## ABSTRACT

People with shared interests are using the Internet to solve problems, accomplish tasks, and create resources that would be well beyond the reach of any one person or organization. The Internet is being used to create virtual libraries, factor large numbers, organize massive volunteer efforts, and filter information in a collaborative fashion. The ability to leverage the efforts of large numbers of networked users has important economic, social, and political consequences. This phenomenon is important to policy makers because it can potentially be used toleverage scarce taxpayer dollars and promote applications of the information infrastructure.

# Leveraging Cyberspace

*Thomas A. Kalil, National Economic Council*

*"Give me a lever long enough and a place to stand, and I will move the Earth."*

*– Archimedes*

The rapid growth in the ubiquity and functionality of the Internet is amazing. The Internet now connects 10 million computers, tens of millions of users, and more than 100 countries. At its current rate of growth, the Internet will connect 100 million computers by the year 2000. Because anyone with a computer and a connection to the Internet can publish, the global information space is also growing rapidly. Developers of search engines such as Altavista and Lycos believe that the Web currently contains 30–50 million pages of information, or 200 to 330 Gbytes of text. At current growth rates, the Web could surpass the 29 Tbytes of the Library of Congress in two years [1].

In addition to allowing anyone to publish, the open architecture of the Internet also allows anyone to add to its functionality. In recent years, researchers and companies have developed new applications such as audio and video streaming, multicasting, downloadable "applets" written in object-oriented languages like Java, voice over the Internet, three-dimensional virtual environments, intelligent agents, electronic payments, and new media types for rich documents. Vendors have scrambled to "Internet-enable" their existing products, such as spreadsheets, groupware, authoring tools, workflow software, and relational databases. Because the Internet serves as its own distribution channel, the diffusion of new standards, protocols, and software can occur very rapidly without the need for shelf space in retail outlets. As Tony Rutkowski, former Executive Director of the Internet Society, is fond of observing, "The Internet is its own revolution."

Although the evolution of the Internet is fascinating, the Clinton-Gore Administration does not view it as an end in itself. The Administration is interested in the Internet and the other information and communications technologies that comprise the national and global information infrastructure because of their potential to contribute to a broad range of economic and social goals. The development of the national information infrastructure (NII) is an Administration priority because of its potential to increase U.S. productivity and economic growth; improve the way teachers teach and students learn; make government more efficient, transparent, and responsive; allow individuals to make better decisions about their health care needs; allow scientists and researchers to collaborate more easily; and enable Americans with disabilities to lead more independent lives.

In my view, one of the more important applications of the Net is its ability to enable "communities of interest" to solve problems, accomplish tasks, or create resources that would be well beyond the ability of any one individual or organization to complete. Steven Whitehead refers to this phenomenon as "cyberspace leveraging," which he defines as "using computer networks to harness the power of a large population of networked users" and leveraging the "small efforts of the many" as opposed to the "big efforts of the few" [2]. Think of the Internet as a distributed, massively parallel supercomputer connecting not only microprocessors but people, information repositories, sensors, intelligent agents, and mobile code.

There are many examples of Internet users leveraging cyberspace, some of which will be described in greater detail below. Six hundred volunteers from five continents used 1600 computers to factor RSA-129 in eight months, a mathematical feat that was projected to take 40 quadrillion years. Users of multi-user dungeons (MUDs) collaborate to build elaborate text-based virtual reality environments. Archives of scientific e-prints and popular indexes of the World Wide Web (WWW) such as Yahoo are maintained primarily through the submissions of thousands of individuals and organizations. Using a bulletin board on Niftyserve (a commercial on-line service), Japanese users of the HP-100 collaborated to create a public domain library of Japanese fonts. The WWW was used to organize NetDay, a grassroots effort by more than 20,000 volunteers to deploy internal local area networks in thousands of California schools. The collective wisdom of many participants in a given newsgroup often distilled into answers to "frequently asked questions," or FAQs.

Leaders in both government and business recognize the importance of this phenomenon. Vice President Gore recently gave a speech in which he argued that "distributed intelligence" and "massive parallelism" are powerful metaphors for our age that help explain why democracy has triumphed over totalitarian and authoritarian regimes, and why "American businesses are pushing power, responsibility, and information away from the center [3]." John Gage, chief science officer of Sun Microsytems, says that companies should try to "harness the collective brainpower of the Net in the service of your product. The old idea was that the only people who could help you invent new things were people inside your company. On the Net, you can invoke the talents of people worldwide, 24 hours a day, who are doing it out of love — doing it just to do it" [4]. Similarly, in his analysis of "business ecosystems," James Moore concludes that major information technology companies no longer succeed or fail as single entities. "Central firms are joined by hundreds or even thousands of other


0163-6804/96/$05.00 

organizations, such as VARs [value-added resellers] or ISVs [independent software vendors] working to bring solutions of total value to customers. The end customer wants to sign up with the most robust community" [5]. Economists such as Hayek have long recognized the power of markets to harness distributed knowledge. As he explains, "every individual has some advantage over all others because he possesses unique information of which beneficial use might be made only if the decisions depending on it are left to him" [6].

> *Think of the Internet as a distributed, massively parallel supercomputer that is connecting not only microprocessors but people, information repositories, sensors, intelligent agents, and mobile code.*

Of course, not all tasks can be completed more rapidly by having more people or organizations work on them. This is certainly the case for large, complex software projects. As Fred Brooks has concluded, "The man-month as a unit for measuring the size of a job is a dangerous and deceptive myth. Men and months are interchangeable commodities only when a task can be partitioned among many workers with no communication among them" [7]. If a task cannot be partitioned because of sequential constraints, or if there are complex interrelationships between parts of the task, assigning more people to a task may not reduce the time required to complete it, and may even increase it. Writing a speech is a good example of an activity that can take longer if more people are involved. For other kinds of tasks, the tradeoffs are more complicated. Making a decision can certainly take longer if more people participate, but the decision may be of higher quality and more legitimate.

## Examples of Leveraging Cyberspace

The Seven Wonders of the Ancient World were massive construction projects like the Great Pyramid, built with 2 million blocks of stone, each weighing several tons. What future wonders of the "Wired World" might evolve from the voluntary association of people who work and play on the Net? By looking at some actual and proposed examples of cyberspace leveraging, we may be able to discern which types of problems can be solved by large numbers of networked users.

### Auto-FAQ

In a paper at the Second International WWW Conference, GTE researcher Steven Whitehead observed that the development of intelligent systems is usually constrained by a knowledge acquisition bottleneck, and that it "typically involves a small number of highly skilled knowledge engineers, is expensive and does not scale." To overcome this bottleneck, he proposes cyberspace leveraged intelligent network agents (CYLINAs). CYLINAs require:

- A widely accessible medium for information exchange, such as the WWW
- A motivated user community that is willing to contribute expertise and time
- An application that allows nonprogrammers to incrementally contribute and maintain knowledge

Whitehead has developed a prototype CYLINA that answers questions, called Auto-FAQ. Users pose questions to the Auto-FAQ system, which maintains a database of question-answer pairs. When a user asks a question that Auto-FAQ cannot answer, this question can be added to a Gap List, which is made available to a population of domain experts who can provide the information. Depending on the application, this population of experts could be a single guru, the staff of a customer service center, or the global Internet community. Users are also given an opportunity to rate the usefulness of the answers. This user feedback is used to filter information which is incorrect, not useful, or out-of-date. In addition, Auto-FAQ is more likely to retrieve records (question-answer pairs) which have high utility scores. In short, Auto-FAQ uses "cyberspace leveraging" to identify the gaps in its database of question-answer pairs, to allow experts to contribute answers to these gaps, and to gather feedback from its users on the utility of its information.

### Cracking RSA-129

In August 1977, three MIT professors — Ronald L. Rivest, Adi Shamir, and Leonard M. Adelman — offered a $100 award to anyone who could decode a cipher they printed in a *Scientific American* column on mathematical games. Because of the strength of the algorithm they used, they estimated that it would take 40 quadrillion years to decode the cipher, which used a 129-digit number as their "public modulus." In April 1994, a team of four computer scientists announced that they had cracked it in only eight months. By using a mathematical attack called a "multiple polynomial quadratic sieve," the team was able to split this job, which required 100 quadrillion computer tasks, into smaller tasks. By sending messages to relevant mailing lists and newsgroups, the team recruited 600 volunteers from five continents, who collectively donated time on 1600 computers [8]. In addition to donating computer time on nights and weekends, volunteers also ported the factoring code to a variety of machines, which ran on computers that ranged from a 386SX PC to a Cray C90. Participants in a project called FAFNER (Factoring via Network-Enabled Recursion) are already working on cracking RSA-130, and intend to work on even larger numbers in the future to evaluate the security of public key cryptosystems such as RSA.

### Creating and Organizing a Global Information Space

The volume of information on the Internet is exploding. Finding a specific piece of information, however, can be frustrating and time-consuming. Much of the information on the Internet may be out of date or of uncertain quality. Some people have observed that the Internet is like a library without a card catalogue.

One example of a project designed to remedy this problem is the World Wide Web Virtual Library. This project was started by Tim Berners-Lee, creator of the WWW, in 1991. In 1993, the WWW project (then based at CERN, the European Laboratory for Particle Physics) began to delegate responsibility for creating and maintaining Web-based subject catalogues. As of 1996, there are over 200 virtual libraries maintained by volunteers from around the world, on subjects from aboriginal studies to zoos. Each of these virtual libraries may have hypertext links to other virtual libraries on progressively more narrowly defined subjects. This system allows the task of creating and organizing a global information space covering all subjects to be divided into more manageable pieces. For example:

- The top-level virtual library on engineering points to 23 domain-specific engineering virtual libraries (e.g., civil, electrical, mechanical).

- The Mechanical Engineering Virtual Library points to WWW servers maintained by hundreds of university departments, vendors, institutes, and research initiatives, such as the Microelectromechanical Systems (MEMS) Clearinghouse.
- The MEMS Clearinghouse maintains an archive which contains announcements of job openings, calls for papers, bibliographies, dissertation abstracts, newsletters, user contributed files, software tools, descriptions of fabrication facilities, and software tools, and points to the MEMS-related servers maintained by vendors, research initiatives, and professional societies.

*The WWW site created by NetDay organizers included a home page for all 13,000 California public and private schools. Volunteers were able to sign up on-line, indicate their level of expertise and commitment, and see who else had agreed to help out.*

In some academic disciplines, the Internet has become the primary means by which scientists communicate their research findings. The Los Alamos "e-Print Archive," started by theoretical physicist Paul Ginsparg in August 1991, now covers 17 disciplines in physics. It is used by more than 35,000 researchers around the world to distribute and archive scientific preprints, and handled 13,000 submissions in 1995. The archive allows research results to be disseminated instantaneously, eliminating the delays of months or years associated with traditional journals. As Ginsparg notes, it has also instituted a "form of democracy in research both by circumventing the boat-mail gap between continents ... and by granting access to new results without discrimination to beginning graduate students and to seasoned operators [9]." The e-Print Archive has a very low centralized overhead. It successfully leverages the "small efforts of the many" by requiring scientists to submit their own preprints. Scientists are generally willing to make their research results freely available because they are interested in communicating their findings to their peers and establishing professional reputations, not in earning royalties.

Yahoo is another high-profile effort to organize the WWW. Started in 1994 as a volunteer effort by Stanford graduate students Jerry Yang and David Filo, this high-tech start-up lists more than 200,000 Web sites under 20,000 different categories, and is used by almost 800,000 people per day. Although Yahoo! is an outgrowth of Yang's and Filo's "hotlists," the majority of the uniform resource locators (URLs) for new WWW sites are submitted from people who want their site listed. In this case, leveraging cyberspace by relying on self-submission from thousands of people each day is necessary but not sufficient. Yahoo would not be as useful if it relied solely on self-submission. Much of Yahoo's value comes from its carefully developed taxonomy of subjects, and the fact that the company has hired 20 people to categorize new sites in a consistent fashion.

### NetDay

Networks can also be used to organize and leverage the small efforts of the many in the "real world," as opposed to cyberspace. For example, on March 9, 1996, President Clinton, Vice President Gore, and more than 20,000 volunteers participated in "NetDay." NetDay was a grassroots effort to install internal wiring in thousands of schools throughout California that used the Internet in creative ways to mobilize volunteers. The President helped to kickstart NetDay, first proposed by John Gage (Sun Microsystems) and Michael Kaufman (KQED), by bringing together high-tech industry leaders for a meeting on educational technology in California schools on September 21, 1995.

Hundreds of companies sponsored individual schools by paying for the NetDay wiring kits, and helping to design and test the networks. Several companies provided free authoring tools, WWW server software, and Internet access to every school in California. Organizers developed workshops, instructional videos, and online manuals to teach thousands of parents and other volunteers how to wire schools.

The WWW site created by NetDay organizers included a home page for all 13,000 California public and private schools. Volunteers were able to sign up on-line, indicate their level of expertise and commitment, and see who else had agreed to help out. An on-line clickable map allowed Californians to "zoom in" to the street level; each school was represented by a dot that was color-coded based on the number of volunteers who had signed up. This has allowed people with a shared interest in upgrading the technology in their local school to discover each other in a self-organizing, bottom-up way.

### Global Cooperative Computing

A group of researchers at the Massachusetts Institute of Technology (MIT) Artificial Intelligence Laboratory have proposed a framework for global cooperative computing to support software evolution, including "development, distribution, execution, feedback gathering, profiling, optimization and patching" [10]. Their motivation for creating this framework is that while "fundamental computing substrate has been dramatically altered to include pervasive computer networks, a large, public programming expertise, and the World Wide Web," systems for software development have not changed much over the last decade.

Although they note that the Internet has already been used effectively for software development such as Linux and the GNU project, they believe that continued progress will require mechanisms and tools for:

- Indexing and documenting code for public reuse
- Cross-organization collaboration on code development
- Acquiring and managing user annotations on code
- Quick, easy software upgrades and bug fixes

### GLOBE

Global Learning and Observation to Benefit the Environment (GLOBE) is an international environmental education and science partnership conceived of and promoted by Vice President Gore. The program is designed to heighten environmental awareness, increase scientific understanding of the Earth, and promote higher student achievement in science and math. Students conduct measurements and observations, such as atmospheric temperature, precipitation, soil moisture, species identification, and surface water Ph, and share the data with scientists and other students. The data is combined using visualization techniques. Currently 2700 schools in 32 countries have registered to participate, and 950 schools have already submitted 156,000 data reports. GLOBE demonstrates the potential of the Net to help parallelize the collection and observation of local data, while still allowing participants to understand the "big picture."

#### Collaborative Information Filtering

Researchers at the MIT Media Lab have created algorithms to deal with information overload that make personalized recommendations by automating "word of mouth [11]." Ringo, for example, was a social information filtering system that made personalized music recommendations. Users of Ringo developed profiles by rating artists and albums on a scale of 1 (pass the earplugs) to 7 (One of my FAVORITE few!). Ringo used these user profiles to generate advice to individual users. It compared user profiles to determine which users have similar tastes in music. By computing an average of all of the ratings given to an album by users with similar tastes, Ringo was able to predict the likelihood that an individual user would like or dislike a given album. Users were also able to review short reviews written by other Ringo users with shared tastes, and to add albums and artists to the database.

This technology is now being commercialized by Agents Inc., whose on-line "Firefly" service asks potential customers to:

> "Visualize the universe of people you know. Reduce that to the handful of those people whose tastes and opinions you trust. These are the people who you rely upon today for recommendations. Now imagine that you could have your own personal and intelligent agent whose task it is to find more people across the globe — who also share your tastes and opinions — and not only learn from their experiences (by communicating with their agents) but introduce you to these people."

## Understanding Cyberspace Leveraging

These are only a few of the existing or proposed examples of cyberspace leveraging. As the number of users, bandwidth, networked computational resources, and sophistication of software (e.g., intelligent agents) all increase, these collaborative projects will become even more complex. Artificial life expert Tom Ray wants to create a "very large, complex and inter-connected region of cyberspace that will be inoculated with digital organisms"; organisms that will increase in diversity and complexity through natural selection [12]. David Gelertner envisions "Mirror Worlds," software models representing some piece of reality (a hospital, a corporation, a city) fed by so much data that "the model can mimic the reality's every move, moment-by-moment [13]."

If distributed, cooperative problem-solving/resource creation is likely to increase in importance over time, we should try to understand what factors will determine whether these efforts will succeed or fail. Based on the case studies, below are some initial thoughts on the important success factors.

*If distributed, cooperative problem-solving/resource creation is likely to increase in importance over time, we should try to understand what factors will determine whether these efforts will succeed or fail.*

***Can Dependencies between Parts of the Task Be Eliminated or Managed?*** If this cannot be done, recruiting more people to work on a project won't help and may even be counterproductive.

***What Will Motivate People to Participate?*** Currently, the "gift economy" is alive and well on the Internet. Many people will give away resources (papers, indexes, computer programs) that may have taken them weeks to develop. Sometimes this motivation is easy to understand. As a scientist, I want my peers to read my work. If I have already developed an index for my own use, the costs associated with sharing it with the world are very low. If I am a programmer at the beginning of my career, I may want the notoriety that comes with authoring a cool Java applet. I may derive some "psychic income" from being part of a larger community working toward a common end. The "opportunity cost" associated with giving away compute cycles at night or during the weekend to factor RSA-129 may be very low. It will be interesting to see whether the introduction of technologies such as "micropayments" will create new markets, and encourage more people and firms to provide intellectual property or services on a commercial basis.

***Is There Part of the Task that Must Be Centrally Administered?*** Most of the examples have at least a portion of the task that must be centrally administered. Yahoo would not work if everyone classified their own Web sites. The Los Alamos e-Print Archive is easier to search and less likely to suffer from URLs that no longer point to anything because all of the papers are centrally archived.

***Does the Initiative Demonstrate Increasing Returns?*** Some of these efforts demonstrate virtuous circles. As more people use Yahoo, the incentive to register one's site with Yahoo increases, which in turn attracts more visitors to Yahoo. Similarly, collaborative filtering schemes increase in accuracy with the number of users. This suggests that reaching a certain "critical mass" may be important.

## Policy Implications of Leveraging Cyberspace

The ability of networks to leverage the small efforts of the many has a number of important policy implications. First of all, governments need to pay attention to this phenomenon because they should always be searching for opportunities to maximize the impact of taxpayer dollars, especially in a period of tight budgets and fiscal austerity. Some federal research and development programs, for example, require that the private-sector participants pay for at least 50 percent of the cost of the project. The Net may create opportunities for leverage that go far beyond one-to-one, and policy makers should look for opportunities to take advantage of it. For example, there is a good deal of political support for shifting responsibility and authority for programs (e.g., welfare reform, training, community development) from the federal government to state and local governments. Policymakers should develop mechanisms (clearinghouses, centers of excellence) to harness the expertise that exists at the local level.

Second, many of the applications of the NII that the Administration is interested in promoting rely heavily on networked information and services. Examples include consumer health information, government information, interactive modules for life-long learning, digital libraries, framework data for geographical information systems, and networked databases for scientific collaboratories. As this article has shown, creating

> *Many professionals with a wide range of backgrounds could each devote several hours each week to answer questions posed by students on the Internet and, in the process, develop a useful database of frequently asked questions.*

and organizing a global information space is a task that can be parallelized. Individuals publish papers or other forms of networked information on the Net. Others develop indexes, meta-indexes, self-publishing archives, search engines, annotations or evaluations, and other mechanisms for adding value to the ocean of raw data and information.

As the government works with academia, the private sector, and other stakeholders to promote these NII applications, it should look for ways to leverage cyberspace. For example, the National Science Foundation has funded an effort to build a Virtual Department of Geography [14]. Over 30 geography professors will work over the next three years to develop peer-reviewed interactive course modules that cover the entire discipline. As part of President Clinton's initiative to promote the use of technology in our nation's schools, the Administration would like to see similar initiatives launched for other disciplines. The Administration is also interested in promoting "virtual office hours." Many professionals with a wide range of backgrounds could each devote several hours each week to answering questions posed by students on the Internet and, in the process, develop a useful database of FAQs.

Finally, if cyberspace leveraging is important, understanding and advancing its technical and social underpinnings are important. This suggests the need for continued investment in research in areas such as distributed computing, distributed AI, computer-supported collaborative work, coordination science, and knowledge-sharing.

## References


[1] S. Steinberg, "Seek and Ye Shall Find (Maybe)," *Wired Mag.*, vol. 4, no. 5, May, 1996, p. 113.

[2] S. Whitehead, "Auto-FAQ: An Experiment in Population Leveraging," *Proc. 2nd Int'l. WWW Conf.*, 1994, pp. 25–38.

[3] A. Gore, "The Technology Challenge: How Can America Spark Private Innovation?" Feb. 14, 1996.

[4] R. Rapaport, "The Network Is the Company," *Fast Company*, Apr.-May 1996, p. 120.

[5] J. Moore, "The Advent of Business Ecosystems, " *Upside*, Dec. 1995.

[6] F. A. Hayek, *The Use of Knowledge in Society*, Menlo Park, CA: Institute for Humane Studies, 1977. p. 8.

[7] F. Brooks, *The Mythical Man-Month: Essays on Software Engineering*, Reading, MA: Addison-Wesley, 1982, p. 16.

[8] K. Leutwyler, "Superhack," *Scientific American*, July 1994, pp. 17–20.

[9] P. Ginsparg, "Proposal to the NSF for Support of E-Print Archives," Feb. 24, 1994, p. 4.

[10] I. Eslick *et al.*, "Global Cooperative Computing: A White Paper."

[11] U. Shardanand and P. Maes, "Social Information Filetering: Algorithms for Automating Word of Mouth," *Human Factors in Comp. Sys., CHI '95 Conf. Proc.*, ACM SIGCHI, pp. 210–17.

[12] T. S. Ray, "A Proposal to Create a Network-Wide Biodiversity Reserve for Digital Organisms," ATR Tech. Rep. TR-H-133, 1995.

[13] D. Gelertner, *Mirror Worlds*, New York: Oxford University Press, 1992, p. 3.

[14] K. Foote, "Building and Linking Geography Curricula Nationally Using the Internet and Worldwide Web," June 5, 1995, available at http://www.utexas.edu/depts/grg/virtdept/about/proposals/nsf95/cover1.html.


## Biography


THOMAS A. KALIL [M] is currently a senior director to the National Economic Council with responsibility for science and technology issues. The NEC is a White House organization created by President Clinton to coordinate economic policy. In addition to his role in shaping the Administration's national information infrastructure agenda, he is also the U.S. National Coordinator for the G-7 Global Information Society pilot projects. Tome served as an advisor to the Clinton-Gore campaign on technology and competitiveness issues, and helped organize the Little Rock Economic Summit. Prior to that, he was a trade specialist at the Washington offices of Dewey Ballantine, where he represented the Semiconductor Industry Association on U.S.-Japan trade issues and technology policy. He received a B.A. in political science and international economics from the University of Wisconsin at Madison, and completed graduate work at the Fletcher School of Law and Diplomacy. He is the author of articles on nuclear strategy, U.S.-Japan trade negotiation, U.S.-Japan cooperation in science and technology, and the NII. Tom is a member of the Council on Foreign Relations and the Association for Computing Machinery. He can be reached at Kalil_t@al.eop.gov.