Computer Science Department Technical Report
University of California
Los Angeles, CA 90024-1596

THE NATIONAL EXCHANGE FOR NETWORKED INFORMATION SYSTEMS: A WHITE PAPER

L. Kleinrock
J. P. G. Sterbenz
N. Maxemchuk
S. S. Lam
H. Schulzrinne
P. Steenkiste

November 1993
CSD-930039

THE NATIONAL EXCHANGE FOR NETWORKED INFORMATION SYSTEMS: A WHITE PAPER

Leonard Kleinrock, Chair
James P. G. Sterbenz
Nick Maxemchuk
Simon S. Lam
Henning Schulzrinne
Peter Steenkiste

Acknowledgement to the High Performance Networking Enterprise group consisting of the authors above as well as:

Nicholas Bambos
Larry Bergman
Ken Calvert
Bilal Chinoy
Henry Dardy
Deborah Estrin
Robert Felderman
Peter Ford
Victor Frost
Mario Gerla
Sumit Ghosh
Joseph Godsil
S. Jamal Golestani
Bill Johnston
Arne Nilsson
Aurel A. Lazar
Raj Singh
Jeff Terstriep
Dan Winkelstein

THE EXECUTIVE SUMMARY

There is a clear need for creating an environment in which the best talent in our country can collaborate to develop the National Information Infrastructure. We must make it possible for this talent to apply their skills in evaluating networked information systems of great complexity, scale, heterogeneity, and dynamic behavior.

To accomplish this, we propose the creation of a National Exchange for Networked Information Systems (the Exchange) which will allow the community of geographically distributed researchers, developers, vendors, and users to interact, to share ideas and results, and to document their findings in a *live and active* library of information. Through the medium of this Exchange, they will be able to model, measure, evaluate, and design the most advanced networked information systems in the world.

In this Exchange, the following kinds of *objects* will be collected: data, tools, and models. Under an acceptance control policy, these objects will be placed into the *Libraries* of the Exchange: the *Data Library* the *Tools Library* and the *Models Library.* The objects will contain full details of the environment in which they were created and in which they operate. Moreover, the Exchange will provide access to Shared Experimental Facilities to enhance the collaboration and interaction of its participants.

The Exchange we propose will provide a reliable, consistent, dynamic, and accessible archive of data, tools, models, and systems. It will allow the work of earlier developers to be captured and reused on the same project or for new projects. No longer will the effort that goes into collecting data, writing a simulation, coding a model, or running these models, be lost when a project ends. The existence of such an Exchange will reduce the duplication of such efforts which is so prevalent in today's research environment. Moreover, it will foster the creation of benchmarks against which measurements and data will be compared in a meaningful way.

The capability to carry out studies which explore as yet unachievable systems in the dimensions of scale, complexity, and heterogeneity is extremely appealing. These Libraries will be made available to the full community of interested parties. It will draw the best talent to the problem, it will allow the collective wisdom of all views to come to bear on the problem, and it will enable the building of a base of results and tools upon which taller structures can be studied. It will generate a sense of community that draws together the spectrum from research through development and deployment, and across the entire set of scientific and engineering disciplines necessary to deploy the National Information Infrastructure.

The usefulness of such an Exchange was amply demonstrated in the early days of the ARPANET when its Network Measurement Center was established. The Center's use in driving the system architecture and design, in identifying weak points and bottlenecks, and in guiding the expansion of the ARPANET to its current manifestation as the Internet, were key contributors to its success. It is important that we learn from this past success as we mount the effort to carry networking to its next great height.

This country has defined a number of national and grand challenge applications employing the National Information Infrastructure (NII). An ideal environment in which to pursue and meet these challenges is through the Exchange. We recommend that a detailed plan be developed as the first step toward the creation of the National Exchange for Networked Information Systems. The Exchange will provide the infrastructure and collaborative environment to provide a long-lasting capability for continued leadership in networked information systems.

1. INTRODUCTION

In order to facilitate the development of a National Information Infrastructure (NII), we advocate the creation of a NATIONAL EXCHANGE FOR NETWORKED INFORMATION SYSTEMS. This Exchange will create that portion of the infrastructure to enable a new way for research collaboration, whose purpose is to advance the state of the art in high performance networked information systems; these advances will take place through the study of key issues such as scalability, complexity, and heterogeneity.

The idea is to create a national capability whereby the community of geographically distributed researchers, developers, vendors, and users will be able to create and access an *active* library of objects which will allow them to interact and conduct modeling, measurement, evaluation, and design of the most advanced networked information systems in the world.

In this Exchange will be *Data, Tools, and Models Libraries* respectively containing the following objects:

- data (measured or generated)
- tools (e.g., simulation languages, mathematical packages)
- models (e.g., analytical, simulation)

In addition the Exchange will provide *Shared Experimental Facilities* consisting of:

- network infrastructure (e.g., testbeds, computer systems, transmission facilities)

- network emulators (e.g., for transmission links and sub-networks)

Under an acceptance control policy, these objects will be placed into the Exchange. They will be automatically indexed and will contain the full details of the environment in which they were created or in which they operate. Objects may be removed from the Exchange and archived if they fail to meet measurable use criteria. In addition the Exchange will provide a sophisticated interface which will permit these objects to be enhanced, modified, or activated, and, in general, used in isolation or in a collaborative interactive fashion.

Our vision is for the Exchange to facilitate the creation of an infrastructure that will have long term benefits for the nation. What is likely to emerge spontaneously is a paradigm shift in the way communities of researchers, developers, vendors, and users can and will interact. Furthermore, what we are proposing is to develop a methodology through which diverse and hitherto disconnected communities that share a common goal will interact henceforth. The timing is exactly right for the creation and growth of this Exchange now. Imagine a shared environment which allows measured or generated data, evaluation tools, and models to be stored in the Libraries according to well-defined formats and standards and then made available to the entire community of researchers, developers, vendors, and users. This community will then be able to use these tools and data to develop shared models and design methodologies. The leverage is significant in terms of avoiding duplication, disseminating knowledge, and establishing benchmarks for good system behavior.

The Exchange we propose will provide a reliable, consistent, dynamic, and accessible archive of data, tools, models, and systems. It will allow the work of earlier developers to be captured and reused on the same project or for new projects. No longer will the effort that goes into collecting data, writing a simulation, coding a model, or running these models, be lost when a project ends. The existence of such an Exchange will reduce the duplication of such efforts which is so prevalent in today's research environment. Moreover, it will foster the creation of benchmarks against which measurements and data will be compared in a meaningful way.

The capability to carry out studies which explore as yet unachievable networked systems in the dimensions such as scale, complexity, bandwidth, latency, and heterogeneity, is extremely appealing. The Libraries and Facilities will be made available to the full community of interested parties. The Exchange will draw the best talent to the problem, allow the collective wisdom of all views to come to bear on the problem, and enable the building of a base of results and tools upon which more advanced network structures can be studied. The synergy is enormous -- we are pooling the talents of experts who will be able to collaborate as never before. The Exchange will generate a sense of community that draws together the spectrum from research through development and deployment, and across the entire set of scientific and engineering disciplines necessary to deploy the NII.

2. RATIONALE

2.1 The Problem

There is a clear need for creating an environment in which the best talent in our country can collaborate to develop the most advanced National Information Infrastructure in the world. We must make it possible for this talent to apply their skills in evaluating an infrastructure of great complexity, scale, heterogeneity, and dynamic behavior, which exceeds their current capabilities. To do this, we must provide a facility, the Exchange, that allows this diverse community to interact, to share ideas and results, and to document their findings in a *live and active* library of information.

This National Exchange for Networked Information Systems will involve measurement, modeling, development, evaluation, experimentation, interaction, and collaboration. Such activities are critical to understanding:

- how to develop the best system design before deployment

- how the system is performing and why it behaves the way it does

- how it will perform if changes are made to parameters and to the architecture
- how it will scale
- how it will perform in other environments and scenarios
- how it can be improved

Clearly, the steps described above will be included in an iterative procedure of continual refinement towards improved system designs. Moreover, we will use the understanding we achieve to develop the principles that underly these networked information systems. This is especially needed in the case of high-speed networked information systems, where the scope, complexity, and scale of the effort required to develop and deploy such systems are unprecedented.

This process is appreciated by relatively few systems designers. However, even in those cases where measurement, modeling, and evaluation are carried out, it is usually done from scratch, without the benefit of tools which have been developed by other systems builders and analysts, and the results are seldom stored in a reusable fashion. Furthermore, results are mainly disseminated through presentations at conferences and publications in journals. Unfortunately, this mechanism often does not allow other researchers to easily incorporate the results in their work, since this often requires the use of specific tools or infrastructure as well as access to unpublished and often deleted data. This is especially problematical for small research groups which often lack the critical mass in funding, expertise, and facilities. The only thing that remains as a reusable object is typically a packaged simulation or management tool (i.e., the commercially available and supported software packages). The loss of the data, tools, and software to the community is vast and cannot be tolerated.

Developing networked information systems is more challenging than many other types of systems, because designing, building, and integrating high-speed wide area networks is extremely expensive. Indeed, networks are inherently an order of magnitude more complex because they are a *system of systems*. We are fast approaching the interconnection of highly heterogeneous gigabit networks whose problems of interoperability will be legion! This looming problem is further exacerbated by a problem of equal magnitude, namely, scalability of these networks.

Whereas advanced testbeds are extremely valuable in advancing the state-of-the-art, they can only explore a small space of architectures due to the complexities which arise from these network problems of scalability, heterogeneity, dynamic traffic, technology dependencies, and configuration dependencies, given limited resources and limited time.

As a result, it is important that we be able to predict the behavior and evaluate the performance of large networks before they are built, based on information obtained through smaller scale experiments. In short, the successful deployment of high-speed wide area networks will require close collaboration between different disciplines in the networking community: analysis, modeling, development, implementation, testing, and measurement.

Moreover, building high-speed networks requires not only expertise in networking, but also in many related areas such as systems theory, operating systems, computer architecture, circuits and systems, and mathematics. Again, an effective way of exchanging information and fostering collaboration among such diverse fields in a timely way is required.

2.2 The Precedent

There is a clear precedent, from the earliest days of networking and running right up to the present, which supports the usefulness of the Exchange. Indeed, the first packet switching network, the ARPANET, was designed with a clear understanding of the need for, and commitment to, modeling, measurement, and experimentation.

A considerable effort went in to the provision of measurement hooks in every switch of the ARPANET. A Network Measurement Center was established at UCLA from the beginning for the purpose of providing for performance measurement and experimentation of the ARPANET. Moreover, the network was analyzed using mathematical modeling as well as simulation long before it was implemented. This turned out to be a key contributor to the success of the ARPANET. Indeed, the network developers would have been hard pressed to identify the source of major network failures (both deadlocks and degradations) had they not been able to make measurements, collect data, trace packets, and run experiments with the tools that were so carefully provided. It was through the use of these tools that the designers were able to guide the development and expansion of the ARPANET in the first few years of its existence.

In the mid-1970's, the ARPANET administrative responsibility was passed from ARPA to the Defense Communication Agency. The activities of the Network Measurement Center that thrived for more than half a decade ceased to exist at this point, and have never been revived. The network continued to grow at significant rates, and the growth was not only unmanaged, but it was also uncontrolled. In this environment, it is surprising that the highly successful Internet evolved. A major step in this development was the introduction of the NSFNET which provided a backbone network for the Internet at megabit rates. The highly successful Internet has impacted the scientific, academic, commercial, and government sectors in remarkable ways. Part of the success of the Internet is due to the fundamental guiding principles of operation that were developed in the early ARPANET days, a time when it was possible to measure, model, evaluate, and interact with that network.

Fortunately, the evolution from the kilobit rates of the ARPANET to the megabit rates of the Internet was just that, an evolution. For this reason, the principles that were uncovered through extensive modeling, evaluation, and experimentation of the early ARPANET operational network carried us reasonably well through the growth of the Internet.

However, although the bandwidth of networks has not increased in revolutionary ways, the size of the Internet has stepped up in an uncontrolled fashion. Most of the problems we now face with the Internet have to do with this increase in size, and these problems are essentially related to scale and heterogeneity. The fact that we lack the appropriate data, tools, models, and methods with which to explore, experiment with, and predict, the impact of protocol and structural changes to the Internet has significantly impaired the ability of the technical community to address these problems properly. It is also clear that these problems will only become worse and increase in importance as we face the design, development, and deployment of the NII.

Furthermore, we are now faced with revolutionary changes in networking as we move into the gigabit per second rates of the NII. No longer can we rely on our previous experience and intuition. The principles of lower speed networks do not easily extrapolate to the emerging networks we are planning for the NII. We must now reestablish the tradition of measurement, modeling, evaluation, and experimentation with these systems as we continue the path into advanced networking.

2.3 Issues and Challenges

Advanced networks present us with a plethora of new issues, both in analysis and implementation. To illustrate, we present a brief discussion of three, of the many, unique high-speed networking challenges.

One important technical issue in gigabit networking, that has only marginal significance in today's networks, is the "latency problem". At present, in terrestrial networks, the time for a source to inject a message (or application data block) into the network is typically much greater than the time it takes the last bit of the message to propagate from the source to the destination. As the link data rate increases, the propagation delay may well exceed the source injection time, and new models for looking at networks will then be required. One way to see this is to recognize that for the transmission of a one megabit block of data across the United States, it will take roughly 650 msec to inject it, and only 20 msec for the last bit to propagate across the country if we use a T1 line (at 1.5 Mbps) for transmission. On the other hand, if we use a 1.2 Gbps line, the injection time drops to less than one msec while the propagation delay remains constant at 20 msec. The latency component is negligible in the former case, but in the latter case the latency due to the speed of light dominates the response time. Clearly latency becomes a central issue in gigabit networks. Moreover, by the time the first bit propagates across the country with a 1.2 Gbps line, the cross-country link contains roughly 25 megabits in transit (this is the bandwidth-delay product); another consequence of the latency problem is the increasing number of bits that must be buffered and may be lost in the case of link errors.

Another manifestation of the latency problem was made evident after one of the first standards for high-speed networks, the IEEE 802.6 standard for metropolitan area networks, was passed. In both the analysis and simulation of this protocol the propagation delay was ignored, which resulted in an unfair access protocol that had to be changed.

Latency is expected to also affect:

- the ability of interactive applications to migrate from the local area to wide area networks
- the ability to control remote sources and processes
- the ability to distribute communicating processes and the memory objects they use and modify

An objective of the Exchange is to provide a means of validating and comparing the new models that will be needed to understand the dynamics of new issues such as these.

Note that, in the context of the "Global Grid" the range of round trip latency that must be supported is at least 8 orders of magnitude: from the tens of nanoseconds for interconnection of multiprocessors, through tens of milliseconds for terrestrial fiber optic links, to significant fractions of a second for satellite links.

A second technical issue of increasing importance is that of network reliability. Our networks are becoming more difficult to manage and maintain as their size and complexity increase and as new services are provided to satisfy increasingly demanding applications. Advances in technology such as the increased transmission capacity of fiber optic links and the switching capacity of fast packet switches allow for the support of higher bandwidth applications. However, we find that the increased scale of the network and performance guarantees required by applications that generate highly diverse traffic characteristics, severely increase the complexity of the system and its control algorithms; this complexity renders our networks more vulnerable to system failures.

At the same time, our dependence on networks is increasing and failures have acquired the potential to be catastrophic. Recent failures in the existing telephone network have not only interrupted service, but have also disrupted the financial and air traffic control operations. Fault resistant networks are more important now than in the past, and the means must be found to take advantage of the economics of new technologies and the demand for new services without compromising reliability. One function of the Exchange is to develop the facilities to simulate or emulate large scale networks and their control mechanisms in order to better predict the effect of new technologies and services on the reliability of the entire network before they are deployed.

A third important research area in future networking is that of protocol development. As the data rates of networks increase to support emerging applications, it is clear that existing protocols without modification are inadequate for the NII. In particular, flow and error control mechanisms that are based on low data rates do not scale well to networks with large bandwidth-delay products. Furthermore, as emerging applications require quality of service guarantees of different types (e.g. throughout, delay, and jitter) the protocols must change or evolve. Balanced against the introduction of totally new protocol mechanisms in a gigabit network is the fundamental requirement of networking to allow systems to interoperate; thus one must balance the desire to introduce new protocols against the need to migrate while maintaining interoperability.

Creating protocol structures for the NII requires a closer linkage between those developing and applying formal methods to protocols, and those evaluating architectural and implementation issues to identify and eliminate system bottlenecks. The facilities proposed in the Exchange will encourage these needed collaborations.

2.4 The Objective

The Exchange will provide a framework for information interchange for the coordination of research, development, and implementation of gigabit networks. It will bring together and closely couple the analytical and experimental researchers, developers, implementors, vendors, network planners and administrators, application developers, and users in order to establish a healthy synergism between all the parties. It will foster a sense of community across the entire set of scientific and engineering disciplines necessary for the development and deployment of the NII.

In conclusion, we advocate a near-term as well as a long-range objective for the Exchange. The near-term objective is to create an infrastructure for disseminating data, tools, models, and end results, in order to identify the best system architectures for implementation and deployment of the NII. We expect the provision of this infrastructure to spontaneously lead us towards the long-term objective: to enable the continuing development of a methodology for analysis, modeling, development, implementation, testing, and measurement of future networked information systems.

3. THE EXCHANGE

The Exchange will serve as the memory and the activator of the network research and development community. It will allow access, retrieval, and evaluation of the collected wisdom of this community.

In this Exchange, the following kinds of *objects* will be collected: data, tools, and models. Under an acceptance control policy, these objects will be placed into the *Libraries* of the Exchange: the *Data Library* the *Tools Library* and the *Models Library* They will contain full details of the environment in which they were created and in which they operate. Moreover, the Exchange will provide access to *Shared Experimental Facilities* to enhance the collaboration and interaction of its participants.

Below we discuss these items in more detail. However, we hasten to emphasize that the Exchange is more than just a collection of data, tools, and models. A further essential ingredient is that the objects in the Exchange operate together, so that they can be used to build complete systems that can be used to measure, analyze, simulate, and model realistic networks.

3.1 The Data Library

In any system design, it is important to understand what demands will be placed on the system. In the case of networks, this amounts to characterizing the traffic that will be carried. Little information is available on the traffic we can expect over gigabit rate networks or the behavior of emerging applications that will generate this traffic, but this information will be invaluable to people designing, simulating, modeling, and evaluating high-speed networks and protocols.

In order to gain such information, it is important to gather measured data from high-speed networks (including current testbeds and the evolving Internet), as well as to collect data generated from simulation or complex analytical traffic generators. Data will be needed from all levels of the network, including:

application behavior (such as address reference traces and procedure call history)

- event traces and overhead measurements of protocols and operating systems
- source traffic characteristics at the network interface

- behavior of network control mechanisms (including resource allocation, congestion control, and address resolution)
- utilization and queueing behavior of switches, gateways, and routers
- traffic characteristics on network links

Data can include aggregate traffic measured during experiments, and characterization of specific isolated traffic sources such as distributed computing applications, scientific simulation, visualization, and multimedia. As applications become more sophisticated and bursty in nature accurate characterizations become more important, since one of the important research areas is how these different types of traffic will interact in a large scale gigabit network, where performance guarantees are required for connection oriented traffic passing through fast packet switches.

To be a useful object of the Exchange, data will have to be represented using standard formats, so that it can be easily used by simulation and modeling tools in the Exchange. There will also be a standard procedure to describe the experiment under which the data was collected or generated, so that it can be used in a meaningful and standard way. This will also make it possible in some cases to replicate and validate the experimental data via other experiments or by simulation.

In the collection of data, it is important to define standard procedures and interfaces for taking measurements in both experimental and production networks. This will allow the use of shared monitoring tools, and will facilitate the use of remote monitoring. This will be an important element of simplifying outside access to experimental networks.

Furthermore, the Data Library will include benchmark algorithms and results. The existence of a standard set of experimental conditions and traffic models will facilitate the direct comparison of network architectures and protocols. Developing these algorithms is a major task, both because evaluation is possible along many dimensions (such as throughput, delay, and ability to deal with congestion), and because there are many factors that affect performance and that have to be controlled while doing benchmarking (such as application behavior, host and operating system architecture, communication protocol design, and network size and topology).

The need for universally available data sets and a standard set of experimental conditions for networking researchers can not be overemphasized. The essence of the scientific method is the reproducibility of experiments. It would be unthinkable to build theories in basic science upon experiments that had only been performed once; however, in networking this is common. Because of the complexity in describing the conditions and traffic sources used in most networking experiments, journal articles rarely contain sufficient information to reproduce and verify an experiment. The Data Library will provide a source for both the inputs and results of experiments so that the experiments can be reproduced and the results interpreted by independent researchers.

Finally, it should be noted that anomalous behavior of network elements and of control mechanisms should be an integral part of the Data Library. Experiments which have generated unexpected and unexplained results can be described in the Data Library, allowing other researchers to collaborate in the discovery of the cause of the anomaly.

3.2 The Tools Library

The second essential element in the Exchange is the set of tools for modeling, simulating, designing, and measuring networks. These include:

- measurement and data collection tools
- simulation packages and languages (for functional verification and performance analysis)

- network engineering and network element design tools
- analytical packages (such as for fitting mathematical models to experimental data and analyzing simulation results)
- visualization packages (for interpreting simulation results and for network engineering and management)

Making existing tools available on a wide scale will immediately be beneficial since it will allow research in high-speed networking without having to spend a substantial amount of time building these tools. However, the payoff will be greatly increased if the tools can work together, and can interact with the objects in the Data and Models Libraries.

Due to license restrictions for commercially available tools, it may not be possible to make these generally available in the Tools Library. This issue will be revisited in the context of the Shared Experimental Facilities subsection, but the provision of the interface between these and other tools will still be an important component of the Tools Library, allowing their use and interaction by those who have license rights or access through the Shared Experimental Facilities.

This ability to interoperate will, for example, make it possible to use a single visualization tool to explore and compare:

- measured data collected by various means
- the output of simulations generated by differing packages
- the output of network engineering and design tools
- the output of an analytical model

This will be particularly useful in verifying new models against collected data before varying the parameters of the model to explore new design choices. Furthermore, it will be possible to use a combination of measured data traces from the Data Library and traffic generators from the Model Library to drive large scale heterogeneous simulations with sub-models using various simulation packages and languages.

Finally, there will be the ability to vertically integrate the modeling process from high level behavior, to detailed component simulation, as never before. It will be possible, for example, to plug a gate level simulation reflecting a network element design into the original high level network simulation model, to confirm that a particular implementation has the same behavioral effect on the network as a whole as originally intended. Moreover, this ability can be extended to hybrid models in which a simulation or analytic model is used to replace part of a physical network; that is, a portion of a network can be disconnected from the rest of the physical network and replaced by a model of itself. This hybrid system can then study how the real system will behave when changes to the model take place.

This vision is not only very desirable, but also quite realistic. The key is to define standard interfaces and data formats that will allow tools to interoperate. These interfaces will hide sufficient internal detail so they can be fairly general and applicable to a wide range of networks, but they have to capture the essential elements of the interaction. Although defining these interfaces is certainly not trivial, we think that it can be done and it will be of very great value. Earlier efforts that have used this approach have shown this to be truly useful (e.g. in the cases of experimental ARPANET and Internet protocols documented in RFCs, and VLSI design tool interaction using standard interchange file formats).

3.3 The Models Library

A final element of the Exchange is the set of models to be used in the analytical and simulation process. These are models of the various network elements, including:

- application behavior and traffic generators
- session control (connection management and congestion control algorithms)
- communication protocols (including error control, flow control, admission control, and sequencing)
- operating systems
- host architectures and host--network interfaces
- switches, gateways, and routers
- network links (e.g., SONET, HIPPI, FCS) and free space channels

In each case, appropriate to a particular modeling effort, the level of abstraction must be carefully chosen to allow simulation runs to complete in a reasonable time or analytic models to be evaluated. This must be done with enough detail in the models to fully understand sensitivity to initial conditions and parameters, and to allow the variation of parameters whose values affect the performance metrics.

For example, a particular network element may be modeled at the various levels such as:

- high level behavioral models of network topology and data flow
- architectural models of systems and protocols at the functional level
- detailed gate level logic models that reflect the hardware design
- detailed analog and optical models of transceivers and transmission lines

The raw data in the Data Library is only a first step towards traffic source characterization. One goal is to develop traffic models for the Models Library that can be used to characterize applications under a variety of conditions, such as networks and hosts of different speeds, different loads, and different traffic conditions. Traffic models already exist for simple applications, e.g., voice, file transfer, remote login, and video distribution, but models for more complex applications such as distributed computing and interactive multimedia are needed. Application and traffic models will be an important part of the Exchange. The presence of accurate source models will make it possible to evaluate and model not just networks, but also the distributed systems in which they are used.

It should be noted that modeling serves two purposes: functional verification (of complex systems where correctness proofs are infeasible) and performance evaluation (where closed form analytical solutions are intractable). While individual network elements with simplifying assumptions may lend themselves to correctness proofs and closed form analysis, the complex systems of the sort we anticipate for the NII will certainly require far more extensive modeling and simulation.

The various models may be constructed by a variety of techniques, including:

- analytical models and traffic generators coded in a programming language or a queueing network language

- stand alone procedures (subroutines) that model specific network elements, custom coded in a programming language such as C

- models written in a discrete event simulation language (such as SIMSCRIPT or GPSS)

- queueing network models constructed with a package (such as RESQ)

- architectural models constructed with a package using a graphical user interface (such as BONeS or OPNET)

- gate level hardware logic models generated by schematic capture or written in a hardware description language (such as VHDL)

- analog and electro-optic models (for example using SPICE)

While having a library of models of systems, sub-systems, and network elements at various levels, constructed by different techniques, is an ambitious goal, the benefit will be enormous. Researchers will be able to duplicate the simulation experiments of others, explore the consequences of changes in simulation parameters, and construct new systems merely by plugging together models that others have built. These researchers will be able to "converse" using this Models Library and improve and extend the components in this library interactively. The potential gains from the resulting synergy of such an infrastructure is vast.

3.4 Shared experimental facilities

Research in networking, as in other branches of engineering and science, consists of a theoretical component that can be worked on by a small group with minimal resources, and a more practical component that requires expensive experimental facilities. Linking the theoretical and practical components in networking research is of prime importance since networking is an engineering discipline, and theory without validation has little meaning. However, the cost and span of experimental high-speed national networks makes it difficult, if not impossible, to perform all of the needed experiments, and to give access to experimental facilities to all of the researchers with the need. One way in which the Exchange can contribute is to establish Shared Experimental Facilities that lower the barriers to experimental networking research, by providing:

- access to shared national facilities

- shared national facilities for network simulation

- shared national facilities for network emulation

- safe, controlled access to individual experimental networks

- linking individual facilities to form larger heterogeneous internetworks

While Libraries will provide the data, tools, and models to allow the simulations of the scale we consider necessary to conduct research towards the implementation of the NII, additional infrastructure is highly desirable. There are two major components to this infrastructure: access and facilities. While we will concentrate on the latter, it is important to note that the links that make the access to these facilities possible (and whose characteristics may be useful for verifying experiments across real network links) are just as important as the facilities themselves. In many cases this access will be provided by the existing Internet, but there are scenarios which may require dedicated or higher bandwidth links than are generally available to individual researchers. We will now describe each of the facilities in greater detail.

Simulation is one of the most cost effective means of studying the operation of large scale networks. However, the size, bandwidth, and complexity of networks is increasing, and more and more often, the primary concerns are relatively rare events rather than only long term average performance measures. Here, approaches such as importance sampling and sensitivity analysis can offer significant insight, and the analyst who understands their application can benefit the developer in such cases. Often, the computer facilities needed to perform accurate simulations that are long enough to provide meaningful measurements will be beyond the means of many university departments and individual researchers in networking. The shared national facilities for *simulation* provide the infrastructure to allow simulations of scope and collaborative component not otherwise possible. These facilities will include:

- compute servers to allow accurate complex simulations to be run by individual researchers on powerful systems, or on instruction set architectures required by a particular software implementation
- specialized simulation engines and supercomputers which are not ubiquitous enough to be generally available to even large research groups
- software license servers allowing the efficient sharing of expensive commercial software

One of the most costly components in high-speed networking research is the cost of the transmission facilities; in addition, experimental networking research often precedes the availability of the transmission facilities. In a network *emulator* the characteristics of a transmission line (transmission rate, propagation delay, errors, outages, etc.) can be modeled by a hardware device to overcome these problems. These link emulators can then be connected together to form a network, and either general purpose computers or specialized hardware can be used as the nodes and switches in the emulated network. Users can be connected to the network emulator and experience virtually the same environment that they would in a national network, and the network emulator can be connected to an experimental network to extend its apparent size.

The difference between a network emulator and the actual network is that the emulator is constructed in a room, and results in less costly transmission facilities. A network emulator can be more easily set up, reconfigured, and torn down, than the actual network, and provides researchers with the opportunity to try many more experiments before deploying a geographically distributed network. In order to make networking research more cost effective and to make it possible for research to precede the deployment of transmission facilities, a network emulator will be constructed as part of the Shared Facilities of the Exchange.

An important part of the research in high-speed networks is the development of *experimental networks* that can be used to evaluate hardware technology, network protocols and algorithms, and applications, in a real system. Although these experimental networks are by their very nature limited in scale and often focused on a particular aspect of networking, they provide invaluable information that cannot be obtained by simulation, emulation, or modeling.

In other scientific disciplines, experimental prototypes are often used only by their designers. The nature of the networking discipline makes this approach both undesirable and impractical. Large scale networks include many different technologies that have to work together correctly. It will be useful to provide controlled secure outside access to these experimental networks, as well as to have the capability to link individual facilities and testbeds together to form a larger heterogeneous internetwork. The ability to try different protocols and applications on different experimental systems is an important part of experimental network research since it increases the confidence that the hardware, protocols, and applications will be able to function in a heterogeneous environment. This type of experiment will become more common as more researchers can gain access to experimental networks. Moreover, experimental networks are extremely expensive to build, so it is important to maximize the number of experiments that are run on each of them for economical reasons.

Sharing of experimental networks is already happening today, but it is often prevented or complicated by the fact that different groups use different (frequently ad hoc) tools and conventions. We expect that the Libraries, which include well defined interfaces for measurement and formats for data and algorithms, will facilitate this sharing of experimental facilities.

3.5 Organizational issues

To make the Exchange an effective mechanism for supporting the development of networked information systems, an organization is needed that: specifies interfaces and formats so that the objects in the Exchange can work together; performs acceptance control for the entry of data, tools, and models into the corresponding Libraries; and manages access to the Shared Experimental Facilities.

The objects in the Exchange will have to operate together so that they can be used to build complete packages for use in networking research. This requires the specification of standard interfaces between and among the tools and models and the definition of the data formats. Defining and developing the interfaces and data formats requires a concentrated effort by the research community. Note that defining the interfaces and formats is itself a research issue.

The Exchange will only have an impact if it used, i.e. if researchers make contributions to the Exchange and use the Data, Tools, and Models Libraries for their work. One of the key elements to success will be the ability to attract high quality data, tools, and models. Producing these contributions is of course the task of the research community, but the Exchange can encourage this in several ways. First, there will be minimum requirements for documentation that explains installation, use, and limitations of each tool or model and the format of data. Second, objects may be removed from the Exchange and archived if they fail to meet measurable use criteria. It will be possible to accommodate not only contributions that are well established, but also those that are in their early stages of development (and hence with a limited track record). Finally, issues related to copyright and access control will have to be addressed. Note that there is a large body of tools and models generated by researchers who are willing to share freely, but there is currently no infrastructure to support their wide dissemination. In fact, it is likely that the submission of a data set, tool, or model to the Exchange will be regarded as a form of publication, which will further encourage such activity.

The above issues are very important since they will have an impact on how attractive it is to use the Exchange, and hence on how much impact it will have. However, there are many existing examples of sharing of tools and data on a smaller scale that indicate that this is a realistic scenario: mathematical libraries, personal computer shareware, X applications for Unix, information exchange through newsgroups and bulletin boards, etc.

The Exchange will not only deliver a broad set of tools, but it will help in the development of a methodology for designing, measuring and analyzing large-scale, heterogeneous networks that will evolve, backed up by a set of widely available tools.

4. WHERE THE EXCHANGE CAN HELP: SOME EXAMPLES

Emerging applications addressing the national and grand challenge problems (including multimedia) combined with new methods of integrating networks and operating systems (such as distributed virtual shared memory) themselves pose new research issues in terms of scaling and heterogeneity. Problems of scaling arise due to increases in latency (or the bandwidth-delay product), link rates, node counts, the number of concurrent connections, and the number of parties participating in a communications context. Current testbeds are limited in their scope and cannot easily address these, in particular the node and participant count. As we begin to understand and characterize application behavior in the context of the emerging high-speed networks, we can feed this information back into application design and optimization to perform well in the NII, in addition to properly engineering the NII infrastructure.

In this section, we examine examples of these problems and how the Exchange will serve to assist in their solution in greater detail, starting at a high level problem (national challenges), proceeding to an application (multimedia), and finally to system level support (distributed virtual shared memory).

4.1 National Challenges

Recently, ARPA has defined a set of national challenges for the NII. National challenges are characterized by their large economic impact, broad applicability, and the potential to introduce new services and processes rather than just automate existing ones. Examples of national challenge applications include design and manufacturing, crisis management, health care delivery, education and training, digital libraries, and environmental monitoring. These national challenge applications demand the high-speed networking infrastructure that the NII will provide, as do the previously defined grand challenge problems, but with the potential for larger scale and penetration into society.

In terms of networking requirements, these applications are characterized by large scale and heterogeneity of end system architectures, transmission media, and rates. They also encompass many services not traditionally provided by data networks, such as real-time services (for example, for remote medical consultation or crisis management) and stringent security and privacy requirements (for example, for medical record transfers). Due to their interdisciplinary nature, it can be expected that a large number of research institutions will be working on these applications, many of which may not have a strong networking or performance evaluation background.

The Exchange can aid in the development of the necessary applications, information and communication services, and host interface and network infrastructure to address these national challenges, as illustrated by the examples in the following two subsections. The Model and Data Libraries will store model applications and sample traffic traces so that researchers working on services and infrastructure will have realistic application requirements to work from, even though they may not be intimately familiar with the applications generating the data. In turn, application developers can make use of base models of commonly needed communication services and infrastructure without knowing all the details of their implementations. Measurement results will provide a baseline against which alternative approaches to building applications and delivering services can be compared.

The Shared Experimental Facilities will provide the common proving ground for these applications and services where interoperability, service interaction, and scaling can be investigated before they are actually deployed. These facilities will offer the venue for collaboration between industrial, academic, and government investigators.

4.2 Multimedia

Distributed multimedia applications are widely regarded as an important use of high-speed networks, including their use as an integral part of national challenge solutions; these applications include teleconferencing, collaborative work, remote visualization, virtual reality, distributed simulation, and interactive video-on-demand. These services share a number of demanding requirements, including real-time constraints; an example of such constraints is the necessity for delivery to the receivers within a specified delay bound and the need for the various streams (video, audio, and data) to be sufficiently synchronized. Furthermore, these sources transmit large amounts of data, either as raw or compressed video, or as data that will be rendered into images. Often video is multicast to a large, widely distributed group of receivers.

The requirements for real-time multimedia differ substantially from the data services currently dominating packet switched networks. A range of open problems remain before distributed multimedia applications can be deployed on a wide scale. They range from problems of insufficient physical bandwidth to problems in the design of operating systems and user interfaces; below we will address some of the network related issues:

- Quality of service guarantees: Current packet switched networks typically cannot make any guarantees to an individual connection as to the quality of service it will receive (such as throughput, delay, and loss). In the case of audio if the delay exceeds a few hundred milliseconds, conversation between two parties becomes difficult, and in the case of video broadcast, packet loss must be low enough so as not to perceptibly degrade image quality; in both cases retransmission of lost packets is generally not feasible. Related to the quality of service requirements are the resource reservation and routing algorithms that are necessary allow the network to provide these guarantees.

- Congestion control: Congestion control measures for data services such as window flow control are not directly applicable to multimedia services, with rate control and rate feedback appearing more promising. The algorithms for resource reservation and routing of multipoint connections to prevent and control congestion targeted at large complex networks must typically be implemented in a distributed manner.

Both of these issues are related to network scale and complexity. It is difficult to predict how proposed algorithms and mechanisms will scale to the complex heterogeneous NII under real conditions of latency, packet loss, and congestion.

- Heterogeneity: Multimedia end systems will consist of various architectures and will be connected by diverse subnetworks and switches, with latency varying over several orders in magnitude in the "Global Grid".

Research into issues (such as those listed) related to distributed multimedia services will be greatly enhanced by the availability of the common Data, Tools, and Models Libraries, as well as the infrastructure provided by Shared Experimental Facilities. Since the range of problems within the topic of distributed multimedia is so wide, common Libraries and Facilities will make it possible for smaller research teams to contribute, without each needing expertise in all the required disciplines (e.g., signal processing and operating systems) and without having to invest in a full set of expensive multimedia production and processing equipment.

Many approaches to the scaling, congestion control, and heterogeneity problems depend on the source characteristics (such as burstiness). However, generating multimedia streams, particularly for interactive video or visualization, is expensive. For example, encoding video through MPEG with motion compensation requires the services of specialized hardware or a supercomputer. Also, experiments and performance evaluation are difficult to compare or replicate as the source streams used are usually not accessible to other researchers and not fully described in publications. In the image coding community, a small set of images is used to indicate the achievable coding quality and compression ratio; similarly, benchmark programs are used to compare the performance of processors and computer architectures. Such images, programs, and video clips (e.g., the first few minutes of the movie *Star Wars* will be included in the Exchange and will be useful for theoretical analysis, simulations, and experiments in test networks, thus providing a means of attacking problems with a common set of assumptions.

Static source data by itself, however, may not be sufficient, as there is an increasing realization that interaction between the network and the source can improve performance. For example, a video source may reduce the image quality and bit rate when the network notifies it of pending congestion. Applications that can generate multimedia streams in real time will speed the research into these interactions. The set of conferencing tools used for Internet audio and video conferences already shows the usefulness to network research of generally available applications for traffic generation.

The Models Library will contain simulation models of multimedia sources, enabling more realistic network scenarios without having to fall back on the usual greatly simplified source models. This will allow easy reference, avoiding the need for elaborate or incomplete descriptions in simulation studies and publications.

4.3 Distributed Shared Virtual Memory

The use of distributed shared memory as a way to provide the shared memory paradigm for interprocess communication has received significant attention in the research community, particularly in the context of distributed memory multiprocessors and local area networks. A promising way to provide distributed shared memory over wide area networks is to use extensions of virtual memory mechanisms to give the application the appearance of shared memory. This is referred to as distributed virtual shared memory (DVSM), and it is the memory mapping and management in the operating system that are extended to deal with pieces of code and data located throughout the network.

Emerging national and grand challenge applications will rely more heavily on multimedia, as described above. Currently, even though systems may provide integrated transport of data, audio, and video, these streams are split at the host--network interface and handled in a very primitive manner. We need to understand how integrated multimedia can be handled in the host and operating system, including how audio and video objects fit into the virtual memory mechanisms, and how frame buffers can be integrated into the host memory hierarchy. Furthermore, a potential benefit of DVSM is related to the efficiencies of a coordinated design between the operating system and transport protocol mechanisms.

There are many operating system, DVSM, and communication protocol policies and design alternatives, as well as competing host memory and network interface architectures that need to be compared by detailed simulations and experimental analysis. In addition to the system design issues, a number of open research issues at the network level remain to be solved in the context of DVSM:

- Latency: One of the hardest problems to overcome when distributing communicating applications over a long haul network is that of the high round trip delay. Practical constraints such as the physical location of specialized facilities frequently prevent the local clustering of communicating objects. Some virtual memory operating systems are based on objects (segments) that are related to the application semantics, providing a natural and automatic grouping of related code and data; by extending these mechanisms to the network environment, a natural prefetching can occur so that the objects are present before use, eliminating the round trip delay.

- By instrumenting a range of current high performance applications and collecting data such as address reference traces and the interaction between communicating processes and their use of objects, the Data Library will contain the necessary information to provide realistic traffic generators to drive simulation models. The interaction of DVSM implementations across the Shared Experimental Facilities will allow proposed solutions to be tested in a realistic network environment.

- Data rate: As application demand, processor power, and network infrastructure all grow in performance, a potential exists for mismatch between the host processing power and network data rates. While it is possible (and desirable) to model the network at a high level of abstraction for some of the DVSM work, it will also be necessary to test the proposed host and operating system models against live network links in the presence of realistic congestion and error conditions, as provided by the Shared Experimental Facilities infrastructure.

- Heterogeneity: In a general application of DVSM, the participation of diverse heterogeneous systems is desirable. While a few specific types of heterogeneity are relatively easy to deal with (such as different page sizes and byte ordering), the ability for completely different system architectures to participate in a DVSM is very much an open research issue. By providing the Shared Experimental Facilities and the ability for differing researchers to cooperate, the Exchange will significantly enable the ability to conduct DVSM heterogeneity experiments.

- Network scale: In DVSM implementations, multiple processes will be communicating among one another, and it will be very important to show that proposed solutions scale to a large number of participants. Furthermore, algorithms will be required that determine optimal object migration policies and the coherence of the data in objects where shared write access is required. The network scale provided by access to the Shared Experimental Facilities to test applications will be essential in showing how these algorithms, along with the DVSM mechanisms in general, scale to extremely large high-speed networks.

The implementation of DVSM and related system level mechanisms, based only on the restricted setting of experimental networks, could result in the premature deployment of fundamentally flawed hardware and software mechanisms. If these mechanisms break down under the scale present in the future NII, the result could be extremely costly, and require changes in thousands of workstation computers. This sort of costly consequence can be avoided by fully exploring scale and heterogeneity effects up front. The Exchange will be very well suited to enabling this type of early exploration to prevent just such consequences; moreover, where they to occur in a deployed network, then once again, the knowledge and wisdom gained from the Exchange as well as the use of the Exchange itself can serve as a very effective mechanism for locating the problem and suggesting its remedy.

The grand scope and detail of the modeling needed for the solution of difficult research and implementation problems from applications to network architecture in the environment of the NII requires collaborative efforts that will be enabled by Shared Experimental Facilities. Similarly, by using and extrapolating data from the current Internet and gigabit testbeds, we will have the basis for creating realistic models of large high-speed networks to be used in simulations, assisted and enabled by the Data and Models Libraries.

5. CONCLUSIONS

This white paper has but one recommendation to make. That recommendation is simply to:

Promote the development of a detailed plan to create this
National Exchange for Networked Information Systems

This is but a first step in creating the Exchange. Once the plan has been created and approved, only then should an implementation phase be launched.

We have provided a vision for creating that portion of the National Information Infrastructure which will enable the community of researchers, developers, vendors, and users to collaborate and interact through a National Exchange for Networked Information Systems. The Exchange will provide the necessary mechanism for this community to address the key issues of scale, heterogeneity, and dynamic behavior of the emerging high performance networked information systems. The ability to address these issues is critical if we are to continue to lead the world in these advanced networks.

Having laid out the vision, the next step is to flush it out with a detailed plan. We propose that such a plan be developed.