chain Formulation	60
5.3	Performance Evaluation	63
5.3.1	Model Validation via Simulation	64
5.3.2	Average Throughput of AF-based CCA	66
5.3.3	The Impact of AF-based CCA on Epoch Time	67
5.3.4	The AACPT Process	68
5.4	Summary	71
<b>6</b>	<b>WINDOW-CORRELATED WEIGHTING FUNCTION BASED CCA FOR LONG-DISTANCE NETWORKS</b>	<b>73</b>
6.1	Introduction	73
6.2	Preliminary Analysis	73
6.3	WWF-based CCA: The Proposed Model	76
6.3.1	Window-correlated Weighting Function (WWF)	76
6.3.2	The WWF-based CCA Overall Behavior	80
6.4	Performance Evaluation of WWF-based CCA	81
6.4.1	Experiments Setup	81
6.4.2	Results and Discussion	81
6.5	Summary	90
<b>7</b>	<b>AN ADAPTIVE AGILE-TCP CCA FOR ANY-DISTANCE NETWORKS</b>	<b>91</b>
7.1	Introduction	91
7.2	Agile-TCP: The Proposed Adaptive CCA	92
7.3	Performance Evaluation of Agile-TCP	92
7.3.1	Simulation Setup	93
7.3.2	Simulation Results and Discussion	94
7.4	Agile-TCP Module in Linux Kernel	99
7.5	Performance Evaluation of Agile-TCP in Linux	100
7.5.1	Test-bed Setup	100
7.5.2	Test-bed Results and Discussion	100
7.6	Summary	102
<b>8</b>	<b>CONCLUSION AND FUTURE WORKS</b>	<b>103</b>
8.1	Conclusion	103
8.2	Future Works	104
	<b>REFERENCES</b>	<b>105</b>
	<b>APPENDICES</b>	<b>113</b>
	<b>BIODATA OF STUDENT</b>	<b>116</b>
	<b>LIST OF PUBLICATIONS</b>	<b>117</b>

## LIST OF TABLES

Table		Page
2.1	The High-speed TCPs Implemented in Real Operating Systems	10
2.2	Comparison of the Main Characteristics of High-speed TCP Variants	23
3.1	Test-bed Parameters Settings	29
3.2	Common Simulation Parameters Settings	33
4.1	Notations Used in AF-based Algorithm	36
4.2	Simulation Parameters Settings for Short-distance	42
5.1	Notations Used for Mathematical Model	57
5.2	Experiment Parameters Setting for Mathematical Model	64
5.3	The Setting of Simulation Parameters for Validation	65
6.1	Notations Used in WWF-based Algorithm	76
6.2	Simulation Parameters Settings for Long-distance	81
7.1	Simulation Parameters Settings for Different Distances	94

## LIST OF FIGURES

Figure	Page
1.1 General Behavior of TCP	1
1.2 The Main Approaches to Congestion Control	3
1.3 Local Area Network	4
1.4 Multi-rooted Hierarchical Data Center Network	5
1.5 Long-distance Network	6
2.1 Classification and Historical Evolution of TCP Congestion Control Algorithms	11
2.2 S-TCP Congestion Window Dynamics	14
2.3 HS-TCP Congestion Window Dynamics	14
2.4 TCP-AFRICA Congestion Window Dynamics	17
2.5 C-TCP Congestion Window Dynamics	19
2.6 TCP-Fusion Congestion Window Dynamics	20
3.1 Research Framework	26
3.2 The Implementation of the Proposed Algorithms in Linux Kernel, NS-2, and Matlab	28
3.3 Dummynet-based Test-bed Topology and Configuration	29
3.4 Congestion-Free Network Topology	31
3.5 The Sequence of Establishments/Terminations of Multiple Flows Scenarios, where ( $n$ ) is the Number of Flows and ( $m$ ) is the Simulation Time	32
3.6 Network Topology with Single Dumbbell Bottleneck	32
4.1 The Concept of Agility Factor ( $\lambda$ ) Mechanism	36
4.2 The $cwnd$ Evolution of AF-based CCA and Standard TCP	37
4.3 Relations Between $cwnd$ , Link Utilization, $\lambda$ , and $\alpha$	38
4.4 The Flow Control Diagram of The AF-based CCA	41
4.5 $cwnd$ Evolution (Buffer Size = 5 Packets)	45
4.6 $cwnd$ Evolution (Buffer Size = 25 Packets)	46
4.7 $cwnd$ Evolution (Buffer Size = 100 Packets)	47
4.8 $cwnd$ Evolution (Buffer Size = 250 Packets)	48
4.9 $cwnd$ Evolution (Buffer Size = 500 Packets)	49
4.10 The First Scenario: Average Throughput vs. Buffer Size	50
4.11 The Second Scenario: Average Throughput vs. Buffer Size	51
4.12 The Third Scenario: Average Throughput vs. Buffer Size	52
4.13 1st Scenario: Loss Ratio vs. Buffer Size ( $10^{-4}$ PER)	53
4.14 2nd Scenario: Intra-Fairness vs. Buffer Size ( $zero$ PER)	53
4.15 The RTT-fairness Among Flows with Different RTTs ( $10^{-3}$ PER)	53
4.16 The Inter-fairness Among the Studied CCAs	53
5.1 Congestion Window Evolution of AF-based CCA	58
5.2 The Evolution of $cwnd$ .	58
5.3 State Transition Diagram for Markov Chain: Example of $N = 5$ , $\beta = 0.75$ , $w_i \in \{2, 3, 4, 5, 6\}$ and $i \in \{1, 2, \dots, N\}$ .	61

5.4	State Transition Matrix for Markov Chain: Example of $N = 5$ , $\beta = 0.75$ , $w_i \in \{2, 3, 4, 5, 6\}$ and $i \in \{1, 2, \dots, N\}$ .	61
5.5	Simulation-based Comparison between AF-based Congestion Control Algorithm (CCA) with Different $\lambda_{max}$ and NewReno, under Different PERs.	64
5.6	AF-based CCA Normalized Average Throughput under Different PERs	66
5.7	Normalized Average Throughput under Different Buffer Sizes and $10^{-8}$ PER, where $\beta = 0.5$ and $\lambda = 5$	66
5.8	Normalized Average Throughput under Different RTTs and $10^{-8}$ PER, where $\beta = 0.5$ , $\lambda_{max} = 5$ and buffer size is only 4 packets.	67
5.9	AF-based CCA vs. NewReno Epoch Time	67
5.10	Impact of $\lambda_{max}$ on the Average Throughput and Epoch Time	68
5.11	The Normalized Average Throughput of AF-based CCA under Different Configurations	70
5.12	The Relation Between $\lambda'$ and $\beta$	70
5.13	The Impact of Using Equation (5.20) on the Average Throughput of AF-based CCA, $\beta = \{0.5 \rightarrow 0.9\}$ , PER= $10^{-8}$ , and buffer=4 packets.	71
6.1	The Impact of RTT on UR	78
6.2	The Impact of RTT on $\Delta$	78
6.3	Comparison among Square-root, Cube-root and Logarithmic Functions	79
6.4	The Epoch Time of WWF-based CCA Compared to NewReno, Cubic-TCP and C-TCP	80
6.5	TCP Congestion Window Evolution over Single-flow Scenario (buffer Size = 6400 packets, packet size = 1kbyte)	83
6.6	TCP Congestion Window Convergence in Multi-flows Scenario (buffer Size = 3200 packets, packet size = 1kbyte)	84
6.7	The First Scenario: Average Throughput vs. Buffer Size	86
6.8	The Second Scenario: Average Throughput vs. Buffer Size	87
6.9	The Third Scenario: Average Throughput vs. Buffer Size	88
6.10	Loss Ratio vs. Buffer Size: the Third Scenario, <i>zero</i> PER	89
6.11	Intra-fairness vs. Buffer Size: the Third Scenario, $10^{-5}$ PER	89
6.12	The RTT-fairness Among Flows with Different RTTs, $10^{-3}$ PER	89
6.13	The Inter-fairness Among The Studied CCAs	89
7.1	Types of Links Established by Computers Connected to Internet	91
7.2	The Scheme of Agile-TCP	92
7.3	TCP Congestion Window Evolution over Sequential Multi-flows Scenario (RTT = 8ms, packet size = 1kbyte)	95
7.4	TCP Congestion Window Evolution over Sequential Multi-flows Scenario (RTT = 256ms, packet size = 1kbyte)	96
7.5	Average Throughput vs. Delay	97
7.6	The Second Scenario: Loss Ratio vs. Delay	98
7.7	The Second Scenario: Intra-fairness vs. Delay	98
7.8	The RTT-fairness Among Flows with Different RTTs	98
7.9	The Inter-fairness Among the Studied CCAs	98
7.10	The Data Structure of Agile-TCP CCA Linux Module	99
7.11	Average Throughput vs. PER	101
7.12	File Transfer Time vs. PER	101
7.13	Loss Ratio vs. PER	101

## LIST OF ABBREVIATIONS

AACPT	Automated Algorithm Configuration and Parameter Tuning
ACK	Acknowledgement
AF	Agility Factor
AIAD	Additive-Increase/Additive-Decrease
AIMD	Additive-Increase/Multiplicative-Decrease
BDP	Bandwidth-Delay-Product
CCA	Congestion Control Algorithm
CORE	Common Open Research Emulator
DCN	Data Center Network
ECN	Explicit Congestion Notification
IP	Internet Protocol
IW	Initial Window
LAN	Local Area Network
MIMD	Multiplicative-Increase/Multiplicative-Decrease
NAD	Network Attached Drive
NetEm	Network Emulator
NEWT	Network Emulator for Windows Toolkit
NS-2	Network Simulator 2
OSI	Open Systems Interconnection
PC	Personal Computer
PER	Packet Error Rate
RTT	Round Trip Time
TCP	Transmission Control Protocol
TOC	TimeOut Counter
WWF	Window-correlated Weighting Function



# CHAPTER 1

## INTRODUCTION

### 1.1 TCP Overview

In the last decade, Transmission Control Protocol (TCP) (Cerf and Dalal, 1974; Cerf and Kahn, 1974; Postel, 1981) is profusely used by most Internet applications such as file transfer, email, World-Wide-Web and remote administration. It becomes one of the two original components of the Internet protocol suite, complementing the Internet Protocol (IP), where the entire suite is known as TCP/IP. One of the major parts of TCP is the CCA, which controls the data transmission rate ( $Tr$ ) between the two ends of any TCP connection. As well-known, CCA slowly increases the transmission rate of data packets to probe the available network capacity while avoiding the congestion. The number of traveling packets “in-flight” over a network link between a sender and receiver is called TCP congestion window ( $cwnd$ ). For better understanding, the following sub-sections explain in brief the main TCP components and Figure 1.1 shows the general behavior of standard TCP.

#### 1.1.1 Slow-Start Mechanism

The main concept behind the slow-start is to progressively probe and estimate the available bandwidth. The estimated bandwidth is used to regulate the  $Tr$ , which is always equal to the amount of data packets ( $cwnd$ ) sent, from source to destination, per Round Trip Time (RTT):

$$Tr = \frac{cwnd}{RTT}, \quad (1.1)$$

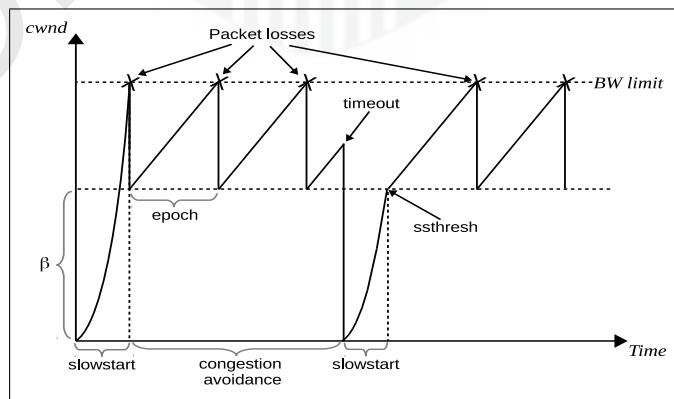


Figure 1.1: General Behavior of TCP



where *cwnd* is initially set to a small Initial Window (IW). At the beginning, IW was set to “one” or “two” segments as in the RFC2581 by Allman et al. (1999). Then, it was modified, in the RFC3390 by Allman et al. (2002), to be a value between “two” and “four” segments or roughly 4Kbytes. Later, Dukkupati et al. (2010), from Google, proposed to increase the IW to “ten” segments or roughly 15Kbytes.

Once TCP starts by *cwnd* equal to IW, it duplicates this *cwnd* every RTT to show an exponential increase. More specifically, it increases the *cwnd* by “one” for each arrival of non-duplicated Acknowledgement (ACK). This stage ends either by an event of packet loss or by reaching the slow-start threshold (*ssthresh*), depends on which event happens first, see Figure 1.1. If a packet loss happens, TCP degrades its *cwnd* by the multiplicative decrease factor ( $\beta$ ), otherwise it ends this stage without degradation. Thereafter, TCP immediately enters another stage, called congestion avoidance, in which it applies the mechanism of linear increase. However, the *cwnd* is reset to the IW to start a new slow-start phase if a TimeOut Counter (TOC) expiration was experienced.

### 1.1.2 Congestion Avoidance Mechanism

The main idea of this mechanism is to avoid the network congestion by gently increasing the value of *cwnd* using the concept of Additive-Increase/Multiplicative-Decrease (AIMD). This mechanism is more conservative than the slow-start, in which the former increases its *cwnd* by  $\frac{1}{cwnd}$  for each arrival of non-duplicated ACK. Consequently, the *cwnd* increases by “one” every RTT, as in the standard TCP Reno and NewReno (Floyd and Henderson, 1999). However, TCP degrades its *cwnd* by the multiplicative decrease factor ( $\beta$ ) if a packet loss is detected during the congestion avoidance stage, then it starts a new epoch using the same mechanism, as shown in Figure 1.1. Moreover, TCP degrades its *cwnd* to the IW to start a new slow-start phase if a TOC expiration was experienced.

## 1.2 Approaches to Congestion Control

TCP provides stable and reliable delivery of data packets without relying on any explicit feedback from underlying network. However, TCP relies only on the two ends of connection, which are the sender and the receiver, for this reason TCP is widely known as end-to-end or host-to-host protocol. In general, there are two main explicit feedbacks used by the CCA sender to regulate its *Tr*, as follows:

1. The signal to packet loss (packet drop), which indicates that a hop in the network path between the sender and the receiver is overloaded (buffer overflow). This signal is detected either by TOC expiration or by receiving three duplicated ACKs.

2. The two-way delay or RTT, which indicates the time taken by a packet to be sent from source to destination plus the time taken for its acknowledgment to be received, where both propagation and queuing delays are inclusive.

Based on the aforementioned explicit feedbacks, there are only three main approaches, as shown in Figure 1.2, to congestion control in the sender-side of TCP. Afanasyev et al. (2010) classified these approaches into three categories, as briefly explained in the following paragraphs:

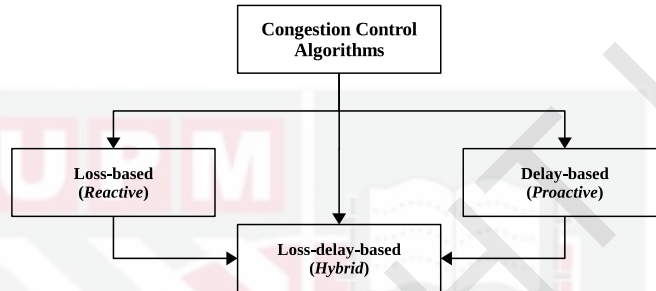


Figure 1.2: The Main Approaches to Congestion Control

### 1.2.1 Loss-based (*Reactive*) Approach

This approach relies only on the signal of packet loss to regulate the  $Tr$ . In other words, the CCA keeps increasing its  $cwnd$  as long as it does not detect a packet loss, otherwise, it degrades the  $cwnd$  using the multiplicative decrease factor  $\beta$ . This approach is suitable for short-distance networks, in which the delay is not functional due to the trivial variation in its value. The main disadvantage of this approach is the unnecessary degradations that happen when the detected losses are not resulted by congestion. This case is very common in wireless, mobile, satellite and long-distance networks, where the packet losses are caused by deferent reasons such as collision, fading and/or interference.

### 1.2.2 Delay-based (*Proactive*) Approach

In this approach, the increment of  $cwnd$  is a function of delay, so that TCP regulates its  $Tr$  only based on the fluctuation of delay. TCP keeps increasing the  $Tr$  as long as the delay is low and relatively decreases it when the delay increases. This approach performs well only when the underlying network has a major delay fluctuation such as that in wireless, mobile, satellite, and long-distance networks. However, this approach becomes unstable if the route-change is common. Moreover, when the fluctuation of delay is trivial, this approach insufficiently utilizes the bandwidth.

### 1.2.3 Loss-delay-based (Hybrid) Approach

This approach combines both loss and delay-based approaches in order to gain high scalability, robustness and efficiency (Katto et al., 2008b,a). Most hybrid CCAs work as a multi-modes switching algorithm, in which they have loss-based mode and delay-based mode. Such CCAs activate the delay-based mode to exploit residual capacity of the bandwidth as long as no packet loss is detected, and they activate the loss-based mode otherwise. Besides, they rely on the observed RTT to switch from loss-based mode back to delay-based mode. Despite that hybrid CCAs perform in a smarter way than legacy TCP methods, they show limited performance in some cases.

### 1.3 Short-distance Networks

In the last decades, the necessity of high-speed and short-distance networks is rapidly increasing due to their wide deployment. Several network applications, such as Local Area Networks (LANs) and Data Center Networks (DCNs), implement this type of networks (Buyya et al., 2008; Armbrust et al., 2010). These LANs and DCNs serve a very wide range of network-based applications such as web hosting, searching engines, social media, multimedia broadcasting, and storage drives. In the environment of LANs and DCNs, as shown in figures 1.3 and 1.4, respectively (Al-Fares et al., 2010; Wu and Yang, 2012; Yoo et al., 2012; Prakash et al., 2012), high-speed and short-distance networks are commonly deployed to connect computing and storage elements to each other in order to provide rapid services.

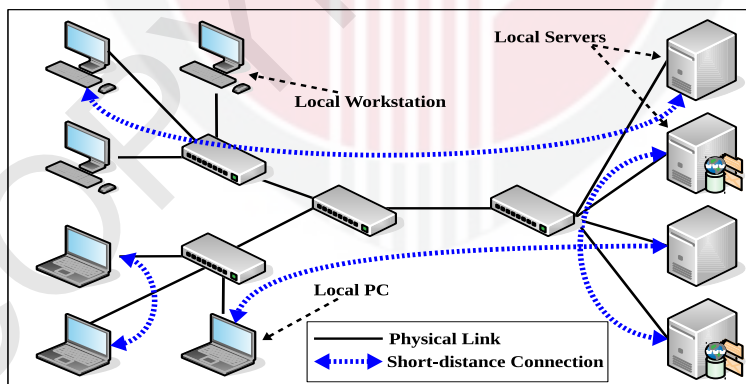


Figure 1.3: Local Area Network

Indeed, this type of networks has its own unique characteristics, in which the link delay and link Bandwidth-Delay-Product (BDP) are very small. In such networks, the link delay can be few milliseconds or even hundreds of microseconds and the used buffer can accommodate only few packets, which result in very negligible delay variation compared to its equivalent in long-distance networks (Tahiliani et al., 2012; Vasudevan et al., 2009). In order to improve TCP perfor-

mance over short-distance networks, the ability of TCP to deal with small buffers and negligible delay variation needs to be extended, by reducing the sensitivity to packet loss.

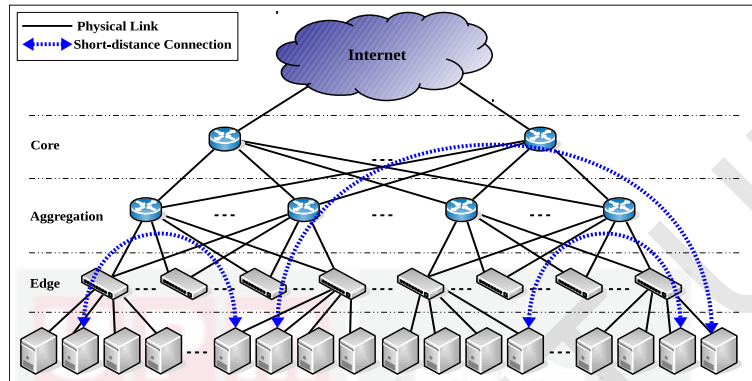


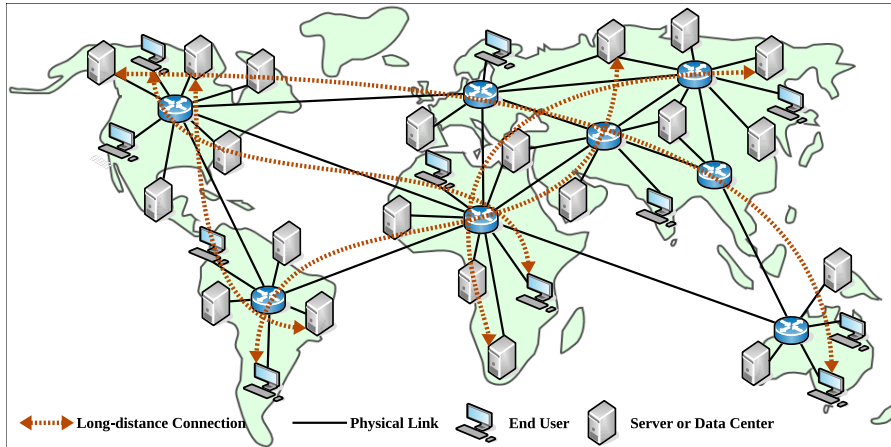
Figure 1.4: Multi-rooted Hierarchical Data Center Network

#### 1.4 Long-distance Networks

Recently, the demand for Internet applications has been increased, which increases the number of data centers across the world. In order to provide a high level of connectivity among these data centers, high-speed and long-distance networks are widely used across many countries and continents, as shown in Figure 1.5. In fact, the characteristics of this type of networks are very unique, where the link delay and link BDP are very large. This networks produce very high level of delay variation caused by the use of big buffers and by the large propagation delay, which can be tens or hundreds of seconds (Afanasyev et al., 2010; Scharf, 2011; Xu et al., 2011; Callegari et al., 2012, 2014). In order to improve TCP performance over such networks, the dependency on RTT needs to be minimized and the value of  $cwnd$  increase needs to be correlated to the link BDP and/or the magnitude of  $cwnd$ .

#### 1.5 Motivation

Despite the wide deployment of TCP, its CCAs still have some critical issues. The stat-of-the-art shows that the existing CCAs are still insufficient, especially for high-speed wired networks, which opens a space for researchers to keep improving these CCAs. Moreover, the fast advances of networks technologies require more improvements for TCP to fulfill the requirements of the latest applications such as cloud computing and big data transfers.



**Figure 1.5: Long-distance Network**

## 1.6 Problem Statement

Although TCP is the predominant transmission protocol used in Internet, its CCA is still unable to adapt to high-speed networks, which are not the typical environment for which TCP CCAs are designed. Thus, the employment of TCP over high-speed networks causes an extreme performance degradation and a very poor bandwidth utilization.

In short-distance networks where the delay variation is negligible, the only practical congestion control approach is the loss-based. This approach relies only on packet losses to detect network congestion, which increases its sensitivity to these losses. The loss-based approach becomes unable to achieve an acceptable level of bandwidth utilization due to the common use of small buffers in such networks, which enforces packets to be frequently dropped.

In order to identify and clarify the relation between TCP and its factors that play an important role of enhancement, it is advantageous to investigate and grasp the behavior of AF-based CCA. The most significant outcome from this operation is to verify and validate the performance of the AF-based CCA and then to optimize the setting of its parameters. Thus, a new mathematical model for AF-based CCA over high-speed networks is highly needed.

In long-distance networks, the environment causes two major unavoidable problems that negatively affect the general performance of TCP. First problem is the long RTT caused by long-distance and long buffering time. Second problem is that when the BDP of a network is high, it requires TCP to expand its *cwnd* to a large number of packets in order to utilize the bandwidth. In the congestion avoidance stage, TCP requires around an RTT to increase its *cwnd* by one and because the RTT in such networks is very long, TCP *cwnd* increase becomes severely slow due to the RTT-dependency of TCP.



Naturally, TCP-based applications require to deal adaptively with any-distance networks. The problem is that TCP is not properly adapted to high variability of delay and buffer size, which often makes TCP either unnecessarily conservative or severely aggressive. Thus, it is worthy to design an adaptive CCA, which combines both loss-based and delay-based approaches, to automatically adapt itself to deal with networks of any distance without the need of manual reconfiguration.

## 1.7 Research Objectives

The main goal of this thesis is to enhance the performance of TCP over high-speed wired networks. In particular, this goal is divided into four important objectives, as follows:

- To propose a novel loss-based CCA, namely AF-based, which reduces the sensitivity to packet losses in order to improve the total performance of TCP over high-speed and short-distance networks.
- To propose a new mathematical model for AF-based TCP in order to investigate and grasp its behavior to identify and clarify the relation between TCP and its factors that play an important role of enhancement. Based on the outcome of this model, an Automated Algorithm Configuration and Parameter Tuning (AACPT) process is used to optimize and automate the setting of AF-based CCA parameters.
- To propose a novel delay-based CCA, namely WWF-based, which correlates the transmission rate to the magnitude of *cwnd* in order to enhance the total performance of TCP over high-speed and long-distance networks.
- To propose a new adaptive CCA, namely Agile-TCP, which combines both AF-based and WWF-based CCAs, in order to be able to automatically acclimate to any-distance networks. In addition, this work prepares the proposed Agile-TCP CCA as a Linux kernel module to get ready for deployment in the real Linux operating system.

## 1.8 Research Scope

This thesis concentrates on studying TCP over high-speed wired networks. In addition, it focuses only on enhancing the performance of CCA at transport layer of end-to-end systems. This improvements are to meet the huge demand of applications, which require very high-speed data transfer, in total isolation from underlying networks, as recommended by the Open Systems Interconnection (OSI) model. In order to facilitate the work and to avoid any extra costs, all algorithms in this thesis are implemented in Linux kernel and tested using the well-known Network Simulator 2 (NS-2) (McCanne and Floyd, 1998), which is a free

and open source simulator. Besides, all experiments are simulated over high-speed wired networks, thus, the implementation of the proposed algorithms over other network technologies lies behind the scope of this thesis. Moreover, the final product of this thesis is prepared as a Linux kernel module to get ready for deployment in the real Linux operating system, while other operating systems are not targeted.

## **1.9 Research Significance**

The significance of this work arises from the need for an efficient TCP, which is able to automatically adapt to high-speed networks based on their characteristics. The challenges are to reduce the sensitivity to packet losses, and to grant the ability to deal with long delays and big buffers in order to boost the throughput, which emphasize the importance of conducting this research. In addition, the proposed algorithms are not only promising for regular file transfer, they are also promising for other applications such as big data transferring and cloud computing.

## **1.10 Thesis Organization**

The rest of this thesis is organized as follows: Chapter 2 presents the literature review and discusses the stat-of-the-art. Chapter 3 generally describes the research methodology used in this thesis including the research framework, experimental setup, network topologies, proposed methods, performance metrics and the evaluation method. Chapter 4 explains the proposed AF-based CCA, which is designed for short-distance networks. Chapter 5 exhibits a mathematical model for the proposed AF-based CCA. Chapter 6 shows the proposed WWF-based CCA, which is designed for long-distance networks. Chapter 7 presents the adaptive Agile-TCP CCA, which is designed for any-distance networks. Finally, Chapter 8 concludes the thesis and shows the future work.

## REFERENCES

- Afanasyev, A., Tilley, N., Reiher, P., and Kleinrock, L. (2010). Host-to-Host Congestion Control for TCP. *IEEE Communications Surveys and Tutorials*, 12(3):304–342.
- Aho, A. V., Kernighan, B. W., and Weinberger, P. J. (1979). Awk-a pattern scanning and processing language. *Softw., Pract. Exper.*, 9(4):267–279.
- Al-Fares, M., Radhakrishnan, S., Raghavan, B., Huang, N., and Vahdat, A. (2010). Hedera: Dynamic flow scheduling for data center networks. In *NSDI*, volume 10, pp. 19–19.
- Alisa, Z. and Qasim, S. (2014). A Fuzzy based TCP Congestion Control for Wired Networks. *International Journal of Computer Applications*, 89(4):36–42. doi:10.5120/15494-4401.
- Alizadeh, M., Greenberg, A., Maltz, D. A., Padhye, J., Patel, P., Prabhakar, B., Sengupta, S., and Sridharan, M. (2010). Data center tcp (dctcp). In *Proceedings of the ACM SIGCOMM 2010 Conference, SIGCOMM '10*, pp. 63–74, New York, NY, USA. ACM. doi:10.1145/1851182.1851192.
- Allman, M., Floyd, S., and Partridge, C. (2002). Increasing tcp's initial window. RFC 2581, IETF Network Working Group.
- Allman, M., Paxson, V., and Stevens, W. R. (1999). Tcp congestion control. RFC 2581, IETF Network Working Group.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). A view of cloud computing. *Commun. ACM*, 53(4):50–58. doi:10.1145/1721654.1721672.
- Baiocchi, A., Castellani, A. P., and Vacirca, F. (2007). YeAH-TCP : Yet Another Highspeed TCP. In *Proc. PFLDnet.*, pp. 37–42, Roma, Italy.
- Bao, W., Wong, V. W., and Leung, V. C. (2010). A model for steady state throughput of tcp cubic. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pp. 1–6. IEEE.
- Bauer, S., Beverly, R., and Berger, A. (2011). Measuring the state of ecn readiness in servers, clients, and routers. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pp. 171–180. ACM.
- Beheshti, N., Ganjali, Y., Rajaduray, R., Blumenthal, D., and McKeown, N. (2006). Buffer sizing in all-optical packet switches. In *Optical Fiber Communication Conference*, pp. 1–3. Optical Society of America.
- Belhaj, S. and Tagina, M. (2008). VFAST TCP: An improvement of FAST TCP. In *Tenth International Conference on Computer Modeling and Simulation (uksim 2008)*, pp. 88–93. IEEE.



- Brakmo, L. S. and Peterson, L. L. (1995). Tcp vegas: End to end congestion avoidance on a global internet. *IEEE Journal on Selected Areas in Communications*, 13(8):1465–1480.
- Buyya, R., Yeo, C. S., and Venugopal, S. (2008). Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In *High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference on*, pp. 5–13. doi:10.1109/HPCC.2008.172.
- Caini, C. and Firrincieli, R. (2004). TCP Hybla: a TCP enhancement for heterogeneous networks. *International Journal of Satellite Communications and Networking*, 22(5):547–566. doi:10.1002/sat.799.
- Caini, C., Firrincieli, R., and Lacamera, D. (2009). Comparative performance evaluation of tcp variants on satellite environments. In *Communications, 2009. ICC'09. IEEE International Conference on*, pp. 1–5. IEEE.
- Callegari, C., Giordano, S., Pagano, M., and Pepe, T. (2012). Behavior analysis of TCP Linux variants. *Computer Networks*, 56(1):462–476. doi:10.1016/j.comnet.2011.10.002.
- Callegari, C., Giordano, S., Pagano, M., and Pepe, T. (2014). A survey of congestion control mechanisms in linux tcp. In Vishnevsky, V., Kozyrev, D., and Lariionov, A., editors, *Distributed Computer and Communication Networks*, volume 279 of *Communications in Computer and Information Science*, pp. 28–42. Springer International Publishing.
- Cerf, V. and Dalal, Y. (1974). Specification of internet transmission control program. RFC 675, IETF Network Working Group.
- Cerf, V. and Kahn, R. (1974). A protocol for packet network intercommunication. *IEEE Transactions on Communications*, 22(5):637–648. doi:10.1109/TCOM.1974.1092259.
- Chu, J., Cheng, Y., Dukkupati, N., and Mathis, M. (2013). Increasing tcp's initial window. RFC 6928, IETF Network Working Group.
- D. Leith, R. S. (2004). H-tcp: Tcp for high-speed and long distance networks. In *Proceedings of PFLDnet*, pp. 95–101.
- Dukkupati, N., Refice, T., Cheng, Y., Chu, J., Herbert, T., Agarwal, A., Jain, A., and Sutin, N. (2010). An argument for increasing tcp's initial congestion window. *Computer Communication Review*, 40(3):26–33.
- Enachescu, M., Ganjali, Y., Goel, A., McKeown, N., and Roughgarden, T. (2006). Routers with very small buffers. In *Proc. IEEE Infocom*, volume 6, pp. 1–11.
- Floyd, S. (2003). HighSpeed TCP for Large Congestion Windows. RFC 3649, IETF Network Working Group.
- Floyd, S. (2008). Metrics for the evaluation of congestion control mechanisms. RFC 5166, IETF Network Working Group.

- Floyd, S. and Allman, M. (2008). Comments on the usefulness of simple best-effort traffic. RFC 5290, IETF Network Working Group.
- Floyd, S. and Henderson, T. (1999). The newreno modification to tcp's fast recovery algorithm. RFC 2582, IETF Network Working Group.
- Floyd, S., Henderson, T., and Gurtov, A. (2004). The newreno modification to tcp's fast recovery algorithm. RFC 3782, IETF Network Working Group.
- Ha, S. and Rhee, I. (2011). Taming the elephants: New tcp slow start. *Computer Networks*, 55(9):2092–2110.
- Ha, S., Rhee, I., and Xu, L. (2008). CUBIC: A New TCP-Friendly High-Speed TCP Variant. *ACM SIGOPS Operating Systems Review*, 42(5):64–74. doi:10.1145/1400097.1400105.
- Hamadi, Y. (2013). Autonomous search. In *Combinatorial Search: From Algorithms to Systems*, pp. 99–122. Springer Berlin Heidelberg.
- Hasegawa, G., Kurata, K., and Murata, M. (2000). Analysis and improvement of fairness between TCP Reno and Vegas for deployment of TCP Vegas to the Internet. In *Proceedings 2000 International Conference on Network Protocols*, pp. 177–186. IEEE Comput. Soc.
- Hassayoun, S., Maillé, P., and Ros, D. (2010). On the impact of random losses on tcp performance in coded wireless mesh networks. In *INFOCOM, 2010 Proceedings IEEE*, pp. 1–9. IEEE.
- Henderson, T., Floyd, S., Gurtov, A., and Nishida, Y. (2012). The NewReno modification to TCP's fast recovery algorithm. RFC 6582, IETF Network Working Group.
- Henderson, T. R., Sahouria, E., McCanne, S., and Katz, R. H. (1998). On improving the fairness of tcp congestion avoidance. In *Global Telecommunications Conference, 1998. GLOBECOM 1998. The Bridge to Global Integration. IEEE*, volume 1, pp. 539–544. IEEE.
- Huh, I., Lee, J. Y., and Kim, B.-C. (2006). Decision of maximum congestion window size for tcp performance improvement by bandwidth and rtt measurement in wireless multi-hop networks. *International Journal of Information Processing Systems*, 2(1):34–38.
- Hwang, J., Yoo, J., and Choi, N. (2012). Ia-tcp: A rate based incast-avoidance algorithm for tcp in data center networks. In *Communications (ICC), 2012 IEEE International Conference on*, pp. 1292–1296. doi:10.1109/ICC.2012.6364079.
- Jacobson, V. (1990). Modified tcp congestion avoidance algorithm. *end2end-interest mailing list*. <ftp://ftp.isi.edu/end2end/end2end-interest-1990.mail>, accessed July 2015.
- Jacobson, V., Leres, C., and McCanne, S. (1987). Tcpcdump/libpcap. <http://www.tcpcdump.org> (accessed June 2015).

- Jain, R. (1989). A delay-based approach for congestion avoidance in interconnected heterogeneous computer networks. *ACM SIGCOMM Computer Communication Review*, 19(5):56–71. doi:10.1145/74681.74686.
- Jain, R., Chiu, D.-M., and Hawe, W. R. (1984). *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*. Eastern Research Laboratory, Digital Equipment Corporation.
- Jamali, S., Alipasandi, N., and Alipasandi, B. (2015). TCP Pegas: A PSO-based improvement over TCP Vegas. *Applied Soft Computing Journal*, 32:164–174. doi:10.1016/j.asoc.2015.03.048.
- Jansang, A. and Phonphoem, A. (2013). A simple analytical model for expected frame waiting time evaluation in ieee 802.11e HCCA mode. *Wireless Personal Communications*, 69(4):1899–1924.
- Jansang, A., Phonphoem, A., and Paillassa, B. (2009). Analytical model for expected packet delay evaluation in ieee 802.11 e. In *Communications and Mobile Computing, 2009. CMC'09. WRI International Conference on*, volume 2, pp. 344–348. IEEE.
- Jin, C., Wei, D., Low, S. H., Buhrmaster, G., Bunn, J., Choe, D. H., Cottrell, R., Doyle, J. C., Feng, W., Martin, O., et al. (2003). Fast tcp: From theory to experiments. *IEEE network*, 19(1):4–11.
- Jingyuan Wang, Jiangtao Wen, Yuxing Han, Jun Zhang, Chao Li, and Zhang Xiong (2013). CUBIC-FIT: A High Performance and TCP CUBIC Friendly Congestion Control Algorithm. *IEEE Communications Letters*, 17(8):1664–1667. doi:10.1109/LCOMM.2013.060513.130664.
- Joel Sing and Ben Soh (2005). TCP New Vegas: Improving the Performance of TCP Vegas Over High Latency Links. In *Fourth IEEE International Symposium on Network Computing and Applications*, volume 2005, pp. 73–82. IEEE.
- Kaneko, K., Fujikawa, T., Su, Z., and Katto, J. (2007). TCP-Fusion : A Hybrid Congestion Control Algorithm for High-speed Networks. In *Proc. PFLDnet, ISI, Marina Del Rey (Los Angeles), California.*, pp. 31–36.
- Katto, J., Ogura, K., Akae, Y., Fujikawa, T., Kaneko, K., and Zhou, S. (2008a). Simple Model Analysis and Performance Tuning of Hybrid TCP Congestion Control. In *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, pp. 1–6. Ieee. doi:10.1109/GLOCOM.2008.ECP.272.
- Katto, J., Ogura, K., Fujikawa, T., Kaneko, K., and Su, Z. (2008b). On Hybrid TCP Congestion Control. In *Proc. ICCCS, TCP*, pp. 1–6.
- Kelly, T. (2003). Scalable TCP : Improving Performance in Highspeed Wide Area Networks. *ACM SIGCOMM Computer Communications Review*, 33(2):83–91.
- Khalil, E. A. (2012). A modified congestion control algorithm for evaluating high bdp networks. *International Journal of Computer Science and Network Security*, 12(11):84–93.

- King, R., Baraniuk, R., and Riedi, R. (2005). TCP-Africa: An adaptive and fair rapid increase rule for scalable TCP. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE.*, pp. 1–11.
- Kittiperachol, S., Sun, Z., and Cruickshank, H. (2009). Integration of Linux TCP and Simulation: Verification, Validation and Application. *Journal of Networks*, 4(9):819–836. doi:10.4304/jnw.4.9.819-836.
- Lar, S.-u. and Liao, X. (2013). An initiative for a classified bibliography on tcp/ip congestion control. *Journal of Network and Computer Applications*, 36(1):126–133.
- LeGrange, J. D., Simsarian, J. E., Bernasconi, P., Neilson, D. T., Buhl, L., and Gripp, J. (2009). Demonstration of an integrated buffer for an all-optical packet router. In *Optical Fiber Communication-includes post deadline papers, 2009. OFC 2009. Conference on*, pp. 1–3. IEEE.
- Liu, S., Başar, T., and Srikant, R. (2008). Tcp-illinois: A loss-and delay-based congestion control algorithm for high-speed networks. *Performance Evaluation*, 65(6):417–440.
- Mascolo, S., Casetti, C., Gerla, M., Sanadidi, M. Y., and Wang, R. (2001). Tcp westwood: Bandwidth estimation for enhanced transport over wireless links. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pp. 287–297. ACM.
- Mathis, M., Semke, J., Mahdavi, J., and Ott, T. (1997). The macroscopic behavior of the tcp congestion avoidance algorithm. *ACM SIGCOMM Computer Communication Review*, 27(3):67–82.
- McCanne, S. and Floyd, S. (1998). Ns network simulator - version 2. <http://www.isi.edu/nsnam/ns> (accessed June 2016).
- Misra, A. and Ott, T. J. (1999). The window distribution of idealized tcp congestion avoidance with variable packet loss. In *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pp. 1564–1572. IEEE.
- MyCurveFit (2015). Free online curve fitting. <http://www.mycurvefit.com> (accessed June 2016).
- NTT.co.jp (2010). World record 69-terabit capacity for optical transmission over a single optical fiber. <http://www.ntt.co.jp/news2010/1003e/100325a.html> (accessed June 2016).
- Oda, H., Hisamatsu, H., and Noborio, H. (2012). Congestion control scheme of Compound TCP+ in wireless LANs. In *Proceedings of the Asian Internet Engineering Conference on - AINTEC '12*, pp. 47–54, New York, New York, USA. ACM Press. doi:10.1145/2402599.2402606.
- Oda, H., Hisamatsu, H., and Noborio, H. (2013). Compound TCP+: A Solution for Compound TCP Unfairness in Wireless LAN. *Journal of Information Processing*, 21(1):122–130. doi:10.2197/ipsjip.21.122.



- Oda, H. and Hisamatu, H. (2010). Compound TCP+ for fairness improvement among Compound TCP connections in a wireless LAN. In *2010 IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR 2010)*, pp. 1–6. IEEE. doi:10.1109/CQR.2010.5619920.
- Optics.org (2013). Nec and corning achieve petabit optical transmission. <http://optics.org/news/4/1/29> (accessed June 2016).
- Padhye, J., Firoiu, V., Towsley, D., and Kurose, J. (1998). Modeling tcp throughput: A simple model and its empirical validation. *ACM SIGCOMM Computer Communication Review*, 28(4):303–314.
- Padhye, J., Firoiu, V., Towsley, D. F., and Kurose, J. F. (2000). Modeling tcp reno performance: a simple model and its empirical validation. *IEEE/ACM Transactions on Networking (ToN)*, 8(2):133–145.
- Postel, J. (1981). Transmission control protocol. RFC 793, IETF Network Working Group.
- Prakash, P., Dixit, A., Hu, Y. C., and Kompella, R. (2012). The tcp outcast problem: Exposing unfairness in data center networks. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, San Jose, CA*, pp. 30–30.
- Prasad, R. S., Dovrolis, C., and Thottan, M. (2007). Router buffer sizing revisited: the role of the output/input capacity ratio. In *Proceedings of the 2007 ACM CoNEXT conference*, pp. 1–15. ACM.
- Rizzo, L. (1997). Dummynet: A Simple Approach to the Evaluation of Network Protocols. *ACM SIGCOMM Computer Communications Review*, 27(1):31–41.
- Rizzo, L. and Lettieri, G. (2016). Tlem, very high speed link emulation. In *AsiaBSDCon, Tokyo, March 2016*, pp. 1–10.
- Scharf, M. (2011). Comparison of end-to-end and network-supported fast startup congestion control schemes. *Computer Networks*, 55(8):1921–1940. doi:10.1016/j.comnet.2011.02.002.
- Shorten, R., King, C., Wirth, F., and Leith, D. (2007). Modelling TCP congestion control dynamics in drop-tail environments. *Automatica*, 43(3):441–449.
- Sing, J. and Soh, B. (2005). Tcp new vegas: improving the performance of tcp vegas over high latency links. In *Network Computing and Applications, Fourth IEEE International Symposium on*, pp. 73–82. IEEE.
- Sivaraman, V., Elgindy, H., Moreland, D., and Ostry, D. (2009). Packet pacing in small buffer optical packet switched networks. *IEEE/ACM Transactions on Networking*, 17(4):1066–1079.
- Srijith, K., Jacob, L., and Ananda, A. L. (2005). Tcp vegas-a: Improving the performance of tcp vegas. *Computer Communications*, 28(4):429–440.
- Stewart, R. R., Tüxen, M., and Neville-Neil, G. V. (2011). An investigation into data center congestion control with ecn. In *2011 Technical BSD Conference (BSDCan 2011), Ottawa, Canada, May 2011*.

- Tahiliani, R. P., Tahiliani, M. P., and Sekaran, K. C. (2012). Tcp variants for data center networks: A comparative study. In *2012 International Symposium on Cloud and Services Computing (ISCOS)*, pp. 57–62. IEEE. doi:10.1109/ISCOS.2012.38.
- Tan, K. and Song, J. (2006). Compound tcp: A scalable and tcp-friendly congestion control for high-speed networks. In *4th International workshop on Protocols for Fast Long-Distance Networks (PFLDNet), 2006*, pp. 80–83.
- Vamanan, B., Hasan, J., and Vijaykumar, T. (2012). Deadline-aware datacenter tcp (d2tcp). In *Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, SIGCOMM '12*, pp. 115–126, New York, NY, USA. ACM. doi:10.1145/2342356.2342388.
- Vasudevan, V., Phanishayee, A., Shah, H., Krevat, E., Andersen, D. G., Ganger, G. R., Gibson, G. A., and Mueller, B. (2009). Safe and effective fine-grained tcp retransmissions for datacenter communication. *SIGCOMM Comput. Commun. Rev.*, 39(4):303–314. doi:10.1145/1594977.1592604.
- Vishwanath, A. and Sivaraman, V. (2008). Routers with very small buffers: Anomalous loss performance for mixed real-time and tcp traffic. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on*, pp. 80–89. IEEE.
- Vishwanath, A. and Sivaraman, V. (2009). Sharing small optical buffers between real-time and tcp traffic. *Optical Switching and Networking*, 6(4):289–296.
- Vishwanath, A., Sivaraman, V., and Rouskas, G. N. (2009a). Considerations for sizing buffers in optical packet switched networks. In *INFOCOM 2009, IEEE*, pp. 1323–1331.
- Vishwanath, A., Sivaraman, V., and Rouskas, G. N. (2011). Anomalous loss performance for mixed real-time and tcp traffic in routers with very small buffers. *IEEE/ACM Transactions on Networking*, 19(4):933–946.
- Vishwanath, A., Sivaraman, V., and Thottan, M. (2009b). Perspectives on router buffer sizing: recent results and open problems. *ACM SIGCOMM Computer Communication Review*, 39(2):34–39.
- Wang, J., Wen, J., Li, C., Xiong, Z., and Han, Y. (2015). DC-Vegas: A delay-based TCP congestion control algorithm for datacenter applications. *Journal of Network and Computer Applications*, 53:103–114. doi:10.1016/j.jnca.2015.03.010.
- Wang, J., Wen, J., Zhang, J., and Han, Y. (2011). TCP-FIT: An improved TCP congestion control algorithm and its performance. In *2011 Proceedings IEEE INFOCOM*, pp. 2894–2902. IEEE. doi:10.1109/INFOCOM.2011.5935128.
- Wang, J., Wen, J., Zhang, J., Xiong, Z., and Huan, Y. (2016). TCP-FIT: An Improved TCP Algorithm for Heterogeneous Networks. *Journal of Network and Computer Applications*, pp. 1—23. doi:10.1016/j.jnca.2016.03.020.

- Wang, R., Valla, M., Sanadidi, S. Y., Ng, B. K. F., and Gerla, M. (2002). Efficiency/friendliness tradeoffs in tcp westwood. In *Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on*, pp. 304–311. IEEE.
- Wang, Z. and Crowcroft, J. (1992). Eliminating periodic packet losses in the 4.3-tahoe bsd tcp congestion control algorithm. *ACM SIGCOMM Computer Communication Review*, 22(2):9–16.
- Wei, D. X., Jin, C., Low, S. H., and Hegde, S. (2006). Fast tcp: Motivation, architecture, algorithms, performance. *IEEE/ACM Trans. Netw.*, 14(6):1246–1259.
- Wu, H., Feng, Z., Guo, C., and Zhang, Y. (2013). Ictcp: Incast congestion control for tcp in data-center networks. *IEEE/ACM Transactions on Networking*, 21(2):345–358. doi:10.1109/TNET.2012.2197411.
- Wu, X. and Yang, X. (2012). Dard: Distributed adaptive routing for datacenter networks. In *Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on*, pp. 32–41. doi:10.1109/ICDCS.2012.69.
- Xu, L., Harfoush, K., and Rhee, I. (2004). Binary increase congestion control (bic) for fast long-distance networks. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 4, pp. 2514–2524. Ieee. doi:10.1109/INFCOM.2004.1354672.
- Xu, W., Zhou, Z., Pham, D., Ji, C., Yang, M., and Liu, Q. (2011). Hybrid congestion control for high-speed networks. *Journal of Network and Computer Applications*, 34(4):1416–1428.
- Yoo, S. J. B., Yin, Y., and Wen, K. (2012). Intra and inter datacenter networking: The role of optical packet switching and flexible bandwidth optical networking. In *Optical Network Design and Modeling (ONDM), 2012 16th International Conference on*, pp. 1–6. doi:10.1109/ONDM.2012.6210261.
- Zhou, W., Xing, W., Wang, Y., and Zhang, J. (2012). Tcp vegas-v: improving the performance of tcp vegas. In *International Conference on Automatic Control and Artificial Intelligence (ACAI 2012)*, pp. 2034–2039. Institution of Engineering and Technology.