

VARUN GUPTA

Computer Science Department
Gates Center for Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
USA

Phone: 412.953.2512
Fax: 412.268.5576
varun@cs.cmu.edu
<http://www.cs.cmu.edu/~varun>

RESEARCH INTERESTS

Stochastic modeling and optimization; queueing theory; applied probability; approximation algorithms; distributed algorithms; game theory with applications to:

- Capacity planning and load balancing for server clusters
- Energy management in data centers and cloud computing
- Automated resource management in content distribution networks
- Distributed optimization of P2P systems
- Influence propagation in social networks

EDUCATION

Ph.D. Computer Science, Expected May 2011.

Carnegie Mellon University, Pittsburgh, PA, USA.

Thesis: *Stochastic Models and Analysis for Resource Management in Server Farms*

Advisor: Prof. Mor Harchol-Balter

Committee Members: David Andersen (CMU); Anupam Gupta (CMU); Alan Scheller-Wolf (Tepper School of Business, CMU); Devavrat Shah (EECS, LIDS and ORC, MIT); Don Towsley (UMass., Amherst)

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

WORK EXPERIENCE

Carnegie Mellon University

Graduate Research Assistant

Prof. Mor Harchol-Balter

Pittsburgh, PA

August 2004–Present

My research goals center around designing smarter, more efficient resource allocation and scheduling schemes for server farms and distributed/P2P systems. To achieve these goals, my work targets developing new and fundamental techniques in many areas of applied probability: e.g., queueing theory, asymptotic analysis of large-scale systems, fluid analysis, stochastic bounds. Some highlights of my research are:

1. We present a new energy-efficient traffic-oblivious capacity scaling policy for data centers with unpredictable demands, and prove its optimality as the system size becomes large.
2. We break the 50-year status quo on approximating the mean response time of the $M/G/K$ queueing system by proposing the first approximation which involves more

than two moments. Further, we open a new research avenue in analysis and probability – Markov-Krein characterization of solutions of stochastic recursive sequences – as a unified approach to obtain bounds on performance of queueing systems.

3. We show that a simple Join-the-Shortest-Queue load balancer is nearly-optimal for Processor Sharing servers, even when the servers are heterogeneous.
4. We prove that the naive policy of operating time-sharing systems at their peak-efficiency is suboptimal when the job sizes have high variance. Via a new stochastic dynamic programming technique, we propose a concurrency control mechanism which can self-adapt to the traffic-demand.
5. We derive the first closed-form approximation for queueing systems with time-varying arrival and service patterns. Further, we present the first analytical results on the effect of time-varying demands on scheduling policies, and prove that optimal policies are different for time-invariant and time-varying demands.

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

Murray Hill, NJ
Summer 2008

We seek policies that minimize access latencies and bandwidth usage in a hierarchical **content distribution network**. We characterize the structure of the optimal centralized **caching scheme**, and propose simple distributed algorithms with provable performance guarantees.

Microsoft Research Ltd.
Research Intern
Host: Milan Vojnovic

Cambridge, UK
Summer 2006

We investigate best **sampling strategies for epidemic-style information dissemination** to improve efficiency of computer patching. We also investigate effectiveness of P2P-based architectures for Windows Update - one of the largest content distribution networks in the world.

PUBLICATIONS

Analysis and Optimization of Multi-server systems

1. “Optimality Analysis of Energy-Performance Trade-off for Server Farm Management.” *Performance Evaluation*. Vol. 67, Issue 11. Pages 1155-1171. V. Gupta, A. Gandhi, M. Harchol-Balter, and M. Kozuch.
2. “On the inapproximability of $M/G/K$: Why two moments of job size distribution are not enough.” *Queueing Systems: Theory and Applications*. Vol. 64, Issue 1, Pages 5-48. V. Gupta, J. Dai, M. Harchol-Balter, and B. Zwart.
3. “Stability of the Bipartite Matching Model.” *Under Submission*. A. Busic, V. Gupta, and J. Mairesse.
4. “On Markov-Krein Characterization of the Mean Sojourn Time in $M/G/K$ and other Queueing Systems.” *Under Submission*. V. Gupta and T. Osogami.
5. “Self-Adaptive Admission Control Policies for Resource-Sharing Systems.” *ACM SIGMETRICS/Performance 2009*. V. Gupta and M. Harchol-Balter.

6. "Analysis of Join-the-Shortest-Queue Routing for Web Server Farms." *Performance Evaluation*. Vol. 64, Issues 9-12, Pages 1062-1081. V. Gupta, M. Harchol-Balter, K. Sigman, and W. Whitt.
7. "The Effect of Higher Moments of Job Size Distribution on the Performance of an $M/G/s$ queueing systems." *ACM SIGMETRICS Performance Evaluation Review*. Vol. 35 Issue 2, Pages 12-14. V. Gupta, J. Dai, M. Harchol-Balter, and B. Zwart.

Storage and Caching in Distributed Systems

8. "Robust and Flexible Power-proportional Storage." *Symposium on Cloud Computing (SOCC) 2010*. H. Amur, J. Cipar, V. Gupta, M. Kozuch, G. Ganger, and K. Schwan.
9. "Distributed Caching Algorithms for Content Distribution Networks." *IEEE INFOCOM 2010*. S. Borst, V. Gupta, and A. Walid.

Influence Propagation

10. "Sampling Strategies for Epidemic-style Information Dissemination." *IEEE Transactions on Networking*. Vol. 18, No. 4, Pages 1013-1025. 2010. M. Vojnovic, V. Gupta, T. Karagiannis, and C. Gkantsidis.

Systems with correlated arrivals

11. "Analysis of Scheduling Policies under Correlated Job Sizes." *Performance Evaluation*. Vol. 67, Issue 11, Pages 996-1013. V. Gupta, M. Burroughs, and M. Harchol-Balter.
12. "Fluid Level in a Reservoir with an On-Off Source." *ACM SIGMETRICS Performance Evaluation Review*. Vol. 36 Issue 2, Pages 128-130. V. Gupta and P. Harrison.
13. "Fundamental Characteristics of Queues with Fluctuating Load." *ACM SIGMETRICS/Performance 2006*. V. Gupta, M. Harchol-Balter, A. Scheller-Wolf, and U. Yechiali.

Scheduling and Mechanism Design

14. "Inducing Optimal Scheduling with Selfish Users." *Under Revision for 2nd Review at Management Science*. L. Debo, P. Enders, A. Gandhi, V. Gupta, M. Harchol-Balter, and A. Scheller-Wolf.
15. "Finding the Optimal Quantum Size: Sensitivity Analysis of the $M/G/1$ Round-Robin Queue." *ACM SIGMETRICS Performance Evaluation Review*. Vol. 36 Issue 2, Pages 104-106. V. Gupta.

CONFERENCE TALKS

1. "Analysis of Scheduling Policies under Correlated Job Sizes." *Performance 2010*. November 17, 2010.
2. "Self-Adaptive Admission Control Policies for Resource Sharing Systems." *ACM SIGMETRICS 2009*. June 18, 2009.
3. "Finding the Optimal Quantum Size: Sensitivity Analysis of the $M/G/1$ Round-Robin Queue." *MAMA 2008*. June 2, 2008.
4. "Fluid Level in Tandem Queues with an ON/OFF Source." *MAMA 2008*. June 2, 2008.
5. "Analysis of the Join-the-Shortest-Queue Policy for web server farms." *Performance 2007*. October 4, 2007.

6. “Fundamental Characteristics of Queues with Fluctuating Load.” *ACM SIGMETRICS 2006*. June 29, 2006.

INVITED TALKS

1. “Energy-efficient Dynamic Capacity Provisioning in Server Farms.” *INFORMS 2010*. November 8, 2010.
2. “Optimal Routing Policies for Heterogeneous Server Farms.” *INFORMS 2010*. November 9, 2010.
3. “Self-Adaptive Admission Control Policies for Resource Sharing Systems.” *INFORMS APS 2009*. July 13, 2009.
4. “Optimizing Resource Sharing Systems.” *University of Washington*. June 12, 2009.
5. “The Effect of Higher Moments of Job Size Distribution on the Performance of an $M/G/k$ Queueing System.” *INFORMS APS 2007*. July 9, 2007.
6. “Fundamental Characteristics of Queues with Fluctuating Load.” *INFORMS 2006*. November 7, 2006.

TEACHING

Teaching Assistant Fall 2010
15-359: Probability and Computing. (Undergraduate Course)
Instructor: Prof. Avrim Blum and Prof. Venkat Guruswami

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

REFEREEING

Operations Research; IEEE Transactions on Parallel and Distributed Systems; IEEE Transactions on Automatic Control; IEEE Transactions on Computers; Performance Evaluation; Annals of Operations Research; INFORMS Journal on Computing; IEEE Transactions on Multimedia.

AWARDS & HONORS

- Intel Research Fellow for Summer 2010.
- 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).

REFERENCES

Prof. Mor Harchol-Balter

Associate Department Head
 Computer Science Department
 Carnegie Mellon University.
 412.268.7893
 harchol@cs.cmu.edu

Prof. Jim Dai

James C. Edenfield Professor
 School of Industrial and Systems Engineering
 Georgia Institute of Technology.
 404.894.9139
 dai@gatech.edu

Prof. Sem Borst

Mathematics of Networks and Systems Group
 Bell Laboratories, USA.
 (also) Professor
 Department of Mathematics and Computer Science
 Technical University Eindhoven, The Netherlands.
 908.582.4295
 sem@research.bell-labs.com

Prof. Alan Scheller-Wolf

Professor of Operations Management
 Tepper School of Business
 Carnegie Mellon University.
 412.268.5066
 awolf@andrew.cmu.edu

Prof. Avrim Blum

Professor of Computer Science
 Computer Science Department
 Carnegie Mellon University.
 412.268.6452
 avrim@cs.cmu.edu