

# George Kola

---

(608)695-9486  
kola@cs.wisc.edu  
<http://pages.cs.wisc.edu/~kola>

2280 High Ridge Trl,  
Fitchburg, WI 53713

## CAREER OBJECTIVE

A position in industry or research lab that involves systems design, performance analysis and systems development.

## EDUCATION

<b>PhD, Computer Science</b> University of Wisconsin Madison	Expected Dec 2007
<b>M.S, Computer Science (GPA: 3.93/4)</b> University of Wisconsin Madison	May 2003
<b>B.E, Computer Science and Engineering (Top 5 of class)</b> College of Engineering, Guindy, Anna University, India	May 2001

## SKILLS

- **Programming Languages:** C, C++, Java, bash scripting, awk
- Knowledge of Linux kernel internals, network protocols implementation
- Experience with profiling servers/applications using CPU performance counters

## Graduate Courses

- Advanced Operating System, Distributed Systems, Advanced Computer Networks
- Topics in Database Management Systems
- Advanced Computer Architecture I & II
- Construction of compilers

## PUBLICATIONS

### Book Chapters

Tevfik Kosar, Se-chang Son, George Kola, Miron Livny, **Data Placement in Widely Distributed Environments**, *Grid Computing: The New Frontier of High Performance Computing*, Editor: L.Grandinetti, Elsevier Press, November 2005

## Journal Publications

Tevfik Kosar, George Kola, Miron Livny, **Building reliable and efficient data transfer and processing pipelines**, *Concurrency and Computation: Practice and Experience* 18(6): 609-620 (2006)

## Refereed Publications

- George Kola and Mary K. Vernon, **Target Bandwidth Sharing Using Endhost Measures**, In *Proceeding of Performance 2007*, Cologne, October 2007.
- George Kola and Mary K. Vernon, **QuickProbe: Available Bandwidth Estimation in Two Roundtrips**, In *Proceedings of ACM SIGMETRICS 2006/Performance 2006 Joint Int'l. Conf. on Measurement and Modeling of Computer Systems*, St. Malo, June 2006, pp. 359-360.
- George Kola, Tevfik Kosar and Miron Livny, **Faults in Large Distributed Systems and What We Can Do About Them**, In *Proceedings of 11th European Conference on Parallel Processing (Euro-Par 2005)*, Lisbon, Portugal, August 2005.
- George Kola, Tevfik Kosar, Jaime Frey, Miron Livny, Robert J. Brunner and Michael Remijan, **DISC: A System for Distributed Data Intensive Scientific Computing**, In *Proceedings of the First Workshop on Real, Large Distributed Systems (WORLDS'04)*, San Francisco, CA (December 2004), in conjunction with OSDI'04, December 2004.
- George Kola, Tevfik Kosar and Miron Livny, **Phoenix: Making Data-intensive Grid Applications Fault-tolerant**, In *Grid 2004*, Pittsburgh, PA, November 2004.
- Tevfik Kosar, George Kola, Robert J. Brunner, Miron Livny, and Michael Remijan, **Reliable, Automatic Transfer and Processing of Large Scale Astronomy Datasets**, In *Proceedings of the 14th Astronomical Data Analysis Software & Systems Conference (ADASS 2004)*, Pasadena, CA, October 2004.
- Tevfik Kosar, George Kola and Miron Livny, **Data Pipelines: Enabling Large Scale Multi-Protocol Data Transfers**, In the *2nd International Workshop on Middleware for Grid Computing*, Toronto, Ontario - Canada, October 2004.
- George Kola, Tevfik Kosar and Miron Livny, **Client-centric Grid Knowledgebase**, In *Cluster 2004*, San Diego, CA, September 2004.
- George Kola, Tevfik Kosar and Miron Livny, **Profiling Grid Data Transfer Protocols and Servers**, In *Euro-Par 2004*, Pisa, Italy, August 2004.
- George Kola, Tevfik Kosar and Miron Livny, **A Fully Automated Fault-tolerant System for Distributed Video Processing and Off-site Replication**, In the *14th ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV 2004)*, June 2004.
- George Kola, Tevfik Kosar and Miron Livny, **Run-time Adaptation of Grid Data-placement Jobs**, In *Parallel and Distributed Computing Practices*, 2004.

- Tevfik Kosar, George Kola and Miron Livny, **A Framework for Self-optimizing, Fault-tolerant, High Performance Bulk Data Transfers in a Heterogeneous Grid Environment**, *Proceedings of 2nd Int. Symposium on Parallel and Distributed Computing (ISPDC2003)*, Ljubljana, Slovenia, October 2003.
- George Kola, Tevfik Kosar and Miron Livny, **Run-time Adaptation of Grid Data-placement Jobs**, *Proceedings of Int. Workshop on Adaptive Grid Middleware (AGridM2003)*, New Orleans, LA, September 2003.

### Technical Reports

- George Kola and Miron Livny, **Diskrouter: A Flexible Infrastructure for High Performance Large Scale Data Transfers**, *Technical Report CS-TR-2004-1518, University of Wisconsin-Madison Computer Sciences*, December 2003.
- Tevfik Kosar, George Kola and Miron Livny, **Building Data Pipelines for High Performance Bulk Data Transfers in a Heterogeneous Grid Environment**, *Technical Report CS-TR-2003-1487, University of Wisconsin-Madison Computer Sciences*, August 2003.

## DISSERTATION RESEARCH

### Title: Design, Implementation, Evaluation and Performance Engineering of High Performance Network Transport Protocols

My thesis research involves three parts

1. Design and development and performance engineering of TCP-Madison, high performance transport protocol which uses only end host measures and achieves the following
  - a. High Bandwidth Utilization (>90%)
  - b. Low average backlog ( < 10 packets)
  - c. Negligible loss (<10<sup>-5</sup>, typically 0)
  - d. Fair bandwidth sharing
  - e. Rapid convergence to bandwidth share
  - f. Prevention of congestion collapse
2. Design and development of a user-level platform for network protocol experimentation. The platform gives the same performance results as kernel level protocol implementation and enables protocol experimentation at user level without need for kernel level modification. The platform supports TCP-Madison, FAST TCP, BIC TCP, TCP-Africa and TCP Reno.
3. Design of a quick available bandwidth estimation tool and evaluation of high performance transport protocols over a wide range of well characterized Internet paths.

## EXPERIENCE

### Dissertation Research under Prof. Mary Vernon

Aug 2004 – present

### Graduate Research Assistant in the Condor Project

Sep 2001 to Aug 2004

- Designed, developed and deployed 'DiskRouter', an application level overlay network where nodes use the available disk space to maximize the throughput of data transfers.

Highlights:

- Deployed at 8 organizations/sites
- Supports optical switching
- Used to move 100's of terabytes of real-world scientific data
- Up to 3-times speedup in end-to-end data set transfer time
- Designed solution for end-to-end processing of data intensive applications (astronomy image processing, educational research video processing) that enabled scientists to accomplish their processing in a fraction ( $< 1/5$ ) of the original time in a fully automated manner.

### Research Intern at Compaq Western Research Lab/HP Labs, Palo Alto

Summer 2002

- Characterized the middle tier (Enterprise Java bean) of the 3-tier web workload
- Analyzed and characterized ECPERF which later became SPECJBB
- Feedback to the product group including bug identification and performance enhancement suggestions.

## SELECTED COURSE PROJECTS

### Linux Process Control via the File System

Spring 2002

The Linux 2.4 kernel has only primitive process control features. It does not allow multiple tools to attach to a process making debugging of tools that attach to a process (e.g. Paradyne) extremely difficult. Further, the Linux 'ptrace' system call used to attach to a process changes the parent pointer to that of the tracing process. This causes problems for processes which care about their children. For examples, this causes problems when processes launched by distributed batch schedulers like Condor are traced. In this project, we address most of the Linux process control limitations. Our enhancement allows multiple tracing processes, gives the ability to detach from a stopped process leaving it stopped, enables tracing a subset of system calls, signals, and hardware faults and it does not change the parent pointer. We profile the overhead this enhanced functionality adds to kernel critical path and find that selective system call tracing adds only one extra instruction to the common case of an untraced process.

Performance gains are impressive, improving upon 'ptrace' by up to 374% by allowing fine-grained system call tracing. This modification was done to Linux Kernel 2.4.17. This work was done in a group of two for the CS739 course under Prof. Barton Miller. A company expressed interest in this project and obtained access to the source.

### **An Analysis of iSCSI for use in Distributed File System Design**

Spring 2003

We did a full system characterization of a client using iSCSI. We identified the overheads associated with iSCSI and compared and contrasted that with NFS file access and local disk access. We then tried to tune the system to optimize iSCSI performance and showed the impact of various parameters. Using the insights from this analysis we designed a distributed file system optimized for use with iSCSI and point out the class of applications that would benefit from such a distributed file system.

### **Comparison of Native XML Databases and XML-enabled Relational Databases**

Fall 2003

With the wide-spread usage of XML, it has become necessary to build databases to store the XML content. There are two approaches to storing XML content: one is to have a fully native XML database and the other approach is to add features to existing relational database to support XML. The former approach has been taken by object database vendors and the later approach has been taken by relational database vendors. In this study we intend to highlight some important features of the two approaches and show the cases where one approach performs better than the other.

### **Scaling SMP Machines through Hierarchical Snooping**

Spring 2002

Symmetric Multi Processors (SMPs) have become a runaway success. Scalability is one of the main problems being faced by symmetric multiprocessors. In this work we examine an approach to increasing SMP scalability by introducing a hierarchical bus with transaction filtering to reduce bus traffic. We implemented our filtering mechanism in Simics/Ruby, a full-system simulator and evaluated the effect of this filtering on commercial and scientific workloads. We also discuss the effect of transaction filtering on memory consistency.

### **Fingerprinting the CPU Scheduler**

Fall 2001

Many cool ideas in Operating Systems do not get introduced into commercial systems because they require kernel modifications and users prefer to use out of the box machines and standard kernels and prefer not to tinker with the OS. [Gray Box techniques](#) try to solve this problem by influencing the OS to behave the way that best suits the application needs. For influencing the OS, gray box software should have a good algorithmic and empirical knowledge of the system. Empirical knowledge is gained by

running carefully designed micro benchmarks. In the project, we fingerprinted the CPU schedulers of [Solaris 8](#), [Linux 2.4](#) and [FreeBSD 4.4](#) and used the gained knowledge to implement a user-level *proportional share scheduler*.

## ACHIEVEMENTS

- Admitted to UW Madison graduate school with four years guaranteed aid
- Best technical cadet ‘Gold medal’ in the national level camp organized by National Cade Corps Naval Division (India)
- District first and a state rank in Tamil Nadu Professional Entrance Examinations (These are similar to SAT II subject tests, but are grouped into two groups: Engineering and Medicine. I topped my district in both and secured a state rank (6/100,000+ in Medicine and 12/100,000+ in Engineering).

## TECHNICAL TALKS

- "Profiling Grid Data Transfer Protocols and Servers", EuroPar 2004, Pisa, Italy, Sep 2004.
- "A Fully Automated Fault-tolerant System for Distributed Video Processing and Off-site Replication", NOSSDAV 2004, Kinsale, Ireland, June 2004.
- "Data Pipelines: Real Life Fully Automated Fault-tolerant Data Movement and Processing", Condor Week 2004, Madison, WI, April 2004
- "Run-time Adaptation of Grid Data-placement Jobs", Int. Workshop on Adaptive Grid Middleware (AGridM2003), New Orleans, LA, September 2003.
- "DiskRouter: A Mechanism for High Performance Large Scale Data Transfers", Condor Week, Madison, WI, May 2003

## SELECTED UNDERGRADUATE PROJECTS

Project	No of Persons	Description
Design, Implementation and Performance Evaluation of a transparent Interworking architecture between Jini and CORBA	3	The project allows CORBA clients to invoke Jini services transparently (as if they are CORBA services) and allows Jini clients to invoke CORBA objects transparently (as if they are Jini services).  The paper based on this work has been presented in the poster session of the 7 <sup>th</sup> international conference

		<p>on High Performance Computing. (<a href="http://www.hipc.org/hipc2000">www.hipc.org/hipc2000</a>).</p> <p>Full detail of the work can be found at <a href="http://www.corbajini.8m.com">www.corbajini.8m.com</a></p>
Transparent Object Request Broker	2	<p>The project allows programmers to use remote objects as if they are local objects. A parser scans the code and finds out the remote invocations and links them to the appropriate stubs. At runtime, the runtime system makes invocations to the appropriate server providing the service. In this system, both the server exporting an interface and the client using a remote interface do not have to do anything special. The client and server code are compiled using our system which parses and generates the stubs and skeletons (the server can provide a file specifying the functions it wants to export – this allows the server to restrict some functions from being exported).</p>
Anna University Proxy Server Project	5	<p>A scalable, high-performance proxy server was designed and implemented for use in our university. It has support for load balancing. As of April 2001, the proxy server was running in the university and serving about 300 users.</p>
'C' Compiler	2	<p>A compiler for the C Language was designed and implemented. We also added support for pass by reference (found in C++).</p>
Operating System	1	<p>A small operating system was designed and implemented to improve my understanding of</p>

		operating system concepts.
Robot Controller	2	The Software for controlling the mobile robot (designed during the 'Robotics' elective) was designed and implemented.
Super-Scalar Processor Simulator	2	A processor was designed and implemented (in software). It takes x86 instructions and splits them into micro-ops that are scheduled on a super-scalar RISC core. It uses register renaming to reduce hazards. It is a functional simulator.
Graphics Package	1	A small graphics package implemented in Java that can be used to teach graphics (2d, 3d, animations)
Medical Image Database	2	Oracle Database was used to store medical images and the database was accessed using Java (JDBC). J2EE was used and a web interface was provided to the medical image database. The User Interface had options for a set of image manipulation operations like zooming, scaling, dithering, rotation etc. The database also had options for storing video clips and the UI had a player to play video.

## PERSONAL INFORMATION

**Visa Status:** F1 (Citizen of India)