1996

A Tale of Two Applications on the NII

Geoffrey C. Fox
Syracuse University, Northeast Parallel Architectures Center

A Tale of Two Applications on the NII

Geoffrey C. Fox
Syracuse University
Northeast Parallel Architectures Center
111 College Place
Syracuse, New York 13244-4100
gcf@npac.syr.edu, http://www.npac.syr.edu

Abstract

We describe the expected capability of the NII (as an evolution of the Internet) interns of five broad service areas—collaboration, multimedia information dissemination, commerce, metacomputing, and Webtop productivity. We illustrate the demands on these services and the technology implications by examination of two application areas—manufacture of complex systems, such as aircraft and crisis management, command and control.

1 Introduction—WebWindows

We expect a revolution in the computing industry with a shift from the current software model built around UNIX, Windows 95/NT, Macintosh, and mainframe operating environments. Rather, future applications will be built for WebWindows [1]. WebWindows is loosely defined as either the distributed operating system of the World Wide Web (WWW), or more operationally, by the use of Web servers and clients, their open interfaces (HTML, VRML, CGI, etc.), and their technologies (Java, LiveMedia, etc.). We believe that WebWindows will prove to be the dominant environment for two linked reasons. Firstly, the Web is a pervasive technology, and not linked to a particular hardware platform. Secondly, the Web naturally supports a heterogeneous distributed computing model—a concept critical in many applications including both manufacturing and command and control.

The National Information Infrastructure (NII) more generally describes the software, services, and digital infrastructure, which will support WebWindows applications. We have previously described [1] how the NII will be used in seven different areas, including society (delivery of digital TV and information systems), education, collaboratory (distributed collaborative research and development), business enterprise systems (often called Intranets), and health care (from modern telemedicine through patient record systems). As shown in Figure 1, we expect NII applications to be built in a layered fashion. In particular, we suggest that one will build and re-use a set of NII services. This is the "multi-use" generalization for software and systems of the old dual use concept of linking military and commercial products. Alternatively, one can consider it as an implementation of the COTS (Commercial Off The Shelf) philosophy where military applications are built on top of base commercial multi-use services. Table 1 lists our seven NII applications, and shows that they can share most of the services. In Section 2, we briefly review the latter multi-use base while in the last two sections, we review our two chosen applications—command and control, and manufacturing. These exercise significantly all five identified service areas, but in different ways.

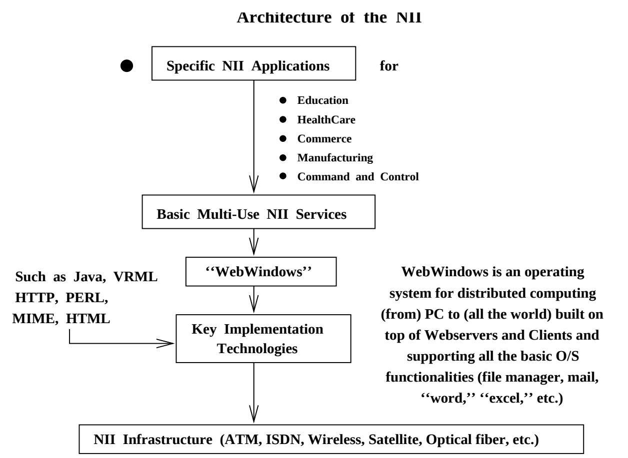

Figure 1: A layered view of Web (NII) software building applications on top of generic services that are in turn built on pervasive technologies

2 Five NII Service Areas

Returning to Web software architecture of Figure 1, we now discuss services and applications. These are not precisely defined, for services are essentially generic applications, and most applications are complex metaproblems [2] built recursively from services and "sub-applications." Thus, there is a grey fuzzy line distinguishing services and applications.

- **WebTop Services**—Publishing, Productivity, Software Engineering

  Here, we put base productivity tools including "WebWord," "WebLotus123," Web linkage to relationship databases, etc. NPAC has built an interesting prototype of such a capability in

Table 1: Services used by seven NII applications

Applications	WebTop Productivity	Info VISiON	Commerce/ Security	Collab- oration	Meta Computing
Society	X	X	X	X	
Education	X	X		X	
Enterprise Systems	X	X	X	X	
Health Care	X	X	X	X	
Command and Control/ Crisis Management	X	X	X	X	X
Manufacturing	X	X	X	X	X
Collaboratory	X	X		X	

Webfoil—a Web based presentation system. Essentially, all applications use this service.

- **InfoVISiON**—Information, Video, Imagery, Simulation, ON demand

  This includes base database storage, management, query, and dissemination of the full range of multimedia archives of the World's distributed digital libraries. We can expect the "digital TV" application to produce hundreds of Terabytes of available information dominated by video data. Note delivery of results of a simulation—such as access on demand to a weather model—is included in this service.

  In manufacturing, InfoVISiON corresponds to delivery of data in a configuration controlled database. In command and control, this service allows commander to index and access videos of battle damage videos or real-time engagements.

- **Commerce**—Digital Cash, Security, Authentication, etc.

  This collection of services enables electronic commerce, including on-line banking and shopping. These services are also essential for the use of the WWW for processing and exchange of proprietary (manufacturing), and classified (military) data.

- **Collaboration**—Real-Time Interactive and "Batch"

  This includes desktop video conferencing, three-dimensional graphics MOOs, geographically distributed CAVEs leading to full televirtual interactions. The emerging VRML 2.0 standards will be very important in building virtual environments. As discussed earlier, a wide variety of other types of interactive information exchange is necessary. This underlies the concepts of collaboratories (virtual research groups or scientific laboratories), and the virtual company of the next century's agile manufacturing environment. In the more static mode, we see workflow and configuration control, which allows tightly integrated projects, such as those needed to build a complex system including an aircraft (see Section 4) or a large software module with a distributed team.

- **Metacomputing**—A worldwide collection of computers organized together as a single computational engine for simulation or information processing [3].

  This service can be used to control remote medical and scientific instruments; search the world for information; simulate the weather expected in a military engagement, or link computers in different companies for a multi-disciplinary optimization of a new vehicle.

Some services listed above can be already prototyped in terms of today's Web technologies. For example, base WebTop or early Collaboration services are now becoming available. Some other services are still waiting for their pervasive enabling technologies, such as physical infrastructure that will enable InfoVISiON or security that will enable Internet Commerce. Finally, the computationally extensive NII services, characterized above broadly as "Metacomputing" require a major extension of the whole Web paradigm, currently still focused on static page services, but already gradually expanding towards computation and interactive simulation via technologies such as Java and VRML.

3 Manufacturing and Computational Fluid Dynamics on the NII

HPCC (High-Performance Computing and Communication) and the NII are used today, and can be expected to play a growing role in manufacturing, and more generally, engineering. For instance, the popular concept of agile manufacturing supposes the model where virtual corporations generate "products-on-demand." The NII is used to link collaborating organizations. Powerful computers are needed to support instant design (or more accurately redesign or customization) and sophisticated visualization and virtual reality "test drives" for the customer. At the corporate infrastructure level, concurrent engineering involves integration of the different component disciplines—such as design, manufacturing, and product life cycle support—involved in engineering. These general ideas are tested severely when they are applied to the design and manufacturing of complex systems such as automobiles, aircraft, and space vehicles such as shuttles. Both the complexity of these products, and in some sense the maturity of their design, places special constraints and challenges on HPCC and the NII.

High-performance computing is important in all aspects of the design of a new aircraft. However, it is worth noting that less than 5% of the initial costs of the Boeing 777 aircraft were incurred in computational fluid dynamics (CFD) airflow simulations—the "classic" Grand Challenge in this field. On the other hand, over 50% of these sunk costs could be attributed to overall systems issues. Thus, it is useful but not sufficient to study parallel computing for large scale CFD. This is "Amdahl's law for practical HPCC." If only 5% of a problem is parallelized, one can at best speed up and impact one's goals—affordability, time to market—by this small amount. HPCC, thus, must be fully integrated into the entire engineering enterprise to be effective. Very roughly, we can view the ratios of 5% to 50% as a measure of ratio of 1:10 of the relevance of parallel (classic MPPs) and distributed computing (i.e., the NII or Web!) in this case.

The maturity of the field is illustrated by the design criterion used today. In the past, much effort has been spent on improving performance—more speed, range, altitude, size. These are still critical under extreme conditions, but basically these just form a given design framework that suffices to buy you a place at the table (on the short-list). Rather, the key design criteria is competitiveness, including time to market, and total affordability. Although the design phase is not itself a major cost item, decisions made at this stage lock in most of the full life cycle cost of an aircraft with perhaps 80% of total cost split roughly equally between maintenance and manufacturing. Thus, it certainly would be important to apply HPCC/NII at the design phase to both shorten the design cycle (time to market) and lower the later ongoing costs of manufacturing and maintenance.

We take as an example the design of a future military aircraft—perhaps 10 years from now. This analysis is taken from a set of NASA sponsored activities centered on a study of ASOP—Affordable Systems Optimization Process. This involved an industrial team, including Rockwell International, Northrop Grumman, McDonnell Douglas, General Electric, and General Motors. ASOP is one of several possible approaches to multidisciplinary analysis and design (MAD) and the results of the study should be generally valid to these other MAD systems. The hypothetical aircraft design and construction project could involve six major companies and 20,000 smaller subcontractors. This impressive virtual corporation would be very geographically dispersed on both a national and probably international scale. This project could involve some 50 engineers at the first conceptual design phase. The later preliminary and detailed design stages could involve 200 and 2,000 engineers, respectively. The design would be fully electronic and demand major computing, information systems, and networking resources. For instance, some 10,000 separate programs would be involved in the design. These would range from a parallel CFD airflow simulation around the plane to an expert system to plan location of an inspection port to optimize maintainability. There is a corresponding wide range of computing platforms from PCs to MPPs and a range of languages from spreadsheets to High-Performance Fortran. The integrated multidisciplinary optimization does not involve blindly linking all these programs together, but rather a large number of suboptimizations involving at one time a small cluster of these base programs. Here we see clearly, an essential role of the NII to implement these set of optimizations, which could well need linking of geographically separated compute and information systems. An aircraft is, of course, a very precise system, which must work essentially flawlessly. This requirement implies a very strict coordination and control of the many different components of the aircraft design. Typically, there will be a master systems database to which all activities are synchronized at regular intervals–perhaps every month. The clustered suboptimizations represent a set of limited excursions from this base design that are managed in a loosely synchronous fashion on a monthly basis. The configuration management and database system are both critical and represent a major difference between manufacturing and command and control, where in the latter case, real time "as good as you can do" response, is more important than a set of precisely controlled activities. These issues are characteristic of the high-performance distributed computing (HPDC) where, although loosely coupled, the computers on our global network are linked to "solve a single problem."

ASOP is designed as a software backplane (the NII or Web) linking eight major services or modules shown in Figure 2. These are design (process controller) engine, visualization, optimization engine, simulation engine, process (manufacturing, productibility, supportability) modeling toolkit, costing toolkit, analytic modeling toolkit, and geometry toolkit. These are linked to a set of databases defining both the product and also the component properties. Parallel computing is important in many of the base services, but HPDC is seen in the full system.

4 Command and Control or Crisis Management on the NII

Command Control (sometimes adding in Computing, Communications, Intelligence Surveillance, and Battle Management with abbreviations lumped together as BMC^{4}IS) is the task of managing and planning a military operation. It is very similar to the civilian area of Crisis management, where the operations involve combating effects of hurricanes, earthquakes, chemical spills, forest fires, etc. Telemedicine was once considered a video conferencing application, but now this is becoming like command control in the interventional informatics concept [4] where the doctor (cf. commander) needs a rich interactive information to win the "battle" i.e., treating the patient. Both the military and civilian cases have computational "nuggets" where parallel computing is relevant. These include processing sensor data (signal and image processing) and simulations of such things as expected weather patterns

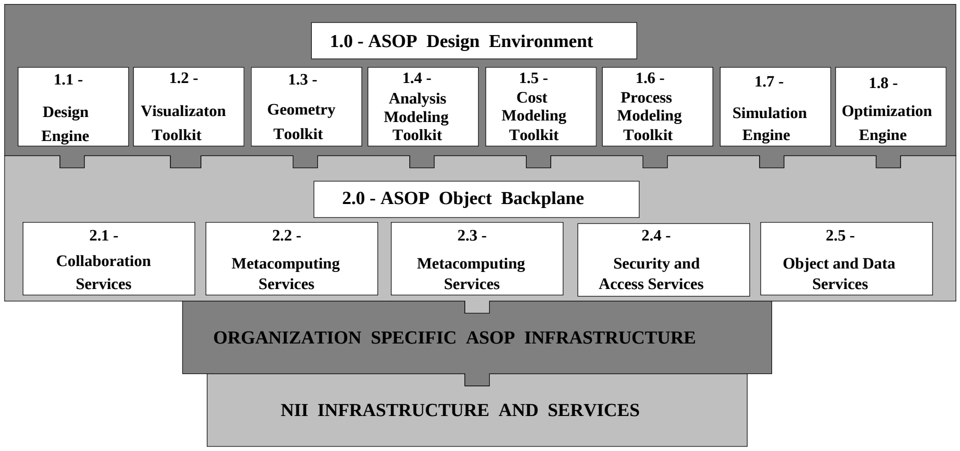

Figure 2: Affordable Systems Optimization Process (ASOP) Implemented on the NII for Aeronautics Systems

and chemical plumes. One also needs large-scale multimedia databases with HPDC issues related to those described for InfoVISiON in Section 2.

The NII is needed to link military planners and decision makers, crisis managers, experts at so-called anchor desks, workers (warriors) in the field, information sources such as cable news feeds, and large-scale database and simulation engines.

A key characteristic of the required NII support is adaptivity. Crises and battles can occur anywhere and destroy an arbitrary fraction of the existing infrastructure. Adaptivity means making the best use of the remaining links, but also deploying and integrating well mobile enhancements. The information infrastructure must exhibit security and reliability or at least excellent fault tolerance (adaptivity). Network management must deal with the unexpected capacity demands and real time constraints. Priority schemes must allow when needed critical information (such as chemical plume monitoring and military sensor data) precedence over less time critical information, such as background network video footage.

Needed computing resources will vary from portable handheld systems to large backend MPPs. As there will be unpredictable battery (power) and bandwidth constraints, it is important that uniform user interfaces and similar services be available on all platforms with, of course, the fidelity and quality of a service reflecting the intrinsic power of a given computer. As with the communications infrastructure, we must cope with unexpected capacity demands. As long as the NII is deployed nationally, computational capacity can be exploited in remote sites. The Department of Defense envisages using the basic NII (GII) infrastructure for command and control, augmented by "theater extensions" to bring needed communications into critical areas. The "take it as it is" characteristic of command and control requires that operating systems and programming models support a general adaptive mix (metacomputer) of coordinated geographically distributed but networked computers. This network will adaptively link available people (using perhaps personal digital assistants) to large-scale computation on MPPs and other platforms. There are large computational requirements when forecasting in real-time physical phenomena, such as the weather effects on a projected military action, forest fires, hurricanes, and the structure of damaged buildings. On a longer time scale, simulation can be used for contingency planning and capability assessment. Training with simulated virtual worlds supporting televirtuality, requires major computational resources. In the information arena, applications include datamining to detect anomalous entries (outliers) in large federated multimedia databases. Data fusion including sensor input and processing, geographical information systems (with perhaps three-dimensional terrain rendering), and stereo reconstruction from multiple video streams are examples of compute intensive image processing forming part of the needed HPDC environment.

A critical need for information management involves the best possible high-level extraction of knowledge from databanks—the crisis manager must make judgments in unexpected urgent situations—we cannot carefully tailor and massage data ahead of time. Rather, we need to search a disparate set of multimedia databases. As well as knowledge extraction from particular information sources, the systematic use of metadata allowing fast coarse grain searching is very important. This is a specific example of the importance of standards in expediting access to "unexpected" databases. One requires access to databases specific to crisis region or battlefield, and widespread availability of such geographic and community information in electronic form is essential. There are very difficult policy and security issues, for many of these databases need to be made instantly available in a hassle-free fashion to the military commander or crisis manager—this could run counter to proprietary and security classification constraints. The information system should allow both network news and warriors in the field to deposit in near real-time, digital versions of their crisis and battlefield videos and images.

As mentioned, we expect that human and computer expertise to be available in "anchor desks" to support instant decisions in the heat of the battle. These have been used in a set of military exercises called JWID (Joint Warrior Interoperability Demonstrations). We note that this information scenario is a real-time version of the InfoVISiON service needed to support the society of the Information Age.

Command and Control has historically used HPDC as the relevant computer and communication resources, are naturally distributed, and not centralized into a single MPP. We see this HPDC model growing into the standard information support environment for all the nation's enterprises, including business, education, and society, as indicated in InfoVISiON in Table 1.

5 Glossary of Concepts and Acronyms

Applets An application interface where referencing (perhaps by a mouse click) a remote application as a hyperlink to a *server* causes it to be downloaded and run on the *client*.

ASOP (Affordable Systems Optimization Process) refers to a process using multidisciplinary optimization to produce more affordable systems

CAVE A room with stereo images on the walls designed to create a televirtual environment.

CFD (Computational Fluid Dynamics) refers to computational solutions of differential equations, such as the Navier Stokes set, describing fluid motion.

COTS (Customer Off The Shelf)—An important concept in Defense Systems, labelling the use of commercial software snf system components rather than specialized limits which are expensive and hard to upgrade, even if initially of higher capability.

Data Fusion A common *command and control* approach where the disparate sources of information available to a military or civilian commander or planner, are integrated (or fused) together. Often, a *GIS* is used as the underlying environment.

Distributed Computing The use of networked heterogeneous computers to solve a single problem. The nodes (individual computers) are typically loosely coupled.

Geographical Information System (GIS) A user interface where information is displayed at locations on a digital map. Typically, this involves several possible overlays with different types of information. Functions, such as image processing and planning (such as shortest path) can be invoked.

Global Information Infrastructure (GII) The GII is the natural world-wide extension of the *NII* with comparable exciting vision and uncertain vague definition.

High-Performance Computing & Communications (HPCC) Refers generically to the federal initiatives, and associated projects and technologies that encompass *parallel computing*, *HPDC*, and the *NII*.

High-Performance Distributed Computing (HPDC) The use of distributed networked computers to achieve high performance on a single problem, i.e., the computers are coordinated and synchronized to achieve a common goal.

Hypertext Markup Language (HTML) A syntax for describing documents to be displayed on the *World Wide Web*.

Hypertext Transport Protocol (HTTP) The *protocol* used in the communication between *Web Servers* and *clients*.

InfoVISiON Information, Video, Imagery, and Simulation ON demand is scenario where *multimedia servers* deliver multimedia information to clients on demand—at the click of the user's mouse.

Integrated Service Data Network (ISDN) A digital multimedia service standard with a performance of typically 128 kilobits/sec, but with possibility of higher performance. ISDN can be implemented using existing telephone (*POTS*) wiring, but does not have the necessary performance of 1–20 megabits/second needed for full screen TV display at either VHS or high definition TV (HDTV) resolution. Digital video can be usefully sent with ISDN by using quarter screen resolution and/or lower (than 30 per second) frame rate.

Internet A complex set of interlinked national and global networks using the IP messaging protocol, and transferring data, electronic mail, and *World Wide Web*. In 1995, some 20 million people could access Internet—typically by *POTS*. The Internet has some high-speed links, but the majority of transmissions achieve (1995) bandwidths of at best 100 kilobytes/sec. the Internet could be used as the network to support a *metacomputer*, but the limited *bandwidth* indicates that *HPDC* could only be achieved for *embarrassingly parallel* problems.

Java A distributed computing language (*Web Technology*) developed by Sun, which has similarities with C++ but supports *Applets*.

MAD (Multidisciplinary Analysis and Design, or Multidisplinary Optimization) refers to the coupling of several areas, such as structural dynamics and fluid flow in a combine tradeoff to produce higher capability vehicles.

Massively Parallel Processing (MPP) The strict definition of MPP is a machine with many interconnected processors, where 'many' is dependent on the state of the art. Currently, the majority of high-end machines have fewer than 256 processors. A more practical definition of an MPP is a machine whose architecture is capable of having arbitrarily many processors—that is, it is scalable. In particular, machines with a distributed memory design (in comparison with shared memory designs) are usually synonymous with MPPs since they are not limited to a certain number of processors. In this sense, "many" is a number larger than the current largest number of processors in a shared-memory machine.

Metacomputer This term describes a collection of heterogeneous computers networked by a high-speed wide area network. Such an environment would recognize the strengths of each machine in the Metacomputer, and use it accordingly to efficiently solve so-called *Metaproblems*. The *World Wide Web* has the potential to be a physical realization of a Metacomputer.

Metaproblem This term describes a class of problem which is outside the scope of a single computer architectures, but is instead best run on a Metacomputer with many disparate designs. These problems consist of many constituent subproblems. An example is the design and manufacture of a modern aircraft, which presents problems in geometry grid generation, fluid flow, acoustics, structural analysis, operational research, visualization, and database management. The Metacomputer for such a Metaproblem would be networked workstations, array processors, vector supercomputers, massively parallel processors, and visualization engines.

Multimedia Server or Client Multimedia refers to information (digital data) with different modalities, including text, images, video, and computer generated simulation. Servers dispense this data, and clients receive it. Some form of browsing, or searching, establishes which data is to be transferred. See also *InfoVISiON*.

Multipurpose Internet Mail Extension (MIME) The format used in sending multimedia messages between *Web Clients* and *Servers* that is borrowed from that defined for electronic mail.

National Information Infrastructure (NII) The collection of *ATM*, cable, *ISDN*, *POTS*, satellite, and wireless networks connecting the collection of 10^8–10^9 computers that will be deployed across the U.S.A. as set-top boxes, PCs, workstations, and MPPs in the future.

The NII can be viewed as just the network infrastructure or the full collection of networks, computers, and overlayed software services. The *Internet* and *World Wide Web* are a prototype of the NII.

PERL An interpreted language particularly targeted at text manipulations and systems programming.

POTS The conventional twisted pair based Plain Old Telephone Service.

Televirtual The ultimate computer illusion where the user is fully integrated into a simulated environment and so can interact naturally with fellow users distributed around the globe.

Virtual Reality Modeling Language (VRML) A "three-dimensional" HTML that can be used to give a universal description of three-dimensional objects that supports *hyperlinks* to additional information.

WebTop refers to the implementation of a set of standard desk top and personal computer tools, which are essential in any computing environment.

Web Clients and Servers A distributed set of clients (requesters and receivers of services) and servers (receiving and satisfying requests from clients) using *Web Technologies*.

WebWindows The operating environment created on the World Wide Web to manage a distributed set of networked computers. WebWindows is built from *Web clients* and *Web servers*.

WebWork [3] An environment proposed by Boston University, Cooperating Systems Corporation, and Syracuse University, which integrates computing and information services to support a rich distributed programming environment.

World Wide Web and Web Technologies A very important software model for accessing information on the Internet based on hyperlinks supported by *Web technologies*, such as *HTTP*, *HTML*, *MIME*, *Java*, *Applets*, and *VRML*.

References

[1] Geoffrey C. Fox and Wojtek Furmanski. The use of the national information infrastructure and high performance computers in industry. In *Proceedings of the Second International Conference on Massively Parallel Processing using Optical Interconnections*, pages 298–312, Los Alamitos, CA, October 1995. IEEE Computer Society Press. Syracuse University Technical Report SCCS-732.

[2] Geoffrey C. Fox. Software and hardware requirements for some applications of parallel computing to industrial problems. Technical Report SCCS-717, Syracuse University, NPAC, Syracuse, NY, June 1995. Submitted to ICASE for publication (6/22/95); revised SCCS-134c.

[3] Geoffrey C. Fox, Wojtek Furmanski, Marina Chen, Claudio Rebbi, and James H. Cowie. WebWork: integrated programming environment tools for national and grand challenges. Technical Report SCCS-715, Syracuse University, NPAC, Syracuse, NY, June 1995. Joint Boston-CSC-NPAC Project Plan to Develop WebWork.

[4] Dave Warner, Jeff Sale, and Eric Viirre. A systems approach for developing an integrative healthcare information distribution infrastructure. Technical Report SCCS-749, Syracuse University, NPAC, Syracuse, NY, November 1995.