**Internet and Grid-Based Computation Systems and their Applications**

Fall 2005

Instructor - J.C. Browne

The Internet is providing the basis for a new generation of systems for computation and for analysis, sharing and transport of data. Thomas Friedman's book, "The World is Flat" shows how the existing capabilities for global scale collaboration and interaction have changed the world. The second generation of "Flattening of the World" requires algorithms and processes for automating cooperation and collaboration among groups (humans and machines) with common interests based on intelligent, adaptive and robust distributed control. This seminar is about both first generation and second generation algorithms and processes for "Flattening the World."

Major computer companies (IBM, Microsoft, Sun, etc.) are developing or adopting systems for implementation of Internet-based systems and services. IBM has some of its laboratories and development organization for grid and web services and computing on demand in Austin. Microsoft's .Net initiative is focused on development of Internet-based systems. Sun and HP also have Grid/Internet based computing initiatives.

There are several models of computation for Internet-based systems. Examples of such systems include Globus which is usually used to couple of a small number of very large-scale parallel machines with a program based on MPI. Globus is essentially a resource management system. Legion (now called Avaki) is a commercially supported system for distributed management and control of data and computations. Another commonly used system is Condor which utilizes "farms of workstations." Condor is also a resource management system. A similar (to Condor) commonly used model is SETI@home where a very large number of personal computers are utilized in a loosely-coupled mode. There are languages and systems based on so-called coordination models and languages. There are also internet-based systems for collaborative computations. Alternative models of Internet-based computation including peer-to-peer organization of computations where computations are organized as systems of autonomous but coordinating agents or systems are emerging as a major topic of research. A critical challenge is to integrate these models of distributed computation. See for example, Parashar and Browne[1].

There are many open research topics in Internet and Grid-based computation. Strategies and protocols for management of heterogeneous resource systems and computations on them is a central issue. Models and languages for formulation of internet-based systems is another central issue. Adaptation of computations to diverse and dynamic execution environments is largely unstudied. Fault-tolerance and robustness on the Internet scale is primitive.

This seminar will be a study of current research on the models, languages and systems for formulating Internet-based systems. The course material will be papers from the current

literature. Examples of papers and lectures can be found on the web page for the Fall 2002 offering of this seminar. http://www.cs.utexas.edu/users/browne/CS395Tf2002/ Lectures will be given by the instructor and by guest lecturers including representatives of TACC, and commercial firms which have developed software for implementation of Internet based applications. Students participating in the seminar will also give presentations on their projects.

Each student participant will be responsible for a project which can be evaluation of the state of the art in one research topic, evaluation of one system for development of Internet-based applications or study of one of the many open and interesting problems in Internet scale systems and computations. The evaluations development systems for applications will be based on study of the conceptual basis for each system, applications of the system and implementation of and measurement of the behavior of an application implemented in the system. One project from the Fall 2002 offering of this seminar will be presented at the 2003 HPDC and published in the Proceedings of that conference. Examples of recent projects are at http://www.cs.utexas.edu/~browne/cs395f2003/.

The class is open to any graduate student in Natural Sciences or Engineering with an interest in internet-based computational systems but an interview with the instructor is recommended before enrollment for students outside of Computer Sciences and Electrical and Computer Engineering. The instructor welcomes inquiries about the course (browne@cs.utexas.edu). The Appendix gives the instructors perspective on the importance of this topic.

[1] Conceptual and Implementation Models for the Grid," M. Parashar and J.C. Browne, *Proceedings of the IEEE, Special Issue on Grid Computing*, IEEE Press, Vol. 93, No. 3, pp 653 – 668, March 2005.

## Appendix A

## "Flattening of the World" – The Next Generation

Jim Browne
Department of Computer Sciences
University of Texas

Friedman's book, "The World is Flat" shows how the very limited capabilities for global scale collaboration and interaction have changed the world. The capabilities which Friedman cites are basic and fairly primitive applications of distributed systems. Collaborations and interactions are largely human and central server mediated. There is little real collaboration among machines or among humans and machines. The second generation of "Flattening of the World" requires algorithms and processes for automating cooperation and collaboration among groups (humans and machines) with common interests based on intelligent, adaptive and robust distributed control.

The grand challenge is to generate and ***apply*** algorithms and processes based on distributed control and which tolerate, heterogeneity, faults and dynamic resource configurations, for a wide spectrum of analytical, computational and collaborative tasks. The bases for formulation of these algorithms and processes begins with fundamental theory of distributed systems and should include network theory, existing parallel and distributed systems, artificial intelligence and machine learning and robotics, game theory, evolution and applications. The technical challenge is to synergistically combine knowledge from multiple disciplines to overcome the intrinsic difficulty of consensus, control and collaboration in distributed systems.

There are significant data points which suggest at least some possibility that a paradigm shift can emerge. Effective and efficient basic computation of Google pageranks with fully distributed control has been demonstrated [1]. The SDIMS [2] abstraction of Yalagandula and Dahlin has demonstrated the possibility of collective operations over thousands of distributed sites. There are plausible simple applications such as extending the SETI@home computational model to general distributed computations and distributed implementations of data analyses which can serve as initial testbeds.

The importance to society is the possible emergence of new algorithms and processes for analysis, computation and collaboration which will enable global scale data sharing, analysis and computation. This research challenge, which if successful will lead to automation of important collaborative processes, can in its boldest statement, be thought of as a significant element in next generation of flattening of the world.

The importance to the distributed systems research community is that this research challenge will make distributed systems a central discipline across all of science and engineering.

References

[1] Karthikeyan Sankaralingan, Madhulika Yalamanchi, Simha Sethhumadhavan and James C. Browne “Page Rank Computation Key Word Search on Distributed Systems in Peer to Peer Networks,” Grid Computing, Volume 1, No. 3, pp. 291-307, (2003).

[2] Praveen Yalagandula and Mike Dahlin, *SDIMS: A Scalable Distributed Information Management System* ACM SIGCOMM 2004, Portland, Oregon.