

**A Scalable Framework for IP-Network Resource Provisioning Through  
Aggregation and Hierarchical Control**

by

Chen-Nee Chuah

B.S. (Rutgers, The State University of New Jersey) 1995  
M.S. (University of California at Berkeley) 1997

A dissertation submitted in partial satisfaction of the  
requirements for the degree of  
Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Randy H. Katz, Chair  
Professor Jean Walrand  
Professor J. George Shanthikumar

Fall 2001

The dissertation of Chen-Nee Chuah is approved:

---

Chair

Date

---

Date

---

Date

University of California at Berkeley

Fall 2001

**A Scalable Framework for IP-Network Resource Provisioning Through  
Aggregation and Hierarchical Control**

Copyright Fall 2001

by

Chen-Nee Chuah

## Abstract

A Scalable Framework for IP-Network Resource Provisioning Through Aggregation  
and Hierarchical Control

by

Chen-Nee Chuah

Doctor of Philosophy in Engineering – Electrical Engineering and Computer  
Sciences

University of California at Berkeley

Professor Randy H. Katz, Chair

There has been an increasing need to make the Internet architecture capable of meeting the diverse service requirements of newly emerging applications such as multimedia conferencing and e-Commerce. This thesis addresses the following question: *Is it possible to deliver latency-sensitive applications (LSAs), e.g., streaming audio and video, with satisfactory quality of service (QoS) in large-scale network without compromising scalability and bandwidth efficiency?*

We have proposed an innovative distributed control architecture, the *Clearing House (CH)*, and a set of adaptive mechanisms to address these issues. Our design rationale is influenced by discussions with two major U. S. Internet service providers and driven by a realistic model of application-level performance requirements. Towards this end, we focus on Voice-over-IP (VoIP) as an example workload and perform subjective experiments to quantify the impact of packet losses and delays on perceived voice quality. Our results show that packet loss rate should be below 1% and per-hop delay should at most be 5 ms to

guarantee high-quality VoIP delivery.

Two key ideas that contribute to the scalability of our CH architecture are: *aggregation* and *hierarchical control*. Our approach exploits the inherent hierarchy of the Internet structure and peering relationships between ISPs. In our model, various basic routing domains are aggregated to form logical domains (LDs), which can then be aggregated to form larger LDs and so forth. This introduces a hierarchical tree of the LDs, and a CH-node is associated with each LD. The processing load and state maintenance required to manage an entire ISP domain are now distributed to various CH-nodes at different levels of granularity.

The CH-nodes establish bandwidth reservations on intra- and inter-domain links for aggregate traffic (trunk), rather than individual flows, so that no per-flow management is required at any routers. We approximate the arrival process of trunks as Gaussian, and measure their corresponding mean,  $\mu$ , and variance,  $\sigma^2$ , during a chosen measurement window. Aggregate reservations are set up based on the measured  $\mu$ ,  $\sigma$ , and the QoS performance goal (e.g., tolerable loss rate). Using VoIP as an example workload, our simulations show that this technique can achieve a loss rate of 0.12% with only 8% over-provisioning.

In addition to resource reservations, two other essential resource control tasks within CH are admission control and traffic policing. Admission control is necessary for limiting the usage of resources by competing flows, while policing is useful for detecting and penalizing malicious flows (i.e., flows that violate their allocated share of bandwidth). For scalability, per-flow admission control is only performed at ingress points of an ISP domain, but it should consider the network-wide congestion level in estimating the impact of admitting new flows. Our scheme, Traffic-Matrix based Admission Control (TMAC), addresses this problem by leveraging the knowledge of the traffic distributions within an ISP and the link capacity constraints to compute the admission thresholds at ingress routers. Our simulation results show that TMAC can achieve 97% utilization level with less than 1% packet loss rate, which is sufficient for most voice applications.

We have also designed a scalable mechanism, called MDAP (Malicious Flow Detection via Aggregate Policing,) for detecting and policing malicious flows without keeping per-flow state at any edge routers. MDAP aggregates admitted flows for group policing without compromising the ability to identify individual malicious flows when necessary. The key insight behind MDAP is a coordinated way of assigning a unique flow-identifier *Fid* to every flow based on its ingress and egress point. As a result, the amount of state maintained by edge routers can be reduced from  $O(n)$  to  $O(\sqrt{n})$ , where  $n$  is the number of admitted flows. We study the performance and robustness of MDAP through ns simulations. Our results show that we can successfully detect a majority (64-83%) of the malicious flows with almost zero false alarms. Packet losses suffered by legitimate flows due to undetected malicious activity are insignificant (0.02-0.9%). The average detection time for correctly identified malicious flows is less than 1/10 of the average flow lifetime.

From the above discussions, we concluded that the CH architecture is capable of satisfying the QoS objectives of LSAs (e.g., < 1% loss rate and 150 ms delay) by exploiting statistical techniques and real-time traffic measurements. We evaluate the practicality and deployment issues of our approach through a lab prototype. Our implementation experience indicates that CH mechanisms introduce very minimal (at most 5%) processing overhead to an edge router.

---

Professor Randy H. Katz  
Dissertation Committee Chair

To my parents,  
my sisters, *Lan, Sim, Teen & Choo*,  
my brother, *Teik*,  
and *Mark*.

# Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Problem Definition . . . . .	4
1.3 State of the Art . . . . .	7
1.3.1 Re-engineering the Internet . . . . .	8
1.4 My Thesis: Scalable, End-to-end Resource Provisioning . . . . .	9
1.4.1 Basic Assumptions . . . . .	10
1.4.2 Contributions . . . . .	10
1.4.3 Key Design Principles: Aggregation and Hierarchical Control . . . . .	14
1.5 Dissertation Overview . . . . .	15
<b>2 Related Work</b>	<b>17</b>
2.1 QoS Control in Packet Networks: An Overview . . . . .	17
2.1.1 Schedulers and Buffer Management . . . . .	18
2.1.2 QoS Control Architecture . . . . .	19
2.1.3 The Big Picture . . . . .	21
2.2 Network Resource Provisioning . . . . .	23
2.2.1 Diff-Serv Bandwidth Broker Architecture . . . . .	24
2.2.2 Virtual Private Networks . . . . .	26
2.2.3 Capacity Planning in Telecommunication Networks . . . . .	27
2.2.4 Advance Reservations . . . . .	28
2.2.5 Dynamic Packet State . . . . .	29
2.2.6 Pricing-based Approach . . . . .	29
2.3 Admission Control . . . . .	31
2.3.1 Measurement Based Admission Control . . . . .	32
2.3.2 End-point Admission Control . . . . .	33
2.4 Traffic Policing . . . . .	34
2.4.1 Stochastic Fair Blue . . . . .	34
2.5 Summary . . . . .	36

<b>3</b>	<b>Methodology</b>	<b>39</b>
3.1	General Framework . . . . .	39
3.2	Workload Modeling . . . . .	41
3.2.1	VoIP Traffic Model . . . . .	42
3.2.2	VoIP Performance Requirements . . . . .	43
3.2.3	Packet Audio Traces . . . . .	48
3.3	Performance Evaluation . . . . .	51
3.3.1	Simulation Framework . . . . .	52
3.3.2	Lab Prototyping . . . . .	54
3.4	Summary . . . . .	55
<b>4</b>	<b>The Clearing House: A Distributed QoS Control Architecture</b>	<b>56</b>
4.1	Introduction . . . . .	57
4.1.1	Design Goals and Assumptions . . . . .	57
4.2	Design Rationale . . . . .	61
4.2.1	Background: The Internet Structure . . . . .	61
4.2.2	Key Design Decisions . . . . .	64
4.3	Clearing House Architecture . . . . .	66
4.3.1	Hierarchical CH-Tree . . . . .	68
4.3.2	An Illustration . . . . .	69
4.3.3	Local, Intermediate, and Parent Clearing House . . . . .	71
4.4	CH Control Flows . . . . .	73
4.4.1	Resource Reservations . . . . .	75
4.4.2	Caching and Aggregate Scheduling . . . . .	75
4.4.3	Admission Control . . . . .	76
4.4.4	Traffic Policing and Malicious Flow Detection . . . . .	77
4.5	Summary and Discussions . . . . .	77
4.5.1	VPN: An Example Application . . . . .	78
4.5.2	Other Resource Control Problems . . . . .	79
<b>5</b>	<b>Intra- and Inter-Domain Resource Control</b>	<b>81</b>
5.1	Introduction . . . . .	83
5.1.1	Motivation . . . . .	83
5.1.2	Our Approach . . . . .	84
5.1.3	Main Contributions . . . . .	86
5.2	Aggregate Reservations based on Gaussian Modeling . . . . .	89
5.2.1	Gaussian Traffic Predictor . . . . .	90
5.2.2	Deploying Gaussian Predictors . . . . .	92
5.2.3	Simulation Study . . . . .	94
5.2.4	Discussions . . . . .	98
5.3	Traffic Matrix Based Admission Control (TMAC) . . . . .	99
5.3.1	Single Domain Case with One CH-node . . . . .	100
5.3.2	Multiple Domain Case with Hierarchical TMAC . . . . .	104
5.4	Performance Study: TMAC Characteristics . . . . .	105
5.4.1	Network Topology . . . . .	105
5.4.2	Traffic Generation . . . . .	106
5.4.3	Performance Evaluation . . . . .	108

5.4.4	Discussions . . . . .	112
5.5	Aggregate Scheduling in the CH Architecture . . . . .	114
5.5.1	Background: Aggregate Scheduling . . . . .	114
5.5.2	RxW Scheduling . . . . .	116
5.5.3	Aggregate Scheduling and the Clearing House . . . . .	117
5.5.4	Simulation Framework . . . . .	117
5.5.5	Performance Evaluation . . . . .	119
5.6	Summary . . . . .	125
<b>6</b>	<b>Furies: Malicious Flow Detection via Aggregate Policing</b>	<b>128</b>
6.1	Introduction . . . . .	130
6.1.1	Motivation . . . . .	130
6.1.2	Traffic Policing and Malicious Flow Detection . . . . .	130
6.1.3	Performance Indexes . . . . .	131
6.1.4	Our Contributions . . . . .	132
6.2	Design Rationale . . . . .	134
6.2.1	Furies Service Model . . . . .	135
6.2.2	Flow-Identifiers and Group Policing . . . . .	137
6.2.3	Assumptions . . . . .	137
6.3	Furies Architecture and MDAP Mechanisms . . . . .	138
6.3.1	Components of Furies . . . . .	138
6.3.2	Fid Assignment and Releasing . . . . .	140
6.3.3	Group Policing . . . . .	142
6.3.4	MDAP Detection Process . . . . .	143
6.3.5	Policing of SLA Traffic . . . . .	145
6.3.6	Other Issues . . . . .	147
6.4	Simulation Study . . . . .	148
6.4.1	Network Topology . . . . .	149
6.4.2	Traffic Generation . . . . .	150
6.4.3	Performance Evaluation . . . . .	150
6.5	Implementation and Prototyping . . . . .	157
6.5.1	Overview of Implementation . . . . .	157
6.5.2	Performance Evaluation . . . . .	158
6.5.3	Discussions . . . . .	161
6.6	Deployment Issues . . . . .	162
6.6.1	Distributed RM Implementation . . . . .	162
6.6.2	Changes to Routers . . . . .	162
6.6.3	Virtual Private Networks . . . . .	162
6.6.4	Multiple ISPs . . . . .	163
6.7	Summary . . . . .	163
<b>7</b>	<b>Conclusions and Future Work</b>	<b>165</b>
7.1	Thesis Summary . . . . .	165
7.1.1	Workload Modeling . . . . .	166
7.1.2	Clearing House Architecture . . . . .	167
7.1.3	Resource Control Mechanisms . . . . .	167
7.1.4	Key Design Principles and Lessons . . . . .	170

7.2	Future Directions . . . . .	172
7.2.1	Signaling for Resource Control/Policy Distributions . . . . .	172
7.2.2	Effect of Routing Instability . . . . .	173
7.2.3	Inter-Domain Traffic Engineering . . . . .	174
7.2.4	Security Issues . . . . .	175
7.3	Conclusions . . . . .	176
	<b>Bibliography</b>	<b>178</b>

# List of Figures

2.1	Distributed administration of Internet infrastructure with multiple ISP (Internet Service Provider) domains and heterogeneous access networks. . . . .	22
2.2	QoS control mechanisms in both control and data planes. . . . .	23
2.3	Two-tier model for Diff-Serv Bandwidth Broker architecture. . . . .	25
3.1	Iterative “Analysis, Design & Evaluation” phases. . . . .	40
3.2	End-to-end delay components. . . . .	45
3.3	Experiment setup to carry out the subjective test that maps human perceived voice quality to different packet loss rates. . . . .	47
3.4	Subjective test results: how packet loss rate affect perceived voice quality. . . . .	48
3.5	An example topology of a first-tier IP backbone network in the United States. . . . .	53
4.1	An example first-tier ISP backbone and its logical network map. . . . .	61
4.2	Multiple-ISP scenario. . . . .	63
4.3	New service model: IE-Pipes( $s, d$ ) between specific ingress router IR- $s$ and egress router ER- $d$ . . . . .	64
4.4	(a) Local Clearing House (LCHs) associated with their basic domains that lie within a single logical domain. (b) An hierarchical CH-tree with multiple levels of logical domains. . . . .	65
4.5	Thesis roadmap: The CH architecture and its various resource control mechanisms. . . . .	67
4.6	An illustration: A hybrid of flat and hierarchical CH-architecture within and across ISP domains. . . . .	70
4.7	A logical view of the Local Clearing House (LCH) and its interaction with edge routers. . . . .	74
5.1	Thesis roadmap: The CH architecture and its various resource control mechanisms. . . . .	82
5.2	Hierarchical Clearing House architecture within large ISPs: the LCH performs resource management and admission control within a local POP, while the PCH maintains inter-POP and inter-ISP reservations. . . . .	85
5.3	An example: Gaussian distribution and Q-function. . . . .	91
5.4	ISP1 is an example of large ISPs that has multiple POPs and an associated local CH architecture. The PCH-nodes of various ISPs maintain peering relationships to coordinate inter-domain resource allocations. . . . .	92

5.5	Gaussian predictors for simulated voice traffic with $T_{\text{mea}} = 1$ and 10 minutes. $\rho = 180$ . . . . .	95
5.6	Gaussian predictors for actual voice traces with $T_{\text{mea}} = 1$ and 10 minutes. . . . .	96
5.7	Average $f_{\text{over}}$ (in %) when reservations are made based on Gaussian predictors for (a) aggregated voice traces, and (b) simulated voice traffic at $\rho=180$ . . . . .	97
5.8	A logical view of the Traffic Matrix Based Admission Control (TMAC) unit. . . . .	102
5.9	Simulation topology. . . . .	106
5.10	<b>Abrupt Traffic Fluctuations:</b> Share of bottleneck bandwidth allocated to each ingress-egress pair that share Link-3: $D(0, 7)$ , $D(1, 7)$ and $D(5, 7)$ , $t_u = 2$ minutes. . . . .	109
5.11	<b>Sensitivity analysis:</b> Link utilization vs. the update interval for upper-bound traffic matrix, $t_u$ , for different source models. . . . .	110
5.12	<b>Trade-offs:</b> Loss-load curves for four different source models as the control parameter $\sigma$ is varied. $t_u=10$ minutes. . . . .	111
5.13	A general framework for aggregate scheduling. . . . .	115
5.14	Simulation model. . . . .	118
5.15	Topology of the IP backbone with 12 basic domains. . . . .	119
5.16	Throughput of a Clearing House node as the traffic load is varied. . . . .	121
5.17	Call blocking rate as the traffic load is varied. . . . .	121
5.18	Mean response time as a function of traffic load. . . . .	122
5.19	Tear-down response time as a function of traffic load. . . . .	123
5.20	Distribution of response time at a load of 2000 requests per second. . . . .	124
5.21	Distribution of response time at a load of 3000 requests per second. . . . .	125
6.1	Thesis roadmap: The CH architecture and its various resource control mechanisms. . . . .	129
6.2	An example logical map of the Internet infrastructure. . . . .	134
6.3	(a) Components of Furies. (b) MDAP mechanisms. . . . .	139
6.4	The logical flow of control messages between an edge router (ER) and a Resource Manager (RM). . . . .	140
6.5	Simulation topology. . . . .	149
6.6	Case 1: Zero Malicious Flows. . . . .	151
6.7	Case 2: Many small homogeneous flows; a small fraction, $\gamma =0.1$ , misbehave. . . . .	151
6.8	Case 4: One large flow ( $r_l$ ) and many small flows ( $r_s$ ); $\gamma$ of small flows misbehave. . . . .	155
6.9	Throughput comparisons between Furies+Click and default Click. . . . .	159
6.10	Response time for processing flow requests for varying loads. . . . .	160

# List of Tables

1.1	Heterogeneous traffic behavior and QoS requirements of Internet applications.	3
2.1	Comparisons between the Clearing House approach and previously proposed architectures.	37
2.2	Comparisons between our resource control schemes and related work.	38
3.1	Summary of traffic traces.	49
5.1	Four traffic source models used to generate different workload for our ns-simulations.	108
6.1	Case 3: One large malicious flow and many small complying flows. $\eta = 5$ , $b_{\text{TBF}} = 6000$ , $\epsilon = 0.05$ .	154
6.2	Case 4: One large flow and many small flows. $\gamma$ of small flows misbehave. $\eta = 5$ , $b_{\text{TBF}} = 6000$ , $\epsilon = 0.05$ .	156
6.3	Comparisons between heterogeneous and homogeneous source models: $\gamma = 0.1$ , $b_{\text{TBF}} = 6000$ , $\epsilon = 0.05$ .	156

## Acknowledgements

I am truly grateful to everyone who has directly or indirectly supported and helped me complete this Ph.D. dissertation. First and foremost, I would like to thank my advisor, Professor Randy Katz, for giving me a chance to work with him in the ICEBERG project about three years ago. Despite his heavy schedule, Randy takes the time to meet with his students every week, sometimes even on weekends. He reviewed my drafts of papers and dissertation with amazing turn-around times and offered many insightful feedback on my research. This thesis would not have been possible without his support and guidance. I am very fortunate to have been able to tap his technical and professional advice.

I thank my dissertation committee members, Professors Jean Walrand and George J. Shanthikumar, for their comments and suggestions during the course of developing my research agenda and completing this dissertation. I also thank Dr. Steve McCanne for chairing my qualifying exam committee and sharing his expertise in networking research.

In addition to my committee members, I benefited greatly from working with Professor Anthony Joseph and the ICEBERG group members. The concept of a clearing house for resource trading over the Internet was first suggested by Anthony, and set the stage for the first part of my dissertation. The design of the Clearing House (CH) architecture has since been greatly improved and refined through discussions with members of ICEBERG. I owe special thanks to my collaborator, Lakshminarayanan Subramanian, for constantly challenging my ideas and helping me understand the system aspects of a networking architecture. I sincerely thank Brian Shiratsuki and Keith Sklower for maintaining our systems, and for being very responsive whenever I ran into problems and called for help.

I have also learnt a lot from interacting with the faculty and student members of the Smart Networks project. I truly appreciate the technical feedback given to me by Professors Jean Walrand, Pravin Varaiya and Venkat Anantharam during my various presentations at the Smart Networks group meetings.

During the course of my graduate education at Berkeley, I was most fortunate to work with Professors Joseph Kahn and David Tse for my M. S. thesis. They both taught me how to be bold in formulating research problems, be thorough in analysis, and be concise in writing. Their guidance helped me survive my first year of graduate school and build a sound foundation for my research career. I also thank them for being extremely patient and supportive while I was searching for a new Ph.D. topic.

I am grateful to Drs. Supratik Bhattacharyya, Lee Breslau, Jennifer Rexford, Scott Shenker, Nina Taft, John Wroclawsky, and Dina Papaqianaki for their technical and career advice. I benefited greatly from discussions with them in person as well as through electronic mails, and I look forward to interacting with them in the future.

Like many other Berkeley system research groups, ICEBERG holds semi-annual retreats where students get the chance to present their work to industry sponsors and solicit early feedback. I have often received useful suggestions and criticism on my work during these retreats. In particular, I would like to thank Drs. Sally Floyd, Bryan Lyles, Reiner Ludwig, and Kevin Mills for their generosity in time and advice.

I spent two summers of graduate school in industry internships, during which I had a chance to interact with experts from the industry. During Summer 1997, I was an intern with Dr. Jerry Foschini at Lucent Technologies, NJ. Since then, Jerry has been a constant source of career guidance to me. I would like to thank him for his encouragement and support. I also thank Drs. Dimitry Chizhik, Mike Gans, Reinaldo Valenzuela, and Jonathan Ling for a most stimulating research and learning atmosphere for exploring the various aspects of the BLAST (Bell-Labs Layered Space-Time) architecture.

In the summer of 1998, I interned again at Lucent Technologies, this time in UK. I thank my supervisor, Dr. Ioannis Kriaras, for making this internship possible. I was fortunate to witness and participate in the debate on what the Quality-of-Service (QoS) architecture and reservation protocol should be for UMTS (Universal Mobile Telecommu-

nications Systems). Drs. Xiao Bao Cheng, Jin Yang, and Luca Salgarelli were extremely helpful to me as I learned my way around Cisco routers, internal traffic generators, and the RSVP protocol stack. Their suggestions and ideas helped shape my research agenda when I returned to Berkeley.

I would not have come to graduate school if not for the positive research experience as an undergraduate at Rutgers University. It all began when Professor David Goodman offered me a part-time position at the Wireless Information Network Laboratory (WINLAB) at the end of my sophomore year. I am grateful to Professors David Goodman, Roy Yates, and Christopher Rose for their patience in mentoring and guiding me throughout my very first research project. I thank the following WINLAB alumni: Drs. Sudheer Grandhi, Ching-Yao Huang, Sanjiv Nanda, Ashwin Sampath and Mohammad Saquib for their generosity in time and advice. I do truly appreciate the WINLAB staff members for their help and support.

My wonderful memories of graduate school were most enriched by my day-to-day interactions with fellow graduate students. I thank the members of Soda 473, the “happy family” of Soda 330, and other EECS comrades – Sharad Agarwal, Andy Begel, Yan Chen, Adam Costello, Wei-Dong Cui, Tom Henderson, Barbara Hohlt, Almudena Konrad, Morley Mao, Giao Nguyen, David Oppenheimer, Bhaskar Raman, Drew Roselli, Angela Schuett, Jimmy Shih, Amoolya Singh, Daniel Tan, Helen Wang, Tina Wong, and Tao Ye. They have helped maintain my sense of humor and make my Berkeley years more fun and colorful.

Gene Cheung, Fang Pei Chen, Tz Yin Lin, Jocelyn Nee, Ching Shang, Da-Shan Shiu, and Ma Yi have been with me from the very start when we first stepped into graduate school. They provided me with constant support and encouragement, especially when things got too difficult, and when I needed some boost of confidence. I also owe special thanks to my friends from my home town – Theen Theen Tan, Kean Hock Yeap, and Niny Khor, for always being there through thick and thin, and for ensuring that I had a balanced and

fun-filled life outside of Berkeley.

I must thank Mary Byrnes, Ruth Gjerde, Peggy Lau, and Sheila Humphreys for helping me survive the bureaucracy at Berkeley. Their friendly attitude and strong dedication have made life so much more pleasant for graduate students in the EECS department. I truly appreciate their efforts and more importantly, their warm friendship. I also thank Bob Miller, Nathan Berneman, and Damon Hinson who managed our research group matters. They were wonderful to work with and always responded promptly to my requests and questions regarding administrative matters.

I cannot express enough of my gratitude to my parents, my sisters – Lan, Sim, Teen, and Choo, and my brother, Teik. Their constant love and support have always been the source of my strength and the reason I have come this far. They taught me the importance of hard work, discipline, and believing in oneself. Last but not least, I thank Mark Spiller for his love, encouragement, and faith in me. Mark's idealism has always been an endless source of inspiration for me. I also thank the Spiller family for their support and generosity.