SUNDAR IYER

724 Bryant St.
Palo Alto, CA-94301
http://www.stanford.edu/~sundaes

Phone: (650) 575-9659
Email: sundar@nemo-systems.com
Alt: sundaes@cs.stanford.edu

PARTICULARS

EDUCATION

Stanford University — Stanford, CA
Ph. D. in Computer Science — *2000-, Defended 2003*

Stanford University — Stanford, CA
M. S. in Computer Science — *June 2000*
Distinction in Research

Indian Institute of Technology (I. I. T.), Bombay — Mumbai, India
B. Tech. in Computer Science and Engineering — *April 1998*

CURRENT STATUS

U.S. Permanent Resident, Citizen of India.

RESEARCH INTERESTS

My research interests span the areas of network algorithms, system architecture and component design. I have a specific interest in algorithms for packet processing, switching, router design, memory architectures and scalable load balancing systems.

DISSERTATION

Title: "Parallelism and load-balancing for the Internet"
Advisor: Prof. Nick McKeown

My thesis develops an analytical framework for the design of networking systems (specifically memory intensive aspects of Router Architecture, Packet Processing and Switching Algorithms) which give performance guarantees via the use of load balancing algorithms and parallelism in design.

ACADEMIC HONORS

- Cisco Stanford FMA Fellowship, Stanford University, 2002-2003.
- Siebel Scholars Fellowship, Stanford University, 2001.
- Paper chosen amongst the best papers from IEEE Hot Interconnects 2001 and selected for publication to IEEE Micro.
- Christofer Stephenson Memorial Award for the best Masters Thesis in Computer Science, Stanford University, 2000.
- Indian National Talent Search Merit Scholarship, 1992-1998.
- Indian National Talent Search Examination in Physics (NSEP) awardee, 1993.
- Selected to participate in the Indian National Maths Olympiad, 1993.
- Indian National Merit Scholarship, 1992-1998.

WORK EXPERIENCE

- **CTO, Co-Founder and Member of the Board, Nemo Systems,** Nov 2003 - Sep 2005. Nemo Systems (acquired by Cisco Systems in Sep 2005) was a fabless semiconductor company, building caching memory sub-systems for networking applications.
- **Consultant, Nevis Networks Inc.**, Sep. 2003 - Mar 2004. Analyzed the performance of the Nevis switch fabric. Suggested theoretically optimal and practical packet switching algorithms for the distributed shared memory switch fabric.
- **Systems Engineer, PMC-Sierra,** Sep. 2000 - Oct. 2001.
- **Senior Systems Architect, SwitchOn Networks,** Jan. 1999 - Sep. 2000. Part of the initial team of six members of SwitchOn Networks (later acquired by PMC-Sierra in Sep. 2000) during its founding. Jointly responsible for the design, architecture, and algorithm evaluation of the ClassiPI chip set. ClassiPI was a networking content co-processor which was released commercially in Feb. 2001. Other joint responsibilities included technical marketing and patent strategy.
- **Consultant, RIMO Technologies, India**, Jun. 1998 - Dec. 1998.

RESEARCH EXPERIENCE

- **Research Assistant, Stanford University**, Sep 1998 - Present.
  Parallel Packet Switches, (Sep 1998 - June 2000): We ask the question as to whether it is possible to build a large switch from multiple smaller switches stacked in parallel, called the parallel packet switch (PPS), such that it behaves exactly like a large switch. I designed an algorithm and derived the first analytical proof of the conditions under which this is possible. I then extended this work so that the PPS can support quality of service and multicast traffic. Later, I derived and analyzed a practical distributed algorithm for the PPS. This research demonstrates how switching capacity can be scaled in an efficient manner.

  Distributed Shared Memory Routers, (Jan 2001 - Mar 2002): From a network designer's perspective a shared memory router is ideal in that the packets are stored in a central location and the memory bandwidth and space is shared across packets from all ports. This sharing helps in conserving memory space and results in low cost and low power routers. However, it is a widely held myth that such shared memory routers are not scalable to higher speeds due to limitations imposed on the speed of a single memory. I worked on the first demonstration and proof to show how to build and scale the capacity of shared memory routers using distributed memories.

  The Pigeon Hole Principle for Routers, (Sep 1998 - Mar 2002): In the course of our work on routers we derived a technique called "Constraint Sets". This was used to analyze a number of router architectures especially those with a single stage of buffering. The "Constraint Set" technique is a generalization of the Pigeon Hole principle. Later, I extended this work to Combined Input Output Queued routers, which are routers with two stages of buffering.

  Scalable, High Performance Packet Buffers, (Sep 2000 - Mar 2002): Packet buffers on Internet Routers have two requirements. They need to be large and have to be accessed at a very high rate. Unfortunately commodity memories such as DRAMs have slow access speeds though they allow large storage, while commodity SRAMs allow fast access but are inefficient in storage. I analyzed a well-known hierarchical buffer architecture that has a small amount of SRAM cache combined with DRAM. However, unlike computer architecture where caching only allows statistical guarantees, the algorithms proposed exploit the characteristics of memory requirements for networking to enable the design of a memory architecture which gives deterministic latency guarantees. I showed how the optimal algorithm can be modeled using difference equations and used adversarial traffic patterns to derive bounds on the minimum size of the SRAM memory. This resulting memory architecture supports the access speeds of SRAM and has the storage capacity of DRAM.

  Deterministic Architectures for Statistics and State Maintenance, (Sep 2000 - Mar 2002): A number of packet processing tasks in the Internet put a large demand on the costs and the power budget for designers. We specifically evaluate two such tasks; namely maintaining statistics counters and maintaining state of connections. We propose architectural solutions and derive optimal algorithms. Our analytical results are based on the use of potential functions to model the system. This research shows how the above packet processing tasks can be scaled to very high line rates.

  Distributed Algorithms for Buffered Crossbars, (June 2002 - Present): Internet core routers have been designed lately using a crossbar switching fabric. While it is possible to build routers using crossbars which give deterministic performance guarantees, the algorithms required to realize this are centralized and have high computational complexity. Hence in practice most routers do not use these algorithms and do not give any guaranteed performance. A colleague of mine, Shang-Tse Chuang, and I analyzed a slight modification to the crossbar fabric called the buffered crossbar, which is simply a crossbar with a small number of buffers in it. We derive a bijection between a crossbar and a buffered crossbar to show that any algorithm on the buffered crossbar can be implemented on the crossbar and vice versa. We derive a suite of distributed algorithms on the buffered crossbar and derive analytically the conditions under which they can give both statistical and deterministic guarantees for both first come first serve routers and routers which support quality of service. Since these distributed algorithms can operate independently on each input and output port of the switch without communicating with each other, they are readily implementable. Our results show that Internet routers built using crossbars can be re-designed in a practical manner using buffered crossbars and give superior performance and deterministic guarantees.

Stability Properties of the Maximum Size Matching, (May 2002 - Present): Contrary to intuition, it is known in queueing theory that a greedy policy that maximizes the instantaneous throughput in a system of queues may not maximize the long term throughput. However, greedy policies are of interest because they are usually easy to implement. A maximum size matching is an example of such a greedy scheduling policy on crossbar switches. I analyzed this greedy policy and showed that with stochastically constrained traffic and the framework of scheduling packets in batches, the greedy policy maximizes the long term throughput. A number of open problems remain in this area and at the time of writing, this work is in progress.

- **Research Intern, ATL SprintLabs**, June - Sep. 2001.
  Internet Measurements & Deflection Routing: Our work involved designing a practical solution for tackling link overload on the network backbone. First, I analyzed how overload occurs on the Sprint backbone. This involved writing a tool to characterize link overload and create statistics from SNMP measurements collected for a period of a year from the Sprint backbone. With the intuition gained from these statistics we concluded that there was a need to alleviate overload using deflection routing. I suggested a method of setting link weights on the Sprint network (taking advantage of the topology characteristics) to enable a loop free deflection routing algorithm.
- **Research Intern, RIMO Technologies, India**, May - July 1997.
  Packet Classification Algorithms: Designed classification algorithms and developed a commercial software packet classifier for Windows NT and the VXL Terminal Server embedded system platform.
- **Undergraduate Research Intern, I. I. T. Bombay**, Apr - July 1996.
  Java Application Software: Developed algorithms and software for Java applications for distributed computing, under Prof. Sharat Chandran.

TEACHING EXPERIENCE

- **Teaching Assistant.** EE384X: Packet Switching Architectures-I, Prof. Balaji Prabhakar and Prof. Nick McKeown, Winter 2002, Stanford University.
- **Teaching Assistant.** EE384Y: Packet Switching Architectures-II, Prof. Balaji Prabhakar and Prof. Nick McKeown, Spring 2001 and Spring 2002, Stanford University.
- **Instructor.** Networking, Center for Development of Advanced Computing (CDAC) at MET Mumbai, Winter 1997, Bombay, India (jointly with Prof. A. Karandikar, I. I. T. Bombay).

PUBLICATIONS

PAPERS

1. Sundar Iyer, Nick McKeown, “Analysis of the Parallel Packet Switch Architecture”, to appear in *IEEE/ACM Transactions on Networking*, Apr. 2003.
2. Sundar Iyer, Nick McKeown, “Using Constraint Sets to Achieve Delay Bounds in CIOQ Switches”, to appear in *IEEE Communication Letters*, 2003.
3. Sundar Iyer, Supratik Bhattacharyya, Nina Taft, Christophe Diot, “An Approach to Alleviate Link Overload as Observed on an IP Backbone”. To appear in *proceedings of IEEE INFOCOM*, San Francisco, March 2003. A more detailed version of this paper appears in two technical reports as, “An Approach to Alleviate Link Overload as Observed on an IP Backbone ”, *Sprint ATL TR02-ATL-071127*, July 2002 and “A Measurement Based Study of Load Balancing on an IP Backbone”, *Sprint ATL TR02-ATL-051027*, May 2002.
4. Sundar Iyer, Rui Zhang, Nick McKeown, “Routers with a Single Stage of Buffering”, *Proceedings of ACM SIGCOMM*, Pittsburgh, Pennsylvania, Sep 2002. Also in *Computer Communication Review*, vol. 32, no. 4, Oct 2002.

5. Sundar Iyer, Nick McKeown, “On the Speedup Required for a Multicast Parallel Packet Switch”, *IEEE Communication Letters*, June 2001, vol. 5, no. 6, pp. 269-271.
6. Sundar Iyer, Ramana Rao, Nick McKeown, “Analysis of a Memory Architecture for Fast Packet Buffers”, *IEEE - High Performance Switching and Routing*, Dallas, May 2001, pp. 368-373. A more detailed version of this work was presented in the *IEEE GBN Workshop*, Alaska, April 2001.
7. Sundar Iyer, Nick McKeown, “Making Parallel Packet Switches Practical”, *Proceedings of IEEE INFOCOM*, Alaska, April 2001, vol. 3, pp. 1680-87.
8. Sundar Iyer, Ramana Rao Kompella, Ajit Shelat, “ClassiPI: An Architecture for Fast and Flexible Packet Classification”, *IEEE NETWORK, Special Issue on Fast IP Packet Forwarding and Classification for Next Generation Internet Services*, Mar-Apr. 2001.
9. Sundar Iyer, Ajay Desai, Ajay Tambe, Ajit Shelat, “ClassiPI: A Classifier for Next Generation Policy Based Engines”, *IEEE Hot Chips*, Stanford University, Aug 2000.
10. Sundar Iyer, Amr. A. Awadallah, Nick McKeown, “Analysis of a Packet Switch with Memories Running Slower than the Line Rate”, *Proceedings of IEEE INFOCOM*, Tel Aviv, March 2000, pp.529-537.
11. “Practical Algorithms for Performance Guarantees in Buffered Crossbars”, with Shang-Tse Chuang, Nick McKeown, *IEEE INFOCOM 2005*.

INVITED PAPERS

12. Sundar Iyer, Nick McKeown, “Maximum Size Matchings and Input Queued Switches”, *Proceedings of the 40th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, Oct 2002.
13. D. Shah, Sundar Iyer, Balaji Prabhakar, Nick McKeown, “Maintaining Statistics Counters in Router Line Cards”, *IEEE Micro,* Jan-Feb, 2002, pp. 76-81. Also appeared as “Analysis of a Statistics Counter Architecture” in *IEEE Hot Interconnects*, Stanford University, Aug. 2001.

PAPERS UNDER REVIEW

14. Sundar Iyer, R. R. Kompella, Nick McKeown, “Designing Packet Buffers for Router Line Cards”, *IEEE/ACM Transactions on Networking*. Presently available as, *HPNG Technical Report - TR02-HPNG-031001*, Stanford University, Mar. 2002.

PENDING PAPERS

15. “Attaining Statistical and Deterministic Switching Guarantees using Buffered Crossbars”, with Shang-Tse Chuang, Nick McKeown. In preparation for *IEEE/ACM Transactions on Networking*.
16. “Maintaining State in Router Line Cards”, with Nick McKeown. In preparation for *IEEE Communication Letters*.

OTHER REPORTS

17. Sundar Iyer, Nick McKeown “Techniques for Fast Shared Memory Switches”, *HPNG Technical Report - TR01-HPNG-081501*, Stanford University, Aug 2001.
18. Sundar Iyer, “The Parallel Packet Switch Architecture”, *Masters Thesis Report*, May 2000, Stanford University.

PATENTS & APPLICATIONS

These patents were applied for when working in Industry. All patent applications are in processing, except when mentioned otherwise.

1. **Network Memory**
   - Multiple patents filed for various algorithms and techniques for the design of network memory. Details of these patents are available on request.
2. **Network Address Translation**
   - *(granted)* Subhash Bal, Raghunath Iyer, Sundar Iyer, “Method and Apparatus for Performing Internet Network Address Translation”, U.S. Patent-6457061. Filed Nov. 1998, Issued Sep. 2002.

3. **Pattern Matching**
   - Sundar Iyer, Ajit Shelat, George Varghese et al., "Fast, Deterministic Exact Match Look-ups In Large Tables". Filed Oct. 2002.
   - Sundar Iyer, Ajit Shelat, George Varghese et al., "Method and Apparatus for Parallel String Searching". Filed May. 2000.
4. **Route Lookup and Packet Classification**
   - Sundar Iyer, Ajit Shelat et al., "Algorithm and Design for Deterministic Parallel Route Lookups". Filed Jul. 2001.
   - Sundar Iyer, Moti Jiandani, Ajit Shelat et al., "Control System for High Speed Rule Processors". Filed Apr. 1999.
   - *(granted)* Sundar Iyer, Ajit Shelat et al.,"Method and Apparatus for High-Speed Network Rule Processing using an Array of Cells - II". U.S. Patent-6510509. Filed Mar. 1999, Issued Jan. 2003.
   - Sundar Iyer, Moti Jiandani, Ajit Shelat et al., "Method and Apparatus for High-Speed Network Rule Processing- I". Filed Dec. 1998.
5. **Deflection Routing**
   - Sundar Iyer, Supratik Bhattacharrya, Nina Taft, Christophe Diot, "A Method to Alleviate Link Overload on an IP Backbone". Filed July 2003.

TALKS

CONFERENCE TALKS

1. *(Pending)* "An Approach to Alleviate Link Overload as Observed on an IP Backbone", *IEEE INFOCOM 2003*, San Francisco, California, Apr. 2003.
2. *(Invited)* "Maximum Size Matchings and Input Queued Switches", *40th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, October 2002.
3. "Routers with a Single Stage of Buffering", *ACM SIGCOMM*, Pittsburgh, Pennsylvania, Aug 2002.
4. "Analysis of a Statistics Counter Architecture", *IEEE Hot Interconnects*, Stanford, California, Aug. 2001.
5. "Analysis of a Memory Architecture for Fast Packet Buffers", *IEEE - High Performance Switching and Routing*, Dallas, Texas, May 2001.
6. "Making Parallel Packet Switches Practical", *IEEE INFOCOM 2001*, Anchorage, Alaska, USA, April 2001.
7. "Techniques for Fast Packet Buffers", *IEEE - GBN Workshop 2001*, Anchorage, Alaska, USA, April 2001.
8. "ClassiPI: A Classifier for Next Generation Content/Policy Based Switches", *IEEE Hotchips*, Stanford, California, August 2000.
9. "Co-processors and the role of specialized hardware", *NETWORLD + INTEROP 2000*, Las Vegas, Nevada, May 2000.
10. "Analysis of a Packet Switch with memories running slower than the line-rate", *IEEE INFOCOM 2000*, Tel Aviv, Israel, March 2000.

INDUSTRY/OTHER TALKS

11. *(various venues)* "Analysis of the Parallel Packet Switch Architecture",
    - *(Invited)* Dept. of Electrical Engineering, I. I. T. Bombay, India - Dec. 2001.
    - SwitchOn Networks, Pune, India - Sep. 1999.
12. "Requirements for a packet classification API", Network Processor Forum (CPIX), Denver, Colorado - Mar. 2001.
13. *(various venues)* "Designing Packet Buffers for Networking",
    - Rambus Inc., Los Altos, California - Apr. 2001, Sep. 2002.
    - Infineon Technologies, San Jose, California - Apr. 2002, May 2002.
    - Juniper Networks, Sunnyvale, California - Apr. 2001.

- University of California Davis, Davis, California - Oct 2002.

14. "Some Observations towards Load Balancing over a Network", Sprint ATL Retreat, San Francisco, California, Aug. 2001.
15. "Shortest Path Routing via Alternate Nearest Neighbor", Sprint ATL Retreat, Lake Tahoe, California, Mar. 2002.
16. "Analyzing CIOQ Switches using the Constraint Set Technique", Stanford University, California, Feb. 2002.
17. "Practical Algorithms for Performance Guarantees in Buffered Crossbars", Stanford Networking Seminar, California, Jan. 2003.

PROFESSIONAL ACTIVITIES

- Technical Program Committee Member, QoS-IP, Milano, Italy, 2003.
- Member Network Processing Forum (Originally CPIX) - 2000-2001.

SERVICE

- Reviewer (Journals) - *IEEE/ACM Transactions on Networking, IEEE Journal on Selected Areas in Communications, IEEE Network, Computer Networks, IEEE Communication Letters, Journal of High Speed Networks.*
- Reviewer (Conferences) - *IEEE INFOCOM 2000, 2001, 2002, 2003, IEEE Hot Interconnects 2001, IEEE Globecom 2000, ACM SIGMETRICS 2003, ISCA 2003, HPSR 2003*

LANGUAGES

Proficient in English, Hindi and Marathi. Working knowledge of Tamil and German.

REFERENCES

FROM ACADEMIA

Prof. Nick McKeown
Associate Professor
Dept. of Electrical Engg. & Comp. Sc.
353 Serra Mall, Room 340
Stanford, CA 94305
Phone: (650) 725-3641
nickm@stanford.edu

Prof. George Varghese
Professor
Dept. of Comp. Sc. & Engg., MS 0114
9500 Gilman Drive, UC San Diego
La Jolla, CA 92040-0114
Phone: (858) 822-0424
varghese@cs.ucsd.edu

Prof. Balaji Prabhakar
Assistant Professor
Dept. of Electrical Engg. & Comp. Sc.
350 Serra Mall, Room 269
Stanford, CA 94305-9510
Phone: (650) 723-6579
balaji@stanford.edu

Prof. Rajeev Motwani
Professor
Dept. of Computer Science
353 Serra Mall, Room 474
Stanford, CA 94305
Phone: (650) 723-6045
rajeev@cs.stanford.edu

FROM INDUSTRY

Dr. Flavio Bonomi
Chief Architect
Cisco Systems
170 West Tasman Dr.
San Jose, CA 95134
Phone: (408) 526-4085
flavio@cisco.com

Mr. Ajit Shelat
Chief Technology Officer
SwitchOn Networks (now PMC-Sierra)
5 Mantri Mansion, Vrindavan Society
Pashan, Pune, 411 008, India
Phone: +91-20-589-8412
ashelat@vsnl.com