LA-UR- 95-0035

Title: The Sunrise Project: An R&D Project for a National Information Infrastructure Prototype

Author(s): Juhnyoung Lee

Submitted to: Korea Information Science Society

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Form No. 836 R5
ST 2629 10/91

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

35

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

The Sunrise Project: An R&D Project for a National Information Infrastructure Prototype

Juhnyoung Lee
Advanced Computing Laboratory
Los Alamos National Laboratory
Los Alamos, NM 87545
U.S.A.

1 Overview

Sunrise is a Los Alamos National Laboratory (LANL) project started in October 1993. It is intended to a prototype National Information Infrastructure (NII) development project. A main focus of Sunrise is to tie together enabling technologies (networking, object-oriented distributed computing, graphical interfaces, security, multimedia technologies, and data mining technologies) with several specific applications. A diverse set of application areas was chosen to ensure that the solutions developed in the project are as generic as possible. Some of the application areas are materials modeling, medical records and image analysis, transportation simulations, and education. This paper provides a description of Sunrise and a view of the architecture and objectives of this evolving project.

The primary objectives of Sunrise are three-fold:

- To develop common information-enabling tools for advanced scientific research and its applications to industry,
- To enhance the capabilities of important research programs at the Laboratory, and
- To define a new way of collaboration between computer science and industrially relevant research.

This paper is an excerpt from the discussion on the Sunrise Project in a web server located at Los Alamos National Laboratory. The URL for Sunrise is http://www.acl.lanl.gov/sunrise. The work described here represents the contributions of many people, indeed, the entire Sunrise team. Researchers who must be mentioned are David Forslund, Richard Phillips, Bob Tomlinson, Pat Kelly, Allen McPherson, Tim Merrigan, Pat Eker, David Kilman, Paul Hinker, John Reynders, Jon Bradley, Jonathan Greenfield, Steve Tenbrink, Niels Jensen, Paul Ginsparg, Mark Doyle, Francisco Reverbel, Mohamad Ijadi, and others.

The basic paradigm being developed involves a document-centric user interface which allows arbitrary object support including embedded applications, multimedia video/voice fragments and links to a wide information space. We will eventually develop an information kiosk based on an ATM network so that all the participants can exchange, publish, or interact with applications and data. This will function on heterogeneous platforms, provide for constrained access to data through security mechanisms, and be extensible. The data mining technology will include the ability to quickly browse large complex image databases with various feature extraction capabilities, provide advanced, selective compression algorithms, and the ability to merge and purge large complex datasets. The focus is on real scale problems, reflecting our belief that only experience with real problems encompassing huge datasets will facilitate true progress. Therefore, our focus will be on developing prototype tools that can be tested in real-sized testbeds. The project is organized into subgroups consisting of people working on the enabling technologies and people working on each of the several representative application areas.

2 Enabling Technologies

The Sunrise project is organized according to an integrated, layered approach. The three basic layers are:

- Hardware, e.g. networking fabric and computer systems,
- Services, e.g. security and database systems, and
- Applications, e.g., medicine and education.

Some aspects of each layer will be discussed briefly.

2.1 High-Performance Computing

Sunrise is making use of several high-performance computing resources as network-accessible computing servers. The most significant of these are a Thinking machines CM-5 and a Cray Research T3D. The CM-5 is a 1056 node machine with 33 GB of memory and 440 GB of disk (accessible at up to 200 MB/s). It has 4 High Performance Parallel Interface (HIPPI) 800 Mb/s channels for sending data to other devices. It has a peak speed of 132 Gflops and has sustained over 60 Gflops on real applications. It is being used for global climate modeling,

materials modeling, gas and oil seismic and reservoir modeling, as well as distributed visualization and medical image analysis.

The T3D is a 128 node machine with 128 DEC Alpha processors and 8GB of memory. It has 90 GB of attached RAID disk and one HIPPI channel for high speed communication. It is primarily being used for medical image analysis and various industrial applications.

2.2 Networking

The Sunrise project will involve the use of several technologies that will enhance the effectiveness of computers in an office environment. A critical part of this effectiveness is communication and information retrieval. Using email and other Internet applications of today as models, we hope to expand the capabilities of a workstation and its network to provide such things as video on demand, teleconferencing, visualization postprocessing, and other tools that free up time for the user to be more effective in other endeavors.

At the present time, a high performance network fabric is not ubiquitous, so not all of the services just enumerated can be provided to all clients. Therefore, Sunrise currently uses whatever connectivity is available and scales services accordingly. Presently we are using local area 10 MB/s Ethernet and 100 Mb/s FDDI, and wide area 1.5 Mb/s (DS1) and 45 Mb/s (DS3) Internet services, and 34 Mb/s SMDS services. A network technology that, hopefully, will support the full service vision is one where various data rates and qualities of service are supported. The emerging Asynchronous Transfer Mode (ATM) appears to provide that support and will eventually be used throughout the Sunrise environment.

Initially, the ATM switches will only support variable bit rate (VBR) service and will be operating at 100 Mb/s. The constant bit rate (CBR) service for support of voice and video will come later as will higher data rates of 155 Mb/s under the Synchronous Optical NETwork (SONET) specifications. The SONET aspect is an important issue because the telephone companies from the local carriers (US West) to the long haul carriers (AT&T, MCI, Sprint, etc.) are now building a SONET infrastructure to support wide area networks using ATM. Their goal is to have the ATM switch located at the central office of the telephone company rather in the computer room. The important point, however, is that the network technology will be the same for local area networks (LANs) as well as wide area networks (WANs), which is definitely not

the case today. Using similar technologies in the WAN and LAN make interconnectivity seamless and much more robust.

2.3 Distributed Objects

The Sunrise project is based on the premise that a distributed object approach will be necessary for a national-scale information system. We are trying to anticipate the future and are using the Object Management Group's (OMG) evolving Common Object request Broker Architecture (CORBA) [1] since it is the emerging industry standard in this area. We are also investigating the use of the OpenDoc compound document standard [2]. The OpenDoc storage model is based on Bento. Both the OpenDoc and Bento efforts are being coordinated by the Component Integration Lab (CIL).

The usefulness of the NII will be greatly enhanced when users of applications can quickly and easily locate and access data and computing resources in the network. Those resources may be, for example, public or private electronic libraries, data conversion services, or various other application specific services. Since thousands of services built by hundreds of organizations will exist on the network, it is important to facilitate ways of building applications by using software components built by different organizations.

We must build an infrastructure that allows systems to be built and modified with interchangeable components in much the same way that we combine stereo components to build an audio system. When building an audio system, we can interchangeably add or subtract components without redesigning the system or even replace a tape player with a CD player or some other device without changing anything else in the system.

Users accessing the NII will be combining desktop machines such as PCs and workstations with various data and computational servers into a widely distributed yet integrated system. Components of the system will often have different CPU architectures, different data communication infrastructure and data rate capabilities, different operating systems, and be widely distributed. Nevertheless, users' applications will need to be able to call on those resources as if the resources were locally available.

This composable component-based system requires further evolution of distributed

computing environments. To accomplish this evolution, we are basing the Sunrise environment on the notion of distributed objects. Properly designed object-based systems have the unique advantage that changes in one area of the system do not affect other areas. A properly designed set of object classes will enable us to link resources in a modular way to manage complexity and yet allow great extensibility. We are building the Sunrise distributed computing infrastructure on Orbix [3], Iona Technologies' CORBA implementation.

Early emphasis in the development of objects has been on the Sunrise *TeleMed* project with the National Jewish Center in Denver. The system allows CAT scan and x-ray images of lungs to be submitted from anywhere in the network to an electronic repository. The image is then be processed to compute a "signature" that can then be used to find other similar images. The *TeleMed* project will be discussed in detail in Section 3.

2.4 Executive User Interface

The overall goal of this activity is to develop a general, media-rich user interface system which has capabilities that meet the needs of all Sunrise application clients. In general, this is an executive user interface (EUI), one that has a set of convenience capabilities common to all application areas, while still providing application clients easy access to their preferred user interfaces. One of the most important convenience capabilities is to facilitate telecollaboration between users. In its simplest form this feature permits peer-to-peer collaboration in a common screen space. More advanced capabilities will include many-to-many collaboration and video teleconferencing. The Sybase product, Gain*Momentum* [4] is being used for interface development.

Among the features of the EUI are:

- Compound (media-rich) document executive; provides a complete multimedia authoring and display capability. Data stored in Bento (the OpenDoc persistent storage system) documents can be dynamically assembled and viewed.
- A logbook facility; will allow a researcher to maintain a time-stamped multimedia record of all activities. Since the logbook will be retained as a Bento file, it can be accessed by other applications.

- Generic Sunrise data navigator; provides a search and retrieval tool, including mixed media search capability. An appropriate viewer is invoked depending on media type.
- Customized, application-specific user interfaces can be defined and easily incorporated. Launch buttons can be added as needed to the basic EUI.
- Other EUI convenience features are a MIME-compliant mailer and viewer, multimedia annotations; digital post-its, sketchers, audio, video, animation, etc. For intra- and inter-document navigation, thumbnails, hyperlinks, and World Wide Web links are provided.

2.5 Security

For the National Information Infrastructure to truly be a valuable national resource, it must support a diverse set of users and usages. The network will therefore have traffic ranging in sensitivity from public information (such as income tax forms) to sensitive medical and financial records. The loss or corruption of either would cause serious (and in some cases irreparable) damage to individuals, companies and government. Computer security is therefore essential to the success of the NII.

The goal of the computer security part of the Sunrise project is to provide host and network security for services and for machines offering those services. To that end, one of the primary projects is to provide the tools to build secure applications easily and in a fool-proof way.

We have adopted an ANSI X9-based certifying authority system for secure public key distribution and a Kerberos v4 and v5-based library for securely communicating with the certifying authority. All security features are implemented as a service in the Orbix system (the CORBA implementation mentioned earlier) and are thus available for security enforcement among all objects. In addition, we provide a library for secure communication (includes DES and RSA encryption routines) as well as connect and accept routines for automatically transferring credentials and session keys securely.

2.6 Data Mining

With the advent of both National Information Infrastructure and technology capable of recording, analyzing, and displaying high-resolution digital data for a bewildering array of applications, an information explosion is rapidly building. Just as the Interstate Freeway System led to the explosive growth of automobile use, an information superhighway will lead to explosive growth in the availability of and demand for electronic information. The viability of such a system will hinge on our ability to accommodate such growth, but managing this information explosion is additionally complicated by the multidimensionality of this information. In the next decade, users will be accessing traditional text databases, digital sound, images, video, numerical archives as well as evolving paradigms for structured data. Many applications will also involve heterogeneous information, both structured documents and raw numerical data, and essentially all users will be submitting their requests from remote locations.

The goal common to all users will be to maximize the amount of useful information extracted while minimizing the loss of productivity resulting from wasted time and resources, prohibitive access or transmission charges, and retrieval of meaningless, incorrect, or corrupted data. To meet this goal, we must develop the capabilities to represent and store large information data sets efficiently, to extract and transmit material across a distributed multiuser network, and to analyze, interpret and display such information effectively and meaningfully.

For the optimal utilization of information, the Sunrise project is developing software tools to perform extraction, storage, and matching of feature vectors from digital imagery. These tools will enable us to demonstrate the value of an image browsing capability. In addition, Sunrise is developing data compression strategies that facilitate progressive coding/ transmission. The objective of progressive coding is to enable retrieval, transmission, and visualization of data at varying levels of feature resolution and/or in a spatially restricted manner to reduce the costs of data access and transmission.

3 Applications

The Sunrise project is looking at several applications so that the enabling technologies we develop will be applicable to a broader range of problems.

3.1 Telemedicine

The most well-developed application in the Sunrise suite is a medical application called *TeleMed*. This application serves as an excellent case study of all the Sunrise technologies previously described and thus warrants description in some detail. *TeleMed* grew out of a relationship with physicians at the National Jewish Center for Immunology and Respiratory Medicine (NJC) in Denver, Colorado. These physicians are experts in pulmonary diseases and radiology, helping patients combat the effects of tuberculosis and other lung diseases throughout the Nation. These individuals are an expensive and scarce resource, who travel around the country to share their expertise with other physicians.

To make the knowledge and experience at the National Jewish Center available to a wider audience, Los Alamos National Laboratory has developed a telemedicine system called *TeleMed* which is based on a national radiographic repository located at Los Alamos National Laboratory. Without leaving their offices, participating doctors can view radiographic data via a sophisticated multimedia interface. With the new system, a doctor can match a patient's radiographic information with the data in the repository, review treatment history and success, and then determine the best treatment. Furthermore, the features of *TeleMed* that make it attractive to clinicians and diagnosticians, make it valuable for teaching and presentation as well. Thus, a resident can use *TeleMed* for self-training in diagnostic techniques and a physician can use it to explain to a patient the course of their illness.

LANL's expertise in the areas of high-performance computing, advanced networking, image processing, and multimedia information systems has proved to be an important asset for *TeleMed*. These technologies come into play in the following ways:

- **Image processing** - As databases are developed to store large volumes of information, new methods for mining huge amounts of data will become increasingly important. Tools will be needed for analyzing text, imagery, 1-D signals, and other numerical information (LDL & HDL cholesterol counts, blood pressure, body temperature, etc.) One such key tool that is being used in the *TeleMed* project is automated data analysis of radiographic images. This tool provides the physician with an agent who browses radiographic data and calls the physician's attention to suspect areas. One typical agent can analyze CT images from patients with lymphangioleiomyomatosis (LAM) disease. Pulmonary CT studies reveals a

primary signature of the disease, namely the presence of many cysts throughout the lung. Sunrise team has created algorithms that automatically locate the cysts in the pulmonary CT data and record their size [5]. One can compute a histogram of the cyst sizes, which is called a cystogram. The cystogram provides a meaningful quantitative measure of the progress of the disease.

- **High-performance computing** - Especially important to the *TeleMed* project is the fact that LANL houses the massively parallel computing systems mentioned earlier, the Thinking Machines CM-5 and the Cray T3D. Since medical image analysis techniques (such as the LAM agent described above) can be computationally intensive, Sunrise acts as a server of massively parallel computing cycles to physicians in the *TeleMed* network. Access to such a compute server allows physicians to do near real-time consultation with image analysis agents.

- **Advanced networking** - Over the last few years LANL has made numerous contributions to advanced networking technology from the invention of the now-ANSI standard HIPPI specification to a pivotal role in the CASA Gigabit testbed project. The importance of this technology to the *TeleMed* project is clear, since multi-megabyte radiographic data files must be transferred rapidly to the diagnostician. Moreover, Sunrise is aggressively exploring the role of ATM communications technology for use both in image data transfer and desktop video telecollaboration.

- **Multimedia information systems** - The July 1994 issue of R & D magazine [6] features, as a cover story, LANL developed MediaView [7], a comprehensive multimedia document authoring and management system. MediaView represents the advanced multimedia expertise Los Alamos is applying to the *TeleMed* project. Through MediaView, the concept of a document-centric user interface was independently developed. This type of interface is used in the *TeleMed* project and affords users access to a complete range of media-rich components. Physicians can, for example, affix digital Post-it notes to a patient record or x-ray. Voice annotations can be dictated and directly attached to a patient document. Also, a physician can digitally sketch on an x-ray, pointing out important features to a collaborator or patient.

TeleMed has recently been deployed via DS1 service between Los Alamos and the National Jewish Center. At this point, physicians at National Jewish are learning the user interface and reacting to its design and shortfalls. As the *TeleMed* system reaches a productive state, other sites will be added to the *TeleMed* network. These will all be sites who are treating and doing research on multi-drug resistant tuberculosis (MDR-TB). Those intending to join are the National Institute of Health in Bethesda, MD, the Center for Disease Control in Atlanta, GA, the Bureau of Tuberculosis Control in New York, NY, and the Department of Health Services in Los Angels, CA. With these sites in place, physicians will truly be able to telecollaborate on MDR-TB cases, using the Los Alamos radiographic repository as a common point of contact.

3.2 Transportation Simulation

The TRansportation ANalysis SIMulation System (TRANSIMS) is a new simulation-based analysis system intended to model architectures of advanced transportation systems. It is an entirely new approach extending beyond traditional transportation system modeling. TRANSIMS is based on bottom-up generation of system activity and behavior using detailed simulations of interacting subsystems and traveler intelligence. In this approach regional microsimulation and representation of traveler-by-traveler plans and travel execution yields emergent behavior patterns to predict transportation issues ranging from land use to air quality.

As the early TRANSIMS effort, a proof-of-principle demonstration was completed that illustrated the approach's features an capabilities using an intermodal pilot study that focused on a notional, metropolitan area. The prototype study was based on an actual sample of 2,100 households in Albuquerque, New Mexico. Based in this sample, TRANSIMS synthesized 218,000 households comprising 400,000 individual travelers. While this synthesized metropolitan area is based on an actual data sample, the notional approach does not disclose confidential information regarding any specific individual or household.

Two million trip plans were generated for the individual travelers over a 24-hour period based on travel requirements, network characteristics, mode availability, and an individual socio-economic cost model. Walking, private vehicle, and public transit mode selection was allowed in this demonstration, along with traveler demanded mode transfer. Individual vehicle and traveler operational characteristics were permitted and the microsimulation element of

TRANSIMS executed the trip plans and aggregated the results to illustrate freeway traffic during rush periods. Based on specific emission profiles for the vehicles as mobile sources, regional air quality studies were represented.

The TRANSIMS architecture is intended to provide a new and unique tool for regional-level, micro-scale system representation that will allow unification of issues in regional transportation planning and analysis, traffic engineering, and environmental impact analysis. Currently, a demonstration prototype of a remote user interface is under development. This demonstration interface will provide remote TRANSIMS users with the following features:

- a graphical network browser,
- a graphical trip plan browser,
- an interactive data plotting capability, and
- an image browser.

3.3 Education

The goals of using NII for education are (1) to illustrate the capability and power of multimedia databases and virtual laboratories in a real world context, using digital video and synchronized sound on the internet via high speed lines, and (2) to facilitate high cognitive level learning over the Internet through student-scientist and student-student collaboration at a distance.

The Sunrise project is putting together a multimedia server that allows for the distribution of digital video and synchronized sound to many clients simultaneously. On the server would reside video and audio information that was previously stored, and accessible through a user-friendly database. This effort is already on-going within LANL. This effort is to test this technology with an educational target population.

Following setting up the server for use in this pilot, Sunrise will select the content for the server, work with collaborating scientists, and perform the liaison with the local schools. This will involve: technical support, curriculum discussions, review of video and selection, training of users at the schools, and collection of usage data at the two schools (Las Cruces and Los

Alamos) in the pilot test. Sunrise will also be responsible for writing a report that documents the pilot test experience at the participating schools.

Sunrise will work together in assisting students and teachers in curricular applications and in exploring alternative, state-of-the-art I/O devices (e.g. PDA). One application Sunrise will facilitate is the use of specific multimedia scientific databases (e.g. from NASA).

Using very large data sets in K-12 Class rooms: How should and can large scientific data be used by K-12 students and teachers? Such data should be common in science class rooms in the same way that recent surveys, new events, and books can be discussed in social studies or language arts classes.

Preprocessing and presentation of this data are critical issues in enabling it to have any educational effect. We plan to generate structured image databases and datasets which can be remotely accessed. Many of the images will be animations so compression and efficiency will be important. Ideas for projects include showing the (non) correlation between maps and aerial photographs, the impact of urbanization over 10-20 years around a large city, or the impact of the floods this year in the Mississippi River basin.

In addition to accessing and displaying the images, we want students to use such images and data in their studies. Therefore, it would be important to give them access to the interactive document readers and editors proposed elsewhere in this proposal. In 10-20 years, such documents will be a fundamental form of knowledge transfer and storage. We need to start now understanding how students and teachers are going to effectively use such capabilities.

3.4 Materials Modeling

As in many fields, materials science is undergoing an information revolution. The development of experimental technologies that probe materials properties on extremely small scales results in tremendous amounts of data. New approaches to modeling materials systems using advanced computational methods also lead to data on an unprecedented scale. One of the great challenges facing materials science is to learn how to maximize the amount of useful information extracted from these data. Equally important is to learn how we can couple the two approaches - experiment and modeling.

Examples of the new experimental methods available in materials science are the variety of microscopy tools, such as transmission-electron microscopy (TEM), atomic-force microscopy, etc. TEM, for example, has advanced considerably over the past few years with the best instruments now reaching accuracies on the order of an angstrom. That kind of data provides atomic-level images of a large variety of materials structures. How to best extract information from those images involves both identification of important features as well as creating databases to better access the images. The ability to effectively and efficiently browse these databases will greatly enhance their value. In this project, we will bring to materials science the advances made in the Data Mining part of the Sunrise project, which is geared towards developing new ways to handle large datasets. Specifically, we are working with researchers in the new Materials Science Laboratory to maximize the information obtained in TEM.

Materials modeling has advanced considerably with the use of massively-parallel computing. Atomistic simulations, for example, can now probe the detailed positions and movement of millions of individual atoms. The amount of information from these simulations is almost overwhelming and extracting information from them poses many of the same problems facing experimentalists.

The optimal approach to materials science is to couple experiment with modeling. The main focus of this project is to develop new ways of sharing information between experimentalists, theorists, and computational scientists. This information will take the form of images, data, graphs, documents, etc. and will be accessible from a common set of tools that is available from a variety of computer platforms. It will involve transmitting and storing large amounts of data as well as developing a Common Graphical Interface between the output of experimental equipment and that from computer-based modeling.

4 Conclusions

The Sunrise project is to develop a prototype NII system involving a suite of applications, including materials, telemedicine, transportation, electronic publication and education, driving the base technologies of distributed computing, user interfaces and telecollaboration, active visualization, data mining, transmission, and navigation, security, and

ATM networking. We are tightly integrating the applications with the enabling technologies by carefully analyzing the common components of each and designing a general system to support them. We are using a compound-document approach building on and extending some emerging industry standards such as OMG's CORBA, and CIL's OpenDoc framework for distributed compound documents. We plan to extend this work by building a more robust framework underneath most of the applications and by adding more navigation tools, an electronic publication framework, and integrating the work on image compression with the image browsing system that has been developed.

References

[1] Object Management Group, Ed., *Common Object Request Broker Architecture*, Wiley-QED, 1991.

[2] Anon., *OpenDoc Overview*, Component Integration Laboratories, San Francisco, CA., 1994.

[3] Horn, Chris, "Focus on Open Systems," *Object Magazine*, Sept-Oct, 1993. Also see http://www.iona.ie/www/Orbix/arch/Summary.html.

[4] Anon., *GainMomentum User's Guide*, Sybase, Inc., Emeryville, CA., 1994.

[5] Everhart, J., Cannon, M., Newell, J., and Lynch, D., "Image Segmentation Applied to CT Examination of Lymphangioleiomyomatosis (LAM)," to be published.

[6] *R&D*, Cahners Publishing, July, 1994.

[7] Phillips, R., "MediaView: A General Multimedia Digital Publication System," *Comm. ACM*, Vol. 34, No. 7, July, 1991.

[8] Phillips, R., "A Network-Based Distributed, Media-Rich Computing and Information Environment," *Digital Media and Electronic Publishing*, Leeds, UK, 6-8 December 1994.