Towards a Theory of Networked Computation

Joan Feigenbaum
http://www.cs.yale.edu/homes/jf

Michael Mitzenmacher
http://www.eecs.harvard.edu/~michaelm

JULY 2009

Supported by NSF grant CCF-0601893

Executive Summary

The increasing prominence of the Internet, the Web, and large data networks in general has profoundly affected social and commercial activity. It has also wrought one of the most profound changes in Computer Science since its inception. Traditionally, Computer-Science research has focused primarily on understanding how best to design, build, analyze, and program computers. The research agenda has now expanded to include the question of how best to design, build, analyze, and operate networks. How can one ensure that a network created and used by many autonomous organizations and individuals functions properly, respects the rights of users, and exploits its vast shared resources fully and fairly?

The Theory of Computation (ToC) community can help address the full spectrum of research questions implicit in this grand challenge by developing a Theory of Networked Computation (ToNC), encompassing both positive and negative results. ToC research has already evolved with and influenced the growth of the Web, producing interesting results and techniques in diverse problem domains, including search and information retrieval, network protocols, error correction, Internet-based auctions, and security. Moreover, the ToC community's influence extends into the commercial IT sector, where algorithmic ideas have contributed in important ways to major companies, including Google and Akamai.

A more general Theory of Networked Computation could influence the development of new networked systems, just as formal notions of "efficient solutions" and "hardness" have influenced system development for single machines. To develop a full-fledged Theory of Networked Computation, the ToC community will build on its past achievements both by striking out in new research directions and by continuing along established directions.

Two NSF-sponsored workshops were held during the Spring of 2006 in order to flesh out the ToNC-research agenda [ToNC]. This report contains the results of those workshops. In it, we describe the state of the art of networked computation, some general research themes that constitute the heart of the ToNC scope, specific open problems in ToNC (not an exhaustive list of such problems, but enough to support our claim that progress can be made in this important area by a large segment of the ToC-research community), important issues that cut across multiple research themes, and recommendations for institutional support of ToNC research. Highlights of the report are given here in the Executive Summary, and details can be found in the following sections.

Research Goals

Workshop participants identified three broad, overlapping categories of ToNC-research goals:

- **Realizing better networks:** Numerous theoretical-research questions will arise in the design, analysis, implementation, deployment, operation, and modification of future networks.

- **Computing on networks:** Formal computational models of future networks will enable us both to design services, algorithms, and protocols with provable properties and to demonstrate (by proving hardness results) that some networked-computational goals are unattainable.
- **Solving problems that are created or exacerbated by networks:** Not all of the ToNC-research agenda will involve new computational models. The importance of several established theoretical-research areas has risen dramatically as the use of networked computers has proliferated, and some established methods and techniques within these areas are not general or scalable enough to handle the problems that future networks will create. Examples of these areas include massive-data-set algorithmics, error-correcting codes, and random-graph models.

We briefly give the flavor of each category here. Sections II, III, and IV below flesh out in detail the broad-ranging research agenda developed at the workshops [ToNC].

Like today's Internet, future networks may be characterized by massive scale, subnetwork autonomy, user self-interest, device heterogeneity, and/or emergent behavior. Given our limited ability to model, measure, predict, and control today's Internet, we will need a more principled approach if we are to "realize better networks." What are the right primitives and abstractions with which to study networks? Is "layering" fundamental, and, if so, what is the optimal set of layers? How should responsibility for essential network functions be assigned to various network components? How should state be allocated among components? What should the relationships be among naming, addressing, and routing; indeed, which objects in the network should have names that are meaningful network-wide? In the systems-research community, these questions are representative of "network-architecture" research. From a ToC perspective, these are the type of questions that must be answered in the process of formally defining various types of networks and rigorously formulating models of networked computation.

With one or more precise definitions of "network" in hand, it will be natural to ask what can be "computed on a network" and how efficiently computations can be done on a network. The Web-searching problem domain perfectly exemplifies both the evidence that networked computation can be tremendously powerful and the tough challenges that lie ahead if it is to be improved. Search engines that handle billions of Web pages and support a dizzying array of economic, scholarly, and social activities are remarkable technological achievements. On the other hand, numerous technical problems (including many of an algorithmic or combinatorial nature) will have to be solved if we are to have "personalized search" (which strongly implicates privacy), defenses against "Google bombing" and other adversarial or strategic behavior by webpage owners, the ability to search for video or audio clips as well as keywords, and many other search capabilities that users clearly want. The existing bodies of theory on parallel and distributed computing may provide partial answers to the questions of what can be "computed on a network" and how efficiently, but the massive scale, subnetwork autonomy, user self-interest, device heterogeneity, and emergent behavior that characterize present and future networks are not satisfactorily dealt with by either of these existing theories.

More generally, a formal complexity-theoretic approach will enable investigation of the *inherent power and limitations* of networked computing. Notions of "resources" and “reductions” will allow us to determine which fundamental networking problems are easy, which are hard, and why. One approach to the development of “complexity theory of networked computation” is the *black-box channel* approach (described in Section III below). In this model, communication channels are described by properties (“bit-hiding channels,” “anonymous channels,” “authenticated channels,” *etc.*); the composition of two channels is a channel, and thus “protocols” that are themselves channels can be built by composing channels. It may be possible to leverage known reductions among properties to prove both upper and lower bounds on the complexity of protocol-design tasks and to develop a useful notion of “universality” in networked computation (analogous to the notions of universality in circuit computation or Turing-Machine computation).

In the third category (“problems created or exacerbated by networks”), the focus is on scaling up and improving existing models and methods (*e.g.*, streaming, sampling, and sketching) to meet the challenges posed by modern networks. For example, given a massive, evolving graph presented as a stream of edge-insertions and -deletions, are there one-pass, space-efficient algorithms to compute (or approximate) key graph properties, *e.g.*, conductance, eigenvalues, and bad cuts? If a (single) computer (that is not a node in the evolving graph under consideration) can compute or approximate these values, can it also efficiently prescribe corrective action when problems are detected?

Cross-Cutting Issues

Several cross-cutting, high-level issues are relevant to all three categories and arose repeatedly during plenary and breakout sessions at both workshops

- **Incentive compatibility:** Perhaps the most important distinguishing feature of modern networks is that they are simultaneously built, operated, and used by multiple parties with diverse sets of interests and with constantly changing mixes of cooperation and competition. Formal models of networked computation and notions of hardness and easiness of computation will have to incorporate subnetwork autonomy and user self-interest in an essential way.

- **SPUR:** Achieving the broadest possible vision of “networked computation” will require substantial progress on Patterson’s SPUR agenda [Patt]. In his words, “we have taken ideas from the 1970s and 1980s to their logical extreme, providing remarkably fast and cheap computing and communication (C&C) to hundreds of millions of people. … [F]or our new century, we need a new manifesto for C&C: ... Security, Privacy, Usability, and Reliability (SPUR).”

- **Build on success:** Although today’s Internet may leave something to be desired with respect to security, privacy, usability, and reliability, it has far surpassed expectations with respect to several important design goals, *e.g.*, flexibility and scalability. Are the new design criteria compatible with the (manifestly successful) old criteria, and, if not, what are our priorities?

- **“Clean slate”:** The phrase “clean-slate design” has become a mantra in networking-research forums and in calls for proposals. Not surprisingly, many people have raised the question of whether anything that requires a “clean slate” could ever be brought to fruition in a world in which networked computation is pervasive and mission-critical. From a research perspective, the crucial point is that clean-slate *design* does not presume clean-slate *deployment*. Part of the ToNC agenda is the evaluation of new technologies, methods, algorithms, *etc.* from the perspective of incremental deployability and paths to adoption.

- **Diversity of "networks":** The scope of the networking research agenda is broader than "next-generation Internet," and thus the ToNC agenda must be broader as well. Interesting theoretical questions arise in the study of special-purpose networks (such as the DoD's Global Information Grid); of moderate-sized but functionally innovative networks; of sensor nets and other technologically constrained networks; of mobile networks; and of P2P and other application-layer networks.

Institutional Support of ToNC

The ToC community will pursue the ToNC-research agenda on many fronts and in many ways. Valuable types of research projects include but are not limited to:

- **Small, single-investigator, purely theoretical projects:** By "small," we mean funded at a level sufficient to pay for one or two months' of PI summer salary per year, one or two PhD students per year, and a few incidentals such as conference travel or commodity computers for the project participants.
- **Medium- and large-sized, multi-investigator projects involving both theory and experimentation:** The distinguishing features of such a project are (1) multiple PIs, at least one of whom is a theorist and at least one of whom is an experimentalist and (2) the inclusion of experimental work on a "real problem" arising in a network that can be built or at least envisioned in the current technological environment. Funding levels for these projects can range from anything that is bigger than "small" up to several million dollars per year.

Several NSF Program Directors have explicitly welcomed this type of medium- and large-sized project proposal, and the "distinguishing feature" text above comes from them. Careful consideration was given at the workshops to whether small, purely theoretical projects are equally important for success of the ToNC agenda, and participants decided that they are, for two basic reasons: (1) The intellectual scope of ToNC should not be limited by networks that can be built or even envisioned in the current technological environment; technologically untethered but mathematically rigorous investigation of networked computation is also worthwhile. (2) Some of the most eminent and productive members of the ToC community have traditionally worked by themselves or in collaboration with other theorists, and they have established broad and deep research track records in the process. Some have no experience working closely with experimentalists; nonetheless, they have built theories (*e.g.*, in distributed computing and in cryptography) that are of interest to practitioners as well as theorists. This subcommunity is unlikely to participate if all funded ToNC projects are medium- or large-sized projects of the type described above; yet, its potential contribution to the ToNC agenda is immense and should not be precluded by lack of funding.

Next Steps

Now that the Network Science and Engineering (NetSE) program has been established, support for ToNC looks promising within the CISE Directorate at NSF. It would be highly desirable to have support from Federal agencies other than NSF and from forward-looking IT companies. Advocacy and outreach will be important in obtaining this type of broad support. ToNC researchers should continue to promote our technical agenda both in our traditional forums (*e.g.*, STOC, FOCS, SODA, and Complexity) and in forums that unite us with other communities (*e.g.*, EC, PODC, CCS, Crypto, SIGCOMM, and NetEcon).

Finally, the ToNC community should continue to coordinate and collaborate with the broader networking community, in advocacy and in research. For example, ToNC researchers can play a vital role in the Global Environment for Network Innovations [GENI] by formulating testable hypotheses about the inherent power and limitations of networks. The architecture-research community is currently wrestling with fundamental questions about the value, costs, and tradeoffs of various networking primitives and abstractions. Very similar questions must be answered in the pursuit of a rigorous Theory of Networked Computation, and GENI presents a unique opportunity to experiment with new networks that have both innovative functionality and rigorous foundations.

Notes on this Report

A preliminary version of this report was released at the end of 2006; at that time, our intention was to get comments from knowledgeable readers both inside and outside of the ToC community and then publish a revised version by the middle of 2007. Shortly after the release of the preliminary version, one of us (Feigenbaum) was asked to join the GENI Science Council and, soon thereafter, to join the NetSE Science Council. We decided at that time that our revised report would be better if it were informed by the conclusions of the NetSE Science Council, and thus we delayed our revision. The NetSE Research Agenda has now been released; its influence on this report was significant, particularly in the area of institutional recommendations.

We also received very useful comments on the preliminary version from David Clark, Hector Garcia-Molina, Tom Leighton, Prabhakar Raghavan, Chris Ramming, and Jennifer Rexford, and we take this opportunity to thank them.

Finally, since we wrote the preliminary version of this report in 2006, the future of the GENI project has become unclear. However, we note that everything that we've written about GENI applies to any large-scale experimental networking platform. Because the NetSE Research Agenda highlights the importance of large-scale experimentation, we have left all of our original text about GENI intact.