

Varun Gupta

Computer Science Department
4212 Wean Hall
Carnegie Mellon University
Pittsburgh, PA 15213
USA

Phone: 412.268.1188
Fax: 412.268.5576
varun@cs.cmu.edu
<http://www.cs.cmu.edu/~varun>

Research Interests

Stochastic modeling and optimization; queueing theory; approximation algorithms; distributed algorithms; asymptotic analysis; game theory with applications to:

- capacity provisioning and load balancing in server farms
- power management in data centers
- schemes for automated resource management in content distribution networks
- concurrency control in web servers and database servers
- distributed optimization of P2P systems

Education

Ph.D. Computer Science, Expected 2010
Carnegie Mellon University, Pittsburgh, PA, USA
Advisor: Prof. Mor Harchol-Balter

B.Tech. Computer Science and Engineering, *President's Gold Medal winner*, August 2004
Indian Institute of Technology, New Delhi, India
Thesis: *Algorithms for Computational Biology: Sequence Analysis*
Advisor: Prof. S.N. Maheshwari

Research Projects and Experience

Carnegie Mellon University
Graduate Assistant
Prof. Mor Harchol-Balter

Pittsburgh, PA
August 2004–Present

My research interests lie in the design of resource allocation and scheduling schemes for performance-oriented operation of server farms and distributed/P2P systems. My work utilizes and contributes to techniques from stochastic modeling and queueing theory to perform analysis of queueing networks and scheduling policies to obtain an intuitive understanding of the effects of the system parameters on various performance metrics. A brief description of a few selected problems follows.

Load-balancing heuristics: All server farm architectures comprise of a high-speed front-end router/load-balancer (such as the IBM Network Dispatcher), and multiple back end servers. Despite their ubiquity, designing good load balancers, and understanding the effect of scheduling at the back end servers on the performance of load balancers is an open question. This project analyzes different load balancing policies and show the near-optimality of a simple Join-the-Shortest-Queue heuristic with Processor Sharing back end servers. This work appeared in Performance 2007.

Resource sharing systems: A critical part of time sharing systems, such as web servers

and database servers, is performing concurrency control by limiting the number of transactions allowed to share the server's resources. The degree of concurrency, also known as the multi-programming limit, is chosen to ensure efficient utilization of resources, while minimizing thrashing overheads. We perform mathematical analysis of the effect of the degree of sharing on the performance of such resource sharing systems, and prove that the common concurrency control approaches are suboptimal. Further, we propose robust traffic-oblivious concurrency control schemes that self-adapt to the traffic pattern. This work appears in ACM SIGMETRICS/Performance 2009.

Systems with time-varying load: While most queueing theoretic work on capacity provisioning for server farms assumes time-invariant arrival rates, real world systems experience arrival rates exhibiting a diurnal pattern. This project analyzes computer system performance under situations where the workload and service capacity vary over time, and proposes capacity provisioning schemes for such systems. Systems with time-varying traffic and service characteristics are a fundamental building block for more complicated resource sharing systems and multi-server systems. In this work, we invent new queueing-theoretic techniques to obtain bounds and approximations for systems with time-varying load. Part of this work appeared in ACM SIGMETRICS/Performance 2006.

Bell Laboratories, Alcatel-Lucent
Research Intern
Host: Sem Borst

Murray Hill, NJ
Summer 2008

The goal of the internship was to find policies to minimize access latencies and bandwidth usage in a hierarchical content distribution network. The project characterizes the structure of the optimal centralized caching scheme, proposes simple distributed algorithms with provable performance guarantees, and presents stochastic analysis of some of the proposed distributed algorithms. A preliminary version of this work appears as a poster paper in ACM SIGMETRICS/Performance 2009.

Microsoft Research Ltd.
Research Intern
Host: Milan Vojnovic

Cambridge, UK
Summer 2006

The main goal of this internship was to investigate best sampling strategies for epidemic-style information dissemination to improve efficiency of computer patching. The project also investigates effectiveness of P2P-based architectures for Windows Update - one of the largest content distribution networks in the world. Part of this work appeared in IEEE INFOCOM 2008.

Indian Institute of Technology
Bachelor's Thesis in Bioinformatics
Prof. S.N. Maheshwari

New Delhi, India
Aug 2003–Jul 2004

For my B.Tech. thesis, I worked under Prof. Maheshwari's supervision in collaboration with Prof. Alok Bhattacharya (School of Life Sciences, Jawaharlal Nehru University) to identify and solve problems in the field of Computational Biology. We developed a heuristic algorithm to find approximate repeats in DNA sequences. We also designed a Fourier-transform based algorithm to find approximate occurrences of a pattern (in terms of *Hamming distance*) in a database string with a general alphabet distance/similarity matrix.

Report available at <http://www.cs.cmu.edu/~varun/docs/report.pdf>

Read-Ink Tech. Pvt. Ltd.
Handwriting Recognition

Bangalore, India
Summer 2003

Prof. Thomas O. Binford (Emeritus Professor of Computer Science - Research, Stanford University)

During my summer internship, I worked with Prof. Binford and his team who are developing the next generation handwriting recognition engine with a target recognition rate of 99%.

Publications

- On the inapproximability of $M/G/K$: Why two moments of job size distribution are not enough
with J. Dai, M. Harchol-Balter, B. Zwart, To appear in *Queueing Systems: Theory and Applications*.
- Self-Adaptive Admission Control Policies for Resource-Sharing Systems
with M. Harchol-Balter, *ACM SIGMETRICS/Performance 2009*.
- Self-Organizing Algorithms for Cache Cooperation in Content Distribution Networks
with Sem Borst and Anwar Walid, Poster paper in *ACM SIGMETRICS/Performance 2009*.
- Fluid Level in a Reservoir with an On-Off Source
with P. Harrison, *ACM SIGMETRICS Performance Evaluation Review, September 2008*.
- Finding the Optimal Quantum Size: Sensitivity Analysis of the $M/G/1$ Round-Robin Queue
ACM SIGMETRICS Performance Evaluation Review, September 2008.
- Sampling Strategies for Epidemic-style Information Dissemination
with M. Vojnovic, T. Karagiannis, C. Gkantsidis, *IEEE INFOCOM 2008*.
- Analysis of Join-the-Shortest-Queue Routing for Web Server Farms
with M. Harchol-Balter, K. Sigman, W. Whitt, *Performance 2007*.
- The Effect of Higher Moments of Job Size Distribution on the Performance of an $M/G/s$ queueing systems
with J. Dai, M. Harchol-Balter, B. Zwart, *ACM SIGMETRICS Performance Evaluation Review, September 2007*.
- Insensitivity for PS Server Farms with JSQ routing
with M. Harchol-Balter, K. Sigman, W. Whitt, *ACM SIGMETRICS Performance Evaluation Review, September 2007*.
- Fundamental Characteristics of Queues with Fluctuating Load
with M. Harchol-Balter, A. Scheller-Wolf, U. Yechiali, *ACM SIGMETRICS/Performance 2006*.

Teaching

Teaching Assistant

Fall 2006

15-451: Algorithm Design and Analysis. (Undergraduate Course)

Instructor: Prof. Avrim Blum and Prof. Manuel Blum

Teaching Assistant

Spring 2005

15-849: Performance Modeling. (Graduate Course)

Instructor: Prof. Mor Harchol-Balter

Awards and Honors

- Awarded the President's Gold Medal for the highest CGPA at the end of undergraduate studies, Indian Institute of Technology, New Delhi, 2004.
- Awarded the R. Vibhakar Memorial Award for the best student of third year during the 2002-03 session, Indian Institute of Technology, Delhi.
- Awarded the Rajiv Bhambawale Trust Award for securing highest CGPA amongst the third year undergraduate students during 2002-03 session, Indian Institute of Technology, Delhi.
- Awarded the Raman Subramanian Award for securing the highest CGPA amongst the second year undergraduates during the 2001-02 session, Indian Institute of Technology, Delhi.
- Nominated as an executive member of the Delhi wing of IEE (UK) for the year 2001.
- Secured All India Rank 22 at the Indian Institute of Technology Joint Entrance Examination (IITJEE 2000) and All India Rank 10 at Roorkee Entrance Test 2000 (now IIT Roorkee).
- Secured 2nd rank (senior level) in Delhi region at the Regional Mathematics Olympiad 1997 organised by National Board for Higher Mathematics (NBHM). Consequently, I was selected for the Nurture Program to receive training for Indian National Mathematics Olympiad (INMO).
- Secured 17th rank (junior level) in Delhi region at Regional Mathematics Olympiad 1996 organised by National Board for Higher Mathematics (NBHM).

Skills

- Languages: English, Hindi, basic French
- Computer Skills: C/C++, Java, SML, Unix, Linux, \LaTeX , Matlab, Mathematica, Perl, HTML, SQL