RESEARCH STATEMENT
(A Self Evaluation with a Statement for Future Work)

Weizhen Mao
Department of Computer Science
College of William and Mary

During my academic years at the College of William and Mary, I have fifty published and peer-reviewed papers, among which fifteen are journal papers and thirty-five are conference papers. I have acquired three external research grants funded by the National Science Foundation, with one recent and two others from a while back. I have graduated three Ph.D. students, with the fourth passing the proposal exam stage.

The focus of my research is on the design and analysis of algorithms, especially online, approximation, and randomized algorithms. I mainly choose the problems of study from real-world applications with an emphasis on resource allocation, i.e., problems that involve allocating scarce resources among numerous requests to optimize performance of systems and quality of services.

Although I often use similar approaches and techniques in my research, I have worked on a variety of problems from different areas in computer science. Here, I summarize my research in four areas: wireless ad-hoc networks, job scheduling, network topology for routing and partitioning, and some other work.

Wireless ad-hoc networks

Wireless ad-hoc network communication technologies have experienced a rapid growth in research and significant advancements in recent times. As a result, new resource allocation problems arise with the goals of maximizing throughput and utilization and minimizing energy consumption and time cost. The study of such networks, particularly RFID networks, sensor network, vehicular and body sensor networks, has become a major focus of my research in the past few years. The projects that I have been involved in this area are all collaboration with my colleagues and some graduate students in the department.

- **RFID networks:**
  (C33/INFOCOM10, C32/INFOCOM10, C28/MobiHoc08)

  Radio Frequency Identification (RFID) technology allows a reader to read information from a tag attached to an item from a distance. We have studied several very important applications related to warehouse inventory control. We have shown that some rather complicated tasks for RFID, such as maintaining an inventory in a warehouse when the inventory changes dynamically, counting the number of tags in a changing environment where tags come and go, and finding popular categories among a large number of tags, can all be carried out in an efficient way. In particular, we have designed randomized algorithms to tackle these problems and proved rigorously for each case that our algorithms achieve high time efficiency and low probability of errors. This work on RFID systems is supported by a recent NSF grant.
• Sensor networks:
  (J14/TMC10, C30/WASA09, C25/MobiHoc07, C24/WASA07, C22/MobiHoc06)

Sensor networks are deployed for pervasive computing, such as sensing environmental conditions, where a large amount of data are collected by the sensor nodes and then sent back to the base station for processing and analysis. We have studied various resource allocation problems arising from sensor networks. Specifically, we have designed exact, heuristic, and approximation algorithms (based on dynamic programming, randomization, and rounding technique, respectively) for strategically placing the more powerful but expensive storage sensor nodes in a sensor network to minimize the energy consumption incurred during the transmission of data back to the base station. Also, we have designed a histogram-based algorithm for detecting outlier data collected by sensor nodes to reduce communication cost. In addition, we have studied the dynamic channel assignment problem for load balancing in sensor networks.

• Vehicular and body sensor networks:
  (J15/TPDS11, C35/INFOCOM11, C31/ICNP09)

Two recent projects that I have been involved in are about the applications of wireless technology in vehicular networks and body sensor networks. In one project, we have studied the problem of associating road-side WiFi access points with mobile users in the Drive-thru Internet scenario. We have designed and evaluated algorithms through several techniques to improve efficiency and fairness for all users. In the other project, we have proposed a novel and radio-agnostic solution for heterogeneous body sensor networks, which allows different data streams to request different throughput and time delay assurances with reduced communication overhead.

Job scheduling

The area of job scheduling lies in the intersection of Operations Research, Mathematics, and Computer Science. This area is especially intriguing to me since it contains a wealth of resource allocation problems. My first research project was on the so-called Liu’s conjecture on the study of preemption versus nonpreemption of jobs, introduced to me by my Ph.D. advisor. Ever since then, I have worked on a variety of scheduling problems, from which I select a few to discuss below.

• Scheduling malleable jobs with overhead:
  (J13/EJOR08, C29/NAS08, C27/PDCS07, C19/ICCSI03)

Malleable parallel jobs can distribute their workload among any number of available processors in a parallel computer in order to reduce their execution time. In contrast, nonmalleable parallel jobs must use a fixed number of processors. The ideal execution time of a malleable parallel job with length $p$ is $p/k$ if it utilizes $k$ processors. However, inherently serial code and parallel processing overhead (from process management, shared memory access and contention, communication, and/or synchronization) often prevent actual execution times from achieving this ideal. It is natural to consider this extra time as a type of setup time, a term commonly used by the scheduling community.
We have derived an execution time function that takes both speedup (i.e., $p/k$) and setup time into account for the parallel execution of malleable jobs. We have tested the validity of our mathematical model through numerous experiments on large parallel systems. We have worked on various algorithms that can efficiently schedule malleable parallel jobs online under the proposed model through competitive analysis.

- **Scheduling jobs with lookahead:**
  (C18/ICCSI03, C17/ICSSI02, J3/C&OR94)

  Online algorithms are popular choices for job scheduling in the realtime environment. However, since online algorithms act without any knowledge of the future, the quality of the schedules given by such algorithms are sometimes not satisfactory. We have proposed a lookahead model for job scheduling, which is online in nature but also allows limited lookahead into the near future to examine the next few in-coming jobs. A fundamental issue about this model is whether lookahead can indeed improve the performance of an otherwise online algorithm. We have incorporated the lookahead capability in various scheduling problems, such as single machine scheduling, multiple machine scheduling, and unrelated machine scheduling. We have also evaluated the benefit of lookahead by the rigorous competitive analysis method and by simulation.

- **Scheduling multi-operational jobs on multiple machines:**
  (C13/ICCSI98, C9/SAC97, C4/HPCN95)

  In traditional shop scheduling, the execution of a job goes through each of the multiple shops (stages) in an arbitrary order or a fixed order, and it is assumed that there is only one machine available in each shop to perform the task. In this project, we have considered an extended model, in which there are multiple parallel machines available in each shop to execute a job’s certain task. In addition, we have included a constraint that disallows certain machines from different shops to be assigned to execute a job’s tasks. We have studied the open shop and flow shop problems in the multi-machine model and proposed optimal and online algorithms to solve the problems.

**Network topology**
(C23/PDCN07, C21/HPC-Asia05, J11/DAM03, J9/PPL96, J8/JOC96)

Many multi-computer systems of parallel processors are connected in network topologies of 2-D meshes, 3-D tori, and, in higher dimensions, n-D tori that are also known as $k$-ary $n$-cubes. Such systems are commonly studied for two purposes: finding routes for efficient communication and creating partitions of the system for job scheduling. For the issue of routing, we have proposed an isomorphic routing model that fits the SIMD architecture and obtained complexity results for several problems within the model to minimize the routing makespan in a 2-D mesh. We have also devised a collision-free routing scheme for one-to-all personalized communication in 3-D torus networks. For the issue of partitioning, we have designed efficient algorithms to partition a multi-dimensional torus of processors into subtori in such a way to produce high-quality job schedules. Further, we have studied the combinatorial properties of the more general $k$-ary $n$-cubes and utilized these properties to construct lower-bounding functions in branch-and-bound partitioning algorithms.
Other work
(C15/ISPAN00, J10/Networks99, C10/CISS97, J7/LS95, J1/JOC93)

This collection of work do not precisely belong to any of the above research areas, but are all about online algorithm design and analysis, a methodology that has been heavily employed in many projects discussed earlier in this document. The problems studied include on-demand data broadcast scheduling, virtual circuit routing, file transfer routing and scheduling, mobile server scheduling, and the classic problem of bin packing. Since these problems often arise from applications that require an immediate decision for processing each data item in the input sequence, online algorithms are obvious choices to tackle them. In addition to the design of efficient online algorithms, we have mainly focused on the evaluation of the algorithms’ performance through the method of competitive analysis, which is a technique aiming to find the worst-case ratio between the solution produced by an online algorithm and the optimal solution.

Future plan

In the long run, I will continue my search and identification of challenging resource allocation problems from new applications and take theoretical approaches to study their complexity and algorithms. In the short run that I am able to foresee, I plan to work on three projects, among which one represents a new direction and two are continuous work of recent projects.

• Multi-core scheduling on chip multiprocessors:
  With the wide adoption of chip multiprocessors (CMP), processor industry has shifted the focus of growth from clock frequency to processor parallelism and concurrency. This new trend has prompted an important question of how to schedule programs to cores residing on chips in CMP to effectively alleviate contention of resource, such as shared cache and memory bandwidth. We have noticed some resemblance between multi-core scheduling and the traditional job scheduling. We are interested in the feasibility of applying job scheduling methodology to the multi-core platform, where many resource allocation problems can be formulated and investigated in the direction of problem complexity and algorithm analysis.

• Wireless ad-hoc networks:
  We will continue our work in wireless ad-hoc networks. For RFID systems, we plan to investigate a few new topics, including making RFID reading protocol more efficient and preserving privacy. We also plan to further our previous work in body sensor networks as there are still a few lingering issues on algorithm performance that have not been resolved.

• Parallel job scheduling:
  We plan to complete our current project of malleable job scheduling. We have defined a new online algorithm for our proposed setup time model and have obtained adequate experimental data to show the better performance of the algorithm than some previous ones. Although we have got some partial results on the competitive ratio of the algorithm, our goal is to prove a tight bound of the ratio.