RESEARCH STATEMENT **Gurpur M. Prabhu**

Over the last several years I have carried out research in the areas of parallel processing and information technology. Some of the problems I worked on are described below. With the advent of Internet-enabling technologies my current interests are to develop semantic-based mechanisms for information retrieval and information processing. While the Internet has connected islands of locally networked computer systems, my focus is on logically bridging isolated islands of data. By making data easy to access and easy to process, I hope to advance the state of the art in the development of information applications for scientific discovery using Grid computing, and decision-making domains such as biological sciences, E-commerce, E-governance, and information analysis for combating crime and terrorism.

Parallel random number generation

An important problem I solved with graduate student Srinivas Aluru was that of generating random numbers for parallel computers. Efficient parallel random number generation is essential for parallelizing the numerous applications that depend on random numbers. Linear congruential generators are easily parallelizable but exhibit regularities that make them unsuitable for a number of applications. We developed an efficient parallel random number generator ($O(1)$ parallel time per generation of one number on each processor) based on the lagged Fibonacci exclusive-OR generator, also known as the generalized feedback shift register generator. We have received requests for this generator from several researchers. Srinivas Aluru subsequently proved that the Exclusive-OR lagged Fibonacci generator is the only one that can be efficiently parallelized, and the additive and multiplicative lagged Fibonacci generators cannot be.

The N-body problem

Another problem that Aluru and I worked on is the N-body problem. The N-body problem consists of simulating the motion of N particles under the influence of mutual force fields based on an inverse square law. This problem has applications in several domains including radiosity methods in computer graphics, astrophysics, molecular dynamics, fluid dynamics, and numerical complex analysis.

Research efforts have focused on reducing the $O(N^2)$ time per iteration required by the naïve algorithm of computing each pairwise interaction. Notable algorithms that have gained wide popularity are the Barnes-Hut and Greengard algorithms. The fastest algorithm is by Greengard, whose Ph.D. thesis won the ACM distinguished dissertation award in 1988. Greengard’s algorithm claims to compute the pairwise interactions in $O(N)$ time per iteration.

We proved that Greengard’s algorithm is not $O(N)$ as claimed. The argument is based on the choice of problem precision as a constant in Greengard’s analysis. Both Barnes-Hut and Greengard’s methods use a data structure whose size depends on the spatial distribution of the particles. This makes it impossible to establish bounds on these algorithms without assumptions of the particle distributions. We have shown that performance claims of these algorithms are invalid even with particle distributions resulting in the smallest running times. A journal paper on this result was published in the *SIAM Journal of Scientific Computing* along with papers in Supercomputing ’94 and the Journal of Supercomputing in 1998.

Interdisciplinary Efforts

Two of my doctoral students, Naresh Nayar and Milton Wikstrom, did their work in computational science by using high-performance techniques in the areas of weather modeling and computational chemistry. I mentored Prof. Charles Wright’s student, Diane Rover, who worked at the Scalable Computing Laboratory and her work on visualization of program performance and the SLALOM benchmark have received national research awards. Another recent doctoral student, Rajat Todi, used the HINT benchmark to evaluate file access patterns for realistic I/O workloads in cluster environments.
Along with doctoral student Babak Forouraghi, I studied the design of learning algorithms for flaw classification of materials used in nondestructive evaluation. Babak Forouraghi, Prof. Lester Schmerr and I developed a novel technique for multiobjective optimization that combined concepts from fuzzy logic, machine learning, and multivariate statistics. Our research papers from this work have appeared in the AAAI national conference on artificial intelligence, the international conference on fuzzy systems and the world congress on expert systems.

Information and Internet Technologies

During 1990 to 1992 I was the principal investigator in an industrial grant from Next Generation Management Systems. The task was to propose an integrated information technology solution for their distribution operations. Dr. Sree Nilakanta of the College of Business was a co-investigator with me and we examined various technologies and recommended an enterprise integration architecture with an open systems software foundation.

Current and Future Projects

At the current time I am involved in Internet-enabling technologies and intelligent distributed computing. Much of the existing work is protocol- and syntax-based. My interest lies in exploring semantic-based search techniques for information retrieval and processing. Distributed Computing is an all-encompassing concept, consisting of, but not limited to, metadata, application logic, interfaces, performance management, database integrity, middleware technology, and recent trends such as autonomic computing, pervasive computing, and proactive computing. The distributed computing model has distributed data sources on one side and distributed users of data on the other side. In order to be effective and efficient, there are three kinds of problems that must be addressed:

- (1) Data and the processing logic are provided entirely by each of the data sources for use by the large set of distributed users.

- (2) Data alone is provided by the data sources but the processing logic is provided by the distributed users.

- (3) The data processing logic provided, partly by the data sources, *and* partly by the distributed users, should work cooperatively through effective, intelligent communication in a secure and scalable environment.

Problems (1) and (2) listed above have been solved elegantly by Semantic Bridge that was designed and developed during 1999 – 2002. This approach employs user-defined *programs* in place of user-defined queries, thereby combining the advantages of the RPC and Query Interface mechanisms. The objectives of Semantic Bridge are to:

- Enable the development of abstract, database schema/implementation-neutral programs to meet the specific needs of remote users.
- Enable the development of abstract, database schema/implementation-neutral programs to service standard, universally accepted semantics such as standard Business Reports, without requiring the data sources to provide the implementation.

The above objectives are accomplished without compromising either the data security, or the system security of the hosts. The main advantages gained are scalability and maintainability.

Self-Aware Computing

The next goal of my research is to address Problem (3) in the above list – to design a system where the data processing logic provided by the data source works cooperatively with the data processing logic supplied by the distributed user. An additional goal of the project which is a direct consequence of the proposed approach is to design a system of writing programs that make computers aware of the actions they take. This will enable computers to dynamically interact with other computers in complex decision-making scenarios, and empower them to make intelligent decisions. The ultimate aim is to enable computers to understand the meaning of computations, communicate with other computers, and act sensibly without the need for external human or program intervention in a wide variety of situations. To achieve this goal, the following objectives have been set to develop the structural framework needed to design intelligent distributed systems:

(1) A mechanism to capture through modeling, a computationally acceptable notion of semantics of every external input which can be utilized by the parallel execution of a 'self-aware' process.
(2) To develop a general decision-making algorithm for any problem-solution space for the parallel self-aware process based on this semantic information and a parameterized value system defined by the developer.

(3) To develop extensions to Semantic Bridge to allow external programs to interact with “intelligent” data sources.
(4) To develop a complete set of tools for implementing systems using the above architecture for designing intelligent systems.
(5) To investigate the performance of the intelligent system in different situations for a given domain.

A number of theories and models from the disciplines of cognitive science, linguistics, psychology, and computer science are used to address the above objectives. Specifically, we use Perceptual Control Theory (PCT), Affect Control Theory (ACT), Head-driven Phrase Structure Grammar (HPSG), Aspect Oriented Programming (AOP), Dooyeweerd’s Aspects, Jackendoff’s Conceptual Structures, Category Theory, Answer Set Programming (ASP), Dialog Game Board, and Semantic Bridge to design and implement our intelligent computer system.

I have been the major professor of 7 Ph.D. students and 31 Masters students and published over 45 papers in journals and conferences. I have served on the graduate committees of 37 Ph.D. students and 111 Masters students. I consider the mentoring of graduate students as a very important facet of research. My role includes assisting students in identifying good research problems, suggesting lines of approach, and emphasizing the need to maintain high scientific standards. In this process, I try to lead my students towards those areas where their strengths lie; recognizing where these strengths are is a gradual process that requires continual re-evaluation. My job as an advisor includes teaching my students how to write technical papers. This is a time-consuming task, often as time-consuming as doing the research itself. I painstakingly revise successive drafts, providing suggestions for improvement while resisting the temptation to write the papers myself.

In many cases I have not included my name on the published paper, allowing students to get motivated by writing and publishing a paper by themselves. I have also followed this principle for students on whose graduate committees I have served on. Although I have contributed several ideas on the computational aspects of the work of interdisciplinary Ph.D. and Masters students, I have resisted from including my name on the final publications arising from their work. I believe this motivates students and instills in them the confidence to pursue research careers in academia or industry. My advising extends to the early stages of career development, where I counsel students on the options available to them.

Since 1994 I have been in charge of the Undergraduate program, carrying out many duties that are normally what an associate chair would do in many departments. From 1998 to 2002 we underwent a transition in leadership with two interim chairs and I took the initiative to be in charge of accreditation and other aspects of the Undergraduate curriculum to provide some stability to our program. The responsibilities and duties of this role took time away from proposal writing and to some extent my research activities. But with my efforts having succeeded in creating a quality undergraduate program worthy of continuous accreditation, I have now started focusing more on research.