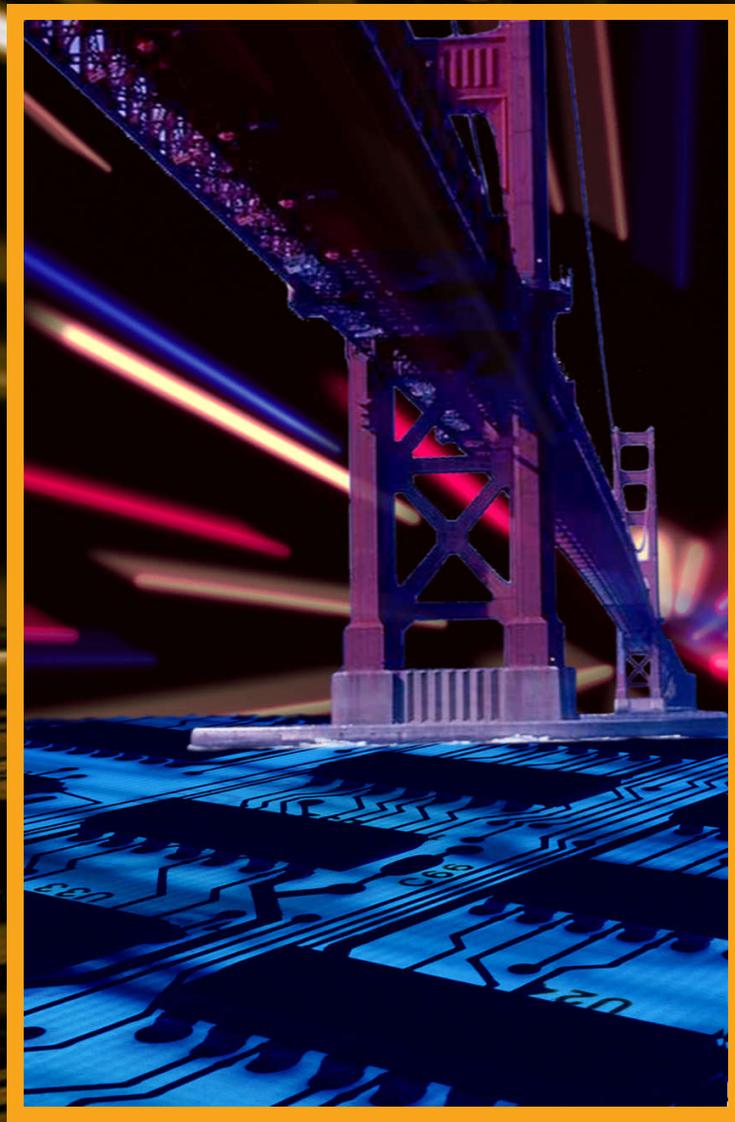


BRIDGING THE GAP FROM
N E T W O R K I N G

technologies ^{TO} applications

1999 WORKSHOP REPORT



N A S A A M E S R E S E A R C H C E N T E R

TABLE OF CONTENTS

1

3 EXECUTIVE SUMMARY

- 3 **Workshop Purpose and Goals**
- 3 **Selection of Technology Topics and Case Studies**
- 4 **Technology Area Outcomes**
- 5 **Next Steps**

7 QUALITY OF SERVICE (QoS) TECHNOLOGY

- 7 **Introduction**
- 7 **QoS Technology Landscape**
 - 7 Data Path Resource Management
 - 8 Resource Allocation, Signaling and Admission Control
 - 9 Path Selection / Routing Technologies
 - 10 Middleware Services / Interface Technologies
 - 11 QoS Service Definition / Construction and QoS Systems
- 12 **QoS Testbed Activities**
- 13 **QoS Technology Road Map**
- 18 **Application Guidelines**
- 18 **Significant Barriers to Development, Deployment and Adoption of QoS Technology**
- 20 **QoS Technology Appendix A - Application Analysis of QoS Requirements / Constraints**
- 20 **QoS Technology Appendix B - Application Questionnaire**

23 MULTICAST TECHNOLOGY

- 23 **Multicast Technology Overview and Current Status**
- 23 **Current Status from the Perspective of Application Developers**
- 24 **NGI Testbed Status**
- 25 **Multicast Technology Roadmap**
- 27 **Application Roadmap**
- 27 **Multicast Technology Appendix A - References**

29 SECURITY TECHNOLOGY

- 29 **Introduction**
 - 29 NGI Application Context
 - 29 Security Requirements for Networked Applications
- 31 **Status of Selected Security Mechanisms**
 - 31 Encryption
 - 31 IPsec
 - 32 Kerberos
 - 32 Public Key Infrastructure (PKI)
 - 32 Secure Socket Layer/Transport Layer Security
 - 32 Generic Security Services
 - 33 Authorization and Access Control
 - 33 Secure Group Communication
 - 33 Integrated Solutions for Large Scale Distributed Application Environments
- 34 **Security Guidelines for NGI Application Developers**
- 35 **Security Technology Road Map**
- 40 **Security Appendix A - References and Notes**

APPENDICES

- 43 **Appendix A: Agenda**
- 45 **Appendix B: Technologies**
- 51 **Appendix C: Application Case Studies**
- 81 **Appendix D: Demonstrations**
- 95 **Appendix E: Acronyms**
- 99 **Appendix F: Attendees (agencies, NGI/I2/PITAC)**
- 101 **Appendix G: Contacts**

WORKSHOP PURPOSE AND GOALS

The “Bridging the Gap from Networking Technologies to Applications” Workshop was held August 11-12, 1999 (see <http://www.nren.nasa.gov/workshop4.html>). The purpose of the Workshop was to facilitate convergence between the research objectives of the networking technologists and the networking requirements of the applications community. The workshop was hosted by the NASA Research and Education Network (NREN) project at the request of two Large Scale Networking Working Group (LSN) teams: the High Performance Network Applications Team (HPNAT) and the Networking Research Team (NRT). These teams, which consist of representatives from the six NGI agencies, the university Internet2 project, and additional Federal agencies interested in advanced networking, are responsible for coordinating NGI programs. The workshop was also supported by another NGI team, the Joint Engineering Team (JET), responsible for coordinating agency and Internet2 research network interconnection and interoperability.

The aim of the workshop was to bring the applications and network technology communities together to discuss the state of the art in selected network technologies, to identify application requirements in the three technology areas, and to identify activities that need to take place during the next one- to three-year timeframe in order to fulfill the promise of the NGI program.

The workshop was aimed at producing four types of outcomes:

- **Technologies** - identify and characterize the promising existing and emerging technology solutions in each of the three selected technology areas: Quality of Service (QoS), Multicast, and Security.
- **Applications** - to identify and describe promising applications, both specific and generic by type, and what new technology capabilities they need, and on what timelines.
- **Testbeds** - to identify the key testbed activities required to demonstrate the applications, whether existing, planned, or identified as needed.

- **Issues** - to identify other disconnects (e.g., program or agency focus, funding, research agendas, etc.) that would inhibit or could promote the deployment and integration of the needed technologies into testbeds and applications.

SELECTION OF TECHNOLOGY TOPICS AND CASE STUDIES

Three technology areas were selected for the workshop: QoS, Multicast, and Security. These three areas are the focus of a significant amount of research by the NGI agencies and collaborating partners, and the emerging technology solutions in these areas are crucial to achieving the application goals of the NGI program within the next one to three years. “Bridging the gap” between the efforts of the technology community and the needs of the application community in these three areas was identified by the HPNAT and NRT teams as the highest priority coordination need for the workshop to address.

The fifteen application case studies (see <http://www.nren.nasa.gov/case.html>) were selected to be representatives of nationally important broad application areas as well as having requirements that need the emerging technologies selected for the workshop. Applications were selected from the areas of Digital Earth (NASA, NOAA, DARPA, DOE), Digital Video (Internet2, NSF, DARPA), Telemedicine (NIH, NASA), as well as two distributed data intensive applications from DOE, and three applications from the NSF Partnership for Advanced Computational Infrastructure (PACI) Program. Brief descriptions of each application case study are given in Appendix C.

The first day was kicked off with a keynote by the DARPA Chief Scientist, Dr. David Tennenhouse, and was spent primarily in plenary session. This session was intended to provide participants with oversight tutorials of the three technology areas as well as brief overviews of the fifteen application case studies. During the workshop breaks and

SELECTION OF TECHNOLOGY TOPICS AND CASE STUDIES cont.

in the evening of the first day, live demonstrations were provided of several of the application case studies (see Appendix D for descriptions of the demonstrations). The second day was spent in technology breakout sessions meant to achieve the needed depth to produce usable and effective technology roadmaps and application guidelines.

TECHNOLOGY AREA OUTCOMES

Each of the three technology area breakout groups—QoS, Multicast, Security—accomplished a great deal in their brief formal session time. Recorders kept notes on the discussions, and in the month following the workshop, the three technology area facilitators—Doug Montgomery of NIST for QoS, Marjory Johnson of NASA/NREN for Multicast, and Bill Johnston of NASA and DOE with the assistance of Matt Chew Spence of NASA/NREN for Security—produced extensive reports on their technology areas. Their three reports follow this Executive Summary, and form the substance of this workshop report. Each technology area report includes a landscape survey of important current activities, a set of roadmaps for the future, and guidelines for applications developers.

In each of the three technology areas, a critical mass of research activities is underway, spanning the near-term technology landscape in each of the three areas. The three technology area reports that follow give one a useful understanding of the scope of each area and the technology activities that are underway.

QoS

Discussions in the QoS breakout group focused on an examination of current directions in QoS technology research, standardization and development efforts, and characterization of the emerging technologies in terms of utility, deployability, and readiness. As an organizational aid, the overall QoS landscape was parsed into key component functions and mechanisms, including data path resource management; resource allocation, signaling, and admission

control; path selection/routing technologies; middleware services/interface technologies; QoS service definition/construction and QoS systems.

Details pertaining to individual QoS mechanisms are presented in the QoS technology roadmaps. Major issues that remain to be resolved within the QoS area include determining how QoS interacts with other emerging technologies, inter-domain/multi-administration issues, determining how to characterize QoS requirements for applications, the difference between QoS in the NGI community and QoS in the commercial world, requirements for QoS in very high bandwidth networks, and the importance of access to a persistent, instrumented testbed infrastructure for QoS experimentation.

Multicast

Multicast technology is more mature than QoS, even though it is not widely used in current NGI applications. The primary message from the multicast breakout group is that the NGI testbed infrastructure can support basic multicast applications now. The major hurdle to deployment of multicast applications is that the campus network infrastructure is not multicast enabled. This situation is likely to be corrected within the next year.

Major multicast research issues include scalability, protocol complexity, standardization of reliable multicast protocols, and TCP-friendly flow and congestion control. Progress towards resolving these issues will come from efforts within the research community to develop appropriate protocols, from the IETF to standardize approaches, and from the NLANR engineering services group to assist with multicast enabling the campus infrastructure and with providing assistance to application developers.

Security

The security breakout group examined security requirements for several broad classes of NGI applications. Core requirements that were discussed included data confidentiality, user and system authentication, authorization and access control, data integrity, and system availability. Somewhat more specialized requirements

TECHNOLOGY AREA OUTCOMES cont.

included non-repudiation, anonymity, and group-oriented security. All of these requirements fall within the scope of the body of current technology within the security community. Specific mechanisms that address these requirements were examined in depth; details appear in the security technology roadmaps. Many of these individual mechanisms are usable today, or will be in the near future.

The working group also examined the Globus distributed computing system, an approach to providing an integrated security environment to application builders by providing a set of security services on top of several independent security technologies. The group concluded that for most application developers in the widely dispersed and multi-organizational environment of the NGI, some sort of an integrating layer like the Globus software will probably be necessary to provide comprehensive security services for applications. However, for applications that only need some specific security services, several of the identified technologies have implementations that are ready now to provide these capabilities.

NEXT STEPS

Following the workshop, the lessons learned and so what? were discussed within the workshop organizing committee and within the leadership of the sponsoring teams. Some informal conclusions from these discussions are presented in this section, to aid the NGI teams and the LSN in implementing the consensus next steps.

In the technology areas of the workshop, the NGI Networking Research Team (NRT) is responsible for coordinating the NGI technology development and deployment plans in QoS and Multicast across agencies and with the Internet2 community, while the NGI Internet Security Team (IST) is responsible for coordinating the security technology area. Using the roadmaps, status updates, and issues contained in this report, NRT, IST and JET should work to encourage more agency deployment of key NGI technologies according to timelines coordinated across agencies. This will enable

demonstrations and interoperability of NGI technologies across Next Generation Internet Exchanges (NGIXs), between agencies, and with Internet2 universities, which are highly important goals of the NGI program.

For applications, the NGI High Performance Network Applications Team (HPNAT) is responsible for cross-agency coordination, demonstration, and reporting of NGI technology-enabled applications. Using the application guidelines provided in this report, HPNAT should encourage agencies to produce plans for getting selected key technologies deployed, integrated into their priority applications, and demonstrated at venues such as the SuperComputing conferences. This will provide the primary program management means for coordination and reporting of the end-to-end application demonstrations required by the end of the NGI program.

It was clear from the application case studies presented that DOE, NSF and DARPA have special integrative roles to play in the NGI program because these three agencies have broad programs in all three of the NGI goal categories: technologies, testbeds, and applications. With their broad scopes, these agencies sometimes seem to do it all, producing complete integrated examples of NGI solutions such as the “corridor” programs within DOE and the PACI program within NSF. These integrative NGI activities are crucially important because it has proven so difficult for the smaller agencies such as NASA, NOAA, and NIH to produce complete NGI solutions. The importance of this workshop is that it provided opportunities for agency NGI activities to be presented to the broader NGI community, including all the NGI agencies as well as Internet2 universities. Of equal importance is the fact that the workshop brought together the technology and applications communities from all agencies and partners, getting them talking to each other in their own languages. The technology landscape surveys contained in this report will be very useful to help the two communities communicate with each other about the functionality and applicability of the emerging technology solutions. The workshop roadmaps and application guidelines will make it possible to develop examples of integrated

NEXT STEPS cont.

NGI solutions across all NGI agencies and Internet2 universities. Different approaches for facilitating this type of integration include incorporating technology pieces from NGI programs such as Globus, involving broad agency application communities such as Digital Earth, or utilizing large testbeds such as the QBone.

The NGI and Internet2 testbeds have a crucial role to play, because the testbeds need to deploy the technology infrastructure to support the applications before the agencies and partners can implement integrated end-to-end application demonstrations. Crossing network domains is a key issue with all three technology areas, so collaboration among the agency and partner testbeds, as well as testbed collaboration with the technologists and application communities, is critical to the successful development, deployment, and end-to-end integration of the NGI technologies and applications.

The Large Scale Networking Working Group (LSN) is responsible for agency coordination of the NGI program as a whole, including the coordination of agency reporting to Congress and to the President's Information Technology Advisory Committee (PITAC). Using the technology activity descriptions contained in the workshop report, LSN can promote improved cross-agency information dissemination and reporting. Whereas this workshop involved only about 120 selected participants, LSN could encourage the NRT to organize more widespread teleworkshops and teleseminars in the NGI emerging technology areas described in the report. LSN could also encourage the JETnets to focus on coordinating the needed interoperability across NGI testbeds in the identified technology areas. LSN could also improve its reporting across agencies by using technology area themes such as QoS or multicast, or application area themes such as Digital Earth or Digital Video, to assemble cross-agency reports of thematic progress, to complement the individual agency reports of agency progress, in the NGI program.

Of equal programmatic importance to the above but outside the scope of the workshop, the LSN needs to encourage

increases in application funding within the agencies, as required to achieve the stated goals of the NGI program. As exemplified by the 15 application case studies presented at the workshop, it has proved problematical and expensive to get NGI technologies integrated into leading-edge applications, especially when running over more than one agency testbed. More programmatic focus needs to be placed by LSN and its agencies on getting deployable NGI technologies implemented within all the NGI agencies and Internet2 partner universities, made interoperable across all the NGIXs and NGI testbeds, and used to demonstrate important end-to-end applications. LSN may need to discuss among its agencies and with Internet2 program management the possibility of an NGI version of the "early IETF style" forum, in which agency and university researchers in application-driven thematic areas such as Digital Earth or Digital Video, or technology-driven thematic areas such as those highlighted in the workshop report, would get together electronically on a routing basis to develop and carry out consensus activities.

This Bridging the Gap Workshop Report gives excellent summaries of current activities in the three NGI technology areas—QoS, Multicast and Security. It also illustrates how the technology pieces should fit together and support the NGI applications. This report, together with the dialog and understanding achieved between the technologists and the applications community accomplish the workshop goal of Bridging the Gap between the two NGI communities.

INTRODUCTION

There is almost universal agreement that Quality of Service (QoS) is a key issue in the development of the Next Generation Internet (NGI). Unfortunately, many of the parties that share this opinion have completely disjoint views of what QoS is, what problems it addresses, and how it will be used. For example:

- *Users and applications developers want QoS to control the behavior of the network beneath their systems.*
- *Network managers want QoS to control how applications utilize the resources of their cloud and to differentiate service levels for specific customers.*
- *Policy administrators want QoS to control and regulate access to the network services that they have paid for.*

Each of these interest groups may well have limited views of the capabilities and applicability of emerging QoS technologies. Likewise, QoS researchers may have limited exposure to the requirements and expectations of the NGI user community. In order to bridge this potential gap, the QoS session focused on fostering a better understanding of:

- *How emerging QoS technologies will mesh with application requirements.*
- *How applications will adopt/adapt to QoS technology capabilities.*
- *How and when QoS technologies and enabled applications will get deployed in leading edge research networks.*
- *What issues (technical, administrative, programmatic, and social) stand in the way.*

The QoS breakout session was organized around presentations on current directions in QoS technology research and development, presentations on selected QoS enabled research networks, and detailed analysis of QoS requirements for selected applications. Technology roadmaps were then developed by capturing the information, insights and issues identified in the technology presentations and application analyses.

QoS TECHNOLOGY LANDSCAPE

Technology presentations during the breakout session focused on describing the current directions of QoS research, standardization and development efforts, and characterizing the emerging technologies in terms of utility, deployability, and readiness. The information in this section was synthesized from these presentations and the ensuing discussion.

The QoS technology landscape is rapidly evolving and is currently the subject of much research and development, standardization, testing, hype and misunderstanding. Many applications in the workshop cited a need for, or plan to experiment with, specific QoS technologies (e.g., Differentiated Services, Integrated Services, RSVP, MPLS). But, before one can understand and evaluate any single “solution,” we must understand the component functions/mechanisms that are integrated and/or overlaid to comprise individual QoS technologies. During the workshop the overall QoS landscape was parsed into the following key component functions and mechanisms.

Data Path Resource Management

The most basic QoS mechanisms are those responsible for resource management in the data path processing (i.e., forwarding) of individual packets. These mechanisms are based on fundamental techniques in scheduling, buffer management and metering.

- **Packet Classification** - *All QoS technologies must be able to organize the flow of data into groups [i.e., forwarding equivalence classes (FECs)] that are to receive specific treatments. The granularity of classification can range from identification of individual application flows to grouping of large flow aggregates in the core networks. Classification is relatively straightforward in QoS technologies that permit explicitly “marking” flows and/or aggregates (e.g., IP TOS/COS, DiffServ, MPLS, IPv6) in packet headers. In other technologies (e.g., IntServ) routers must perform multi-field (MF) classification by matching flow templates against other protocol fields in the data stream. Critical issues in packet classification schemes are the ability to support aggregation/*

Data Path Resource Management cont.

deaggregation of flows and the ability to scale to a large number of groups. Recent research and development has focused on the design of high performance MF classifiers (inspired in part by requirements for high performance firewalls) and allocation/aggregation issues in DiffServ codepoints (DSCPs) and MPLS labels.

- **Metering, Policing, and Shaping** - In most QoS technologies the ability to ensure some level of performance characteristics is dependent upon the input traffic conforming to a specific profile or service level specification [e.g., committed information/access rates (CIR/CAR), flowspecs]. Metering, policing and shaping mechanisms are used to ensure that flows/aggregates conform to these agreements. Metering mechanisms measure input traffic, policing mechanisms handle traffic that violate the agreements by either discarding packets, marking them (e.g., rewriting the DSCP) as being “out of profile” (thus getting different treatment from buffer management and scheduling mechanisms), or by reshaping the traffic (e.g., using a leaky token bucket filter) so that it does conform. Typically such mechanisms are employed at boundaries between administrative domains and/or flow aggregation points within a network.

- **Buffer Management** - Buffer management mechanisms are responsible for ensuring that sufficient memory is available for packets queued for transmission. Buffer management mechanisms may interact with QoS resource allocation schemes and policing and shaping mechanisms. In situations in which queue sizes exceed allocations, congestion management schemes must discard packets and scale back contributing sources. Basic mechanisms focus on fair scaling of TCP sources through early packet discard [e.g., Random Early Detection (RED)]. Such techniques can be coupled to priority/scheduling/policing mechanisms [e.g., Weighted RED (WRED), and RED with In and Out (RIO)]. Other approaches signal congestion by means other than packet drops, but require modification to end-to-end protocols to react to this [e.g., Explicit Congestion Notification (ECN)].

- **Scheduling** - Fundamentally, QoS technologies classify data flows into separate queues based upon their service requirements. Scheduling mechanisms determine the order in which packets from multiple queues are transmitted over a physical interface. Given that queuing delays are the primary source of QoS variance in typical internets, the behavior of specific scheduling mechanisms dominates the performance properties of individual flows. Used in isolation, common scheduling mechanisms [e.g., Class Based Queuing (CBQ), Weighted Fair Queuing (WFQ), Priority Queuing (PQ), Weighted Round Robin (WRR)] are capable of implementing simple prioritization and isolation of distinct flows/aggregates and insuring bandwidth allocations among classes of traffic. Scheduling mechanisms must be integrated with resource allocation, signaling, and other data path management components to realize end-to-end QoS service models (e.g., Intserv, DiffServ).

Resource Allocation, Signaling and Admission Control

The data path resource management mechanisms allow control over access to the queue buffers and the interface bandwidth of network nodes so as to implement specific QoS behaviors for selected groups of traffic. The issue of establishing which groups of traffic get access to what resources and when are addressed by resource allocation (RA) and admission control (AC) mechanisms. Fundamentally, resource allocation can be achieved by either: (1) provisioning new physical resources to meet projected demands, (2) employing out-of-band or manual configuration of data path mechanisms, or (3) through dynamic signaling protocols coupled with admission and policy control mechanisms.

The resource allocation problem can be characterized by: (1) the scope over which allocations will be made (e.g., end-to-end, across a single network, on a single node); (2) the granularity of allocations (e.g., individual application data flows, aggregate traffic classes); and (3) the time scale of the allocation (e.g., lifetime of individual instances of

Resource Allocation, Signaling and Admission Control cont.

communication, network engineering cycles). Scope, granularity and time scale are system design choices that can have a significant impact on the complexity and overhead of the RA/AC mechanisms used and the resulting services that can be implemented.

- **Offline Allocation Mechanisms** - Basic static, QoS managed services (e.g., CARs, coarse bandwidth allocation and priority schemes) can be accommodated using simple configuration mechanisms. Commercial tools are available to enable off-line design and implementation of engineered QoS services.
- **Dynamic Allocation Mechanisms** - Dynamic resource allocation protocols represent the control plane for advanced QoS systems. These protocols perform one or more of the following functions: (1) convey QoS requests among end-nodes, router/switches, links, and network clouds; (2) return admission control results; (3) collect path attribute information for the user; and (4) convey information between intermediate network elements necessary for service construction.
- **Resource Reservation Protocol (RSVP)** - RSVP is a soft-state signaling protocol for the establishment of flow state information in intermediate nodes along a given path. Typically, the state in question is the allocation of resources to a QoS controlled flow. RSVP may be used either end-to-end to allocate resources to specific application flows, or within a single network to control intra-domain path segments. RSVP was originally designed as an end-to-end signaling protocol and as such, supports each of the four functions defined above. It also supports multicast sessions among multiple heterogeneous participants. Recent extensions have capitalized on the protocol's extensibility to add support for explicit routing and label distribution for MPLS.
- **Bandwidth Brokers** - The bandwidth broker (BB) model abstracts and isolates resource allocation and access control functions within individual domains. Service

construction is implemented by signaling among the BBs in each domain along a path. Each BB performs admission control and configures data path management functions in the necessary routers (e.g., leaf and edge routers in DiffServ clouds). The methods and means by which these internal nodes are configured is a local issue within each domain. This approach allows individual domains to implement services in distinct ways.

- **Admission Control Mechanisms** - When processing dynamic resource allocation requests, nodes must decide if a new flow can be admitted without violating the QoS constraints of already established or reserved flows. Admission control [AC or call admission control (CAC)] mechanisms take as input (1) a traffic profile that characterizes the requested new flow, and (2) characterizations of the already established flows. The traffic profile consists of models and parameters that describe the traffic workload that will be offered by a flow. Overly simplistic traffic profiles lead to loose performance bounds and/or under-utilization of network resources. Overly complex profiles allow for tighter performance bounds and higher utilization, but are often difficult for application designers/users to provide. Admission control mechanisms can be classified as either parameter or measurement based. Parameter based mechanisms make admission decisions based on the specified traffic profiles of established flows, while measurement based mechanisms measure the characteristics of existing flows.

Path Selection / Routing Technologies

Typically, data path mechanisms and resource allocation systems are designed to be independent of path selection and routing technologies. While this independence allows for incremental design and deployment of QoS systems over the existing routing infrastructure, it is not the case that path selection has no potential impact on QoS. Routing technologies that are integrated with, or support explicit QoS mechanisms can enhance overall QoS system capabilities by: (1) achieving better utilization of network resources and avoiding contention, (2) computing routes to meet specific

Path Selection / Routing Technologies cont.

QoS requests, (3) enhancing the stability of QoS services through pinning and dynamic reconfiguration.

Recent advances in processing capabilities, signaling and forwarding technologies make it feasible to design QoS routing technologies. RSVP extensions for explicit routing make it possible to carry QoS requests over non-default paths. Forwarding technologies such as MPLS and efficient MF classification reduce the data path complexity of supporting multiple routes to a single destination. Technology developments in the area of routing and path selection include:

- **QoS Aware Routing Protocols** - Recent advances in QoS routing technology include: (1) the development of optimized multi-path extensions to intra-domain routing protocols to support load distribution and increased utilization in resource constrained networks; and (2) research in more generalized constraint based routing protocols to compute routes that satisfy explicit QoS requirements.
- **Multi-protocol Label Switching (MPLS)** - MPLS is primarily a forwarding technology that enables path establishment for arbitrary aggregates of traffic. Explicit labels are used to group flows to follow label switch paths (LSPs) across networks. LSPs can be established based upon (1) traditional best effort routing protocols; (2) routing schemes that support traffic engineering; and (3) explicit constraint based criteria. While not strictly limited to being a QoS technology, MPLS does potentially leverage QoS systems by providing an ability to associate arbitrary flow aggregates with individual data path resource management schemes and explicit paths across a network. Recent developments in coupling RSVP signaling with explicit, MPLS enabled routes provides a means for experimenting with integrated QoS routing, signaling and forwarding issues.

To date most of the efforts in enhanced routing have focused on a *traffic engineering* model where the objectives are to select paths and assign traffic flows to optimize the

interior performance of individual clouds. In such applications, dynamic path selection operates at the level of optimizing an aggregate traffic matrix that varies in time scales driven by gross changes in resource allocations and topology. It is an open question if there is a demand and if it is feasible to apply QoS routing techniques to on-demand establishment of paths for individual flows.

Middleware Services / Interface Technologies

As new network technologies provide richer and more complex services, the requirements for applications to be able to discover, interface with, manage, and monitor these new capabilities become more complex. In addition, many new network technologies require supporting infrastructures and services (e.g., policy managers, PKI) to make their deployment feasible. In some situations, access and control of network services is just one component of a larger issue/system (e.g., system level resource allocation). These issues are common to most high performance complex systems under development in the NGI community.

Recent efforts in middleware technologies have focused on avoiding application specific solutions to these issues by creating architectures, services, protocols and interfaces that:

- Provide interfaces for the specification of QoS requirements and receiving QoS notifications.
- Implement services for the active discovery, monitoring, and profiling of available QoS services.
- Provide bi-directional QoS translation between application and system services.
- Integrate the functions of advanced reservation and resource management across all components (e.g., computation, storage, instruments) of a distributed system.
- Support dynamic QoS adaptation and reconfiguration.
- Provide QoS service brokerage and coordination.
- Provide persistent, integrated infrastructures for authentication, access control and policy management.

Middleware Services / Interface Technologies cont.

There is a recognized need for understanding how advanced applications interact with complex networking services. In some NGI application communities, example middleware architectures and services (e.g., Globus) are emerging and are being extended and integrated with emerging QoS technologies (e.g., DiffServ). In addition, some standardization efforts have begun to address infrastructural services required for QoS technologies (e.g., policy frame works/ distribution, authentication/access control/accounting).

QoS Service Definition / Construction and QoS Systems

In the preceding sections we identified several classes and examples of individual QoS mechanisms. Complete QoS systems consist of the definition of specific services and the integration of one or more specific mechanisms used to implement the services on network nodes. The nature of complete QoS systems can be quite varied. The sets of services they implement may include both quantitative (i.e., providing specific bounds on observed performance metrics) and/or qualitative (i.e., providing assurances on the performance of a flow relative to others) QoS controls. The scope of a service may be end-to-end (i.e., from application to application), intra-domain, or inter-domain. Several individual QoS services may be concatenated or overlaid to provide multi-domain or end-to-end controls. Within individual domains multiple services/techniques may be employed at the same time to support different types of traffic or administrative requirements.

The full extent of today's QoS landscape is reasonably complex. Examining all possible systems (and their hybrid interconnections) was well beyond the scope of this single workshop session. The following sections focus on three "QoS systems" that were mentioned during the workshop as being the focus of significant research and development or pilot deployment activities.

- **Static, QoS Managed Services** - Collections of individual data path resource management techniques can be used in relatively static configurations to provide coarse management of QoS in individual networks. Technologies to control traffic profiles (e.g., CAR), manage buffer resources (e.g., RED), and assign relative forwarding treatments (e.g., CBQ, WFQ) to flows can be combined with offline resource allocation and provisioning schemes to allow "engineered services" to be implemented. Some proprietary tools exist to enhance the ability to design such services. While generally viewed as highly static (i.e., changes made in network engineering cycles) and non-scalable, these techniques are commonly the first step in evolving towards more dynamic QoS systems.

- **Integrated Services (IntServ)** - IntServ represents the IETF's first mature effort in defining a QoS system that integrates resource allocation, admission control and data path services. The key concepts/components of this approach include:

- Support of end-to-end QoS for individual applications.
- Support for multicast sessions among receivers with heterogeneous requirements/capabilities.
- Use of RSVP to support receiver driven soft-state resource allocation and admission control end-to-end across networks.
- Unmodified data protocols, requiring multi-field classification in routers to map packet to services.
- Decoupled resource allocation from routing.

IntServ architecture, standards and technologies are reasonably mature. Two end-to-end QoS services have been defined and standardized using IntServ techniques. The guaranteed service (G) provides a quantitative bound on throughput and delay. The controlled load service (CL) provides a predictive effect of having an unloaded network path between the sender and receiver.

QoS Service Definition / Construction and QoS Systems cont.

• **Differentiated Services (DiffServ)** - *The DiffServ effort is focused on providing QoS systems that scale well with the size and administrative complexity of today's inter-networks. The key concepts/components of this approach include:*

- Service differentiation in the data path through the definition of simple per hop behaviors (PHBs) that enable the construction of a variety of services.
- Implementation flexibility in the mapping of PHBs to specific data path resource management schemes.
- Simple packet classification of flows and aggregates by explicit marking of packets (i.e., DSCP).
- Simplified core routers through isolating admission control, metering, policing and shaping functions to the edges of DiffServ domains.
- Support of a wide variety of services, administrative policies, and implementation mechanisms. Services should be applicable end-to-end, across a domain, between domains, or along selected path segments.
- Decoupled control-plane signaling from data path service implementation. This allows incremental deployment and experimentation starting from the basic forwarding services outward.

Initial DiffServ research, development and standardization have focused on the definition of the basic architecture, a packet marking classification scheme, and some initial PHBs. The expedited forwarding (EF) PHB defines a "leased line" like behavior that minimizes forwarding delay and jitter while requiring strict policing (i.e., discard violations) of flow profiles. The assured forwarding (AF) PHB defines four classes of forwarding behavior with three drop precedences within each class. Data path resources (e.g., buffers and scheduled bandwidth) are allocated to each AF class. The AF PHB allows flows that are out of profile to be remarked to lower drop precedences.

Another area of current DiffServ activity is the research and development of BB technology to support service level provisioning and configuration across multiple administrative domains.

QoS TESTBED ACTIVITIES

A part of the QoS breakout session was dedicated to a discussion of QoS pilot and experimentation activities in the Internet2 QBone and the DOE/NGI testbeds. Activities were characterized in terms of: (1) planned QoS capabilities; (2) technology infusion plans; (3) general time lines; and (4) identified significant barriers/issues. QoS activities within other NGI testbeds were not discussed.

• **Internet2 QBone** - *The QBone activity was launched in October 1998 to provide an inter-domain testbed for I2 DiffServ technologies. The QBone strikes a balance between being a testbed for network research and providing services to participating organizations. Its initial DiffServ efforts are focused on the definition and experimental deployment of the QBone Premium Service (QPS). QPS is built using the DiffServ EF PHB and provides near-zero packet loss and low (bounded) jitter. In order to support experimentation, the QBone will support an integrated measurement infrastructure capable of disseminating continuous reports of active and passive measurements of the infrastructure. Another significant thrust of the QBone effort is the research and development of inter-domain BB technology. The QBone BB working group is developing requirement documents, experimenting with BB prototypes and protocols.*

- **Phase 0** - (0-6 months) QBone architecture specification completed. Initial recommendations on BB protocols. Phase 0 testbed rollout will support static QPS reservations from campus edge to campus edge. Resource allocation and admission control will be done manually.
- **Issues/Barriers** - Significant issues identified for the future development of the QBone include:
 - Completion of the design and testing of BB technologies and inter-BB protocols.
 - Evaluation and selection of end-to-end signaling protocols.

• **DOE NGI Testbed** - *The DOE NGI testbed activities are focused upon the research and development of a persistent networking infrastructure to enable NGI applications across networks that include ESNet, Abilene, and MREN. DOE NGI planned capabilities include: (1) uncongested best effort services; (2) DiffServ EF-based premium services; (3) inter-*

QoS TESTBED ACTIVITIES cont.

domain resource control/allocation and scheduling; and (4) an instrumentation infrastructure to support design and analysis of adaptive applications. Uncongested best effort services will be implemented using today's data path resource management technologies (e.g., CAR, WFQ, RED). QoS services will focus on resource allocation and implementation issues associated with DiffServ. Particular focus will be given to integrated resource allocation technologies that can address scheduling of multiple assets (beyond network resources) including access to instruments, computational and storage devices.

- **Phase 0** - (0-6 months) - Application analysis, testbed establishment. Initial resource manager deployment with measurement based admission control and distributed static resource allocation.
- **Phase 1** - (6-12 months) - Implement dynamic resource allocation, enhanced resource manager and address ISP/ multi-domain interconnection issues.
- **Issues/Barriers** - Significant issues identified for the future development of the DOE NGI testbed include:
 - *Establishing a persistent testbed infrastructure.*
 - *Reducing the requirements for manual configuration/ engineering.*
 - *Characterization and analysis of application QoS requirements.*
 - *Establishment of authentication and access control capabilities.*
 - *Last-foot issues in establishing end-to-end QoS.*

QoS TECHNOLOGY ROAD MAP

In this section we provide a brief road map, consisting of a set of tables, that attempts to capture the status and future directions of the technologies addressed in the workshop. The roadmap is by force incomplete and speculative in some sense given the rate of innovation and change in Internet technologies. Still, it attempts to provide value by tersely characterizing the maturity of each technology, the potential for its deployment and use, the service/value it provides, and the key issues in further development and adoption.

QoS TECHNOLOGY ROAD MAP cont.

RESOURCE ALLOCATION, SIGNALING AND ADMISSION CONTROL TECHNOLOGIES

	1 Year	3 Year	5 Year	Key Issues
RSVP – (see IntServ also)				
Technology Readiness	Core standards in place and commercial implementations available.	Policy management and distribution mechanisms. Extensions to support QoS routing and MPLS being defined as well as interaction/integration with DiffServ.		Scalability of hop-by-hop signaling mechanisms. Integration with policy management frameworks.
Deployability/ Usability	Intra-domain signaling.			Use in signaling for DiffServ and MPLS.
Utility/ Effectiveness	Support of IntServ.	Determined by role in support of DiffServ and MPLS.		
Bandwidth Brokers – (see DiffServ also)				
Technology Readiness	Initial design and requirements specifications in I2 community.	Definition of signaling protocols, interface to policy management systems, relationship to other resource managers.		Design of inter-BB signaling protocols. Selection of host QoS signaling protocol.
Deployability/ Usability	Early research prototypes available.			Interface and integration with policy management systems.
Utility/ Effectiveness		Experience within multi-domain testbeds (e.g., QBone, DOE NGI) will determine viability.		Interface and integration with general resource managers. Performance as a dynamic resource allocation system.
Admission Control				
Technology Readiness	Parameter based technologies part of initial IntServ and DiffServ resource allocation mechanisms.	Research in measurement-based admission control and resource engineering.		How to manage resources for large scale aggregate traffic. Loss of individual flow information, inability to characterize aggregate.
Deployability/ Usability	Initial AC technologies appropriate for intra-domain deployment.	Experience in scalability and performance of measurement based technologies.		How to design measurement based admission control schemes.
Utility/ Effectiveness	Useful for small scale (intra-domain) applications.			

QoS TECHNOLOGY ROAD MAP cont.

PATH SELECTION/ROUTING TECHNOLOGIES

	1 Year	3 Year	5 Year	Key Issues
QoS Routing				
Technology Readiness	Specifications exist for simple traffic engineering technologies (e.g., OMP extensions).	Research into constraint based routing technologies that are coupled with broader QoS resource control and signaling systems.		Requirements for intra-domain traffic engineering? Coupling of TE to QoS systems (e.g., IntServ, DiffServ).
Deployability/ Usability	Experimentation with OMP extensions.	Experimentation with constraint-based routing for traffic engineering.		Requirements for capabilities beyond intra-domain traffic engineering? Inter-domain TE/QoS routing? Methods for inter-domain advertisement aggregation.
Utility/ Effectiveness		Determined by need/ability to address intradomain traffic engineering requirements.		Is there a requirement for on-demand QoS routing? Requires more direct coupling between routing and policy/administrative systems (e.g., AAA, billing).
MPLS				
Technology Readiness	Specifications for basic architecture, label based forwarding mechanisms, and interfaces to existing routing protocols and link technologies are maturing. Some pre-production implementations of basic functions exist. Emerging work on signaling protocols [e.g., RSVP, Label Distribution Protocols (LDPs)].			Scalability and stability of LSP establishment mechanisms.
Deployability/ Usability	Early experimentation with basic label switching mechanisms using ad-hoc/static establishment procedures.			Integration with other QoS systems (e.g., IntServ, DiffServ).
Utility/ Effectiveness		Experience with MPLS coupled with existing dynamic routing protocols and QoS/constraint based technologies will define utility.		Ability to leverage practical implementation of constraint based routing.

QoS TECHNOLOGY ROAD MAP cont.

MIDDLEWARE TECHNOLOGIES

	1 Year	3 Year	5 Year	Key Issues
Middleware Technologies				
Technology Readiness	Domain specific research and development of middleware technologies (e.g., Globus). Initial discussions of standardization issues.	Research in requirements for, and design of, generalized middleware infrastructures that accommodate: service discovery and composition; application and management software interfaces; and monitoring, measurement and control of underlying network technologies.		<p>Definition of middleware, agreement on its scope.</p> <p>Architectural tradeoffs of network vs middleware vs application implementations of QoS signaling and adaptation.</p> <p>Integration and abstraction of policy management, resource management, and access control technologies.</p>
Deployability/ Usability	Specific to application domains.		Availability of more general purpose systems.	Definition of APIs and establishment of software reuse patterns.
Utility/ Effectiveness			Determined by the ability to integrate and abstract common infrastructural network services (including QoS).	

QoS TECHNOLOGY ROAD MAP cont.

QoS SYSTEMS AND SERVICES

	1 Year	3 Year	5 Year	Key Issues
Static, QoS Managed Services				
Technology Readiness	Many individual mechanisms (e.g., CBQ, WFQ, RED, CAR) are developed and available in commercial router implementations. Some commercial tools are available for network level engineering and configuration of mechanisms. In general, such mechanisms are not subject to standardization.			Complexity in integrating access control, policing, buffer management, and scheduling mechanisms into a coherent network service.
Deployability/ Usability	Individual mechanisms usable today.			Scalability of configuration and management mechanisms/costs.
Utility/ Effectiveness	Individual mechanisms effective for implementing very coarse static services that enhance typical best effort networking.			Ability to compose multi-domain managed services.
IntServ				
Technology Readiness	IETF standards are in place, implementations are available in commercial routers and hosts.	Advances in the integration of IntServ with other QoS mechanisms in hybrid systems.		Scalability of per-flow state in core routers. Aggregation techniques and integration with other mechanisms (e.g., DiffServ, MPLS) must be explored.
Deployability/ Usability	Pilot deployments in place in some research networks. Concerns of scalability limit widespread deployment use. Guaranteed service hard to implement across heterogeneous environments.			Mapping/matching application requirements to service definitions.
Utility/ Effectiveness	Need more experience in application support and mapping.			
DiffServ				
Technology Readiness	Standards for basic architecture, classification, and initial PHBs maturing. Initial experimental implementations of data path mechanisms appearing.	Advances in service definitions and signaling standards and development. Research and development of BB/policy technologies.		Pilot deployment and experimentation with basic PHBs. Control plane diversity - which (if any?) signaling protocol(s) to use for end-to-end service construction? Role of RSVP and MPLS?
Deployability/ Usability		Deployment and experimentation with basic PHBs and services in research networks.	Deployment and experimentation with signaling and BB technology.	Definition and analysis of services based on PHBs.
Utility/ Effectiveness	Initial PHBs provided for wide range of services from strict leased line services, to flexible assured relative performance services. Must wait for experimentation to determine.			Mapping/matching application requirements to service definitions. Evolution of BB architecture/technologies.

APPLICATION GUIDELINES

During the workshop three applications, selected as representative of a variety of system types and QoS requirements, were used as case studies to explore the process by which one could understand and characterize QoS requirements, identify potential mappings to technologies, and understand how applications could adopt and adapt to their capabilities and services.

Application QoS requirements typically include tight bounds on jitter, latency control, and bandwidth control. Issues that were raised during the breakout session include the need to develop metrics for digital video quality so that the effects of QoS on application performance can be quantitatively measured and the need for network instrumentation and tools to help characterize workloads and debug QoS issues.

Most important, application developers today do not have a free rein to choose the QoS technologies that best match their requirements. Instead, it is the availability of any QoS mechanisms in the networks that they must attach to that determines the mapping. The infrastructure investment required to deploy large-scale QoS testbeds precludes support of several competing technologies.

Application developers are directed to the tables in the previous section for information about technological developments that are anticipated in various timeframes.

SIGNIFICANT BARRIERS TO DEVELOPMENT, DEPLOYMENT, AND ADOPTION OF QoS TECHNOLOGY

Over the course of the session several issues were identified as recurring questions for, or significant barriers to, the development, deployment and adoption of QoS technologies. Some of these items represent open research and development issues, others represent programmatic and administrative barriers, and some are issues that can only be resolved by gaining significant experience

with initial approaches. The issues and questions identified include:

1. QoS interactions with other emerging technologies -

While QoS research is primarily focused on enhancing today's best effort services, we must keep abreast of parallel developments in other areas. How QoS technologies interact with emerging developments in security and multicast must be considered early. Examples of issues that must be addressed include: (1) how end-to-end encryption interferes with packet level classification schemes; (2) how multicast affects DiffServ provisioning among heterogeneous domains; (3) how QoS specific policy control mechanisms integrate with broader mechanisms and systems.

2. Inter-domain/multi-administration issues - *Multi-domain issues pose a significant challenge to almost all aspects of QoS, including: resource allocation, admission control, routing and signaling. In each of these areas the issues of policy management, authentication/access control, and accounting/billing can pose significant technical, administrative and social barriers. While there are numerous research and development activities focused on individual technology developments in these areas, the potential range of requirements/solutions space is vast. It is important for the community to identify what level of administrative management, pricing and cost recovery models are appropriate for near-term experimentation and use of QoS technologies. Pilot deployment and experimentation with basic QoS mechanisms should not be overly delayed while designing/debating longer-term administrative control system(s).*

SIGNIFICANT BARRIERS TO DEVELOPMENT, DEPLOYMENT, AND ADOPTION OF QoS TECHNOLOGY cont.

3. Application characterization/QoS requirements - While almost all applications at the workshop stated the requirement for QoS, the typical request could be summarized by asking for “the most bandwidth and least latency/jitter possible.” The workshop highlighted the fact that there are numerous approaches to providing QoS control in networks and that the complexity, resource requirements, scalability and manageability of these approaches can vary greatly. Understanding, in detail, the true range of application requirements and operating constraints can significantly influence the choice of where (application vs network vs middleware) and how (selection/design of mechanisms) QoS control is realized. Unfortunately, for many applications we do not have a deep understanding of these requirements and constraints. More effort needs to focus on understanding NGI applications in terms of: (1) detailed QoS requirements and sensitivity; (2) workload characterization of traffic; and (3) adaptability to variations in service.

4. Network vs application vs middleware design tradeoffs - There are fundamental trade-offs in where one designs QoS control and adaptation functions. Applications that require rigid QoS services and absolute guarantees place higher demands on the capabilities of QoS mechanisms. Applications that can tolerate, adapt to, and renegotiate QoS profiles result in more flexible systems that can operate over a wider range of deployment scenarios. Where and how one provides adaptive services is an open question. Further research and experience in the design of QoS adaptive middleware and applications is required.

5. Persistent, instrumented testbed infrastructure - In order to gain some understanding of the tradeoffs and achieve some convergence among application design

paradigms and QoS mechanisms, the NGI application community must be given access to a persistent testbed infrastructure that supports incremental experimentation with QoS capabilities and services. The testbeds must support multi-domain topologies and provide instrumentation infrastructures to enable application QoS feedback, diagnostics and analysis. Such testbeds must balance networking research with providing a network for application research. Research and development activities that focus on supporting either of these goals in isolation will not provide a productive environment to explore the tradeoffs of the design space. The most effective way to foster such environments may require programmatic models in which application development, networking research, wide area testbed deployment, and local access networking are more tightly coupled than they are typically to date.

6. NGI QoS vs commercial QoS - This workshop focused on the QoS requirements and technical plans of the NGI community. By definition this community has very unique requirements for high performance computing and communications infrastructures. Further consideration should be given to how, if at all, the requirements/approaches to QoS for the class of NGI applications/networks differ from those of the commercial world.

7. Very high bandwidth networks - As the availability and value of very high bandwidth technology continues to increase, we must consider the cost tradeoffs in designing and operating complex QoS control systems. More than one application presentation mentioned relative costs of advanced network engineering versus over provisioning as a question. Research is needed to systematically examine the importance/viability of QoS technologies in very high bandwidth networks. What services are required and which mechanisms are needed?

QoS APPENDIX A - REFERENCES

Presentations during the breakout session

1. QBone: a Testbed for IP Differentiated Services, Ben Teitelbaum
2. QoS Plans for the DOE/NGI Testbed, Linda Winkler
3. A DiffServ Snapshot, Kathleen Nichols
4. Quality of Service Support: The Role of Signaling Protocols, Lixia Zhang
5. QoS Routing Status, Promises, and Challenges, Roch Guerin
6. QoS Gap between Application and Network, Klara Nahrstedt and Rick Schantz
7. NGI Digital Video, Joe Mambretti
8. Virtual Room Videoconferencing, Philippe Galvez
9. Teletomography, Mark Ellisman

General references and useful links:

Quality of Service - Delivering QoS on the Internet and in Corporate Networks, P. Ferguson and G. Huston. John Wiley & Sons, Inc. 1998

http://www.qosforum.com/white-papers/Need_for_QoS-v4.pdf
The Need for QoS - A White Paper, QoS Forum

http://www.qosforum.com/white-papers/qosprot_v3.pdf
Protocols and Architectures - A White Paper, QoS Forum

<http://www.qosforum.com/white-papers/qos-glossary-v4.pdf>
QoS Glossary of Terms - A Technology Backgrounder, QoS Forum

<http://www.qosforum.com/docs/faq/>
QoS FAQ - Frequently Asked Questions about IP Quality of Service, QoS Forum

<http://www.internet2.edu/qos/wg/>
Internet2 QoS Working Group

<http://www.internet2.edu/qos/qbone/>
Internet2 QBone Testbed

<http://www.ietf.org/html.charters/intserv-charter.html>
IETF - Integrated Services (intserv) Working Group

<http://www.ietf.org/html.charters/issll-charter.html>
IETF - Integrated Services over Specific Link Layers (issll) Working Group

<http://www.ietf.org/html.charters/diffserv-charter.html>
IETF - Differentiated Services (diffserv) Working Group

<http://www.ietf.org/html.charters/rsvp-charter.html>
IETF - Resource Reservation Setup Protocol (rsvp) Working Group

<http://www.ietf.org/html.charters/rap-charter.html>
IETF - Resource Allocation Protocol (rap) Working Group

<http://www.ietf.org/html.charters/policy-charter.html>
IETF - Policy Framework (policy) Working Group

<http://www.ietf.org/html.charters/mpls-charter.html>
IETF - Multiprotocol Label Switching (mpls) Working Group

<http://www.ietf.org/html.charters/tewg-charter.html>
IETF- Internet Traffic Engineering Working Group (tewg)

<http://www.ietf.org/html.charters/ippm-charter.html>
IETF - IP Performance Metrics (ippm) Working Group

QoS APPENDIX B - APPLICATION QUESTIONNAIRE

The following questionnaire was developed to help assess the QoS requirements and constraints of applications.

Bridging the Gap - Application QoS Requirements Analysis

1. Application overview.
 - What is your application? (refer to day 1 presentation if appropriate)
 - What are the sites are involved/networks to be used?
 - What are the rough time lines of its development? (1, 3, 5 years?)
2. General application QoS requirements.
 - What are the application's general QoS requirements/expectations?
 - How rigid are the applications QoS requirements?

QoS APPENDIX B - APPLICATION QUESTIONNAIRE cont.

What percentage of total available resources will be requested by the application?

How dynamic are the resource demands? How do they change during the course of one session?

3. Specific data flow QoS requirements.

What specific data flows are QoS sensitive?

What types of flows are these?

Synchronous (time-sensitive)/Interactive/Isochronous (time-critical)/Bulk Data Transfer?

Why is this flow QoS sensitive?

How does its performance (bandwidth, delay, jitter, packet loss) effect the application?

Number of such flows? Multicast/Unicast?

Duration/Bandwidth/Latency/Jitter/Loss Rate/Availability required?

4. Traffic Profiling/Policing/Adaptation

How precisely (number of parameters, tightness of bounds) can you characterize the traffic generated by the application?

What behavior(s) (discard/delay/reroute) do you want from the network when the application exceeds its profile?

Would you like this behavior to be dependent upon some relative ordering of the importance of the data?

5. Traffic Profiling/Policing/Adaptation...

Is (or can) the application designed to dynamically adapt its behavior to compensate for variances in network service?

Is (or can) the application be instrumented to detect these situations and/or receive feedback from the network?

6. QoS Technologies

What QoS technologies are you planning/envision using to achieve these requirements? What testbeds/services will you use?

What factors motivated these choices? (matches application requirements, availability)

How do you expect these technologies to benefit your application?

7. Inhibitors/Roadblocks

What do you perceive as the most significant inhibitors/issues (technical, administrative, programmatic, social) in achieving these QoS requirements?

- *Application/Host Issues?*
- *Local/Access Network Issues?*
- *WAN Issues?*
- *Multi domain/administration Issues?*

MULTICAST TECHNOLOGY OVERVIEW AND CURRENT STATUS

Traditional networking applications utilize unicast transmission (i.e., a single sender transmits information to a single receiver). Even with client/server applications, the server typically maintains a separate unicast connection to each individual client. An alternative type of transmission, which is useful for certain applications, is broadcast (i.e., information is sent from a single sender to all receivers within a network domain). In contrast to either unicast or broadcast, multicast is one-to-many (one sender and multiple receivers) or many-to-many (multiple senders and multiple receivers) transmission. Multicast benefits include savings in both network bandwidth and processing power within the sender. These savings can be substantial and provide the motivation for multicast-enabling applications and network infrastructure alike.

Multicast is appropriate for applications in which the same data is transmitted to several receivers. Typical multicast applications include bulk data transfer, multimedia streaming, software distribution, etc. A sender transmits to a multicast group by specifying a multicast address, distinguished by having 1110 as the first four bits of the address. A distribution tree connects a sender to all the receivers. Messages are transmitted along the distribution tree, with routers replicating packets and forwarding them on multiple links when the path in the distribution tree diverges. Several routing protocols have been developed to construct distribution trees; pros and cons of various protocols can be found in the literature. The protocol set that is currently most widely deployed is PIM-SM (Protocol Independent Multicast - Sparse Mode) as the tree-building protocol, BGP4+ (MBGP, Multicast Border Gateway Protocol) for route exchange between network domains, and MSDP (Multicast Source Discovery Protocol) for discovering multicast sources across network domains. Problems of scalability, address allocation, and protocol complexity remain as key research issues and are expected to be the subject of further investigation. Future protocol development to support multicast will be discussed in the *Multicast Technology Roadmap* section.

To experience the full benefit of multicast, native multicast must be used. This means all routers must be multicast enabled (i.e., the routers must be able to recognize and handle multicast traffic), in contrast to the use of tunneling to connect multicast-enabled islands in the network. The biggest hurdle to widespread utilization of multicast is enabling both the wide-area and the campus infrastructure to support native multicast.

It is recognized that multicast can currently be difficult to use. Both engineering issues and application-design issues were identified at the workshop and are presented in the next two sections.

CURRENT STATUS FROM THE PERSPECTIVE OF APPLICATION DEVELOPERS

A limited number of application developers participated in the multicast breakout session at the Bridging the Gap Workshop. We had anticipated that this would happen, based on conversations prior to the workshop. Accordingly, to ensure that all the application developers had an opportunity to provide input regarding their use of multicast, we distributed a questionnaire prior to the workshop. The most interesting information collected via this questionnaire was the identification of special concerns that might make application developers reluctant to incorporate multicast into their applications. Some of the indicated concerns are valid; others reflect a lack of understanding of multicast technology and its current status. Valid concerns include the potential for network congestion, the availability of multicast on a wide scale (including internationally), and concerns for reliability of data delivery. On the other hand concerns regarding high-bandwidth server requirements or regarding large quantities of data are misdirected, since multicast technology was designed explicitly to alleviate such problems. Another issue that was raised was whether or not multicast is appropriate for relatively small groups, involving for example only 4 or 5 receivers. The answer in

Current Status from the Perspective of Application Developers cont.

this case depends on how much data is involved; for large volumes of data, the use of multicast can provide significant savings on bandwidth, even for relatively small groups.

During the multicast breakout session we acknowledged the fact that development of multicast applications for the Internet is slow. We identified several possible reasons for this apparent reluctance to incorporate multicast. One reason is that writing an application to use multicast requires a fundamental design shift from writing unicast applications, as developers must think in terms of one-to-many or many-to-many communication rather than one-to-one communication. Hence, it is understandable that it will take some time for developers to become comfortable using multicast.

However, the major hurdle to deployment of multicast applications is that the campus infrastructure is not multicast enabled. This issue is addressed more fully in the next section.

Another issue related to the difficulties of incorporating multicast into applications was raised during the breakout session. One participant spoke of the dangers of application developers not understanding the implications of utilizing multicast. Specifically, multicast is inherently unreliable, since it is based on the UDP protocol, rather than TCP. This means that packets may be lost, and the sender and the receivers are not informed when this happens. It seems that some application developers have become so accustomed to using TCP/IP protocols that they are unaware that UDP does not provide the same reliable infrastructure. This could be devastating for some applications, and highlights the importance of educating users regarding the basics of multicast technology.

NGI TESTBED STATUS

The table below depicts the status of deployment of multicast protocols on the NGI testbeds. While it is recognized that some of the protocols are temporary fixes until more scalable solutions are developed, nevertheless this set of protocols is sufficient to enable native multicast across network domains. Hence, it is clear from Table 1 that native multicast is available now on the NGI testbeds.

Special note was made of the fact that the deployment of native multicast on the NGI testbeds has been a recent development. In fact, the impetus for this development was preparation for a prototype demonstration of NASA's Virtual Collaborative Clinic (VCC) application during early May 1999. This demonstration, featuring high-rate multicast up to 30 Mbps, utilized the Abilene, vBNS, and CalREN2 testbeds as well as NREN.

Table 1: Status of Protocol Deployment on NGI Testbeds

Test Bed	PIM-SIM	BGP4+ (MBGP)	MSDP
Abilene	X	X	X
DREN*	X	X	X
ESNet	X	X	X
NREN	X	X	X
vBNS	X	X	X

** Protocols enabled only at the network boundary.*

The three protocols, PIM-SM, MBGP, and MSDP are also now being deployed on the community Internet, as well as on some international networks.

The strong consensus of the multicast breakout group was that the weak link in infrastructure support for multicast applications is the campus infrastructure. Very few campus infrastructures are multicast enabled at present. Reasons for this include lack of personnel to do the work and lack of knowledge regarding how to deploy multicast protocols and how to handle multicast traffic so that it does not degrade performance of other network traffic. This need is currently being addressed by the NLANR Engineering Services group. Campus network engineers are encouraged to contact this group for assistance.

MULTICAST TECHNOLOGY ROADMAP

Future development and deployment of multicast and multicast applications is dependent not only on technological advances, but on changing people's perceptions of multicast as well. Multicast is still considered by some to be a toy technology for researchers, rather than a technology for general use. Hopefully the Bridging the Gap Workshop and this report will help to clarify the situation for application developers and will encourage them to consider utilizing multicast. Engineering assistance, such as that offered by the NLANR group, can be useful to enable the campus infrastructure to support multicast. Once multicast is enabled to the end user, application developers are likely to be more motivated to incorporate multicast into their applications.

Technology issues include protocol development, standardization of reliable multicast solutions, flow/congestion control, access control, QoS multicast, and secure multicast. The discussion during the multicast breakout session focused primarily on multicast protocol issues and reliable multicast; QoS multicast and secure multicast were only briefly mentioned.

- **Protocol development** - As indicated earlier the current set of multicast protocols that is most widely deployed is PIM-SM, BGP4+, and MSDP. These protocols enable inter-domain native multicast. However, many problems remain. For example, scalability of PIM-SM and MSDP is limited, global dynamic multicast address allocation remains a major problem, and protocol complexity is an equally important issue. Protocols currently under development to address the issues of scalability and address allocation include BGMP (Border Gateway Multicast Protocol) for scalable inter-domain shared-tree multicast forwarding trees, and MASC (Multicast Address Set Claim) for global dynamic multicast address allocation. SM (Simple Multicast) is an approach to simplify multicast, reduce router overhead, and eliminate the need for coordinated multicast address allocation across network domains.

- **Reliable multicast solutions** - Reliable multicast is a topic within the Internet Research Task Force (IRTF). There

are many approaches for achieving reliable multicast, different approaches for different types of applications. While some reliable multicast products are currently commercially available, work on standardizing protocols will continue in the near future.

- **Flow/congestion control** - Multicast can create serious traffic engineering problems, since it is UDP based and UDP lacks the flow/congestion control that is an inherent part of the TCP protocol. When the network becomes congested, TCP flows will back off, but UDP flows will not. Consequently, congestion may worsen and TCP flows may suffer virtual starvation. There is considerable current research to determine how to make multicast TCP friendly. The consensus of the breakout group was that it is impossible at this point to conjecture when the issue will be satisfactorily resolved.

- **Access control** - Access control (i.e., controlling who is permitted to send to a particular multicast group), is another important research topic. This is important to prevent denial of service by a malicious sender spewing data to a group.

- **QoS multicast** - Some IP QoS mechanisms (e.g., IP TOS and DiffServ) are directly applicable to multicast; others will require additional work. Some research has been done in QoS-based multicast routing, wherein the objective is to utilize QoS requirements of an application during the construction phase of the multicast tree.

- **Secure multicast** - Secure multicast is a current research topic. Key management is difficult.

Table 2 lists the multicast issues that were discussed at the Workshop and presents a timeline for progress in each area.

MULTICAST TECHNOLOGY ROADMAP cont.

Table 2: Multicast Road Map

Issues	Now	1 year	3 years	5 years	Comments
Basic protocol development	Current set of deployed protocols is PIM-SM, BGP4+, MSDP; protocols under development include MASC, BGMP, SM.		Set of deployed protocols not clear; depends on experience with today's protocols.	Protocol simplification.	Impact of IPv6 deployment impossible to conjecture.
Reliable multicast	Widespread deployment in private enterprise intranets; proprietary solutions commercially available; multiple protocols being researched within IRTF.		Protocol standardization.		
Infrastructure issues	NGI testbeds sufficiently native multicast enabled now.	Campus infrastructure multicast enabled.			
Network engineering issues	Multicast difficult to use now; network interoperability poor; NLANR Engineering Service group available for assistance.	Multicast tools available.			
Flow control/congestion control	Current approach is sender rate-based flow control.		Significant progress toward general solution.		Some felt that it was impossible to conjecture time frame for solution to problem.
Access control	Currently can use closed groups for access control.		Issue resolved.		
QoS multicast	Active research area; QoS-based multicast routing is an area specific to multicast.				Progress largely dependent on advances in general QoS technology.
Secure multicast	Active research area; key management is especially difficult for multicast applications.				Progress largely dependent on advances in general security technology.
Application issues	One sender, multiple receivers (one-to-many) applications deployable on NGI testbeds; NLANR assistance available for application developers.		Standard application tool kits available.	Support for many-to-many applications.	

APPLICATION ROADMAP

Because the wide-area NGI testbeds are multicast enabled now, the important message for application developers is that the testbed infrastructure is ready to support basic multicast applications now.

Two types of multicast applications were identified for deployment during the next year: distribution of massive data sets and distance learning. These two types of applications have somewhat different requirements. Massive data distribution requires only the basic multicast functionality, with no special requirements for delay constraints or interactive support. If reliability is important, several approaches are available for use. The number of receivers need not be especially large; if bandwidth requirements are large, multicast can make a significant difference in bandwidth savings and in savings of processing resources within the sender, even for relatively small sets of receivers. An example of a data-distribution multicast application that could be deployed over the NGI testbeds within the upcoming year is distribution of weather data from NOAA. For this application the receiver group includes approximately 130 universities and data must be distributed every 6 hours, with the volume of data approaching 280 GB per day.

Distance learning will likely involve larger groups, and might require some level of interaction. The distributed seminar series being planned by the LSN-NRT is an example of this type of application. The Bridging the Gap Workshop was actually the first such seminar; plenary sessions of this workshop were multicast to a limited audience. The message from the multicast breakout group is that the NGI testbeds are now ready for distribution of future seminars in the LSN-NRT series.

Applications with more complex multicast requirements, such as applications with multiple senders as well as multiple receivers, will be supportable on the NGI testbeds within the 3-5 year time frame.

MULTICAST TECHNOLOGY APPENDIX A - REFERENCES

Presentations during the breakout session

- IP Datagram Multicast, Radia Perlman
- Multicast Applications in the Commercial Market, Ken Miller

General references and useful links:

<http://ale.east.isi.edu/RMRG/>

IRTF Reliable Multicast Research Group web site

<http://www.irtf.org/charters.secure-multicast.htm>

IRTF Secure Multicast Group Charter and contact information

<http://www.internet2.edu/multicast>

Internet2 Multicast Working Group site

<http://www.ietf.org/html.charters/mboned-charter.html>

IETF Mbone Deployment Working Group web site

<http://www-mice.cs.ucl.ac.uk/multimedia/software/>

Contains a list of available multicast-capable freeware for conferencing, archiving, and delivering stored multimedia content.

INTRODUCTION

As the world becomes more and more dependent on interconnected groups of networks and as applications increasingly become distributed across these networks, security becomes a larger concern. Often security is dealt with as an afterthought, rather than designed into an application from the outset. This practice limits the range of security technologies that can be brought to bear. What needs to be done is to simultaneously increase the sophistication, availability, and ease of use of security services, as well as educating application developers in the use of these services. In networked systems there are several security aspects: computing platforms, communications infrastructure, and distributed applications and services. The security breakout session of the Bridging the Gap Workshop focused on security services and capabilities that could be employed directly by application developers in order to achieve application-level, or end-to-end security.

NGI Application Context

NGI applications can be quite different from traditional networked applications. Collaboration with academic and industrial partners as well as other NGI agencies is not uncommon. NGI applications involve many data sources and make use of the diverse assets of many stakeholders. They may operate in distributed computing environments that extend beyond the NGI testbeds to encompass semi-open networks such as Internet2, state research networks, and industrial testbeds. There are many types of NGI applications, but they tend to share some common characteristics.

NGI applications will primarily use IP as the base internetworking protocol. Such applications may often require multicast, may have multiple synchronized data flows, often span multiple network domains and national boundaries, and may make use of large distributed network based shared storage systems. Such application flows running over NGI testbeds will often be combined into very high speed (OC48-192, i.e., 2.5 Gigabits/sec to 10 Gigabits/sec) aggregate transmission rates. Other characteristics may include sensitivity to QoS parameters such as latency, jitter, and loss,

user mobility requirements, and the ability to handle very long delay in the case of space-to earth connectivity.

Security Requirements for Networked Applications

Within the security breakout session at the Bridging the Gap Workshop, the participants began by outlining high-level security requirements for applications and then discussed what tools/services were available for each area, and the readiness state of these. The group as a whole felt that it was very important for application developers to take advantage of available security mechanisms and APIs rather than trying to implement them from scratch. The high-level network security requirements identified by the breakout session include data confidentiality, authentication, authorization and access control, data integrity, non-repudiation, and resource availability.

- **Data confidentiality** - Ensuring that sensitive data is not compromised is one of the cornerstones of information security. This is true for both data in transit over a network and data residing in network-accessible storage sites. Encryption of data is the most widely used method of preserving confidentiality, as it can render intercepted data incomprehensible to an attacker. Tools available to provide data confidentiality over a network include IPsec [12] and SSL/TLS [1], [2], [3]. Considerations include where within the communication path the data is encrypted, the cost of encryption, and the degree of transparency to the user community.

Application-level management of encryption allows the least exposure of unprotected data, since encrypted data can be read from disk files, decrypted in application libraries, placed in memory and operated on, re-encrypted and communicated to collaborators or other parts of distributed applications. This is arguably the most secure approach as data is only exposed in application memory; however it is also the least transparent. In order for multiple entities

Security Requirements for Networked Applications cont.

(e.g., human users, processes, or code components) to share encrypted data, the keys that encrypt and decrypt the data must be managed so that they are only available to authorized parties at appropriate times. To do this, significant infrastructure is required to coordinate authorization, maintain state information and codified trust relationships, and implement access control and resource usage policies.

Secure application-to-application communication channels such as SSL/TLS [2], [3] are easier to manage, involving only management of a secure identity. However, this approach does nothing to protect data outside of the communication channel.

IPsec [12] protects data at the packet level. Since the application is not involved in the process, there are no application design or modification issues. An important advantage of IPsec is that IP header information (e.g., source and destination addresses) are protected, whereas all higher-layer methods leave it in the clear. Sensitive applications might demand the anonymity that this provides. A disadvantage is that the application has no direct involvement in specifying the trust relationships, and therefore is once or twice removed from managing the security of its data.

- **Authentication** - Authentication is the art of verifying that a user or program involved in an interaction with an application or system is who they purport to be. Use of external, server based authentication is preferable to home-grown systems, especially in cross-domain scenarios. Two mechanisms for user and device identification are Kerberos [7] and Public Key Infrastructure (PKI) [10]. X.509 certificates are the part of a PKI that uses the authority of trusted third parties to produce assured credentials suitable for online authentication. The GSS API [16] can be a useful way to incorporate such mechanisms into an application.

- **Authorization and Access Control** - Authorization has two aspects: establishing policy for access to a resource, or permitted actions on a resource, and designating trusted parties to attest if the attributes of entities requesting access

match the policy for accessing a resource. Access control follows authorization, and enforces the decision of the authorization mechanism. Once an entity presents its identity credentials and they are compared with the resource usage policy, an access control mechanism must permit access, or ensure denial if access policy criteria were not met. Access control is generally applied after authentication of the user identity. The GAA Control API [13] is a tool intended to address this process.

- **Data integrity** - Data integrity is concerned with ensuring that information transmitted over a network or on a storage volume has not been subject to corruption or unauthorized changes. Data integrity methods can be combined with encryption to preclude data substitution and can be combined with authentication methods to verify the data origin.

Data integrity is typically managed by computing a “unique signature” over a file or a block of data being transmitted so that a given byte sequence produced a signature that is shared by no other sequence. Such signatures are called one-way or cryptographic hashes. Widely used algorithms include MAC, HMAC, SHA-1, and MD5 [9].

Hashing algorithms can be applied to data as it is prepared for transit and/or while it is stored on disk or in an archival storage system. The basic techniques are the same, but the management issues for long-term storage are quite different than for communication channels. The Tripwire system [21] can be used to check the integrity of data stored on a network accessible volume by managing the hashes separately from the datasets.

Another common form of data integrity for files that are to be published is called “digital signature.” Files are digitally signed as follows: the author computes a cryptographic hash over the file, and then encrypts that hash using their private key to create the “digital signature.” This “signature” can be added to the file as an appendix. The recipient can verify the integrity of the file by retrieving the author’s public key from an X.509 certificate, recomputing the file hash and comparing it to the hash in the appendix. If they match, then the document is identical to the one originally signed by the author.

Security Requirements for Networked Applications cont.

- **Non-repudiation** - *Non-repudiation is the assurance that once a transaction has occurred, the sender cannot deny transmitting a message and the receiver cannot deny receiving it. This can be of concern in scientific data (e.g., “electronic laboratory notebooks”), medical applications, and financial transactions. Non-repudiation requires strong data origin authentication plus trusted timestamping and, sometimes, a counter-signature by a trusted authority.*

- **Availability** - *Denial of service attacks prevent authorized users from accessing a resource by exploiting system vulnerabilities to tie up resources such as system CPU, memory, or network bandwidth. As more applications implement QoS mechanisms, denial and degradation of service becomes much more of a concern. Using the combination of IPsec and secure DNS to provide host identity credentials represents a partial solution, as this combination can verify packet origin, thereby reducing the impact of bogus packets by rejecting them before they can exploit any vulnerabilities.*

STATUS OF SELECTED SECURITY MECHANISMS

Encryption

Encryption is the basis for most network-related data security. It falls into two basic types: shared (secret) key and public key. With shared key encryption, the same key is used to both encode and decode the encrypted information. This key needs to be known by both sender and receiver yet kept secret. In public key encryption, each party has a private key that is kept secret and a public key that is distributed. Data encrypted with the public key can only be decrypted with the private key and vice versa. Public key encryption is significantly more computationally expensive than shared key, so is usually not used for bulk encryption. Often public key will be used initially to set up the connec-

tion, authenticate the user, and distribute a secret “session key” to be used for actual data stream encryption. Standard algorithms may be too slow for multi-gigabit data streams even when implemented in hardware; software encryption in hosts can be too slow to keep up with real-time application demands.

Software such as SSL is readily available to encrypt data channels. Publicly available SSL implementations such as OpenSSL [1] can encrypt data streams on high end workstations at somewhat less than 100 Mbit/sec.

IPsec

IPsec is a group of protocols developed by the IETF that uses encryption to support secure IP layer packet exchange over potentially insecure networks. IPsec can sign the source and destination address within the IP packet header so that the packet may be authenticated in combination with a host identity credential that might be maintained in a secure DNS server. IPsec can also encrypt the packet payload in order to provide confidentiality. IPsec will be a key part of many Virtual Private Network (VPN) implementations. IPsec has two modes: Tunnel and Transport. Tunnel mode is typically used between router-like elements, and encrypts both the header and data payload. Transport mode operates between host end nodes. There are various IPsec implementations available on the market today, but they tend to be single vendor solutions that do not interoperate well. In the past most IPsec implementations were for gateways/firewalls. However most major computer vendors now have, or have announced, support for IPsec. Hardware IPsec solutions are fast, but tend to be expensive. Software IPsec implementations are significantly cheaper, but are slow, fragile and finicky. Recently several major computer OS vendors, including Sun and Microsoft, have announced IPsec capability in operating systems already released, or to be released in the near future.

Kerberos

Kerberos is a server-based authentication, integrity, confidentiality, and authorization system. Applications can be modified to take advantage of the services available from a Kerberos server. Both Secure Shell (SSH) and TLS (but not SSL) can do this. Kerberos and PKI can be complementary. Group ID extensions to Kerberos should be available soon. For more information see [7].

Public Key Infrastructure (PKI)

PKI refers to the various components and services needed to support robust, widely deployable, and scalable public key based services. PKI allows an entity to prove its identity independent of location or system by signing a token with its private key and handing the signed token to a Certification Authority (CA) system. By way of background:

Public-key cryptography involves two keys, whereby data encrypted with one key can only be decrypted with the other, and visa versa. The public key is freely available and the other is kept private. Material decryptable by the public key must have been originated from the holder of the private key. A CA generates a certificate containing the X.500 distinguished name of an entity and that entity's public key. The CA then signs this "certificate" and publishes it, usually in an LDAP directory service. The recipient can verify the signer's identity by obtaining the identity certificate, extracting the entity's public key, and verifying the signature. The X.509 certificate is in turn verified by obtaining the CA's public key to verify the contents that the CA has signed by using a digital signature. See the data integrity requirements section for an example of digital signature use.

The basic components of PKI include:

- **Certification Authorities (CA)** - CAs provide the mechanism for trusted third parties to accept requests for X.509 identity certificates, verify the identity of the requestor, and then construct, sign, and issue the certificate.
- **Certificate Servers** - These servers are typically associated with CAs, and are used to publish the certificates. They typically present an LDAP interface for Internet access.

- **Application Libraries** - These libraries provide the basic mechanisms for verifying certificates and/or signatures based on X.509 certificate identities.
- **Certificate Revocation** - Mechanisms for revoking certificates once they have been issued or cached.

The first two components mentioned are readily available, the last two are not yet readily available except in some vendor specific implementations. For more information on PKI and X.509, see [8], [9], and [10].

Secure Sockets Layer / Transport Layer Security

SSL is a protocol developed by Netscape Communications that uses the private key associated with a host and/or server public key certificate to encrypt data sent over a TCP connection. Although not an IETF standard, SSL has become a de facto standard as it is supported in ubiquitous web browsers such as Netscape Navigator and Microsoft Internet Explorer. SSL is widely available, but embedding SSL in an application requires binary distribution due to export control rules. As of September 1999 this may no longer be an issue as it appears that the U.S. government has decided to substantially relax export restrictions on encryption software [22].

TLS is the proposed IETF version of a socket based security protocol. Although TLS is based on SSL, there are enough differences that SSL and TLS cannot interoperate. Both of these protocols require the TCP transport protocol, which may not be the ideal transport mechanism for some next generation applications. The performance of SSL/TLS is likely to lag behind IPsec. There are high quality commercial and public domain implementations of SSL [1] and [2].

Generic Security Services

The IETF GSS API provides for confidential messaging and assured integrity messaging. The application interface is very simple, and the underlying implementation can be based on any suitable security service. The initialization of

Generic Security Services cont.

GSS (e.g., the specification of identity) depends on the nature of the underlying security service. Implementations have been defined for using Kerberos and a “simple public key” infrastructure. See [16] for more information. Available implementations include SECUDE public domain and commercial source code [4] and the Entrust commercial libraries [5].

Authorization and Access Control

Access control is the process of enforcing an authorization to make use of a resource. Most access control today is centralized: a central server or file contains an access control list, sometimes keyed to named resources. This works acceptably for small or centrally administered communities such as the users of a single computing facility, but is very clumsy when the resources, the resource stakeholders/policymakers, and the users are geographically and organizationally dispersed.

Authorization is the process of determining whether a requester has the right, as defined by the usage policy, to access or act upon a given resource. This is also sometimes called policy based access control. There is some current IETF work in policy based access control, but it is largely specialized to management of network QoS resources.

There is some early work in generalizing standard mechanisms for policy based access control in the GAA control API [13]. The GAA API is similar to the GSS API in that it does not assume a particular underlying security service, and early implementations are likely to be available for Kerberos and PKI.

Akenti is a prototype R&D authorization system that is addressing the authorization and access control in NGI-like environments. Akenti’s model is that identity, resource use conditions (policy), and user attributes are all represented in certificates that are managed by authorities for the content information, conceptually similar to how a PKI CA is an

identity authority. By design this leads to a system that addresses the issue of policy makers, attribute certifiers, and users who are geographically and organizationally dispersed, and who are only related by various trust relationships. A prototype of this system is being used in DOE distributed scientific systems and collaborations. See [14].

Secure Group Communication

Group communication occurs in many different settings from low-level network multicasting to conferencing and other groupware applications. Group communication is often crucial in scientific and engineering collaborations. Regardless of the environment, security services are necessary to provide communication privacy and integrity when multiple parties are involved. This is not possible without secure and efficient key establishment, authentication and other security mechanisms geared for operational groups with dynamic (constantly changing) membership. Moreover, new (joining) members must be authorized and dynamic membership revocation must be supported. Addressing these issues is within the realm of “secure group communication” or “secure multicast.” There is preliminary IETF work in this area [11]. CLIQUES [15] is an R&D prototype system that also addresses these issues.

Integrated Solutions for Large Scale Distributed Application Environments

Globus [17] is a collection of services intended to facilitate widely distributed, multi-organizational, large-scale computing. It is the basis of NASA’s Information Power Grid project [18]. Because Globus was designed to operate in an open environment, a good deal of attention was paid to designing and implementing security services that were effective and easy for applications programmers to use. The Globus Security Infrastructure [19] is the result of this work, and is being adopted by several supercomputing centers that have

Integrated Solutions for Large Scale Distributed Application Environments cont.

interests in incorporating their resources into widely distributed systems. From [19]:

Increasingly, independent institutions with similar goals and interests are forming loosely coupled virtual organizations for collaboration and resource sharing. The construction of virtual organizations is hampered, however, by two conflicting goals: all members of the organization should have access to a resource as if it was their own, but participating institutions must not be required to change local security mechanisms or surrender control over their access control policies. We describe our experience designing developing, and deploying the Grid Security Infrastructure (GSI), and authentication and authorization infrastructure that meets these requirements. GSI capabilities include single sign-on, no plaintext passwords, proxy credentials, mapping to local security mechanisms (including Kerberos), site control over access control policies, and user-controlled delegation.

GSI provides a relatively complete set of security capabilities that are specifically aimed at NGI-like application environments. Since the reference cited in [19] was written, a lot of work has been done to have GSI more fully take advantage of PKI by adding such features as the ability to support multiple Certification Authorities at multiple organizations. There are experimental integrations with the GAA authorization approach.

SECURITY GUIDELINES FOR NGI APPLICATION DEVELOPERS

There are many different mechanisms available to developers desiring to implement security in NGI applications. The ones most appropriate for a particular application may depend on the nature of the application, the security mechanisms available in the intended operating environment, and the intended user base. Some basic guidelines include:

- **Determine application security priorities based on nature of application.**

The security mechanisms appropriate to financial transactions and medical data (data integrity and confidentiality) can be quite different from those for a distributed collaborative visualization environment (denial of service prevention). Will a simple allowed/denied model of access control suffice, or are different levels of access required for different users? Be sure to include the appropriate security requirements as part of the initial application design.

- **Do not re-invent the wheel.**

Wherever possible, use middleware and APIs to implement security functionality rather than building it from scratch, especially if your application will operate in an environment that implements a security infrastructure such as Kerberos or PKI. This is generally more secure than creating a custom system, reduces implementation vulnerabilities, and may facilitate scalability and future compatibility.

- **Minimize vulnerabilities within your program.**

An improperly designed or configured network application has the potential for allowing access to a user shell on the host system. Most systems have well-known vulnerabilities that can be exploited to allow an attacker to obtain superuser access from a user shell. Strong authentication and access control can be used to partially ameliorate such vulnerabilities, but the application should not introduce other vulnerabilities. A developer can minimize this risk by:

- *Minimizing the use of `setuid`.*
- *Ensuring the application handles exceptions such as buffer overflows gracefully.*
- *Taint-checking the application using tools such as the `perl taint` module.*
- *Minimizing the potential of application misconfiguration through clear documentation and straightforward configuration interfaces.*

SECURITY TECHNOLOGY ROADMAP

REQUIREMENT: Authentication of Remote Users

TECHNOLOGY		1 Year		3 Years
		DESCRIPTION	COMMENTS	
KERBEROS	Application Use	Remote login and access control for "standard" services.	RPCs are not nearly as widely used as socket based stream communication.	
		Client to server authentication, typically using DCE remote procedure calls.		
		Authenticated and encrypted messages via GSS.		
	Deployability/ Usability	Kerberos is widely deployed, DCE somewhat less so.	Kerberos has been most successful in centrally administered, single organization/single trust domain scenarios. Cross-realm use of Kerberos is possible, but painful.	
		GSS is used, but not widely. GSS is the basis of the Globus security infrastructure, and there is a mature Kerberos version of GSS.		
	Utility/ Effectiveness	Kerberos provides good access control for "standard" services, including the Andrew File System.	Some versions of Kerberos can make use of PKI certificates for user identity. Revocation of rights is relatively easy. Not integrated with Web browsers.	
For program communication, what many applications need is secure streams (e.g., secure sockets).				
PUBLIC KEY INFRA-STRUCTURE AND X.509 CERTIFICATES	Application Use	Remote login and access control for "standard" services.		
		Client/server and server/client authentication.		
		Authenticated and encrypted messages via GSS.		
		Authenticated and encrypted streams via SSL and TLS.		
		Authenticated and encrypted Web server access via https.		
	Deployability/ Usability	Standards conforming commercial products exist to provide the basic infrastructure: Certification Authorities and Certificate Directory servers.	The operation of the "infrastructure" is not trivial, being roughly equivalent to, e.g., running a DNS server, or an X.509 directory server. Revocation of rights is not easy without reference to a supporting access control mechanism.	
		Well integrated with Web servers, both open source (e.g., Apache) and commercial.		
		Most major Web browsers (e.g., Netscape Navigator) provide fairly sophisticated user identity management and trust management (through managing the identity certificates of multiple Certification Authorities).		
	Utility/ Effectiveness	Well integrated with the Web environment.	X.509 certificates are usable with SSL, etc., but current implementations are not integrated with PKI. (That is, unlike Web browsers, SSL libraries that an application might use do not manage multiple CAs or have access to CA directory servers for certificates).	
		Some commercial libraries (e.g., Entrust) support certificate management.		
		Globus uses PKI for most of the infrastructure that underlies the Globus Security Infrastructure (GSI). GSI provides a wide range of security services, including authentication.		
			Much better integration with a variety of services can be expected.	

SECURITY TECHNOLOGY ROADMAP cont.

REQUIREMENT: Authorization and Access Control

TECHNOLOGY		1 Year		3 Years
		DESCRIPTION	COMMENTS	
1) GAA 2) KEYNOTE 3) AKENTI 4) KERBEROS	Application Use	Once a user/client is authenticated, authorization determines if this entity is permitted to perform the requested action on the controlled resource. This is frequently part of the enforcement mechanism (access control).	General authorization is much more complex than authentication, involving, in the general case, interpretation of policy. However, some form is required beyond centrally administered access control lists.	
	Deployability/ Usability	GAA: A general API with a philosophy similar to GSS (i.e., a simple interface that can work with various underlying mechanisms).	Currently there is only a prototype implementation (whose policy is based on access control lists), however there is an effort to integrate GAA with Globus.	The integration with Globus is likely to spur development of GAA.
		Keynote: A trust management system with mechanisms for specifying application security policies and credentials via a standard language. It determines if requested actions are compatible with local policies.	An Internet-Draft based on Keynote version 2 was released in June 1999. Code should be available within the year.	
		Akenti: In its simplest form Akenti focuses on decentralized management of access groups.	An operational prototype exists and is being used in several scientific application environments.	
		Kerberos: Provides sophisticated centrally administered authorization. Access control is based on "tickets" (cryptographic credentials) that are issued after authorization. The tickets are quite versatile.	As noted above, Kerberos seems to work best in a single (though potentially large) organization context.	
	Utility/ Effectiveness	GAA: Should provide an easily used interface.	Like GSS, all of the complexity of GAA is hidden from the application. This is good. Like GSS, all of the complexity is in the implementation. There is currently very little implementation experience with GAA.	
		Keynote: Keynote is implemented via libraries. Compatible with X.509 names.	There is currently little implementation experience with Keynote.	
		Akenti: The application interface is quite simple, being that of the Apache and Netscape Web server "htaccess" module interface. Akenti is effective in an environment where the users/clients and stakeholders/policymakers are widely dispersed.	The very fact that Akenti can be effective in a widely dispersed arena also means that there is a fair bit of machinery that has to be put into place to make it work.	
		Kerberos: Integrates into applications via secure remote procedure calls. There is also a Kerberized version of GSS which provide a messaging interface to applications.	To make use of authorization in a large and complex environment involves a fair bit of knowledge and administration.	
				Akenti is an ongoing project, and it is expected that its maturity and capabilities (e.g., to easily manage multi-domain policy) will grow.
				There may be a version of GAA that uses Kerberos, thus providing an authorization function separate from access control.

SECURITY TECHNOLOGY ROADMAP cont.

REQUIREMENT: Single Sign-on

TECHNOLOGY		1 Year		3 Years
		DESCRIPTION	COMMENTS	
1) SSH 2) KERBEROS 3) GLOBUS	Application Use	<p>Single sign-on provides secure access to systems from "anywhere" by providing the user with a "single" cryptographic identity that can be used to establish a secure channel to remote systems.</p> <p>This is not something that would typically be used from within an application, but is obviously an important component of a secure application environment.</p>		
	Deployability/ Usability	<p>SSH: Widely available and easily deployed. Easily used.</p>		
		<p>Kerberos: See Authentication and Authorization Requirements</p>		
		<p>Globus: A suite of services that includes the Globus Security Infrastructure (GSI). A fair bit of infrastructure must be deployed and operated to enable the use of the security services.</p>	<p>There is a sizable and growing Globus developer community, and the GSI services have been adopted by several super-computing centers.</p>	
	Utility/ Effectiveness	<p>SSH: Widely used and quite effective at eliminating passwords in the clear. Can be used either just to provide an encrypted channel for passwords, or with public-keys that are used in access control lists.</p> <p>In the public-key mode ssh obviates the need for Unix passwords.</p>	<p>Not integrated with PKI/X.509.</p>	
		<p>Kerberos: See Authentication and Authorization Requirements</p>		
		<p>Globus: The security services are based on GSS, and both PKI and Kerberos based Globus versions exist.</p> <p>A range of services has been integrated with GSI, including SSH and ftp. The GSI version of the services does use an X.509 credential.</p>	<p>X.509 certificate management is still a bit "rough," though improving.</p>	

SECURITY TECHNOLOGY ROADMAP cont.

REQUIREMENT: Data Integrity

TECHNOLOGY		1 Year		3 Years
		DESCRIPTION	COMMENTS	
1) PGP 2) TRIPWIRE	Application Use	Data integrity is the use of "signed cryptographic fingerprinting" that ensures that the current version of a data block or file is the same as when it was signed. (See "digital signature" section of main report.)	This requirement is concerned with data integrity of "stored" files. Data integrity is almost always a requirement of data in transit, and is almost always provided within a secure communication channel. (See SSL, TLS, etc.)	
	Deployability/ Usability	PGP: Easily deployed.		
		Tripwire: Fairly widely used for integrity of system files. (e.g., for detecting maliciously modified files, esp. executable binaries.)		
	Utility/ Effectiveness	PGP: Useful for data integrity for single or small numbers of files.	There is no mechanism for automatically managing the database of files and their signatures.	
		Tripwire: A potentially general way to provide data integrity for files within a system.	Generally used only as a system management tool. No mechanism to manage files vs signatures after the file leaves a system. (There needs to be a way to provide a globally unique file name and to widely publish the signature for that file).	

SECURITY TECHNOLOGY ROADMAP cont.

REQUIREMENT: Protection Against Denial of Service

TECHNOLOGY		1 Year		3 Years
		DESCRIPTION	COMMENTS	
End system IPSec in combination with secure DNS	Application Use	Authenticating incoming packets so that the host of origin is assured can protect against several types of attacks that will disable the application or its platform: denial of service attacks that result from flooding the end system and consuming all available resource, or by injecting "poison" (packets that can, e.g., crash the application platform) can be ameliorated by rejecting packets that are not from known hosts.		
	Deployability/ Usability	Secure DNS is transparent to the application.	The host identity certificate use of secure DNS is barely deployed anywhere.	Secure DNS serving host identity certificates will probably get its impetus from the deployment of IPSec. Therefore, we should see this in the three year timeframe.
		IPSec is transparent to the application.	IPSec is just barely starting deployment. It is hard to get today, but most of the major vendors claim that it will be a routine part of their next major OS release. (i.e., in CY 2000).	IPSec should be a routine part of all OS releases in this timeframe.
	Utility/ Effectiveness		Limited performance because the header signatures are generated and checked in software.	Expect to see IPSec as a routine component of most Network Interface Cards. The chip sets have existed a year or more.

SECURITY TECHNOLOGY APPENDIX A - REFERENCES AND NOTES

[1] OpenSSL

"The OpenSSL Project is a collaborative effort to develop a robust, commercial-grade, full-featured, and Open Source toolkit implementing the Secure Sockets Layer (SSL v2/v3) and Transport Layer Security (TLS v1) protocols with full-strength cryptography worldwide. The project is managed by a worldwide community of volunteers that use the Internet to communicate, plan, and develop the OpenSSL toolkit and its related documentation."

<http://www.openssl.org/>

[2] SSL

More information on SSL, including various SSL toolkits for application developers, can be found at

<http://www.consensus.com/security/ssl-talk-faq.html>

[3] TLS

See <http://www.ietf.org/html.charters/tls-charter.html>

[4] SECUDE

"The SECUDE development kit is a library that offers well-known and established symmetric and asymmetric cryptography for popular hardware and operating system platforms. The development kit consists of a set of functions which allows the incorporation of security efficiency in practically any application (e.g., client/server, email, office applications) and a documentation in Hypertext Markup Language (HTML) which describe in detail the C programming interface. There are also various commands collected in a security command shell to ensure an immediate deployment of security."

<http://www.darmstadt.gmd.de/secude/>

[5] Entrust

This company offers PKI products and application libraries that include, e.g., the GSS API. See <http://www.entrust.com/>

[6] RSA

See <http://www.rsa.com/rsalabs/faq> for summary information on public key encryption, hash functions, signature functions, etc.

[7] Kerberos

See <http://gost.isi.edu/info/kerberos> for more info on Kerberos, including information on how to "kerberize" an application. In the United States and Canada, Kerberos is available via anonymous FTP from athena-dist.mit.edu.

[8] Ford

Warwick Ford. Computer Communications Security: Principles, Standard Protocols and Techniques. Prentice-Hall, Englewood Cliffs, New Jersey, 07632, 1995.

[9] Schneier

Bruce Schneier. Applied Cryptography: Protocols, Algorithms, and Source Code in C, Second Edition. John Wiley & Sons, 1995.

[10] PKI

See documents <http://csrc.nist.gov/pki/documents>, <http://gits-sec.treas.gov/gits-sec-home.htm> and the homepage of the IETF Public-Key Infrastructure (X.509) (pkix) Working Group at <http://www.ietf.org/html.charters/pkix-charter.html> for more information on PKI.

[11] Group security

An Internet draft on secure multicast key management protocols can be found at <http://www.ietf.org/internet-drafts/draft-ietf-IPsec-gkmframework-01.txt>

[12] IPsec

See <ftp://ftp.isi.edu/in-notes/rfc2411.txt> for the "IP Security Document Roadmap," <ftp://ftp.isi.edu/in-notes/rfc2401.txt> for "Security Architecture for the Internet Protocol," <ftp://ftp.isi.edu/in-notes/rfc2408.txt> for "Internet Security Association and Key Management Protocol (ISAKMP)" and <http://www.ietf.org/html.charters/IPsec-charter.html>

[13] GAA

"The GAA API facilitates authorization decisions for applications. An application invokes the GAA API functions:

- *to determine if a requested operation or set of operations is authorized or if additional checks are necessary.*
- *to request access control information about a particular resource.*

The API supports the needs of most applications, thus not forcing the developers to design their own authorization mechanisms. The API will allow better integration of multiple mechanisms with application servers, e.g., the GSS API can be used to obtain principal's identity." See http://gost.isi.edu/info/gaa_api.html

SECURITY TECHNOLOGY APPENDIX A - REFERENCES AND NOTES cont.

[14] Akenti

Akenti is a security model and architecture that is intended to provide scalable security services in highly distributed network environments. The project goals are:

- *to achieve the same level of expressiveness of access control that would be accomplished through a local human controller in the decision loop.*
- *to accurately reflect the existing policy: authority, delegation, and responsibility present in these environments.*

The approach makes use of:

- *digitally signed certificates capable of carrying:*
 - *user identity authentication*
 - *resource usage requirements (“use-conditions”)*
 - *user attribute authorizations (“attribute certificates”)*
 - *delegated authorization*
- *authorization decisions split among online and offline entities*

See <http://www.itg.lbl.gov/Akenti/>

[15] CLIQUES

CLIQUES' main objective is to fill a gap in the area of secure group communication by investigating group security services, designing a family of flexible and efficient cryptographic mechanisms, realizing it in a general-purpose toolkit and demonstrating its functionality by integration with diverse group-oriented applications.

CLIQUES is concerned primarily with group-oriented security services such as:

- *Key Agreement in Peer Groups*
- *Authentication*
- *Group Membership Changes*
- *Membership Non-repudiation*

See <http://www.isi.edu/~gts/CLIQUES/>

[16] GSS

Generic Security Service Application Program Interface (GSSAPI).

See the IETF Common Authentication Technology (CAT) Working Group homepage at <http://www.ietf.org/html.charters/cat-charter.html> and <ftp://ftp.isi.edu/in-notes/rfc2078.txt>.

[17] Globus

See <http://www.globus.org/>

[18] IPG

See <http://www.nas.nasa.gov/IPG>

[19] GSI

“Design and Deployment of a National-Scale Authentication Infrastructure.” R. Butler, D. Engert, I. Foster, C. Kesselman, S. Tuecke, J. Volmer, V. Welch. (Submitting). “Describes our experience designing, developing, and deploying the Grid Security Infrastructure.” See <http://www-fp.globus.org/documentation/papers.html>.

[20] Non-Repudiation

IETF standards being developed for non-repudiation related topics:

- trusted time stamp authorities <http://www.ietf.org/html.charters/stime-charter.html>
- data certification servers <http://www.ietf.org/html.charters/xmlsig-charter.html>

[21] Tripwire

“Tripwire is a straightforward tool with a single purpose: detect any variance in file integrity. This means that Tripwire can absolutely, unequivocally determine if a protected file has been altered in a way that violates the policy set by the administrator. Tripwire can also determine if files have been added to or deleted from protected system directories. Tripwire also has a powerful and flexible policy language used to define exactly what Tripwire should pay attention to, allowing for minimal ‘noise’ or ‘false positives’ in the reports. Starting with a template policy file appropriate for your operating system, Tripwire makes it very easy to update your policy files as often as you like. This results in reports so concise that you will be able to quickly determine the state of your systems.” See <http://www.tripwiresecurity.com/prodintro.html>.

See “U.S. To Allow Export Of Encryption Products” at http://dailynews.yahoo.com/h/nm/19990916/tc/exports_2.html.

FIRST DAY, MORNING, AUGUST 10

- 0745-0845 LEADERSHIP BREAKFAST MEETING**
Organizers, Breakout leads, LSN/NCO/HPNAT/NRT/JET/I2 leads
- 0730-0900 CONTINENTAL BREAKFAST & REGISTRATION**
All participants
- 0900-0920 INTRODUCTIONS**
Bessie Whitaker, NASA/NREN Project Manager
- WELCOME*
Jack Hansen, Deputy Director for Research NASA Ames
- WORKSHOP PURPOSE AND RATIONALE*
Dick desJardins, NASA/NREN

- 0920-0950 KEYNOTE: “It’s Time to Get Physical”**
Dr. David Tennenhouse, DARPA Chief Scientist

- 0950-1030 CASE STUDIES 1: Digital Earth**
- TerraVision* - Yvan Leclerc, SRI International
- Earth System Grid* - Dean Williams, LLNL
- Distributed Image SpreadSheet* - K. Palaniappan, University of Missouri
- Interactive Space Communications for Remote Investigators* - Thom Stone, NASA/NREN
- Climate Data Access and Visualization* - Don Denbo, NOAA/PMEL

- 1030-1045 BREAK**
All participants

- 1045-1120 TECHNOLOGY 1: QOS**
Facilitator: Doug Montgomery, NIST
Presenter: John Wroclawski, MIT

- 1120-1150 CASE STUDIES 2: DIGITAL VIDEO**
- NGI/Internet2 Advanced Digital Video* - Joe Mambretti, iCAIR/MREN
- High-Quality High-Bandwidth On-Demand Video* - Amy Philipson/ Letcher Ross, Research TV/University of Washington
- Virtual Room Videoconferencing System* - Philippe Galvez, CalTech

- 1150-1250 LUNCH & DEMO WALKAROUND**
All participants

FIRST DAY, AFTERNOON, AUGUST 10

- 1250-1325 TECHNOLOGY 2: MULTICAST**
Facilitator: Marjory Johnson, NASA/RIACS
Presenter: Don Towsley, Univ. of Massachusetts

- 1325-1350 CASE STUDIES 3: TELEMEDICINE**
- Visible Human* - Mike Gill, NIH/Haruyuki Tatsumi, Sapporo Medical University
- Virtual Collaborative Clinic* - Marjory Johnson, NASA/RIACS

- 1350-1405 BREAK**
All participants

FIRST DAY, AFTERNOON, AUGUST 10 CONT.

- 1405-1440** **TECHNOLOGY 3: SECURITY**
Facilitator: Bill Johnston, NASA and DOE
Presenter: Steve Kent, BBN
- 1440-1505** **CASE STUDIES 4: CHINA CLIPPER**
Distributed Data Intensive Applications
Bob Lucas/Brian Tierney, LBNL
- 1505-1540** **CASE STUDIES 5: NPACI**
PACI Overview - Steve Elbert, NSF
Terascale Computing/Global Mass Storage -
Anke Kamrath, UCSD/SDSC
Data Grids - Reagan Moore, UCSD/SDSC
Telescience - Mark Ellisman, UCSD
- 1540-1600** **BREAK**
All participants
- 1600-1700** **TESTBEDS & NGI PROGRAM OBJECTIVES**
Globus - Ian Foster, ANL
JET - Javad Boroumand, NSF
Internet2 - Guy Almes/Ted Hanss
PITAC - Raj Reddy, Carnegie Mellon University
Instructions to Breakout Groups
- 1700-1800** **BREAKOUT GROUPS GET-
ORGANIZED SESSIONS**
All participants
- 1800-2000** **BUFFET DINNER & DEMO WALKAROUND**

SECOND DAY, AUGUST 11

- 0730-0830** **CONTINENTAL BREAKFAST**
All participants
- 0830-1230** **BREAKOUT GROUPS - QOS,
MULTICAST, SECURITY**
All participants
- 1230-1330** **LUNCH**
All participants
- 1330-1430** **BREAKOUT GROUPS REPORTING &
WORKSHOP WRAP-UP**
All participants
- 1500-1700** **LEADERSHIP MEETING**
Organizers, Breakout leads, LSN/NCO/HPNAT/
NRT/JET/I2 leads

QoS LANDSCAPE WITH APPLICATIONS

John Wroclawski, Massachusetts Institute of Technology

The concept of Quality of Service, or QoS, is often interpreted differently by different communities. The networking community uses the phrase “QoS” to describe attributes of the network service received by an application in normal operation. “QoS control” is the ability to explicitly control these attributes and their variation over time. QoS is traditionally expressed, and requested, in network-level parameters such as bandwidth, delay, and packet loss probability.

Recent research is turning to the problem of more effectively bridging the gap between traditional network-level QoS control and the needs of application designers. Two goals of this work are to allow application designers to express their requirements in a more natural manner, and to develop network-level mechanisms that address advanced application requirements such as graceful degradation under load and dynamic construction of performance relationships between multiple data streams.

Simple to Sophisticated

Early QoS technologies typically revolved around a simple model, in which “the application” or “the user” requested a certain level of performance for its data stream, and “the network” granted or refused the request. This model is appropriate for a centrally managed network with a narrow class of telephony-like applications, but is inadequate for the more demanding NGI environment.

Today’s evolving model adds two key concepts:

- *Many parties will wish to exert QoS control. Users and application designers are one group. Information systems managers and those charged with allocating resources among applications are another. Core network operators who wish to differentiate among their customers are a third. Further, there may be more than one “network” in the path between application hosts. A fully developed QoS control environment will allow all of these players to exert significant control over resource allocation, with differing objectives reconciled through administrative or economic mechanisms.*
- *Application usage models vary widely. Some QoS-aware applications will support the signaled long-lived flow model envisioned by traditional QoS technologies. Two other categories are of interest. The first is QoS augmentation of “traditional” best-effort applications, controlled by the application or the network policy administrator. The second is QoS control for applications such as web services where each “flow” is a very short transaction.*

QoS and Adaptive Applications

Widespread deployment of best-effort Internet protocols has led to significant advances in the technology of adaptive applications. These applications share the property that they are designed to operate as effectively as possible over a wide range of network performance levels. A simple example of adaptivity is an FTP application based on TCP that moves data exactly as fast as the network allows. Another example is a multimedia conferencing application that adapts its data coding formats and playback point to network throughput, loss, and delay; giving the best possible perceptual results over a wide range of operating conditions.

Adaptive applications are valuable even in the presence of network QoS control.

- *The combination of adaptive applications and QoS control puts performance choices in the hands of the user rather than the application designer. A single application, operating with different QoS control parameters, can meet the varying needs of a wide range of users, or of the same user in different circumstances. Users are able to trade off application performance and the cost of network resources according to local priorities.*
- *Adaptive applications combined with correctly designed QoS control mechanisms can assure a minimum level of application performance, while offering better performance when network resources are not constrained. Users receive the benefits of controlled performance without losing the incentive to shift workload to less utilized resources.*

QoS LANDSCAPE WITH APPLICATIONS cont.

QoS Tradeoffs

Network QoS control is a process of tradeoffs. The application designer that understands these tradeoffs will be able to develop applications that are usable with the widest range of network technologies and conditions.

- *How rigid are the application's QoS requirements? Applications that can tolerate occasional deviations from their requested service level, or can adapt to variations in service level, impose significantly less demand on the network QoS mechanisms, and can be supported with fewer resources and simpler mechanisms.*
- *What percentage of the total available network resources will be requested by the application? Applications that require a high percentage of the available resources demand more complex resource allocation*

and management mechanisms, leading to higher cost and lower scalability.

- *How dynamic is the resource demand? Applications with relatively static demands require simpler signaling and network management than more dynamic situations. Applications with very short demand bursts are the most difficult to handle, because demand is highly variable but signaling becomes impossible.*
- *How secure must the application data stream be? QoS control requires that components within the network be able to differentiate between types of traffic based on various criteria. The more information hidden by end-to-end encryption and other techniques, the less is available to perform this function.*

TECHNOLOGY ROADMAP

FUNCTION	TECHNOLOGY	DEPLOYMENT	STANDARDIZATION
Traffic scheduling and buffer management	Classical scheduling tech. (WFQ, CBQ, RED, etc.)	Widespread	Generally inappropriate
	Per-flow Intserv Schedulers	Some	Standards in place
	Diffserv PHB's	Some, momentum	Early standards in place
Per-network resource allocation and control	Static resource allocation	Commercial tools	Limited work, specific needs not yet identified
	RSVP for per-cloud aggregate dynamic b/w management	Early commercial development	Work in progress
	Alternative bandwidth managers	Research, prototype development	Too Soon
End-to-end service construction	RSVP for end-to-end signaling	Some (routers) Starting (hosts)	Standards in place
	Alternative inter-provider bandwidth broker protocols	Research	Too Soon
Policy capture and control	Proprietary packages	Several, growing use	N.A.
	Std. policy comm. protocol	Not yet significant	Nearing completion
	Common policy	Not yet significant	Initial work in progress

MULTICAST TECHNOLOGY IN THE NEXT GENERATION INTERNET

Don Towsley, U. Massachusetts

Many applications that will run over the next generation Internet (NGI) will involve three or more participants. Examples of such applications include:

- *Data set transfers to large user populations*
- *Continuous sensor data feed to large user populations*
- *Group teleconferences*
- *Dissemination of video and/or audio streams to large user populations*
- *Distance learning/lecturing*
- *Distributed interactive simulation*

All of these applications can benefit from using multicast, a technology that has been developed and refined over the last ten years on the current Internet. Briefly, the benefits of using this technology are twofold:

- *Substantial reductions in bandwidth usage and end-host processing requirements*
- *Reduction in application development effort*

Consider a video lecture to an audience of 2,000 over a wide-area-network at reasonable quality. A simple calculation indicates that multicast can reduce the bandwidth requirements from nearly one Gbps to several hundred Kbps. Furthermore, the requirements for the source processor (for protocol processing) is reduced from 2,000 MHz (which is not yet available) to one MHz.

The current multicast technology is based on IP (Internet Protocol) multicast, which provides a “best effort” service. IP multicast is a group-oriented architecture where an address is associated with a multicast group. A sender sends IP packets to that group. Anyone interested in receiving packets sent to that group explicitly joins that group and, once joined, receives any packet destined for that group. Any host (user) is permitted to join the group and any host is permitted to send to that group. There is no network-layer mechanism that identifies members of the

group to any user. This architecture requires an infrastructure to deliver multicast-addressed packets to all hosts that have joined that multicast group. Such an infrastructure has been developed and works reasonably well in a single network. However, there are still some difficult issues to be addressed before a reasonable infrastructure is in place for a true Internet. These deal with a limited group address space and difficulties in performing multicast routing across networks.

In order for multicast to be useful, a variety of higher-level services are required. These include:

- *Reliable delivery*
- *Real-time*
- *Access control/security*
- *Management/debugging*

We examine each in turn.

Reliable delivery:

Considerable progress has been made in the development of protocols for providing reliable delivery of data from a single source to many receivers. In order to deal with large numbers of receivers and varying network conditions, these protocols make use of forward error correction, placement of loss recovery responsibilities at receivers, and the use of local recovery techniques. Several protocols have been developed which have been used to deliver data reliably to groups consisting of hundreds to thousands of receivers. To date, multiple source applications requiring reliable data delivery are usually treated as a collection of single source applications. There are currently several protocols available and the IETF is in the process of developing standards in this area. The most prominent outstanding technical issue relates to how to incorporate congestion control into these protocols. This is currently being explored within the confines of the IRTF.

MULTICAST TECHNOLOGY IN THE NEXT GENERATION INTERNET cont.

Real-time:

The standard protocol for real-time data is RTP/RTCP, which was developed with both point-to-point and multicast applications in mind. Full reliability is usually not required; however combinations of retransmissions and FEC can be used to enhance the quality at the receiver. It is expected that support for higher quality will come out of the efforts on developing differentiated services architecture. The key issue that arises when delivering real-time streams to large user populations has to do with the heterogeneous receiver population. This has not been completely resolved, however a promising approach is the use of layering which permits receivers to choose the quality of the stream that matches its capabilities.

Access control/security:

Multicast poses significant and unique problems in the security arena for two reasons. First, the architecture has been designed to permit any host to send to a group and any host to receive packets destined to that group. Second, many applications are characterized by dynamic changes in the group membership. Encryption can be used to secure the transfer of data. However, there is no adequate way to deal with a malicious user bombarding a group with unwanted data. Data encryption, combined with group dynamics, introduces the most significant multicast security problem, namely, how to distribute keys every time the group changes with the departure of a group member.

Management/debugging:

Last, multicast networks and multicast application development introduce new problems related to monitoring and troubleshooting networks and applications. Although several tools/protocols exist to aid the network administrator and/or the applications developer, much still needs to be done in this arena.

In summary, a wide variety of applications have been successfully demonstrated using multicast in the current Internet. As multicast technology further develops and evolves, and higher-level services are refined and created, we should see an explosion in the number, diversity, and size of applications based on this technology.

References and useful links

<http://ale.east.isi.edu/RMRG/> IRTF Reliable Multicast Research Group Web site.

<http://www.irtf.org/charters/secure-multicast.htm> IRTF Secure Multicast Group Charter and contact information

<http://www.internet2.edu/multicast> Internet2 Multicast Working Group site

<http://www.ietf.org/html.charters/mboned-charter.html> IETF Mbone Deployment Working Group Web site

<http://www-mice.cs.ucl.ac.uk/multimedia/software/> Contains a list of available multicast-capable freeware for conferencing, archiving, and delivering stored multimedia content.

THREATS, VULNERABILITIES, AND SECURITY STRATEGIES FOR APPLICATIONS

Stephen Kent, BBN Technologies (GTE)

The Internet is a hostile environment. This mini-white paper begins by introducing the terminology used to discuss security issues. It then provides an overview of threats, examples of attacks that illustrate the nature of threats, and explores the changing nature of threat. It explains a security philosophy that seeks to minimize the points at which attacks can be effectively mounted against distributed applications. Strategies for implementing this philosophy, based on standard security protocols and security infrastructures, complete the mini-paper.

Vulnerabilities, Attacks, and Threats

A vulnerability is a security flaw in a system. All systems exhibit vulnerabilities, when viewed from a system perspective. An attack is a means of exploiting a vulnerability. Usually there is a many-to-one relationship between attacks and vulnerabilities, i.e., several different attacks may be used to exploit the same vulnerability. A countermeasure is a security mechanism of procedure employed to counter one or more types of attacks. A threat is a motivated, capable adversary. The adversary is capable of mounting attacks against the target, and is motivated to do so. Some examples help illustrate this terminology.

Consider a Web-based application relies on passwords to authenticate a user; this is a vulnerability. An adversary may engage in a passive wiretapping attack to intercept passwords, and then pose as authorized users. Alternatively, an adversary may try to guess passwords as a means of gaining unauthorized access to the application. Use of the Secure Socket Layer (SSL) protocol counters passive wiretapping, but does nothing to prevent password guessing, illustrating how a countermeasure may thwart an attack but not remedy an underlying vulnerability. Use of client public key certificates, an optional feature of SSL, counters password-guessing attacks, providing strong user authentication.

The Web server probably executes on a common operating system platform, e.g., Unix or Windows. These OSs typically embody a number of vulnerabilities that could be exploited

by a knowledgeable adversary, giving him access to the application, even if one makes use of the best user authentication technology. Many of the known OS vulnerabilities can be remedied through the application of patches available from OS vendors. One might assume that if the system administrators for the Web site diligently apply the patches received, the system would be secure against such attacks. However, a sufficiently capable and motivated attacker might create a CD-ROM and distribute a legitimate security patch that also introduces new vulnerabilities into the OS. Such malicious software is termed a Trojan Horse.

Vulnerabilities not intentionally introduced into systems by attackers tend to arise from three sources: design errors, implementation errors, and management errors. Unless one is well trained in security technology, it is hard to design a secure system component, a fact that application designers should keep in mind. Implementation errors often arise due to poor software engineering practices; they are bugs that happen to have adverse security implications. Management vulnerabilities may be a side effect of security systems that are just too difficult to manage, or of poor procedures, inadequate training, inattention to detail, etc.

There is a wide range of adversaries. Most attackers are not technically sophisticated, but instead make use of attack tools that are developed by a small cadre of very capable designers who are motivated to produce such tools. Hackers (more properly crackers) get the most press, but disgruntled employees are a more serious problem in many contexts. Industrial and national spies, terrorists, special interest groups, criminals, and even investigative journalists are all potential adversaries. Each has different skill sets, different motivations, and different levels of aversion to detection. Whether any of these becomes a threat depends on the target system.

Finally, one can distinguish between two distinct types of attacks: targeted and opportunistic. The latter is what one typically sees, the result of a hacker or disgruntled employee

THREATS, VULNERABILITIES, AND SECURITY STRATEGIES FOR APPLICATIONS cont.

taking advantage of vulnerabilities in a system that may be selected on a whim, or out of curiosity. In contrast, an adversary who targets a specific system because it is perceived to be valuable (in some dimension), may employ a more extensive set of tools and a well thought-out strategy for achieving his goal.

A Security Philosophy and Implementation Strategies

It is a truism that the greater the number of system components that must be relied upon for security, the harder it will be to achieve security. This observation motivates the principle of least privilege, which calls for each component of a system to have access to only that data needed to perform its task. In a distributed system environment, this principle motivates the concept of end-to-end security for communication, so that only the source and destination, not intermediate nodes, have (authorized) access data. TCP/IP largely embodies this principle, although support for differentiated service and some forms of firewalls conflicts with the principle. With regard to storage and processing of data, this principle argues for encrypted, integrity protected file servers, and against “outsourcing” processing to servers. (Distributed system architectures that rely on ORBs, directories, etc., increase opportunities for security compromises, including denial of service.)

To secure an application, one begins by establishing its security requirements in terms of standard security services: confidentiality, origin authentication, integrity, access control, non-repudiation, and availability. In general, applications should start by making use of OS security mechanisms, since no application will be more secure than the OS on which it executes. OS security may fall short of what is required for an application, e.g., due to limitations in the granularity of access control, prompting the need for application-specific safeguards. Applications should take advantage of network layer communication security mechanisms, e.g., IPsec, if possible. In addition to the obvious desire to avoid “reinventing the wheel,” this advice

is motivated by experience indicating that an application developer is likely to design and implement security facilities that contain vulnerabilities.

However, application-specific security mechanisms may be required when the semantics of the application are not well matched to standard, available security protocols. For example, e-mail and directory services, due to their staged delivery nature, require their own security protocols. Secure E-mail also embodies semantics unique to its environment, e.g., signed messages and signed receipts, that are not supported by lower layer security protocols. Under such circumstances, lower layer communication and OS security mechanisms, are inadequate to support application security requirements. Where possible, such safeguards should be based on standard security mechanisms (e.g., algorithms), infrastructures (e.g., public key certificates), and APIs (e.g., GSSAPI, PKSC 11, etc.). Even making use of such facilities, the design and implementation of application-specific security features will be fraught with danger.

TERRAVISION

Yvan Leclerc, SRI International

TerraVision is a real-time terrain visualization application that SRI has developed over the last seven years as part of the DARPA-sponsored MAGIC project. The initial requirements were to design and implement a visualization application that used remotely stored terrain/image data accessed over a high-speed network at rates approaching 1 Gbps. Though TerraVision typically achieves rates closer to 100 Mbps, it still enables high-quality visualization of very large databases (tens of gigabytes in size) over cross-country networks.

The ultimate objective of our application is to allow end-users to virtually visit and navigate through high-resolution 3-D representations of any part of the world. This objective is implemented as a client (TerraVision) that requests 3-D data from datasets stored on remote servers in real time as the user navigates around the world. Currently, the user selects a limited number of datasets from a table of available datasets stored on the servers. This application can be used on networks with throughputs in the range of 0.5 to 100 Mbps. TerraVision currently supports three classes of servers: Distributed Parallel Server Systems (from LBL), HTTP servers, and NFS.

The application, as described above, is currently functional and is available for download at <http://www.ai.sri.com/TerraVision>. Extensions to the client and distributed storage systems will be made in the following two and one-half years. The primary extensions to the storage systems will be the creation of an open DNS hierarchy of servers that will allow any application to request information (metadata and pointers to data) about a given geographical area in the world.

We are currently proposing a hierarchy of the form `minutes.degrees.tendegrees.geo`. For example, a server with DNS name `10e20n.geo` is responsible for a 10-degree x 10-degree cell of the world. The service area of the cell spans from longitude 10 degrees East and latitude 20 degrees North to longitude 20 degrees West and latitude 30 degrees North. Similarly, server `1e5n.10e20n.geo` is responsible for a 1-degree x 1-degree cell since it is at the third level of the hierarchy `degrees.tendegrees.geo`. The name `1e5n` indicates that the location of the cell is longitude 1 degree East and latitude 5 degrees North relative to its parent server `10e20n.geo`, or longitude 21 degrees East and latitude 25 degrees North. This is described in more detail at <http://www.ai.sri.com/digital-earth>.

TerraVision has been used over the MAGIC network (ATM with one or more OC-3 connections), the ACTS satellite network, other high-speed ATM networks, and over the Internet. Although it provides the most satisfying end-user experience over high-speed, low-latency networks, it is still quite useable over the Internet with T1 connectivity.

PROTOTYPING AN EARTH SYSTEM GRID

Marla Meehl, NCAR

The need to evaluate climate change scenarios under the Kyoto accord makes climate modeling a mission critical application area. DOE's Accelerated Climate Prediction Initiative (ACPI) seeks to address this need through the creation of an advanced climate simulation program, which will accelerate the execution of climate models one hundred-fold by 2005 relative to the execution rate of today's climate models [ACPI98]. High-resolution, long-duration simulations performed under ACPI will produce tens of petabytes of output. The output in turn will be made available to global change impacts researchers nationwide through a network of diagnostic and regional climate centers [ACPI98, GATE99]. These distributed centers, users, models, and data will be connected in a virtual collaborative environment called the Earth System Grid (ESG). The Earth System Grid will provide scientists with virtual proximity to the distributed data and resources comprising this collaborative environment.

Creating an effective and efficient ESG in support of ACPI is challenging at multiple levels, but above all it is a Next Generation Internet (NGI) problem. A large community of global change researchers at laboratories and universities around the nation will need to access significant fractions of the data. User requests for data products will be translated into appropriate combinations of accesses to data caches, requests to central data archives, and new large-scale simulations. The effective management of the required data movement operations will tax even the highest performance and most advanced networks.

We propose a research and development project that will take a first step towards the creation of an Earth System Grid. Specifically, we propose to prototype a system that will support:

- *The rapid transport of climate data between centers and users in response to user requests. The focus is on end-user accessibility and ease of use, which will be accomplished through both the modification of existing*

applications and tools and the development of new tools as needed to operate seamlessly in the Earth System Grid.

- *Integrated middleware and network mechanisms that broker and manage high-speed, secure, and reliable access to data and other resources in a wide area system.*
- *A persistent Earth System Grid testbed that provides virtual proximity and demonstrates reliable high-performance data transport across a heterogeneous environment.*

In constructing this system, we will build on a substantial base of software and experience that includes parallel climate models, high-performance networking, climate model analysis tools, and advanced networked middleware. We also leverage substantial existing investments in supercomputers, servers, mass storage systems, and high-speed networking.

The proposed research and development activities will be performed by a partnership between four DOE Laboratories (ANL, LANL, LBNL, LLNL), a NSF center (NCAR), and two universities (Wisconsin, USC). This uniquely qualified team—most of whom have worked together closely over many years—includes experts in applications, middleware, and networking. Working together, this team will construct an outstanding driver and showcase for DOE NGI research and networks. In addition, making it possible for the research community to readily access distributed computers, simulation models, and data for scientific discovery will also accelerate climate research.

DISTRIBUTED IMAGE SPREADSHEET (DISS)

*K. Palaniappan, Univ. of Missouri-Columbia
A. F. Hasler, NASA Goddard Space Flight Center*

Digital libraries of geophysical datasets are now terabytes in size, and with the advent of a new generation of Earth Science Enterprise high data rate satellite sensors, exponential growth will continue. The Distributed Image Spread-Sheet (DISS) is an interactive scientific visualization and analysis tool that provides a novel multicell spreadsheet interface for constructing, organizing, and intercomparing gigabyte-sized geophysical datasets in collaborative environments.

The DISS uses high-performance networks to enable scientists to study and compare the large volumes of data collected by the next generation satellite systems often in a near real-time mode. An OpenGL version of the DISS is being developed for use with EOS direct broadcast MODIS instrument data from the Terra satellite. The DISS software tool uses a multidimensional spreadsheet arrangement of cells and frames for manipulating gigabytes of multichannel image and model data, and supports image and grid analysis algorithms. Highly interactive visual browsing tools, such as synchronized animation, roam and zoom, stereoscopic display, navigated data probing, surface rendering, flight paths and volume visualization for interacting with arbitrary-sized multidimensional data in each frame of the DISS have been demonstrated to be highly successful for quickly inspecting thousands of separate datasets.

High-speed network access using http and ftp methods, combined with novel image and geometry data compression schemes for efficient bandwidth utilization is supported. High-performance network access to supercomputing resources is necessary for manipulating datasets that require radiometric calibration, geolocation and remapping, regriding, generation of scientific products and data mining

within the interactive visualization framework. Enhanced collaborative capabilities will require incorporating QoS and multicast features combined with multimedia (i.e., MPEG-4) streaming and synchronization features over ATM.

Multisource geophysical datasets from satellites, aircraft instruments, ground-based radar, synoptic measurements, geographical map layers, and assimilated numerical model data in different cartographic projections can be remapped and readily intercompared both qualitatively and quantitatively. Gridded data (NCSA HDF, Vis5D, GrADS), image formats (SGI, JPEG, GIF, TIFF, PNM, SLCCA), video formats (MPEG, VSLCCA), and gzip compression are currently supported in the DISS.

The DISS has been used as the primary tool for presentations of the NASA/NOAA/AMS Earth Science Electronic Theater at the annual American Meteorological Society meetings and numerous public venues such as museums, international scientific meetings, etc. The Distributed Image SpreadSheet (DISS) has been chosen as a testbed application to demonstrate the potential of the Next Generation Internet (NGI), the National Science Foundation Very high speed Backbone Network Service (vBNS), and Internet 2.

INTERACTIVE SPACE COMMUNICATIONS FOR REMOTE INVESTIGATORS

Thom Stone, NASA/NREN

Dealing with science on space missions has always been a problem. In the past it has been accomplished by:

- *Forcing science users to uproot their lives and research to go to wherever the mission was operated, sometimes for years and at great expense to the government and taxpayers*
- *Sending the downlinked data to nine-track tape to be locked away until Congress funds the science*
- *Creating air gaps and expensive proprietary networks, thus ensuring that the Principal Investigator (PI) gets the data at the greatest possible cost and when it is too late to correct problems*
- *Creating backdoors to the Internet and accepting its limitations of unpredictable performance and security vulnerabilities*

More complex and longer duration missions where science payloads can change over time, such as the International Space Station and the Earth Observing System, have made these methods impractical. Modern instrumentation requires expert interaction and monitoring from researchers operating at their home institutions with their usual tools. Next Generation Internet technologies will make such remote investigations practical.

The following functional capabilities would allow remote users (e.g., those not “inside” the payload operations center) to interact with spacecraft instrumentation and data.

- *Receipt of near real-time space data feeds at investigator’s site on the platform of their choice in a format that is meaningful*
- *Planning support (running simulations, managing timelines, generating command sequences)*
- *Access to historical data (and provision of appropriate analysis tools)*

- *Collaboration tools for use with the Payload Operations Center and others in the science/engineering community of interest, including:*
 - *Messaging*
 - *Voice*
 - *Video conferencing*
 - *Planning tools*
- *Access to spacecraft voice loops*
- *Video feeds, especially for manned flights*
- *High-resolution images*
- *Telescience capabilities for manned flights*
- *Support for operational simulations, testing and training*

Two major impediments to achieving true PI/mission interaction are security and bandwidth issues. The first priority in providing proper security is to prevent the distribution of science data from being an entry point for mischief to the Operations Center or the spacecraft. Protecting sensitive or proprietary data must also be addressed.

Spacecraft instrument support is a bandwidth-intensive application. Bandwidth requirements could be as high as 5-20 Mbps per investigator station, and payloads might require 5-20 stations. The 5-20 Mbps per station would not be a sustained rate, as payload management applications are interactive and usage would be stochastic. The peak load would strain existing local area networks and wide area infrastructures.

Peak data-rate considerations are not the only bandwidth issues. The more complex issue is to ensure timely and accurate delivery of multimedia data (near real-time feeds, video and voice) along with traditional Internet traffic. Multimedia data has rigid constraints on factors such as delay, jitter, and packet loss and also requires sustained bandwidth.

INTERACTIVE SPACE COMMUNICATIONS FOR REMOTE INVESTIGATORS cont.

IP Quality of Service (QoS) provisioning is in the design and test stage. QoS addresses proper transmission of Internet-provided multimedia content.

More than one remote station may have to receive the same real-time data feed or channel of compressed digital video. If a copy had to be sent out for each user, even the planned high-speed backbones would be congested. Multicast technology, where a single copy can be routed to multiple users, can provide relief for this congestion.

The challenge to be met in providing access for remote science over the Next Generation Internet is to combine high-speed technologies, security, QoS and multicast.

CLIMATE DATA ACCESS AND VISUALIZATION IN NOAA

Donald W. Denbo, NOAA/PMEL

Emerging technologies in NOAA include networked access to distributed data and databases, virtual reality, immersive and tele-immersive visualization, collaboration and collaborative virtual environments (CVEs). Successful integration of fast networks and emerging technologies in visualization, analysis and collaboration software will lead to improved understanding of NOAA's large and complex environmental data sets.

NOAA has successfully prototyped systems for networked access to distributed data sets using CORBA, Java RMI, Java graphics and a Java Virtual Reality Modeling Language (VRML) DataExplorer. Habanero has been applied for fully functioned collaborative application sharing of the distributed data access software. NOAA's HPCC program is establishing fast Internet2 connections, and test beds are being established to develop Quality of Services to support NGI applications. In addition to networked access to heterogeneous, distributed environmental data sets, NOAA is developing centralized access to distributed databases of budget and financial information, where security is an additional concern.

Virtual reality techniques explored in NOAA span the range from low-cost, readily available, Web-accessible VRML, to our first, recently acquired ImmersaDesk™. VRML is an excellent, approachable desktop methodology for exploring data in 3-D and in stereo, and it can be shared in to create a very accessible collaborative virtual environment. The ImmersaDesk™ is a semi-immersive, tele-immersion device.

NGI/INTERNET2 ADVANCED DIGITAL VIDEO

Joe Mambretti, International Center for Advanced Internet Research,
Northwestern University

With its research partners, iCAIR is active in a number of advanced digital video projects that are directed at enhancing the state-of-the-art as well as enabling better science through the development of new visualization tools. Projects include the development of a comprehensive DV capability for the next-generation Internet, and include efforts related to:

- a) *an architecture that will provide a general digital video "dial tone"*
- b) *access technologies for edge devices, such as DV portals*
- c) *three specific modalities utilizing advanced digital video technologies—video conferencing, video-on-demand, and live transmission*
- d) *digital media asset management (including through various server technologies, digital libraries, video jukebox technologies, metadata, automatic DV metadata extraction and indexing, RDF/XML, media object designation, etc.)*
- e) *DV infrastructure architecture and prototyping, especially with regard to the Grid, e.g., multicasting, DiffServ/QoS scalable streams via massively parallel supercomputer, file systems, etc,*
- f) *digital production studios*
- g) *content issues, such as channel design and development, DV variations—VR movies, simulations, animation, etc.*
- h) *integration with other applications*
- i) *replication services, and*
- j) *API links to middleware, especially Globus, including components defined in the recent IETF draft on middleware.*

Scientific objectives of application; implementation as an end-to-end system over NGI networks:

The general objective is to expand DV from its current restricted usage on the Internet. That will allow it to be utilized more as a common data type, especially with regard to core architecture, access, integration with other applications, infrastructure scalability, differentiated services, quality of services, and interfaces with other Internet technology components. Implementation over NGI networks as an end-to-end solution will occur initially through prototype capabilities. These capabilities will be established at specialized facilities (e.g., Advanced Internet Application Facility) and national test beds, especially through the NGI Emerge project, and through the Internet2 I2-Digital Video Network (I2-DVN project) over the VBNS and Abilene, along with related projects such as QBone.

Status; current and planned functionality; time line for applications development:

The majority of the development projects have currently been established, some architecture developed, and prototype capabilities have been presented at national conferences, e.g., GiDVN (part of iGRID) and NASA Astrophysics VR demo at SC'98, Internet2, INET'99, etc. A prototype International Virtual Institute with DV capabilities and two prototype Digital Video Portals have been created, which provide common PC end-station access to DV capabilities. One Portal, I2 VideoSpace, has been persistently available since April 1999, based on a new facility center on a massively parallel supercomputer. Increasingly over the next year, additional capabilities derived from the listed projects will be developed and deployed in prototype, including those based on digital library capabilities and DV and QoS integration.

NGI/INTERNET2 ADVANCED DIGITAL VIDEO cont.

Networking technologies used or planned; how they enhance or support the application; issues or difficulties with planned implementation:

API signaling, DiffServ/QoS, core resource management, initially via PVCs and eventually through Bandwidth Broker mechanisms, integration with middleware as delineated in the IETF Middleware draft. Additional performance tracking and reporting mechanisms are required.

Requirements or desirements for technological enhancements; technologies being investigated; help sought from the technology community:

DiffServ/QoS implementation in national backbones, enhanced implementations of QoS capabilities in routers, common interfaces with components as designated in the IETF Middleware draft.

HIGH-QUALITY, HIGH-BANDWIDTH ON-DEMAND VIDEO

Amy Philipson, Letcher Ross, Research TV, University of Washington

ResearchTV is a collaborative partnership of research universities and corporate research divisions dedicated to broadening the access to, and appreciation of, our individual and collective activities, ideas, and opportunities in basic and applied research.

One of the major goals of ResearchTV is to use content, content creation, and manipulation processes as a workbench to test materials for our future analog and digital broadcast and on-demand multimedia offerings, thus providing an unusual opportunity to experiment with new methods of distribution and interaction on a global basis.

ResearchTV is now using Internet2 to experiment with enabling real-time sharing of, and direct individual, institutional, and even “head-end” access in various demand and broadcast modes to the growing and continuously created advanced video and broadcast-based resources of the nation’s leading research universities.

Our experiments seek to exercise, test, and refine demand-real-time, broadcast-real-time, and batch distribution of the full range of quality, speed and frame sizes of high-quality source materials. This will provide our participants with an invaluable opportunity for CoS/QoS testing and comparison with regard to very demanding objects for which there are well-understood standards of quality.

The results of ResearchTV experiments contribute to the evolution of improved user interfaces for accessing, invoking, and using full motion video-based objects across a network with the full range of performance and CoS/QoS options and to desktop servers and devices with a broad range of capabilities and performance characteristics.

In addition our collaboration stimulates and enables other research institutions around the country to participate in the creation and sharing of an even larger array of state of the art content and common tools for storage, access, distribution and use.

In February, using the content and technological expertise of ResearchTV members, ResearchTV demonstrated full-frame, full-scale, broadcast quality MPEG-2 video with CD quality audio on-demand over Abilene from a video server at the University of Washington in Seattle delivered to Union Station in Washington, DC. Simultaneously a server at Union Station served video on-demand to clients in Seattle. In both cases the data streams were 5.6 Mbps MPEG-2 streams supplied by Microsoft Netshow Theater servers.

We are now moving to the next stage: building a substantial archive of original ResearchTV materials available on-demand at high quality and multiple bandwidth to multiple users.

Related URL(s): For a detailed description of ResearchTV goals see: <http://www.washington.edu/researchtv/whatis/background/background.html>

VIRTUAL ROOM VIDEOCONFERENCING SYSTEM

Philippe Galvez, California Institute of Technology

The “Virtual Room Video System” (VRVS) has been developed since 1995, in order to provide a low-cost, bandwidth-efficient, extensible means for videoconferencing and remote collaboration over networks within the High Energy and Nuclear Physics communities (HENP).

The VRVS system is based on a Virtual Videoconference Room concept. A series of IP servers/reflectors (unicast and/or multicast) connects users to a virtual room by setting up a series of interconnected IP tunnels, so that they form a private video group.

Since it went into production service in early 1997, deployment of the Web-based system has expanded to include 1330 registered hosts running the VRVS software from more than 37 different countries, and 21 “reflectors” that manage the traffic flow at HENP labs and universities in the US, South America, Europe and Asia.

So far there are four Virtual Rooms for worldwide conferences (involving more than one continent), and four Virtual Rooms each for intra-continental conferences in America, Europe and Asia.

The use of Web technology allows any authorized user, from any location, to access a wide range of services for packet-based videoconferencing. The tools provided in the LBNL and UCL videoconferencing applications suite (*vic*, *vat/rat*, *wb*) are currently used by the system, but since the system is IETF protocols compliant, the next generation of high-end video applications will be easy to integrate. The developed Web-based user interface provides a schedule manager, a Directory Name Service with a point-and-click option to initiate a point-to-point videoconference, a loop-back facility, an administrator’s interface (monitoring, statistics, etc.), a record/playback facility, a full documentation set (including a tutorial), a full application repository and installation instructions as well as other features.

Development and use of this system for international meetings has relied on the use of a minor part of the bandwidth on the Transatlantic link that is managed by the Caltech group and the CERN External Networking group. Some QoS tests using different proprietary techniques available on Cisco routers and ATM switches are continuously performed in order to prioritize real-time traffic.

Future plans for the system include support and deployment of VRVS on the next generation of regional and national backbones (Internet2, ESnet), integration of several new modules, shared workspaces (VRML and Java-based collaborative applications as they become available), integration of high-quality video and audio (MPEG1, MPEG2), H.323 I.T.U Videoconferencing standard, security and confidentiality. In parallel, several QoS tests and monitoring will be performed at a variety of bandwidths, in association with the system deployments foreseen for both DoE/NGI and Internet2-related (I2-DV) R&D projects

As an outgrowth of this work, we recently began collaboration with scientists and engineers in geology, biology, civil engineering, architecture and other fields who wish to use the VRVS technology.

VISIBLE HUMAN ANATOMICAL COLLABORATORY

*H. Tatsumi, Sapporo Medical University,
Michael Ackerman, George Thoma, Michael Gill, National Library of Medicine*

Scientific objectives of application; implementation as an end-to-end system over NGI networks:

The proposed experiment will attempt to prove a model which enables interactive biomedical image segmentation, labeling, classification, and indexing to take place using large images. The application can show different sections of a human body, and enables a researcher to make an interactive segmentation in order to recognize each anatomical object. Also, it calculates and fills areas in the segment with metaballs, and renders them. This would be followed by the attachment of anatomical terms to the objects working with the National Library of Medicine's (NLM) Unified Medical Language System and creating a multilingual object database. Visible Human (VH) data would be transferred to and from the researcher.

The VH dataset is an information rich dataset not existing in private sector datasets because commercial subsets of the VH dataset are often compressed by lossy techniques and hence information reduced. By maintaining a centralized repository, management of the resulting database will be more easily done. Updates would be in one place, ensuring authenticity and reliability.

Biomedical image libraries (in number and size) are sure to grow. Currently licensees of the VH dataset number 1000+ worldwide. Due to the size and international importance of the dataset, multilingual labeling of the dataset has been proposed. Therefore various researchers are needed to provide image segmentation and labeling. The first such researcher will work on a lower extremity subset of the Visible Human dataset. Other potential off-site collaborators exist in Europe. In the future online access to an anatomical segmented human anatomy atlas will be a vital resource for biomedical researchers worldwide. One model involves having NLM developed client software with browser access to the VH dataset selecting a cropped volumetric subset (e.g., the heart). The client software would receive the volume of interest and all labels. A client will do the

rendering with a (future) generic rendering tool possibly being supplied.

Status; current and planned functionality; time line for applications development:

The application is currently in a beta form and consists of the Viewer module which displays, sagittal and longitudinal, coronal sections of a human body. It runs under Apple Openstep environment on several platforms. A second Annotation Module is part of the architecture as well. An SGI-based version with enhanced features including viewing at any angle is planned. Timeline for the SGI version is for it to be in beta form in about six months.

Networking technologies used or planned; how they enhance or support the application; issues or difficulties with planned implementation:

NFS and FTP are currently used. NFS is used to provide the file system access, and FTP will be used to transfer image files that require annotation, segmentation indexing, etc., between NLM and a remote anatomical labeler. Slow disk copies to/from systems remotely are one current problem.

Requirements or desirements for technological enhancements; technologies being investigated; help sought from the technology community:

The most efficient methods for data transfer and file system access beyond FTP and NFS for over the wide-area. Other tools which would facilitate communication in an interactive collaborative environment.

VIRTUAL COLLABORATIVE CLINIC

*Marjory Johnson, Research Institute for Advanced Computer Science
Muriel Ross, MD, NASA Ames Research Center*

Scientific objectives of application; implementation as an end-to-end system over NGI networks:

The Virtual Collaborative Clinic (VCC) demonstrates advanced high-fidelity 3-D imaging and interactive virtual environment technologies across high-performance wide area networks. The scientific objective is to bring the clinic to the patient rather than the patient to the clinic. The ultimate objective is to provide medical service to astronauts on the Space Station and beyond.

The NASA Research and Education Network (NREN) project at ARC developed the wide area network infrastructure for an application demonstration in May 1999. The complex images resulted in massive data streams of up to 40 megabits per second (Mbps), approximately five to ten times the traffic leaving a major research institution on a normal day. Several high-performance networks connected participating end-sites with the application server at the Numerical Aerospace Simulation (NAS) facility at Ames. These networks included NREN, Abilene (the Internet2 research network), and the California Research and Education Network (CalREN2).

Status; current and planned functionality; time line for applications development:

The first VCC demonstration conducted by NASA on May 4 brought together doctors in five sites around the country to view, rotate and discuss 3-D stereo virtual-reality images. High-performance networks transmitted the high-fidelity images in real time, allowing the physicians to consult as if they were in the same room. NREN successfully demonstrated the application in a network environment where there was minimal contention for network resources from other traffic. We are currently able to do high-bandwidth multicast (between 25 and 30 Mbps) and plan to investigate approaches for achieving reliable multicast and experiment with QoS mechanisms. We have not yet begun to address delay and synchronization issues that will be present when we use satellites to reach remote sites. The work on reliable multicast and QoS has started. There is no specific timeline.

Networking technologies used or planned; how they enhance or support the application; issues or difficulties with planned implementation:

Distribution of the 3-D Virtual Collaborative Clinic images among widely dispersed sites requires high-performance networking. Major technological challenges include providing data transmission to multiple sites, minimizing latency, synchronizing the displays of large 3-D image data sets at the end sites, and accommodating satellite/terrestrial networks on disparate platforms. Both QoS and advanced multicast technologies are critical for successful prototyping of this application. Future research will focus on refining the use of these technologies to enhance application performance.

Requirements or desirements for technological enhancements; technologies being investigated; help sought from the technology community:

NREN is investigating QoS and reliable multicast as well as hybrid networks, because satellites must be used to reach remote sites and eventually the Space Station. We are seeking help from router vendors to implement QoS mechanisms in their routers, and from vendors with regard to reliable multicast. NREN is partnering with other testbeds to deploy the application.

DISTRIBUTED DATA INTENSIVE APPLICATIONS

Brian Tierney, Lawrence Berkeley National Laboratories

Scientific objectives of application; implementation as an end-to-end system over NGI networks:

The objective of the China Clipper Project is to develop technologies required for distributed data-intensive applications. The China Clipper environment will provide a collection of independent but architecturally consistent service components that will enhance the ability of a variety of applications and systems to construct and use a distributed, high-performance infrastructure. Such middleware supports high-speed access and integrated views of multiple data archives; resource discovery and automated brokering; comprehensive real-time monitoring and performance trend analysis of the networked subsystems, including the storage, computing, and middleware components. Clipper is not viewed so much as a “system” but rather as a coordinated collection of services that may be flexibly employed by a variety of applications (or other middleware) to build on-demand, large-scale, high-performance, wide-area, problem-solving environments.

Status; current and planned functionality; time line for applications development:

The first Clipper “proto-application” is STAF (Standard Analysis Framework), which is an HENP data analysis application. STAF has been modified to access data remotely over NGI networks from a Distributed-Parallel Storage System (DPSS). This application was carefully instrumented and tuned using the NetLogger Toolkit, which resulted in the ability to demonstrate throughput from a DPSS system to the STAF application of:

- LBNL-SLAC: 57 MB/s over the NTON Network
- ANL-LBNL: 35 MB/s over ESNet

We have also demonstrated the following:

- A prototype of advance reservation capabilities using Globus
- A global namespace over multiple HPSS and DPSS systems using SDSC’s SRB (Storage Resource Broker)
- Distributed cache support for object-oriented databases (OODB) using Objectivity via Object Oriented File System

(OOFS). OOFS is middleware among Objectivity and ASM (Advanced Storage Management) and HPSS. Objectivity provides CORBA and C++ access to HPSS resident objects.

The Clipper project is currently working on:

- Resource brokering for just-in-time construction of application environments. This includes implementing mechanisms to identify and acquire scheduling commitments for all of the resources needed to support an application.

The Clipper architecture is currently being extended to other application domains. These projects include:

- Babar (HEP data)
- Amanda (Astronomy data)
- DOE NGI Applications: (Climate, Combustion, Physics, Visualization)

Networking technologies used or planned; how they enhance or support the application; issues or difficulties with planned implementation:

We are experimenting with the following network technologies:

- IP differentiated services (DiffServ) and interfaces to DiffServ
- MPLS-based QoS
- Using CAR for marking packets based on a policy

Requirements or desirements for technological enhancements; technologies being investigated; help sought from the technology community:

We plan to experiment with and take advantage of the following network technologies, as they are implemented/ deployed:

- DiffServ /RSVP
- Bulk data QoS service
- Global naming services
- Security/access control services

THE NSF PACI PROGRAM

Stephen Elbert, National Science Foundation

PACI (Partnerships for Advanced Computational Infrastructure) is the follow-on to the successful NSF Supercomputer Centers program. PACI is the Centers' program . . . and more.

There are six elements to the mission:

1. provide, facilitate, and enhance access to high performance computational infrastructure for the U. S. academic, scientific, and engineering communities by partnering with universities, states, and the private sector
2. promote vigorous early use of experimental and emerging high performance computational and associated communications technologies that offer high potential for advancing science and engineering
3. enable the effective use of such infrastructure and technologies through education, training, consulting, and related support services, including appropriate software development, experimentation, and support
4. foster interdisciplinary research in science and engineering
5. facilitate the development of the intellectual capital required to maintain world leadership in computational science and engineering; and
6. broaden the base for the nation's advanced computational and communications infrastructure

PACI activities are structured along four components:

1. access to a diverse set of advanced and mid-range compute engines and data storage systems and experimental machine architectures
2. enabling technologies (ET), by developing both software tools for parallel computation and software to enable use of the partnership's widely distributed architecturally diverse machines and data sources to effectively use the partnership's very large distributed systems;

3. application technologies (AT), by engaging groups in high-end applications to develop and optimize their discipline specific codes and software infrastructures and to make these available to the program as a whole, as well as to researchers in other areas; and
4. education outreach and training (EOT), building growing awareness and understanding of how to use high performance computing and communications resources, and broadening the base of participation to help ensure the nation's continued world leadership in computational science and engineering. The EOT function is a unified, PACI-wide effort.

PACI consists of two partnerships: the National Computational Science Alliance (the Alliance) led by the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Champaign-Urbana and the National Partnership for Advanced Computational Infrastructure (NPACI) led by the San Diego Supercomputer Center at the University of California at San Diego. The Alliance is composed of 60 partner institutions and NPACI is composed of 46 US institutions and four international affiliates. Together they support about 850 projects in 280 universities.

Alliance ET and AT teams are working to define, implement, and exercise a distributed metacomputing infrastructure (GRID). Themes being pursued are large-scale distributed parallel computing, collaborative virtual environments, real time data acquisition and control of remote instruments, and discipline-specific computational workbenches for platform-independent (Web-based) seamless integration of informatics, analysis, and simulation. Application areas include chemical engineering, cosmology, environmental hydrology, molecular biology, nanomaterials, and scientific instrumentation.

NPACI is pioneering software development and integration through its thrust areas. The ET thrusts include metasystems, programming tools & environments, data-

THE NSF PACI PROGRAM cont.

intensive computing, interaction environments. The AT thrusts are molecular science, neuroscience, earth systems science, and engineering. Each thrust is a collection of projects, 63 in all. To further stimulate cross-thrust interaction, five alpha projects have been established: bioinformatics & infrastructure, protein folding, telescience, multi-component models, and scalable visualization.

Notable Alliance resources consist of 1536 SGI Origin 2000 processors and a 256-processor NT cluster at NCSA, and a 128-processor Linux PC cluster at UNM. Notable NPACI resources will soon include an 1184 Power3 processor IBM SP, making it the first Teraop resource available to the US academic community. NPACI also supports a 272-processor T3E-600, a 144-processor IBM SP (P2SC), a 14-processor Cray T90, and a 256-processor HP X2000 (at Caltech). Additional PACI resources are in Michigan, Wisconsin, Texas, Kentucky, Massachusetts, Virginia, Hawaii, and Berkeley.

DIRECTIONS FOR NPACI NETWORKING

Anke Kamrath, Mark Ellisman, Marty Hadida Hassan, Reagan Moore; University of California at San Diego

Overview

The National Partnership for Advanced Computational Infrastructure (NPACI) is pleased to participate in the Bridging the Gap workshop. NPACI's mission is to advance science by creating a ubiquitous, continuous and pervasive national computational infrastructure—the Grid—supporting the computational and related needs of the scientific community. This infrastructure will place significant demands on future networking resources to support distributed science by multidisciplinary teams. A number of strategic NPACI thrust areas have significant networking requirements, including terascale computing, grid-based computing (also called metacomputing), data-intensive computing, large-scale visualization, remote instrumentation, and applications in Education, Outreach and Training (EOT). Each will require access to and the manipulation of massive volumes of scientific and other data stored at distant locations.

Robust high-performance networking is a critical component of the PACI infrastructure. The computational science community and PACI partners require access to robust, secure, high-speed networks; quality-of-service (QoS) support, and improved performance-analysis tools.

The availability of these networks and features is altering the way researchers interact and the way science is conducted. A few key NPACI infrastructure activities and applications with specific WAN requirements are highlighted below.

In summary, many components of NPACI's overall goals are pushing the envelope of today's WAN infrastructure. To achieve NPACI's goal in advancing scientific discovery, effective and efficient utilization of distributed computing is critical. A few key areas that will help facilitate this are the following:

- *End-to-end performance tools (and necessary instrumentation)*
- *Support for QoS (across multiple ISPs outside the R&E network fabric)*

- *Support for data management across distributed network data caches*
- *Efficient implementation of data encryption support across the WAN*

An important question for the network community: Will it restrict itself to support for communication infrastructure, or will it also provide support for additional global infrastructure components such as network caches, inter-realm authentication, and monitoring?

We are looking forward to collaborations with the broader networking community to test, demonstrate, and implement novel networking technologies and applications.

CASE STUDY #1: TERASCALE COMPUTING AND GLOBAL MASS STORAGE

NPACI's high-end computing resources place numerous demands on the general WAN infrastructure. They are distributed across four key partner sites and a collection of additional data cache sites. To fully leverage the partnership and provide integrated access to resources, NPACI is attempting to implement a WAN file system and global mass store across its resource partners. Additionally, it is investigating general strategies to integrate discipline-specific data caches into this infrastructure. In particular, this storage system will provide the following:

- *Offsite backups of critical data and metadata*
- *Automatic offsite (second) copies of stored data*
- *A common interface to distributed mass storage systems (HPSS, ADSM, DMF)*
- *Wide-area file systems (e.g., DFS, AFS)*
- *WAN integration with Storage-Area-Networks (SANs)*

NPACI expects to move large amounts of data between its resources. In particular, the terascale system planned for implementation at the San Diego Supercomputer Center this

DIRECTIONS FOR NPACI NETWORKING cont.

year will increase—significantly—the amount of data generated. Conservatively speaking, we assume that this system will average 150 billion floating-point calculations per second, each computation requiring 2 words (16 bytes) of input and generating 1 word (8 bytes) of data output per second. This level of output easily could result in 3.6 TB of data into and out of memory per second, which easily could require 2.4 GB of data into and out of disk per second—just to feed the processors. The network then will have to feed the disk with input and transport the output to tertiary storage! Assuming 10% of this data goes to tertiary storage via the WAN, that is over 100% of an OC48. This demonstrates the need for two OC-48 interfaces into such a system and ignores other requirements, such as real-time visualization and on-line instrument data generation via microscopes and telescopes. Furthermore, requirements such as QoS and multicast capability must be considered. It is clear to us that the network (bandwidth, latency, and services such as QoS and multicast) is a critical component of the terascale computer resources becoming available to the research community.

To effectively implement this system requires enhanced end-to-end networking technologies:

- *Network based caches—The transaction rates and low latencies required by distributed supercomputing will realistically be met through the use of distributed caches, with data moved from remote instruments or archives into the local cache before processing. The control of data distributed across network caches will be require global I/O management tools.*
- *Data Encryption—Current data encryption methods are limited in their use due to performance considerations and export controls. To address the performance considerations, we recommend that the networking community investigate more tightly coupling of encryption methods with the networking layer.*
- *Performance analysis tools—End-to-end performance analysis tools are needed to tune and manage a distributed system effectively. In particular, this may require*

additional instrumentation at the network level (and related peripherals) to provide the needed information.

Additionally, we recommend that network technology developers work closely with the SAN standards/protocols and WAN file system standards communities to ensure adequate integration with new technologies and standards. Additionally, we encourage the network community to participate more actively beyond just the WAN requirements to ensure effective and efficient end-to-end utilization of the network.

TERASCALE COMPUTING

Scientific objectives of application; implementation as an end-to-end system over NGI networks:

- *Development of a production global mass store system and wide-area file system that can support the data requirements of terascale scientific computing*
- *Building scientific applications that effectively utilize advanced computation across the WAN*

What are the status, current and planned functionality, and timeline for applications development?

We have already started this project but are still at the evaluation stage given the infancy of the various technologies involved.

- *DFS/DCE w/ HPSS across the wide-area (test system underway)*
- *HPSS metadata being backed up across the WAN to another HPSS system (underway)*
- *Migrate metadata backups to CalREN (OC12) in late-FY99*
- *Automate second copy backups of data between HPSS systems via the WAN (FY01 and beyond)*
- *Investigating Storage-Area-Networks across the WAN (FY01 and beyond)*

Networking technologies used or planned; how they

DIRECTIONS FOR NPACI NETWORKING cont.

enhance or support the application; issues or difficulties with planned implementation:

- *Data Encryption (concerns about performance degradation when data encryption is enabled)*
- *Improved performance and analysis tools (and possibly more “instrumentation” of the system to analyze end-to-end performance)*
- *Security (Authentication)*

Requirements or desires for technological enhancements; technologies being investigated; help sought from the technology community:

- *Lighter weight WAN file systems (e.g., DFS too heavy weight).*
- *Storage-Area-Networks that operate in the wide area.*
- *How will WAN filesystems and SANs be integrated with new technologies and standards?*

CASE STUDY #2: DATA GRIDS

NPACI is supporting the development of discipline-specific scientific data collections. For a given discipline, multiple data collections are federated into a data grid, promoting the interchange of information between researchers with common interests. Data grids are inherently distributed applications that tie together data and compute resources. Researchers rely on the grid to support many aspects of information management and data manipulation. An end-to-end system provides support for the following:

- *Information discovery— the ability to query across multiple information repositories to identify data sets of interest.*
- *Data handling— the ability to read data from a remote site for use within an application.*
- *Remote processing— the ability to filter or subset a data set before transmission over the network.*
- *Publication— the ability to add data sets to collections for use by other researchers.*

- *Analysis— the ability to use data in scientific simulations, for data mining, or to create new data collections.*

Data grids are implemented through the integration of data resources (archives, databases, file systems) by data-handling systems. The major components are the data model management software for supporting access to a data set that is retrieved via a data-handling system from a remote storage system. Identification of the data set is done through an information discovery system, and the data set is processed before transmission to the application through execution of remote procedures.

Infrastructure to support data grids has been developed by the NPACI partnership. The central components are the SDSC Storage Resource Broker (SRB) for distributed data handling, the Grid Security Infrastructure (GSI) for inter-realm authentication, the SDSC Metadata Catalog (MCAT) for information discovery, and the Globus remote execution environment for procedure execution, accounting, and scheduling. An emerging protocol to support information discovery is the Stanford Simple Digital Library Interoperability Protocol (SDLIP). One planned development effort is to integrate the MCAT information Catalog with the SDLIP protocol.

The data grid infrastructure requires middleware services for the following:

- *Authentication*
- *Distributed information discovery*
- *Distributed data handling*
- *Remote procedure execution*
- *Accounting*
- *Scheduling*

One of the major questions for the network community is whether middleware will be supported as part of global network functionality. There are many examples of the need for such support. Network data caches are provided by Internet2 and NLANR, inter-realm authentication systems tie together independent administration domains, and dynamic

DIRECTIONS FOR NPACI NETWORKING cont.

resource monitoring is done by systems such as the Network Weather Service, which maintains global information about the availability of grid resources and network headroom. Each of these services represents a global environment that encompasses resources outside the local administration domain. Thus, they are strong candidates for integration with the global network resource. The central question is whether the network community will continue to restrict itself to support for communication infrastructure or be subsumed by the group that assumes responsibility for all global services.

The major network difficulty experienced by data grids is the limited ability to achieve high performance. The sustainable bandwidth is typically a small fraction of the rated bandwidth of the network. The factors that contribute to performance degradation include poorly tuned TCP/IP transmission parameters (window size, message size), system buffer sizes that are too small, router congestion, lost packets, etc. To guarantee delivery and the ability to sustain high-bandwidth performance, data grids are being implemented with data caches located at each data resource point. Data can be moved into the cache overnight, with high performance access then provided by the local area network. The data sets that are retrieved from remote collections are cached in either high-performance network-based caches (such as the Distributed Parallel Storage System) or local attached disk cache through either the GASS or Active Data Repository. The net effect is that the applications are able to sustain the access rates and low latency needed for data-intensive computing. Note that the amount of data that is moved can still be substantial—on the order of hundreds of gigabytes to terabytes.

DATA GRIDS

Scientific objectives of application; implementation as an end-to-end system over NGI networks:

- *Development of a production global mass store system and wide-area file system that can support the data requirements of terascale scientific computing*

- *Building scientific applications that effectively utilize advanced computation across the WAN*

Status; current and planned functionality; time line for applications development:

The central components of the NPACI Data Grid are the SDSC Storage Resource Broker (SRB) for distributed data handling, the Grid Security Infrastructure (GSI) for inter-realm authentication, the SDSC Metadata Catalog (MCAT) for information discovery, and the Globus remote execution environment for procedure execution, accounting, and scheduling. The SRB and GSI have been integrated and will be released in version 1.1.5 of the SDSC SRB. The GSI release is planned for July 1999. The integration of the Globus Access to Secondary Storage (GASS) system with the SRB is planned for third quarter of 1999.

Federation of collections using the Data Grid infrastructure is planned for FY2000. This is proceeding through the installation of the SDSC SRB at each of the data collection sites. Federation of the collections will require the ability to support distributed information discovery across multiple information resources. An emerging protocol to support information discovery is the Stanford Simple Digital Library Interoperability Protocol (SDLIP). One of the planned development efforts is to integrate the MCAT information Catalog with the SDLIP protocol.

Networking technologies used or planned; how they enhance or support the application; issues or difficulties with planned implementation:

The major network difficulty experienced by data grids is the limited ability to achieve high performance. The sustainable bandwidth is typically a small fraction of the rated bandwidth of the network. The factors that contribute to the degradation in performance include poorly tuned TCP/IP transmission parameters (window size, message size), system buffer sizes that are too small, router congestion, lost packets, etc. To guarantee delivery and the ability to sustain

DIRECTIONS FOR NPACI NETWORKING cont.

high-bandwidth performance, data grids are being implemented with data caches located at each data resource point.

Requirements or desirements for technological enhancements; technologies being investigated; help sought from the technology community:

- *Lighter weight WAN file systems (e.g., DFS too heavy weight).*
- *Storage-Area-Networks that operate in the wide area.*
- *How will WAN filesystems and SANs be integrated with new technologies and standards?*

CASE STUDY #3: TELESCIENCE FOR ADVANCED TOMOGRAPHY APPLICATIONS

This project is developing end-to-end solutions to enable tomographic data collection and analysis of biological specimens. This system will integrate use of remote imaging instrumentation, distributed heterogeneous parallel computing, federated and distributed databases and image archives, and component-based remote visualization tools. As a result, it has unique and demanding networking requirements. This NPACI project is led by Mark Ellisman of the National Center for Microscopy and Imaging Research (NCMIR) at UCSD.

This high-performance distributed application has significant networking requirements, including PVCs (Private Virtual Circuits), QoS guarantees, and reliable multicast that are not available to the general Internet community. This project will build on experience gained through NCMIR's Collaboratory for Microscopic Digital Anatomy (CMDA) project founded in 1992 and more recent collaborations with Osaka University on trans-Pacific telemicroscopy. For example, remote operation of instruments (or Remote Instrumentation [RI]) across a WAN has specialized requirements. RI network traffic can be classified into at

least two categories: time-critical control information and video/image streams. Control information consists of short messages that travel from the remote user's GUI to the instrument site. Fast delivery of these messages results in more interactive, fine-grain control of the remote instrument. Currently, there is no way to prioritize this traffic and, as a result, latency of instrument control information is unnecessarily high due to contention with the less important—but more bandwidth-intensive—video/image traffic. Prioritized traffic and QoS guarantees would provide a means to solve this problem by granting higher priority (and better service) to time-critical network flows. Such networking technologies have the potential to transform these types of telescience projects and centralized research facilities such as NCMIR into effective international resources.

In addition to being moved to remote users, data is also moved to compute resources to perform large-scale parallel tomographic reconstruction and 3-D data visualization tasks. To manage the interactions between distributed compute resources and the instrument, grid-based computing tools (e.g., Globus, AppLes, NWS) are being implemented. Data sets for computation are 10s of GB and will grow by well over an order of magnitude in the near term. With high-performance computing and high-speed networking in place, these compute-intensive tasks can be accomplished quickly enough that the researcher can view 3-D representations of their specimens while using the microscope. An important goal of this project is to provide this level of feedback to improve and refine data collection.

TELESCIENCE FOR ADVANCED TOMOGRAPHY APPLICATIONS

Scientific objectives of application; implementation as an end-to-end system over NGI networks:

DIRECTIONS FOR NPACI NETWORKING cont.

The Telescience for Advanced Tomographic Applications project will develop an end-to-end, Web-based solution to enable telemicroscopy on biological specimens. This system will integrate use of remote imaging instrumentation, distributed heterogeneous parallel computing, federated and distributed databases and image archives, and component-based remote visualization tools.

In this project, data-acquisition instruments will be linked to distributed parallel supercomputers for data refinement. The data sets will be inserted into a database structure within a framework being developed for a multiple-scale brain-mapping project. The environment will enable users to locate, render, display, and analyze surfaces in augmented reality visualization environments.

The project's most ambitious goal is to enable intelligent, remote steering of imaging instrumentation. In this scenario, quantitative information from graphical models of refined data (based on results from remote computations and comparisons with related entries in the database) will be fed back to a data-acquisition device to acquire data more accurately from a specimen.

Status; current and planned functionality; time line for applications development:

This project has just been created. We are evaluating existing technologies to determine the best way to design our system. This project will build on experiences gained through NCMIR's Collaboratory for Microscopic Digital Anatomy (CMDA), an NSF National Challenge Project started in 1994 following the pioneering development of telemicroscopy in 1992. It will also build on recent collaborations with Osaka University on trans-Pacific telemicroscopy, the GLOBUS and AppLes metaseystems, and data-handling and visualization techniques developed at Universities of Texas and Tennessee. We plan to showcase our first deliverable—automated tomographic data acquisition and reconstruction—at the SC99 conference this November. Database and visualization efforts will follow.

Networking technologies used or planned; how they enhance or support the application; issues or difficulties with planned implementation:

We strive to leverage high-speed research networks such as the vBNS, Abilene, STARTAP, and TransPAC. Fine-tuning of TCP/IP with selective acknowledgement is underway as a better mechanism for video streaming, i.e., to reduce latency. QoS guarantees will become paramount to achieve the feedback loop mentioned above, which requires efficient delivery of large data sets across the network.

Requirements or desirements for technological enhancements; technologies being investigated; help sought from the technology community:

During the near term, we will evaluate different video codecs and video-streaming technologies. Our current implementation lacks intra-frame compression, which is deemed an effective way to improve video quality and frame rates while requiring less throughput from the network. Technologies such as MPEG2 and multi-resolution video will provide this level of support. Also, we will use network monitoring technologies as a means to better schedule processing tasks across the Globus metaseystem.

Through our affiliation with NPACI and NSF, we are receiving guidance from NLANR with regards to networking technologies. They have helped us measure and analyze network traffic during our trans-Pacific telemicroscopy experiments. Selected NLANR individuals will assist in improving our networking protocols and leveraging new networking standards.

TERRAVISION

Yvan Leclerc, SRI International

TerraVision is a real-time terrain visualization application developed by SRI International over the past six years as part of the DARPA-sponsored MAGIC project. The initial requirements were to design and implement a visualization application that uses remotely stored terrain/image data accessed over a high-speed network at rates approaching 1 Gbps. Although TerraVision typically achieves rates closer to 100 Mbps, it does enable high-quality visualization of very large data sets (tens of Gigabytes) over very large networks (cross-country).

The ultimate objective of TerraVision is to allow end-users to virtually visit and navigate through high-resolution 3-D representations of any part of the world. This objective is implemented as a client that requests 3-D data from data sets stored on remote servers in real time as the user navigates around the world. Currently, the user selects a limited number of data sets from a table of available data sets stored on the servers. The data sets can be tens of gigabytes in size or larger and the application can be used on networks with throughputs in the range of 0.5 to 100 Mbps.

Related URL(s):

www.ai.sri.com/TerraVision

Application Partner(s) The MAGIC consortium:

DARPA, Sprint, LBL, U. Kansas, USGS

DISTRIBUTED IMAGE SPREADSHEET (DISS)

*K. Palaniappan, University of Missouri - Columbia
F. Hasler, Goddard Space Flight Center*

The Distributed Image SpreadSheet (DISS) is a collaborative scientific visualization and analysis tool that uses high-performance networks to enable scientists to organize and intercompare gigabyte-sized geophysical datasets collected by the next generation satellite systems. Multi-spectral land and ocean satellite products from distributed geophysical archives will be visualized using the DISS. Timevarying volumetric data from numerical weather models, earth system data assimilation and magnetohydrodynamic space weather models will be intercompared using multicell displays. High-speed network access using http and ftp methods, combined with novel image and geometry data compression schemes for efficient bandwidth utilization will be demonstrated.

Related URL(s):

<http://meru.cecs.missouri.edu>, <http://rsd.gsfc.nasa.gov>

Application Partner(s):

NASA Goddard Space Flight Center, NOAA

NGI/I2 ADVANCED DIGITAL VIDEO DEMONSTRATION

Joe Mambretti, iCAIR

The iCAIR demonstration in advanced digital video provides a view into the future of visual communication. Although people acquire over 80 percent of their information through the visual sense, the current Internet does not adequately support presentations of visual information. To provide support for high-quality digital video, simulations, animations, virtual reality movies, high-definition images, etc., the next-generation Internet must implement an integrated suite of new technologies. This demonstration shows an experimental scalable integrated system for managing and distributing media from remote locations—a digital video portal that integrates a number of key technologies. These technologies include:

- *an access mechanism with multiple channels for distributed media assets (regional, national or international) based on a digital video jukebox (for checking assets in and out of the system),*
- *a digital library,*
- *a video server,*
- *two integrated automatic video metadata indexing functions, which detect major content changes and translate audio tracks to text that is indexed to mapped to the video, and*
- *Quality of Service provided through sufficient provisioning.*

To ensure scalability, especially for multiple streams, the system is being developed on a massively parallel supercomputer. The demonstration utilizes NGI/I2 infrastructure, which soon will also incorporate Quality-of-Service guarantees for media.

Related URL(s):

<http://www.icaair.org>

VIRTUAL ROOM VIDEOCONFERENCING SYSTEM

Philippe Galvez, CalTech

The Virtual Room Videoconferencing System (VRVS) system is based on a Virtual Videoconference Room concept. A series of IP servers/reflectors (unicast and/or multicast) connects users to a virtual room by setting up a series of interconnected IP tunnels, so that they form a private video-group. The VRVS provides a low-cost, bandwidth-efficient, extensible means for videoconferencing and remote collaboration.

The use of Web technology allows any authorized user, from any location, to access a wide range of services for packet-based videoconferencing. The developed Web-based user interface provides a schedule manager, a Directory Name Service with a point-and-click option to initiate a point-to-point videoconference, a loop-back facility, an administrators' interface (monitoring, statistics, etc.), a record/playback facility, a full documentation set (including a tutorial, a full application repository and installation instructions).

Related URL(s):

<http://vrvs.cern.ch>

Application partner(s):

Caltech, CERN

VISIBLE HUMAN ANATOMICAL VIEWER

*Michael Gill, National Library of Medicine
Haruyuki Tatsumi, Sapporo Medical University*

The proposed experiment will attempt to prove a model that enables interactive biomedical image segmentation, labeling, classification, and indexing to take place using large images. Its primary focus is a Biomedical Image Collaboratory between Dr. Haruyuki Tatsumi of Sapporo Medical University (SMU) and NLM. This relies on a Visible Human Viewer program developed on OpenStep (an Apple Applications Programming Interface). This application can show sagittal and longitudinal, coronal sections of a human body, and enables a researcher to make an interactive segmentation in order to recognize each anatomical object. Also, it calculates and fills areas in the segment with metaballs, and renders them. This would be followed by the attachment of anatomical terms to the objects working with NLM's Unified Medical Language System (UMLS) and creating a multilingual object database. Visible Human data would be transferred to and from the researcher via FTP or via NFS (Network File System) with other methods to be determined.

Related URL(s):

<http://archive.nlm.nih.gov/proj/bitatrans-pacific.html>

Application partner(s):

National Library of Medicine and Sapporo Medical University, Sapporo, Japan

VIRTUAL COLLABORATIVE CLINIC

Matt Chew Spence, NASA/NREN

The Virtual Collaborative Clinic (VCC) application developed by the NASA Ames Research Center (ARC) Center for Bioinformatics under Dr. Muriel Ross combines highly sophisticated medical imaging with high-performance, high-speed networking. Doctors can receive and rotate 3-D high-resolution 24-bit color stereo medical images, collaborating in near real-time with remote colleagues for consultation, diagnosis and treatment planning. Using a “CyberScalpel,” doctors can also “cut” into images and move “bone” around for surgical simulation. The 3-D images are constructed from serial sectional images of tissues and organs obtained from various types of medical imaging data such as electron microscopy, CT, MRI scans, and others.

In May 1999 NREN successfully demonstrated VCC running using a 32-Mbps multicast stream between NASA Ames, NASA John Glenn Research Center, Stanford Medical Clinic, UC Santa Cruz, and the Northern Navajo Medical Facility in Shiprock, NM. Using multicast to efficiently distribute the image is only part of the networking challenge in the evolution of VCC from proof of concept to a production application. Another major challenge is how to ensure the functionality of time-sensitive VCC data streams over potentially congested networks. Network Quality of Service (QoS) mechanisms provides a possible solution. Due of the bursty nature of the VCC stream, a priority queuing mechanism is preferable to bandwidth reservation. Currently NREN is examining the effect of different queuing mechanisms on VCC application performance as the first step towards implementing QoS.

Related URL(s):

<http://www.nren.nasa.gov>

<http://biocomp.arc.nasa.gov>

NPACI TELEMICROSCOPY

Martin Hadida-Hassan, UCSD

This demonstration will exemplify the high-performance networking requirements of NPACI by showcasing NCMIR's Telemicroscopy software. Our Telemicroscopy tools provide for interactive, remote control of a powerful electron microscope for the purposes of data acquisition and analysis. One component of the system is a Digital Video stream of the specimen under observation that is delivered to the remote user to better enable steering and focusing of the microscope. Adequate measures of latency and frame rates achievable with high-speed networking infrastructure enable remote users to investigate their specimen and collect high-resolution 3-D from the comfort of their own laboratory. Such use of network-based computing increases access to a scarce and valuable resource, the electron microscope; stimulates new research interactions; and increases the feasibility of long-term research projects that involve multiple sessions on the microscope.

Related URL(s):

<http://www-ncmir.ucsd.edu/CMDA>

<http://www.npaci.edu/Alpha/Telescience/>

Application Partner(s):

NCMIR, NPACI/SDSC

AC	Admission Control	CORBA	Common Object Request Broker Architecture
ACPI	Accelerated Climate Prediction Initiative (DOE)	COS	Class of Service
ACTS	Advanced Communications Technology Satellite	CPU	Central Processing Unit
ADSM	(distributed mass storage system)	CVE	Collaborative Virtual Environments
AF	Assured Forwarding	DAAC	Distributed Active Archive Center
AFS	(wide-area file system)	DARPA	Defense Advanced Research Projects Agency
ALTQ	Alternate Queuing	DCE	Distributed Computing Environment
ANL	Argonne National Laboratory	DFS	(wide-area file system)
API	Applications Programming Interface	DiffServ	Differentiated Services
ARC	Ames Research Center	DISS	Distributed Image SpreadSheet
ARM	Application Response Measurement	DL	Digital Libraries
AS	Autonomous System	DMF	(distributed mass storage system)
AT	Application Technologies	DNS	Domain Name Server
ATM	Asynchronous Transfer Mode	DOE	Department of Energy
BB	Bandwidth Broker	DPSS	Distributed Parallel Storage System
BGMP	Border Gateway Multicast Protocol	DRA	Distributed Routing Algorithm
BGP	Border Gateway Protocol	DREN	Defense Research and Engineering Network
BNL	Brookhaven National Laboratory	DS3	Digital Signal 3 (44.7 Mbps)
CA	Certification Authority	DSCP	DiffServ Code Point
CAC	Call Admission Control	DV	Digital Video
CalREN2	California Research and Education Network	DVTS	Digital Video Transport System
CAR	Committed Access Rate	Ebnet	EOSDIS Backbone Network
CAT	Common Authentication Technology	ECN	Explicit Congestion Notification
CBQ	Class-Based Queuing	EDOS	EOS Data and Operations System
CBR	Constant Bit Rate	EF	Expedited Forwarding
CERN	European Laboratory for Particle Physics	EOC	EOSDIS Operations Center
CIR	Committed Information Rate	EOS	Earth Observing System
CL	Controlled Load service	EOSDIS	Earth Observing System Data and Information System
CMDA	Collaboratory for Microscopic Digital Anatomy	EOT	Education Outreach and Training
Codecs	coder/decoders		

ERDoS	End-to-End Resource Management for Distributed Systems	IntServ	Integrated Services
ESDIS	Earth Science Data and Information System	IP	Internet Protocol
ESG	Earth System Grid	IP(v4/v6)	Internet Protocol (versions 4 and 6)
ESNet	Energy Sciences Network	IPG	Information Power Grid
ET	Enabling Technologies	IPsec	Secure IP
FEC	Forward Error Correction	IPv6	Internet Protocol version 6
FTP	File Transfer Protocol	IRTF	Internet Research Task Force
G	Guaranteed Service	ISAKMP/Oakley	Internet Security Association and Key Management Protocol/Oakley
GAA_API	Generic Authorization and Access Control API	ISP	Internet Service Provider
GASS	Globus Access to Secondary Storage	IST	Instrument Support Terminals
Gbps	Gigabits per second	ITO	Information Technology Office (DARPA)
Gbps	Gigabits per second	JET	Joint Engineering Team
GRC	Glenn Research Center	JPEG	Joint Photographic Experts Group
GRID	A distributed metacomputing infrastructure	JSC	Johnson Space Center
GSFC	Goddard Space Flight Center	Kbps	Kilobits per second
GSI	Grid Security Infrastructure	KSC	Kennedy Space Center
GSSAPI	Generic Security Service API	LAN	Local Area Network
HDF	Hierarchical Data Format	LANL	Los Alamos National Laboratory
HDTV	High Definition Television	LBNL	Lawrence Berkeley National Lab
HECC	High-End Computing and Computation	LDAP	Lightweight Directory Access Protocol
HENP	High Energy and Nuclear Physics	LLNL	Lawrence Livermore National Laboratory
HMAC	(Data integrity algorithm)	LSN	Large Scale Networking
HPNAT	High Performance Network Applications Team	LSP	Label Switch Path
HPSS	(distributed mass storage system)	MAC	(Data integrity algorithm)
HTML	HyperText Markup Language	MAGIC	Multidimensional Applications and Gigabit Internetwork Consortium
I2	Internet2	MASC	Multicast Address Set Claim
iCAIR	International Center for Advanced Internet Research	MBGP	Multicast Border Gateway Protocol
IETF	Internet Engineering Task Force	MBone	Multicast Backbone
iGRID	International Grid	Mbps	Megabits per second

MCAT	Metadata Catalog	NSF	National Science Foundation
MD5	(Data integrity algorithm)	NSI	NASA Science Internet
MF	Multi Field	NTON	National Transparent Optical Network
MIB	Management Information Base	NTSC	National Television Standards Committee
MJPEG	Motion Joint Photographic Experts Group	OC-3	Optical Carrier 3 (155 megabits per second)
MPEG2	Moving Picture Experts Group Phase 2	OC-12	Optical Carrier 12 (622 megabits per second)
MPLS	Multi Protocol Label Switching	OC-48	Optical Carrier 48 (2.5 gigabits per second)
MREN	Metropolitan Research and Education Network	OMP	(Complete modeling tool)
MSDP	Multicast Source discovery Protocol	OODB	Object Oriented Data Bases
MSFC	Marshall Space Flight Center	OOFs	Object Oriented File System
NAP	Network Access Point	ORB	Object Request Broker
NAS	Numerical Aerospace Simulation	OS	Operating System
NASA	National Aeronautics and Space Administration	OSPF	Open Shortest Path First
NCAR	National Center for Atmospheric Research	PACI	Partnership for Advanced Computational Infrastructure
NCMIR	National Center for Microscopy and Imaging Research	PHB	Per Hop Behavior
NCSA	National Center for Supercomputing Applications	PI	Principal Investigator
NCSA	National Computational Science Alliance	PIM-SM	Protocol Independent Multicast – Sparse Mode
NFS	Network File System	PITAC	President's Information Technology Advisory Committee
NGI	Next Generation Internet	PKI	Public Key Infrastructure
NIH	National Institutes of Health	PMEL	Pacific Marine Environmental Laboratory (NOAA)
NISN	NASA's Integrated Services Network	PQ	Priority Queuing
NLANR	National Laboratory for Applied Networking Research	PVC	Permanent Virtual Circuit
NLM	National Library of Medicine	PVC	Private Virtual Circuit
NOAA	National Oceanic and Atmospheric Administration	QA	Quality Assurance
NP	Network Performance	QBone	Quality of Service Testbed
NPACI	National Partnership for Advanced Computational Infrastructure	QoS	Quality of Service
NREN	NASA Research and Education Network	QPS	QBone Premium Service
NRT	Networking Research Team	RA	Resource Allocation
		R&D	Research and Development

R&E	Research and Education	TLS	Transport Layer Security
R&E	Research and Engineering	TOS	Time of Service
RED	Random Early Detection	TransPAC	U.S./Asia Pacific Consortium
RIO	RED with In and Out	UBR	Unspecified Bit Rate
RMI	Route Method Invocation	UCLA	University of California at Los Angeles
RPC	Remote Procedure Call	UCSD	University of California at San Diego
RSVP	Resource Reservation Protocol	UDP	User Datagram Protocol
RT	Real Time	UIUC	University of Illinois, Urbana-Champaign
RTP	Real-time Transport Protocol	UMLS	Unified Medical Language System
SAN	Storage Area Network	URL	Uniform Resource Locator
SCF	System Control Facility	USC	University of Southern California
SDLIP	Stanford Simple Digital Library Interoperability Protocol	USGS	U.S. Geological Survey
SDSC	San Diego Supercomputer Center	vBNS	very High-Performance Backbone Network Service
SHA-1	(Data integrity algorithm)	VC	Virtual Circuit
SLAC	Stanford Linear Accelerator	VCC	Virtual Collaborative Clinic
SM	Simple Multicast	VH	Visible Human
SM	Sparse Mode	VPN	Virtual Private Network
SMU	Sapporo Medical University	VR	Virtual Reality
SNMP	Simple Network Management Protocol	VRML	Virtual Reality Modeling Language
SRB	Storage Resource Broker	VRVS	Virtual Room Videoconferencing System
SSH	Secure Shell	WAN	Wide Area Network
SSL	Secure Socket Layer	WFQ	Weighted Fair Queuing
STAF	Standard Analysis Framework	WRED	Weighted Random Early Detection
STARTAP	Science, Technology, and Research Transit Access Point	WRR	Weighted Round Robin
Tbps	Terabits per second	XML	Extensible Markup Language
TCP	Transport Control Protocol		
TCP/IP	Transport Control Protocol/Internet Protocol		
TE	Traffic Engineering		
TIFF	Tagged Image File Format		

Deb Agarwal	deba@george.lbl.gov	Ray Gilstrap	rgilstrap@mail.arc.nasa.gov
Ian Akyildiz	ian@ee.gatech.edu	Roch Guerin	guerin@ee.upenn.edu
Kevin Almeroth	almeroth@cs.ucsb.edu	Marty Hadida-Hassan	marty@sdsu.edu
Guy Almes	almes@internet2.edu	David Hajazin	david.hajazin@msfc.nasa.gov
Jules Aronson	aronson@nlm.nih.gov	Ted Hanss	ted@internet2.edu
Javad Boroumand	jborouma@nsf.gov	Sue Hares	skh@merit.edu
Heather Boyles	heather@internet2.edu	Bob Hinden	hinden@iprg.nokia.com
Jim Brandt	brandt@ca.sandia.gov	Scott Huddle	huddle@MCI.NET
Rich Carlson	carlson@er.doe.gov	John Jamison	jjamison@mci.net
Helen Chen	hycsw@ca.sandia.gov	Marjory Johnson	mjohnson@mail.arc.nasa.gov
Larry Chao	lchao@mail.arc.nasa.gov	William Johnston	wej@nas.nasa.gov
Rich Carlson	carlson@er.doe.gov	Kevin Jones	kjones@mail.arc.nasa.gov
Eli Dart	eddart@ca.sandia.gov	Youki Kadobayashi	youki@center.osaka-u.ac.jp
Don Denbo	dwd@pmel.noaa.gov	Anke Kamrath	akamrath@sdsu.edu
Barbara Denny	denny@3com.com	Pat Kaspar	pkaspar@mail.arc.nasa.gov
Rob Densock	robert.densock@nist.gov	Steve Kent	kent@BBN.COM
Dick desJardins	rdesjardins@mail.arc.nasa.gov	Carl Kesselman	carl@isi.edu
Mark Ellisman	mhellisman@ucsd.edu	Chan Kim	chan.kim@lerc.nasa.gov
Dino Farinacci	dinof@dinof.com	Hugh LaMaster	hmaster@mail.arc.nasa.gov
Bob Fink	RLFink@lbl.gov	Yvan Leclerc	leclerc@ai.sri.com
Mark Foster	mafoster@mail.arc.nasa.gov	Barry Leiner	bleiner@riacs.edu
Ken Freeman	kfreeman@mail.arc.nasa.gov	Bill Lennon	wjlennon@llnl.gov
Philippe Galvez	Philippe.Galvez@cern.ch	Paul Love	epl@internet2.edu
Pat Gary	pgary@ndadsb.gsfc.nasa.gov	Bob Lucas	rflucas@lbl.gov
Douglas Gerdin		Teresa Lunt	tlunt@parc.xerox.com
Mario Gerla	gerla@cs.ucla.edu	Joel Mambretti	JMambretti@mren.org
Andy Germain	andy.germain@gsfc.nasa.gov	Mark Matties	markm@sled.gsfc.nasa.gov
Mike Gill	gill@nlm.nih.gov	Janise McNair	mcnair@ece.gatech.edu

Grant Miller miller@ccic.gov
Ken Miller miller@starburstsoftware.com
Sally Miller smmiller@mail.arc.nasa.gov
Doug Montgomery dougm@nist.gov
Reagan Moore moore@spsc.edu
Craig Moyers cmoyers@mail.arc.nasa.gov
Klara Nahrstedt klara@cs.uiuc.edu
Vishy Narayan vnarayan@arc.nasa.gov
Cliff Neuman bcn@isi.edu
Michael Oehler mjo@tycho.ncsc.mil
Mike Olsen mike.olsen@msfc.nasa.gov
K. Palaniappan palani@cecs.missouri.edu
Radia Perlman Radia.Pperlman@East.Sun.COM
Sheila Piesko spiesko@mail.arc.nasa.gov
Amy Philipson amy@cac.washington.edu
Robert Poston reposton@us.ibm.com
Mike Rechtenbaugh ... rech@edcmail.cr.usgs.gov
Raj Reddy
Letcher Ross letcher@cac.washington.edu
Rick Schantz schantz@bbn.com
Mary Anne Scott scott@er.doe.gov
William Semancik wjseman@afterlife.ncsc.mil
George Seweryniak seweryni@er.doe.gov
Steve Shultz shultz@mail.arc.nasa.gov
Jeff Smith jsmith@rattler.gsfc.nasa.gov
Willard Smith smith@coe.tsuniv.edu
Matt ChuSpence matt@nren.nasa.gov
Gary Stone stone@cs.nps.navy.mil
Thom Stone tstone@mail.arc.nasa.gov
Nina Taft nina@sprintlabs.com
Peter Tam ptam@mail.arc.nasa.gov
Haruyuki Tatsumi tatsumi@sapmed.ac.jp
Ben Teitelbaum ben@internet2.edu
David Tennenhouse ... dtennenhouse@darpa.mil
Brian Tierney bltierney@lbl.gov
Don Towsley towsley@cs.umass.edu
Gene Tsudik gts@isi.edu
Bill Turnbull wturnbull@hpcc.noaa.gov
George Uhl uhl@mamba-e.gsfc.nasa.gov
Gary Veum veum@cobra.gsfc.nasa.gov
Bessie Whitaker bwhitaker@mail.arc.nasa.gov
Dean Williams williams13@llnl.gov
Bill Wing wrw@cosmail1.ctd.ornl.gov
Linda Winkler winkler@anl.gov
John Wroclawski jtw@lcs.mit.edu
Geoff Xie xie@cs.nps.navy.mil
Matt Zekauskas matt@advanced.org
Lixia Zhang lixia@cs.ucla.edu
Xinhua Zhuang

CO-LEADS OF THE ORGANIZING COMMITTEE

Richard desJardins
NASA/NREN
650.604.4764
rdesjardins@arc.nasa.gov

Doug Montgomery
NIST
301.975.3630
dougm@nist.gov

Marjory Johnson
NASA/RIACS
650.604.6922
mjohnson@arc.nasa.gov

William Johnston
NASA and DOE
650.604.4365
wej@nas.nasa.gov

CO-HOSTS OF THE WORKSHOP

Bessie Whitaker
NASA/NREN
650.604.6152
bwhitaker@arc.nasa.gov

Ken Freeman
NASA/NREN
650.604.1265
kfreeman@arc.nasa.gov

KEYNOTE SPEAKER

David Tennenhouse
DARPA
dtennenhouse@darpa.mil

PRESENTATIONS: TECHNOLOGIES*QoS*

John Wroclawski
Massachusetts Institute of Technology
jtw.lcs.mit.edu

Multicast

Don Towsley
University of Massachusetts
towsley@cs.umass.edu

Security

Steve Kent
BBN Technologies
kent@bbn.com

PRESENTATIONS: APPLICATION CASE STUDIES*Digital Earth*

Yvan Leclerc
SRI International
leclerc@ai.sri.com

Marla Meehl
NCAR
marla@ucar.edu

K. Palaniappan
University of Missouri-Columbia
palani@cecs.missouri.edu

Thom Stone
NASA/NREN
tstone@arc.nasa.gov

Don Denbo
NOAA/PMEL
dwd@pmel.noaa.gov

Dean Williams
LLNL
williams13@llnl.gov

PRESENTATIONS: APPLICATION CASE STUDIES CONT.*Digital Video*

Joe Mambretti
iCAIR/MREN
j-mambretti@nwu.edu

Amy Philipson
Research TV/Univ. of Wash.
amy@cac.washington.edu

Letcher Ross
Research TV/Univ. of Washington
letcher@cac.washington.edu

Philippe Galvez
CalTech
Philippe.Galvez@cern.ch

Telemedicine

Mike Gill
NIH
mike_gill@nlm.nih.gov

Haruyuki Tatsumi
Sapporo Medical University
tatsumi@sapmed.ac.jp

Marjory Johnson
NASA/RIACS
mjohnson@arc.nasa.gov

China Clipper

Bob Lucas
Lawrence Berkeley National Lab
rlucas@lbl.gov

Brian Tierney
Lawrence Berkeley National Lab
btierney@lbl.gov

NPACI

Steve Elbert
NSF
selbert@nsf.gov

Mark Ellisman
Univ. of California San Diego
mhellisman@ucsd.edu

Anke Kamrath
University of California San Diego/
San Diego Supercomputing Center
kamratha@sdsc.edu

Reagan Moore
Univ. of California San Diego/
San Diego Supercomputing
Center
moore@sdsc.edu

PRESENTATIONS: TESTBEDS AND NGI PROGRAM OBJECTIVES

Ian Foster
Globus/ANL
foster@mcs.anl.gov

Javad Boroumand
JET/NSF
jborouma@nsf.gov

Guy Almes
Internet2
almes@internet2.edu

Ted Hanss
Internet2
ted@internet2.edu

Raj Reddy
PITAC/Carnegie Mellon University
RR@cmu.edu

Additional NGI Agency Representatives

DARPA

Mari Maeda
mmaeda@darpa.gov

Bert Hui
bhui@darpa.gov

DOE

Rich Carlson
racarlson@anl.gov

George Seweryniak
sereryni@es.net

NCO

Grant Miller
miller@ccic.gov

Sally Howe
howe@ccic.gov

NIH

Jules Aronson
aronson@nlm.nih.gov

NIST

Robert Rosenthal
rrosenthal@nist.gov

NOAA

William Turnbull
wturnbull@hpc.noaa.gov

NSA

William Semancik
wjseman@afterlife.ncsc.mil

Grant Wagner
gmw@tycho.ncsc.mil

NSF

William Decker
wdecker@nsf.gov

Anne Richeson
aricheso@nsf.gov

Karen Sollins
ksollins@nsf.gov

Logistics

Sally Miller
NASA/NREN
650.604.5411
smmiller@arc.nasa.gov

Pat Kaspar
NASA/NREN
650.604.5391
pkaspar@arc.nasa.gov