Modeling and Simulation of Ultra-Large Networks:
Thirteen Recommendations for New Research Directions

Formulated During the NSF-Sponsored Workshop on
Modeling and Simulation of Ultra-Large Networks:
Challenges and New Research Directions

Nov. 19-20, 2001, Tucson, AZ, USA

Prepared under the NSF Grant ANI-0135530

by

Bernard P. Zeigler
Hessam Sarjoughian
Saurabh Mittal

Arizona Center for Integrative Modeling and Simulation
www.acims.arizona.edu
ECE Dept., University of Arizona,
Tucson, AZ 85721

List of Participants and Observers:

Participants	**Organization**	**Website**
Agarwal, S.	Lucent Technologies	Link
Claffy, K.	San Diego Super Computing Center, UCSD	Link
Colbourn, C.	Arizona Statree University	Link
Cox, P.	Lockheed Martin	Link
Floyd, S.	Lawrence Berkeley Nat'l Lab/ACIRI/UCB	Link
Fujimoto, R.	Georgia Tech	Link
Gelenbe, E.	Univ. of Central Florida	Link
Ghosh, S.	Stevens Institute of Technology	Link
Hariri , S.	University of Arizona	Link
Hui , J.	Arizona State University	Link
Joslyn, C.	Los Alamos National Laboratory	Link
Kim, T.G.	Korea Advanced Inst. of Science & Tech.	Link
Martinez, R.	University of Arizona	Link
Ng, H.	Naval Research Laboratory	Link
Meyer, D.	Sprint	Link
Nicol, D.	Dartmouth College	Link
Orman, H.	Nortel	Link
Raghavendra, C.	University of Southern California	Link
Reily, G.	Georgia Tech	Link
Sanders, W.	University of Illinois, Urbana Champaign	Link
Sarjoughian, H.	Arizona State University	Link
Wainer, G.	Carleton University	Link
Wilsey, P.	University of Cincinnati	Link
Woodcock, B.	Zocal	Link
Zeigler, B.	University of Arizona	Link
Znati, T.	NSF/University of Pittsburgh	Link
Cam, H.	Arizona State University	Link
Gupta, S.	Arizona State University	Link
Krunz, M.	University of Arizona	Link
Louri, A.	University of Arizona	Link
Reisslein, M.	Arizona State University	Link
Rozenblit, J.	University of Arizona	Link
Sen, A.	Arizona State University	Link
Xue, G.	Arizona State University	Link

CONTENTS

Introduction

A workshop on “Modeling and Simulation of Ultra Large Networks: Challenges and New Research Directions” was held on Nov. 19-20, 2001 in Tucson, AZ, USA. It was organized by The Arizona Center for Integrative Modeling and Simulation (http://www.acims.arizona.edu) and The Society for Modeling and Simulation International (http://www.scs.org). The purpose of the gathering was to bring together for a short, but intense, period some of the world’s leading researchers in the networking area to meet with counterparts with expertise in modeling and simulation of networks and of systems more generally. Invited participants were tasked with elucidating the unknowns of ultra-large networks and with new directions of research that can address these unknowns. The results are a set of specific finding of gaps in our knowledge of the behavior of ultra-large networks and how to deal with their design, management, and control. Participants considered whether current approaches can be evolved to deal with the large increases in scale or whether different, revolutionary paradigms are required. They addressed the need for new techniques and approaches for building models of ultra-large networks and developing simulation environments for studying their behaviors. Suggestions for borrowing points of view form other areas such as complex adaptive systems and from basic theory of modeling and simulation were encouraged. This report documents the definition of Ultra-large Networks and the recommendations that were formulated during the workshop and associated discussions. The recommendations, as well as the deliberations from which they emerged were captured during a morning session of the workshop using the Group Systems laboratory at the Center for Management of Information, University of Arizona. The recommendations were rank ordered using Group Systems anonymous voting capability and the results are listed at the end of this report. The recommendations were subsequently posted on the ACIMS web site with access restricted to workshop participants for comment. We wish to thank all participants for their inputs during the meeting as well as the significant number that provided comments and critiques in subsequent discussions. This report includes a bibliography of articles and research papers, in part supplied by the participants, in part gathered from the literature, that may help the reader to evaluate the recommendations. The workshop was sponsored by The National Science Foundation Advanced Networking Infrastructure Research Program (NSF/ANIR).

This report is one of two prepared under the NSF Grant ANI-0135530 that supported the workshop. The second report was prepared in the year following the workshop proper and is based in part on the workshop discussions and recommendations, and in part on subsequent study of the literature as well as exchanges with workshop participants and others (please see, “Modeling and Simulation of Ultra-Large Networks: A Framework For New Research Directions”). Whereas the recommendations in this report are the result of participant collaboration as described above, the conclusions of the second report are solely the responsibility of its authors.

What is an Ultra-large Network?

The Internet, as we know it, is the collection of hardware and software that enables computers from all over the world to communicate with each other. Most people know of this infrastructure through its support of the World Wide Web. This is a set of programs (one of which is the hypertext transport protocol, well known through its acronym “http”) that support graphically attractive dissemination of commercial, scientific, governmental, and many other kinds of information. The current Internet grew from a number of independent smaller experimental academic and defense networks that were linked together initially in the US and later expanding to other parts of the globe. A computer that has a network connection to an Internet Service Provider is called a network node. Although certain basic administrative functions (such as assigning Internet addresses to nodes) are centralized, there is no overall governing authority that coordinates, much less controls, the broad range of infrastructure and activities that the Internet entails.

In the next few years, as large new blocks of users in the emerging countries sign-on, the Internet is expected to connect upwards of one billion nodes. This alone qualifies it as the largest man-made communication network in history and the one and only Ultra-large Network (ULN) that we know of this sort. However, broad new categories of “users” are developing. In the Internet’s expansion to home or factory automation, Internet nodes can be small computer chips that run home appliances. A home might be networked and the computer that controls the environment may set the temperature and ambient lighting. But it also might be on the ‘net’ communicating with the local gas and electric company and setting its clock based on instructions from a computer in Greenwich, England. In further expansion to space, nodes can be the chips on communications satellites or tiny robotic computers traveling far into deep space. Nodes might also be dropped into place or secured to orbiting objects, forming interplanetary communication networks. With such potential all-pervasive, ubiquitous and universal reach, it is not clear whether there is only one ultra-large network or whether it is better to consider many such possible networks.

Thus, among the many unknowns of ULN, is the size and nature of a network that qualifies for the designation. Some participants felt that application of the term “ultra-large” should be restricted to the current Internet and should not, for example, include interplanetary networks since they have special needs distinct from those of today’s Internet. For this line of thought, ULN is a technical term referring to what we now call the Internet, a collection of a large number of autonomous networked systems.

Other participants felt that a more liberal and generic definition is warranted. One definition offered was that a ULN is a network that can support multimodal communication between any set of entities located anywhere on the globe or in flight around the globe. Thus, an Interplanetary Internet, and large-scale sensor networks should be included in one large universal Internet. Such ULNs are at least 2 orders of magnitude larger than the current networks, will include a large number of services and must support heterogeneous links and capabilities. Just as an Internet is defined as a collection of networks, with the

Internet (upper case) defined as a specific Internet, one could say that a ULN is a very large-scale connection of networks, with each component consisting of many nodes. Clearly, as the Internet evolves, it will be become a ULN, but it may not be one now. Although the notion of very large is vague, it need not be precisely defined, just as the "very" in VLSI is not really defined, but there is a general understanding of what is meant.

Some participants also felt that to encompass the full scope of the problem, ULN need to be defined as a combination of science and technology that includes theory, engineering, and other disciplines such as social sciences. Due to their sheer size, diffuse control, and heterogeneity, ULNs are networks that defy design and analysis by contemporary methods. Since they are capable of growth, ULNs should not be classified in terms of static properties, such as for example, having a billion nodes or any other specific physical or hardware characterization. Instead we should define them by the properties that they exhibit.

Besides their technical challenges, opinions were expressed that some effort should be expended to address social engineering issues in the definition of ULN. For example, what are the ethical implications of ULN? How can the ULN accommodate ethical considerations that vary geographically and demographically? Can the ULN be defined to accommodate broad economic diversity? While there was also reluctance to include such extra-technological considerations in the definition of ULN itself, it was countered that this definition should include input from more segments of the global community than just "network engineering" groups. Social impact of technology is often overlooked, and could be a useful source of data that might influence the traversal across design decision points. What do users (individuals, industry, other social institutions) expect from the ULN? Perhaps, there needs to be one or more "business models" of the ULN. Appropriate metrics (economic, social, technical) need to be identified.

Finally, the question was raised: Will there be a single ULN adopted by all people (countries?) Should we be concerned with putting in place a basis, with some national consensus, that might not account for other views?

Thirteen Recommendations Formulated During the Workshop

The thirteen recommendations, as ranked by the Group Systems session mentioned above, are shown below:

Rank	Recommendation	# of votes
1	Multiplicity, scalability and heterogeneity are difficult, but important issues	190
2	Progress in ULN depends critically on acquiring, characterizing and archiving real-world internet data	187
3	New mathematical approaches may be needed, either borrowed from other areas, or specifically developed for ULN	182
4	Effective models and simulators, properly described, need to be promulgated widely to support ULN research	177
5	Interfacing M&S with other tools is critical to deal with the many aspects of ULN design/operation	172
6	Technologies for model validation are needed and go hand-in-hand with progress in Rec. 2	157
7	Various approaches are needed to deal with unanticipated or emergent phenomena, including rare events	152
8	ULN poses particular challenges to moving from an art, to a science, of multiresolution modeling.	152
9	Common repository of tools needed. Stop reinventing the wheel, but facilitate truly new invention	140
10	Real-time M&S is important in itself and as a crucible for other M&S issues	133
11	Extrapolating from where we are	124
12	Evolutionary and blank slate approaches should coexist	120
13	We need to understand the evolution of the Internet, other researchers such as historians, economists and biologists may be helpful.	116

We now present summaries of these recommendations based on the Group Session data captured during the workshop. The summaries are related to the following broad categories of ULN concerns: modeling, simulation, theory, measurement, analysis, and design/operation.

Modeling Related Findings and Recommendations

- Technologies for model validation are needed and go hand in hand with progress in acquiring realistic Internet data
- Common repository of tools needed. Stop reinventing the wheel, but facilitate truly new invention.
- Multiplicity, Scalability and Heterogeneity are difficult, but important issues
- ULN poses particular challenges to moving from an art, to a science, of multi-resolution/multi-aspect modeling

- **Technologies for model validation are needed and go hand in hand with progress in acquiring realistic Internet data.**

Currently, relatively small-scale simulations are employed to demonstrate improvements in new designs and new protocols. However, there is no assurance that such small-scale models accurately capture network behaviors that may be arising when enormous numbers of heterogeneous nodes are interconnected to form ULN. Thus new technologies are needed for validating models that specifically support simulation-based design tools for ULN. Such tools should allow results of experimental infrastructures (see Recommendation 1) to be effectively employed to evaluate ULN models and design approaches. Research is needed to develop models that are sufficiently accurate and detailed to answer specific questions about system performance without attempting to mimic the entire large-scale system. Modeling languages need to be developed to permit defining computer processable models. These specifications would be automatically implemented as a simulations and also support techniques to explore various types of analytic validations, such as state space exploration and model checking. In a similar vein, simplification or extension techniques relating one model to another, for example relating high and low level resolution representations, might be able to allow for inferring verified properties from one model to many others.

- **A common repository of models and tools is needed to obviate the tendency to reinvent the wheel and thereby to facilitate truly new invention.**

A consortium consisting of members from industry, academia and government research agencies should develop a common component simulation architecture/tools that enable users to design, modify, validate and share these components to build ULN models and the associated interfaces. While sharable collections of libraries and dissemination of standards has aspects community development, currently it is not well understood how to develop tools, techniques and methodologies that enables researchers to use other researchers' work, especially in the context of ultra-large systems. Research is needed to develop the meta-modeling frameworks, protocols, environments, and standards that are required to leverage researchers' modeling and simulation efforts. Although such reuse runs a risk of inhibiting “out of the box” thinking, the prospects are greater that the ULN community can more rapidly advance by encouraging some to travel a well-traveled road farther, without everyone starting a new road.

- **Multiplicity, Scalability and Heterogeneity are difficult, but important issues; “System-level” approaches and families of models are needed to address them.**

As suggested in the earlier characterization, ULNs are distinguished from other networks by their sheer size, diffuse control, and heterogeneity of components, thereby defying design and analysis by contemporary methods. Several approaches to adequate understanding and methods for ULN analysis and design should be pursued. Analogies with biological, chemical, and physical systems of large complexity, and mathematical models of "organized complexity" may suggest fresh perspectives. Both reductionist and holistic approaches may be needed to obtain a "big picture" understanding of ULNs. To illustrate, think of ULN behavior as a “thunderstorm,” a global pattern of events that is not the same as the raindrops (devices, users, routers, and so on) that compose it. Generally, complex ULN systems need to be treated at multiple levels of description to achieve useful understanding and the search for global “laws” requires the ability to combine different types of models (both analytic and simulation) at multiple resolutions. It may be that such combination is a simple matter of programming and validating the combination is a simple extension of individual model validation. However, experience in other areas has shown that composing multiple models that were developed with different tools is not just a matter of programming. Research efforts should be started in this area, due to the multiplicity of existing tools, the complexity of interoperability solutions proposed, and the inadequacy of current tools for the particular application domain of ULN.

- **ULN poses particular challenges to moving from an art, to a science, of multi-resolution/multi-aspect modeling.**

ULNs are too complex to tackle in a monolithic manner. To deal with the complexity of such systems, multi-resolution approaches attempt to put detail in a model only where it is necessary, as dictated by the purposes and system aspects under examination. Since parts of the overall model that are not under immediate scrutiny are allowed to be more abstract, multi-resolution approaches combine models at different levels of detail and abstraction. Most often, finding the appropriate abstraction is the key to a problem. Right now working with multiple abstractions is an art that few people can do very well. Moreover, it is hard to teach. Appropriate methods to guide modelers in such methodologies as well as tools to support their work are needed to transition multi-resolution modeling from an art to a science. To develop such a science, research is needed to generate, relate, and cross-validate abstractions at the appropriate levels of resolution at various levels of granularity in ULNs. Such research would address, for example, ways to quantify errors in resolution changes and mismatches and the scalability limitations in extrapolating from small networks to ULNs that arise in single-level approaches. (Semi)automatic tools are also needed to generate useful abstractions from higher resolution models, characterize their domain of validity and error propagation.

Simulation Related Findings and Recommendations

- Real-time M&S is important in itself and as a crucible for other M&S issues
- Effective models and simulators, properly packaged and annotated to be promulgated widely to support ULN research and application

- **Real-time M&S is important in itself and as a crucible for other M&S issues.**

Real-time M&S is one of the most challenging areas of network research. It poses the intriguing challenge of understanding how a simulation or mathematical model can actually impact the course of events in the network being modeled. Real-time systems are often distinguished as “hard” or “soft” depending on whether deadlines for responses are strictly or loosely obeyed. Examples of hard time-constrained applications are on-line remote surgery and intrusion detection systems. Examples of soft systems include reliability in protocols for VoIP (voice over Internet Protocol), video/music streams, and other entertainment applications where not all packets need to make their assigned deadlines for the listener to enjoy the presentation. Demanding constraints on the execution time of models and their simulators are imposed when such models are used to characterize the state of a network and predict the consequences of management or control actions in its current state. Full-detail, packet level simulation of any reasonably large network (not to talk of ULN) presently have no chance to operate in real-time. Although hard real-time operation may not be requred, even in off-line use, models must execute in a time commensurate with the time-scale of management actions that are being evaluated. Research is needed to develop models that allow exploration of a possibly vast design space of parameter settings and protocol combinations to find acceptable, or possibly, optimal operating configurations, designs and tunings of a network. Thus the objectives of real-time M&S accentuate the general requirements to develop better abstraction methods, viz. those that enable more accurate and faster models and simulations.

- **Effective models and simulators, properly packaged and annotated to be promulgated widely to support ULN research and application**

Simulation tools for network analysis are not as widespread as they could be because they are difficult to use and may require expertise and/or computing facilities (e.g., supercomputers) that are not readily accessible to potential users. Such end users include network operators configuring a network, protocol designers, systems administrators troubleshooting network performance, and even networking researchers other than the originators of the tools. Complexities in using existing simulation technologies for ULN investigations arise since they often rely on parallel and distributed infrastructures that are needed to obtain the requisite performance. Fortunately, end users do not need to know all such underlying details of modeling and simulation tools to use them effectively. Unfortunately however, appropriate interfaces to hide such details do not yet exist. Thus research is needed to increase the usabilty of tools to help researchers and network operators ask questions in their own technical semantics that then translate into appropriate experiments to be executed automatically. Helpful as well would be tools to define and identify various behaviors and patterns in the voluminous data sets that can be generated in Internet measurements and in simulations of ULNs. While certain details can and shoul be hidden, some must be brought to light. Specifically, the concepts, abstractions, assumptions and limitations, of the models underlying widely used simulators, need to be exposed explicitly so that users become aware of the potential pitfalls in using encapsulated tools. To expose such information requires that is known in the first place. So this reinforces the need for research to develop the science of multi-resolution modeling discussed in an earlier recommendation. Workshops sponsored by NSF and/or industry should be conducted to educate researchers and designers about the potentialities of newly introduced models and tools.

Theory Related Findings and Recommendations

- Various approaches are needed to deal with unanticipated or emergent phenomenon, including rare events.
- New mathematical approaches and metrics are needed, either borrowed from other areas, or specifically developed for ULN.

- **Various approaches are needed to deal with unanticipated or emergent phenomenon, including rare events.**

As very complex systems, ULNs are very likely to exhibit many forms of previously unseen, and/or unforeseen, behavior. In the absence of perfect understanding or models ot it, a network may well go into failure "modes" that are not anticipated by a designer or implementer. For example, the Internet has experienced accidents like "BGP STORM" that can span the whole of Internet at the node level and whose origin is untraceable. This storm (July 19, 2001) traversed from the core of the Internet (/8 CIDR addressing) to the remotest of the host (/24 CIDR addressing). As it percolated from the core to the hosts, the effect became more pronounced and uncontrollable.

Modeling and simulation is a critical element in learning from such anomalous events, when they occur, and reducing the likelihood of their reoccurrence. The airline industry offers an apt analogy. Over time, their design, supported extensively by CAD (computer-aided design) backed by M&S, has improved to the point that the large majority of flights take place without incident. Nevertheless, with extreme rarity, disastrous failures do occur and these give rise to intensive diagnosis and simulation-supported timeline reconstruction that results in recommendations to eliminate the root causes of the failure. Likewise, it is critical to follow the pattern of the airline industry by developing the appropriate design, analysis and diagnostic methodologies and tools, aided by the powerful models and simulations that enable learning and cumulative improvement over time. Not all unanticipated behavior need by catastrophic -- emergent behaviors known in other complex systems, may be exploited in beneficial ways. Here again, appropriate models and simulations can help direct the exploration and the understanding of such phenomena.

- **New mathematical approaches and metrics are needed, either borrowed from other areas, or specifically developed for ULN.**

Although engineering attempts to make do with prevailing mathematics, ultimately ULN design, as a discipline that deals with complex systems, will face the need for real mathematical advances. Several areas of concern that may require new approaches can be enumerated. For example, currently we lack good mathematical models of the behavior of TCP (Transport Control Protocol) under large-scale aggregation that are needed to represent background traffic (competing traffic not of direct interest for the investigation at hand). In contrast to the simplistic representations often employed in current simulators, such models would provide valid, reliable, abstractions for background traffic in multi-resolution simulations (as mentioned in another recommendation). Developing such abstractions is definitely non-trivial, since interactions involving queue-lengths along each path, packet loss rate, and speed of each link must be accounted for. Further, to account for usage patterns and human responses to social phenomena, mathematical models should include representations of the human and social institutions that interact with the network. Although, the application of existing mathematical tools may suffice, it seems more likely that ULNs demand the development of new kinds of mathematics. While tools from queuing theory, stochastic systems, and combinatorial approaches all afford a very useful view of small networks; it is not clear how to scale these approaches to very large networks in a meaningful way. Therefore, research is needed to examine the potential applications of mathematical approaches that have been developed in related large scale systems areas such as mathematical systems theory, complexity theory of biological systems, and systems control theory. Indeed, it should be noted that there are relevant theories of modeling and simulation that can be applied, and deepened, in application to ULNs. Likewise, progress may depend on identifying appropriate metrics that allow useful high level characterization of ULN states and dynamics (analogous to temperature, pressure and volume that characterize the state and give rise to the relationships encoded in the ideal gas law). New research is needed to identify these metrics, perhaps employing experimental test beds, and to use them to build scalable models.

Measurement Related Findings and Recommendations

- Progress in ULN depends critically on acquiring, characterizing and archiving real-world data

- **Progress in ULN depends critically on acquiring, characterizing and archiving real-world data.**

Real-world data is crucial for future model development. There is a strong need for realistic Internet data to develop, verify, evaluate, test novel protocols using simulation-based design tools. Given the masses of data that could become available, there will be a need for interpretation and categorization of real-world data into scalable data-sets that captures both the user-intent and the nature of the traffic flow. Such data could also serve such purposes as management, control and reconfiguration of the Internet in the advent of any catastrophic breakdown. And just as a vigorous market economy relies on free flow of information and pricing signals, the Internet should have easy-to-understand data-derived indicators that make allocation or its resources more efficient in the prevailing unregulated environment. A major source of data, that of commercial sources such as hardware vendors and ISPs, is of limited value due proprietary restrictions. Thus mechanisms, such as university/industry collaboration, motivating the networking industry to provide useful data should be developed and introduced.

To resolve unknowns in studied phenomena, science turns toward acquiring data from the real world. Such data is particularly needed to construct the models and simulations that will inform design tools for ULN. One might imagine that such data is easily at hand when the source is a man-made phenomenon such as the Internet. In actuality, given the earlier-mentioned dispersed and free-wheeling authority structure, there is no central location that can collect data on Internet operation in some coherent fashion. However, there are many distinct sources of data that could be put together to get a bigger picture. Researchers have in the last several years begun to collect data on Internet traffic and behavior through the interfaces accessible from their computers. By themselves, such external data are not sufficient, since this is like measuring a person’s temperature without having access to X-Ray photos of the internal organs. In relation to such higher resolution data, it is true that many ISPs collect data online to adapt their services to changing demand patterns and to assess the performance of their fielded equipment. Also, network infrastructure manufacturers and vendors collect, or pay laboratories to collect, data on the response times and other attributes of their products such as routers and switches. Unfortunately, a major problem is that commercial data at the ISP and vendor levels, are often proprietary and closely guarded from public view. Companies are concerned that data shared with researchers could find its way to the competition, giving away critical advantages in the marketplace.

The participants felt that real-world data are crucial for future model development for ULN. Any conclusion drawn without this real data could be highly misleading. Thus mechanisms motivating the ISPs to provide useful data should be developed and introduced. There might be direct benefits to the ISPs that counteract the risk of informing the competition. For example, data collection strategies might be formulated in such as way that an ISP might benefit directly from research that categorizes and archives the collected data. Another, somewhat oblique, way to obtain realistic data is simply to ask industry how it would like to see the ULN evolve to serve its needs. One source of such contact is through industrial advisory panels to academic information technology institutions.

The workshop featured more in-depth discussion of the above issues:

Is Internet Data Collection as Research Issue for ULN?

There was some sentiment that collecting data by itself is not necessarily a research issue. In response it was pointed out that designing measurements for sophisticated hardware and software systems is not trivial and may indeed require research and development. Moreover, "clean" collection of any kind of data from the Internet is not straightforward. It was suggested that data collection and data analysis need to go hand in hand, oftentimes the specific method of collecting data has an impact on the statistics of the data and needs to be take into consideration when interpreting the data.

What sorts of data are needed? How do they scale for ULN?

Besides getting or generating the right data there is the question of how to use them. Assuming we could generate a complete profile of exactly what happens on the current Internet, what benefits could it afford? Having the data should assist us in discovering the right questions to ask. Perhaps data mining approaches can help to examine such massive amounts of data to help formulate the right questions for future research. However, it was also pointed out that before blindly collecting data we should answer the question, what categories of data are there? For example, there are packet trace data, internodal connectivity data, application traffic traces, etc. Initial research might therefore be necessary to reveal which data sets are needed for research in which areas. For example, it was pointed out, that network data (e.g., packet traces) are limited in that they often do not capture higher-level aspects such as what was the sender’s intention in generating the data. Also, trace data has well known limitations from a simulation perspective, for example, the trace is invariant to things like delay in the model. It was suggested that good models of user behavior are needed that capture the user's intent, and that these models can be used to generate realistic data.

Further, since it deals with unknown future demands, ULN research needs to anticipate the kinds of traffic we will see on these future networks. Imagine homes and offices hooked up by optical fibers at the computers' internal clock speeds to the Internet. What will people do with this bandwidth? Watch HDTV quality video via the web, or stereo-scopic 3D video? What will this traffic look like? Will it be Long Range Dependent as we see it now for MPEG coded video? At the same time there will be many small (in terms of bandwidth) flows e.g., from networked toasters. How will these flows interact with the large flows? As a basis for this research we need a database of application layer traffic, especially multimedia streaming traffic, that reflects typical quality and stream bandwidth levels that the different users would send.

Once a categorization and a purpose for the different types of data sets have been found, we need to address the following critical question in the context of ULN: Do the different types of data sets "scale"? For instance what happens when scaling up a topology map to billions of nodes? There was also concern for the exchangeability of data sets. Can we establish a standard for data sets to make sure other researchers (one or more steps removed from the original data collectors) can make use of the data. In other words how can we ensure that the collected data can leverage the efforts of others?

Is a new Internet needed to provide a source of measurement? Will smaller testbeds do?

An alternative to measurements on the Internet that have commercial restrictions is the development of new networks that would serve more for experimental purposes than commercial and other applications. While there was some interest in this possibility, it was pointed out that such networks require massive funding and might not reflect the true characteristics of the Internet in its rough and wild state. It was suggested that reasonable experimental test beds, of 100-200 machines/network devices, could be used in a controlled environment to obtain relevant data. This data can then feed into appropriate modeling and simulation environments to answer specific questions, such as what is the expected reliability, availability, vulnerability, performance, quality of service, etc. Once that is done, another research effort is needed to extrapolate the results for a larger network that could have millions of nodes. Research is needed to help make such scalability extrapolations – indeed, to discover whether they are possible.

Other uses of Internet data

Real world data is not just for modeling, it is necessary to ensure the healthy functioning of the Internet itself. The only way to prevent a catastrophic collapse of the Internet is to have sufficient and accurate data collected so that the network could be reconfigured quickly.

In a vigorous economy, it is necessary to have free flow of information and pricing signals. Likewise, the Internet should have easy-to-understand parameters in order to be able to achieve allocative efficiency. Efficiency here includes transmission and buffering but also processing, mass storage, responsiveness and privacy. Large-scale data collection and dissemination has great societal value through the promotion of both economic and technological transparency. It's the only way we can learn from the mistakes, and build upon the work of our predecessors.

Analysis Related Findings and Recommendations

- We need to understand the evolution of the Internet, other researchers such as historians, economists and biologists may be helpful

- **We need to understand the evolution of the Internet, other researchers such as historians, economists and biologists may be helpful.**

The Internet infrastructure, in common with other large human-built infrastructures, such as economies and political structures, is only partly understood in terms of its intentionally designed-in technological properties. The current Internet was much more successful than its inventors could have imagined. Today, after decades of its growth, we do not understood what were the principles employed in design of the Internet Protocol that allowed it to scale to the global system it is now and to handle with ease many applications that were not dreamed of when IP was invented. Understanding these design principles, and encoding them in modeling and simulations, would be invaluable in the design for a new ULN. A related issue is understanding the evolution of the Internet in its full technological, historical, and social context. Research is needed to develop models that predict the potential impacts of economic and social/political decisions. Such models should be aimed at understanding the possible future paths of evolution, and the forces, as expressed in conscious human intention, that would have to be applied to end up on one path rather than another.

Design/Operation Related Findings and Recommendations

- Evolutionary and 'Blank Slate' approaches should coexist.
- Interfacing M&S with other tools is critical to deal with many aspects of ULN design and operation.
- Extrapolating from where we are

- **Evolutionary and 'Blank Slate' approaches should coexist**.

  Research in both evolutionary and "blank slate" approaches can be useful in developing future ULN. In the evolutionary approach, we can learn from deployment of new ideas or artifacts in the current Internet environment. Research is needed to develop tools that support the full iterative cycle going from design and simulations to deployment, using measurement to learn from deployment, and using analysis and design to progress to the next stage. Although research in continuing the evolution of the Internet infrastructure is critical, research that starts from a blank slate, can investigate fundamentally different architectures and protocols than have so far evolved and can provide fresh understanding of possibilities, and of the limitations, of the current infrastructure. The blank slate approach might even invent or enable a useful paradigm shift that otherwise might not emerge. Blank slate approaches also may respond to opportunities to design new ULN such as private corporate networks, sensor networks, and interplanetary networks with results that can then influence the mainstream Internet's future development. It is important that NSF to support the pursuit of both approaches, while vigilant for the more promising one or a new synthesis.

- **Interfacing M&S with other tools is critical to deal with many aspects of ULN design and operation.**

  Research is needed to develop methods to interface network M&S tools with tools developed in other areas such as security, vulnerability analysis, and system design tools such as UML. Precedent has been set by designers of network protocols making use of modeling and simulation tools for specification and validation. However, research is needed to develop M&S tools will be needed to address the emerging issues of security and vulnerability in ULNs. Although is some situations, the method of interface is self-evident, in many cases there will be mismatches in the levels of abstraction, resolution, or underlying purposes that need to be resolved to allow useful inter-operation of disparate tools. Integrating system modeling issues (operating system, file system, computer type, etc.) with network modeling issues is essential to better understand end-to-end issues; discover vulnerability in operating system to virus attacks, security problems, etc. Without taking on system issues, modeling and simulation will fail to analyze and project problems in ULN operations, reliability, security, performance, just to name a few.

- **Extrapolating from where we are.**

  Extrapolating from the Internet of today to the ULN of tomorrow raises questions such as whether success obtained in moving from 10 to one million nodes will continue to hold in moving from to the later to ULN. Scalability challenges, such as these raised for network modeling and simulation may be addressed in several ways. One approach, that appears not to be practical with existing tools, is to apply current day simulators at the packet-level to networks of one billion nodes. Another approach is to apply such simulators to smaller packet-level simulations and to try to understand to what extent these results will extrapolate to the ULN. For example, we need to understand what inferences can be drawn from burst level simulations, employing the packet level as a baseline approach for small networks to check how accurate the other coarser methods are. A third approach, addressed in another recommendation, is to try to develop abstractions that aid our understanding of the dynamics in ULN even if such abstractions cannot and should not faithfully reproduce these dynamics in their entirety. Combinations of one or all of these approaches may form the basis for research that must be undertaken to develop new methods that can be used for billion node network simulations. These may include mathematical models, large scale parallel/distributed simulation, fluid flow models, or other methods we have yet to think of.

Bibliography

Adamic, L., and Huberman, B., “The Web ’s Hidden Order ”, CACM 44:9,2001

Aggarwal, S., Courcoubetis, C., and Wolper, P., "Adding Liveness Properties to Coupled Finite State Machines," ACM Trans. on Programming Languages and Systems, Vol. 12-2, April 1990, pp. 303-339.

Aggarwal, S., Paul, S., Massey, D. and Calderaru, D., "A Flexible Multicast Routing Protocol for Group Communication," Computer Networks, Volume 32, Issue 1, January 2000, pp. 35-60.

Callaway, D. S. et al,. “Are Randomly Grown Graphs, Really Random?”, Phys. Rev. E 64,041902,2001

Chapman, W., Rozenblit, J.W. and Bahill, A. T., “System Design is an NP-Hard Process”, Systems Engineering Journal, 4(3), 222-229, 2001.

Christen, G., Dobniewski, A., Wainer, G., "Defining DEVS models with the CD++ toolkit”, In Proceedings of European Simulation Symposium. Marseille, France.

Clark, G., Courtney, T., Daly, D., Deavours, D., Derisavi, S., Doyle, J. M., Sanders, W. H., and Webster, P., “The Möbius Modeling Tool “, *Proceedings of the 9th International Workshop on Petri Nets and Performance Models*, Aachen, Germany, September 11-14, 2001, pp. 241-250.

Davis, P.C., Fishwick, P.A., Overstreet, C.M., Pegden, C.D., “Model Composability as a Research Investment: Responses to The Featured Paper”, Winter Simulation Conference, pp. 1585-1591, 2000

Deavours, D., Daly, D., and Sanders, W. H., “Möbius: Framework and Atomic Models”, *Proceedings of the 9th International Workshop on Petri Nets and Performance Models*, Aachen, Germany, September 11-14, 2001, pp. 251-260

Fitzek,F. H. P., Reisslein, M., ”MPEG-4 and H.263 Video Traces for Network Performance Evaluation”. IEEE Network, 15(6): 40-54, November/December 2001

Floyd, S. and Paxson, V., “Difficulties in Simulating the Internet”, February 2001. IEEE/ACM Transactions on Networking

Glinsky, E., Wainer, G., "Analyzing overhead of Real-time DEVS models", accepted for publication in AIS'2002. Lisbon, Portugal. 2002.

Ghose, K., Aggarwal, S., Vasek, P., Chandra, S., Raghav, A., Ghosh, A., and Vogel, D., "ASSERTS: A Toolkit for Real-time Software Design, Development and Evaluation”, 9th EuroMicro Workshop on Real-Time Systems, June 1997, pp. 224-232.

Goguen, J., “An Introduction to Algebraic Semiotics,with Applications to User Interface Design ”, in Computation for Metaphor, Analogy and Agents, ed.Chrystopher Nehaniv, springer LNCS,v.1562,pp.242-291,1999.

Hild, D., Sarjoughian, H.S., Zeigler, B. P., "DEVS-DOC: A Modeling and Simulation Environment Enabling Distributed Codesign," IEEE SMC Transactions, Vol. 32, No.1, pp. 78-92.

Joo, S.S., Chung, Y. S., Razouqi, Q., Ghosh, S., “Directions in IP Router Architecture Design And its Validation," 2001, IEEE International Conference on Software Telecommunications and Computer Networks (SOFTCOM), Split, Dubrovnik (Croatia), October 09-12, 2001, pp. 495-503.

Kampis, G., “Self-Modifying Systems”, Pergamon,Oxford 1991

Kleinberg, J. M., “Navigation in a Small World ”, Nature v. 406, 2000

Kasputis, S., Ng, N., “Composable Simulations”, Winter Simulation Conference, pp. 1577-1584, 2000

Lindgren. K., “Evolutionary Phenomena in Simple Dynamics ”, In Artificial Life II Langton et al. (Eds), Addison-Wesley, 1991

Morihama, L., Pasuello, V., Wainer, W., "Validating the behavior of DEVS models", Accepted for publication in SISO Spring Interoperability Workshop, Orlando, FL. U.S.A. 2002.

Newman, M.E., “Models of the Small World ”, J. Stat. Phys.101, 819-841,2000

Paxson, V. and Floyd, S., “Wide-Area Traffic: The Failure of Poisson Modeling”, IEEE/ACM Transactions on Networking, Vol. 3, No. 3, pp. 226-244, June 1995

Pillalamarri, S., and Ghosh, S., “The Impact of Source Traffic Distribution on Quality of Service (QoS) in ATM Networks." Accepted in IEEE International Conference on Communications, June 11-15, 2001, Helsinki, Finland, pp. 2855-2859.

Razouqi, Q., Ghosh, S., ``Incremental, Dynamic, Virtual Circuit Connection (IVCC): A New Paradigm for Routing in Future High-Speed Networks," Accepted in IEEE International Conference on Communications, June 11-15, 2001, Helsinki, Finland, pp. 2578-2582.

Riley, G. F. and Fujimoto, R. M., and Ammar, M.H., "A Generic Framework for Parallelization of Network Simulations", Proceedings of Seventh International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS'99), October, 1999

Riley, G. F. and Fujimoto, R. M., and Ammar, M.H.,, “Stateless Routing in Network Simulations", Proceedings of the Eighth International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, August, 2000

Riley, G. F., and Ammar, H. M., "Simulating Large Networks: How Big is Big Enough?", Proceedings of First International Conference on Grand Challenges for Modeling and Simulation, January 2002

Rozenblit, J.W., “Systems Design: A Simulation-Based Modeling Framework”, in Discrete Event Modeling and Simulation: A Tapestry of Systems and AI-based Theories and Methodologies. (Eds.F. Cellier and H. Sarjoughian), Springer Verlag, pp. 107-127, 2001.

Sanders, W. H., “Integrated Frameworks for Multi-Level and Multi-Formalism Modeling”, *Proceedings of PNPM'99: 8th International Workshop on Petri Nets and Performance Models*, Zaragoza, Spain, September 8-10, 1999, pp. 2-9.

Schulz, S., Rozenblit, J.W., Mrva, M. and Buchenrieder, K., “Model-Based Codesign“. IEEE Computer, 32(8), 60-68, 1998.

Stewart , T., “Six Degrees of Mohamed Atta ”,Business 2.0, http://www.business2.com/articles/mag/print/0,1643,35253,FF.html

Zeigler, B.P., Praehofer, H., Kim, T. G., *Theory of Modeling and Simulation: Integrating Discrete event and Continuous Complex Dynamic Systems*, Second edition, Academic Press, 2000.