

ISTeC Retreat on the Scientific and Engineering Foundations of Information Science and Technology

Saturday, May 8, 8:15am to 3:00pm, Lory Student Center

Report on Large-Scale Networks for Sensing, Computing, and Communicating

Participants

Chair: Edwin Chong (ECE, Math)

Participants: Ron Butler (Statistics), Tom Chen (ECE), Jan Hannig (Statistics), Anura Jayasumana (ECE, CS), Sanjay Rajopadhye (CS, ECE), Colleen Webb (Biology, Math), Darrell Whitley (CS), Fang Yao (Statistics)

Theme

Our session focused on topics broadly related to large-scale networks, by which we mean large numbers of entities existing in a medium that supports interaction. Such entities include computers, sensors, software agents, and biological organisms. The kind of media that support interaction could be wireless, the Internet, or ecological. Such networks arise in a variety of application areas: the Internet, wireless and ad hoc networks, sensor networks, large computing grids, and biological systems. We focused on the issues of modeling, analysis, design, optimization, control, and construction of large-scale networks. We were also interested in bridges with the other two sessions, including large data sets associated with large-scale networks, and the modeling of large computing grids.

Research Issues

Many research issues arise in the study of large-scale networks. Some are theoretical and others are practical. Here, we give a sampling of some research issues discussed in our session.

- How do we model very large networks? Depending on the type of network and the kinds of questions we wish to answer about the network, different modeling techniques can be brought to bear on the problem. For example, we may be interested in network performance metrics like throughput and delay. Alternatively, we may be interested in the dependence of “robustness” properties on system parameters. We describe some modeling techniques in the next section.
- In large networks of sensors, a key problem that arises is sensor management. Here, we are interested in how to control the sensors so that we optimize some objective criterion. For example, we may be performing the task of tracking a target, and we are interested in deciding, in real time, which sensors to activate to trade off tracking performance for sensor usage costs. This problem can be

formulated as a Markov decision process. However, the state-space in this case is typically too large to apply standard methods for computing an optimal policy.

- Researchers in biological systems are interested in networks that adapt and evolve. For example, we may be interested in how selection at the individual level affects the extinction or persistence of populations or ecosystems. Such questions are also of interest in engineered systems, such as the Internet; however, adequate models for answering them in the context of the Internet have yet to be developed.
- Some models that have traditionally been used in the context of communication networks are now being applied to biological systems. For example, some recent work on ion channels have made use of models of stochastic networks with charged customers.
- In the design of large networks, how important is it to optimize the design before building the system? Some have argued that the performance of the Internet could have been vastly improved if more understanding and optimization took place before it was built. On the other hand, some believe that the ubiquity and success of the Internet is the result of the flexibility and the lack of tight restrictions in its design and deployment.
- Societal values play a central role in engineering and ecology, but are not easy to incorporate into research studies. Although models of social behavior are available, they are not easily integrated with models for systems like the Internet

Technical Approaches

A vast spectrum of approaches exists to deal with large-scale networks. These may be characterized as follows.

- **Analytical methods.** Such methods have been used for many years in the study of networks. CSU has a strong group of researchers with expertise along these lines, including Ron Butler (frequency domain methods), Edwin Chong (communication networks), Jan Hannig (stochastic processes), Fang Yao (functional data analysis), and Colleen Webb (theoretical evolutionary ecology).
- **Computer simulation and numerical methods.** As networks increase in size and complexity, analytical methods lose their power in providing a complete characterization of their behavior. The use of computer simulation as a modeling tool is becoming increasingly common. CSU's research expertise along these lines includes Darrell Whitley (simulation of evolutionary search), Edwin Chong (Monte Carlo sampling methods for Markov decision processes), Anura Jayasumana (Internet performance analysis), Sanjay Rajopadhye (automatic parallelization, high performance computing), and Don Estep (numerical solution of PDEs).
- **Hardware emulation.** This approach provides the possibility of speeding up software simulation models by building hardware "emulation" devices. CSU's expertise here includes Tom Chen (VLSI systems), Sanjay Rajopadhye (silicon compilation, reconfigurable computing), and Wim Bohm (embedded systems). A question that arises in the use of hardware emulation in dealing with large-scale networks is whether or not a speed-up in simulation capabilities alone significantly impacts our ability to deal with large-scale systems. The point was

raised that increases in computation speed, while temporary, often does give rise to insights not readily available otherwise.

- **Continuum Models.** Networks have commonly been treated as discrete collections of entities. As networks scale to the size of hundreds of millions (such as the Internet), it becomes increasingly difficult to describe networks in terms of discrete entities. Modeling networks as a “spatial continuum” provides a way to model very large networks at an aggregate level. Tools for such continuum models include partial differential equations (PDEs), which have been applied in biology but not so much in engineering (to model networks like the Internet). We will have more to say about this below.

Funding Sources

All the mainstream federal funding agencies support work in the area of large-scale networks: NSF, DARPA, ARO, ONR, and AFOSR. Indeed, many of the researchers present at the session already have funding from these agencies.

Collaborations

Even though the participants in our session come from a diverse range of research disciplines, several natural points of collaboration emerge. Although collaborations should not be mandated, they can and should be facilitated. Such facilitation can take several forms. The availability of seed funding for speculative collaborations is one example. Another example is a close examination of our reward structure, which discourages speculative collaborations, especially for junior faculty.

Below, we will describe one particular collaborative effort between three researchers from different departments.

An Example Collaborative Effort: Continuum Models for Large Networks

Overview

We describe this project as an example collaborative effort within the area of large-scale networks. This effort brings together researchers from three departments: Don Estep (Mathematics), Jan Hannig (Statistics), and Edwin Chong (ECE).

This project is concerned with the modeling and simulation of very large networks using time-dependent partial differential equations. In many applications, numerical simulation is the tool of choice for the design and evaluation of large networks. However, the computational overhead associated with direct simulation severely limits the size and complexity of networks that can be studied in this fashion. Performing numerical simulations of large networks has been widely recognized as a major hurdle to future progress in understanding and evaluating large networks.

Our thesis is that global characteristics (what might be called “macroflow behavior”) of networks consisting of upwards of millions of components can be captured by continuum modeling. Continuum models consider the behavior of the components on the scale of the aggregate rather than of the individual, while the individual characteristics of the components enter the model through the form of the equations and the parameters defining the model. Continuum modeling of large networks by partial differential

equations would provide a powerful way to deal with the number of components in large networks, and would open up the use of highly developed technologies such as adaptive mesh refinement in the finite element method.

Motivating Problem Domains

Below we describe several examples to illustrate the need for new approaches to modeling large networks.

- **The Internet.** Simulation modeling has become the primary tool in the performance analysis of the Internet, with large research efforts directed toward enabling simulation of increasingly larger networks (e.g., DARPA's Network Modeling and Simulation program). Such simulation tools are used in a variety of ways, including in the evaluation of protocol designs and for network resource provisioning. High-performance computing techniques and technologies, such as parallel computing, have been exploited to advance the frontier of network simulator performance. Yet, the current state-of-the-art in simulation tools does not support the simulation of networks with a hundred million nodes, roughly the current size of the Internet. Indeed, a recent news release from Georgia Tech reports the creation of “the world's fastest detailed computer simulations of the Internet,” capable of simulating “networks from over 1 million web browsers in near real time.” At numbers exceeding millions of nodes, we should expect continuous approximations in the topological domain to provide appropriate models of the network, at least at some levels of abstraction.
- **Wireless ad hoc networks.** Interest in wireless ad hoc networks has flourished in recent years. At the same time, the analysis of such networks has proven to be challenging. A seminal paper by Gupta and Kumar (200) show that under certain modeling assumptions, it is possible to characterize the *scaling law* of the transport capacity in the asymptotic regime where the number of nodes grows to infinity. Roughly, they show that the throughput grows as $O(\sqrt{n})$. This result has spawned numerous papers that characterize asymptotic scaling laws for wireless networks in different settings. Despite the groundbreaking nature of these scaling-law results, they do not provide any means to calculate the *actual* throughput of a network with, say, a million nodes. Simulation remains the only method of answering such questions. However, as is the case in Internet simulation, simulation tools do not currently support network sizes of millions of nodes.
- **Sensor networks.** The possibility of networking not only computers and communication devices but also sensors has led to significant current interest in sensor networks (e.g., NSF's Sensors and Sensor Networks Program). Like wireless ad hoc networks, sensor networks consist of sensors connected together via wireless links. However, sensor networks exhibit some unique features, including limitations in energy, bandwidth, and processing power. Analysis of such networks has closely followed that of wireless ad hoc networks, characterizing asymptotic scaling laws as the number of sensors grows. The performance of large-scale sensor networks is still not possible with state-of-the-art simulators.

- **Other Network Problems.** Although our main motivation stems from problems in communication networks, we believe that other problems areas have similar needs, and would benefit from our work.
 - Dense traffic networks: One example is the simulation of dense traffic networks. This problem has become important in recent years because of the development of intelligent highway systems. Most recently, traffic simulation has become an important tool in the assessment of the impact of terrorist attacks in dense metropolitan areas. The state-of-the-art in traffic network simulation appears to be consistent with that of communication networks. The techniques that are brought to bear in advancing the state-of-the-art also appear to be similar: e.g., parallel processing and fluid traffic models. We believe that our work on continuous models for network simulation will be relevant to the development of simulation tools for traffic networks in dense metropolitan areas.
 - Network disease models: In 2001, over fifty million head of livestock were transported across state boundaries in the United States. Pathways over which livestock are moved form a complex network, with important and unimportant nodes. Such a network system is vulnerable to disease outbreaks that can spread more rapidly than traditional diffusion models allow. Individual states have management plans for dealing with disease outbreaks, but coordination across states is inconsistent and transport modeling has not been used to assess responses. There is a critical need to strengthen our capability to model the network for livestock transportation and how diseases can be spread through this system.