Convergence of telecommunications with computing

Department of Electrical Engineering and Computer Sciences

University of California

Berkeley, California 94720

by
David G. Messerschmitt

Invited paper in special issue on “Impact of Information Technology”, *Technology in Society*, Elsevier Science, Ltd., to appear.

ABSTRACT

Both computing and telecommunications are experiencing a technological revolution. The networked computer is transcending its traditional focus on numerical computations and information organization and retrieval by serving as a communications platform. Telecommunications is transcending its traditional focus on telephony by leveraging the computer’s ability to manage other media, like video, graphics, animation, and data. The two technologies (as well as their respective industries) are converging. This convergence gives considerable new capabilities to the individual as well as society, and the universe of networked computers will serve as an important telecommunications infrastructure for the future. However, this convergence also raises significant issues, such as information and communication overload, authentication, historical archiving, and information ownership, which we also outline.

1.0 INTRODUCTION

The term *telecommunications* is derived from “tele”, meaning at a distance, and “communications”, meaning exchanging of information. The dictionary definition of telecommunications is “communication at a distance (as by telephone)” [1], and the term is most commonly applied to the telephone. At its origin, the *computer* was envisioned as a machine to perform massive numerical calculations. Indeed, this is the origin of the term “computer”, as “something that can compute”. Later, with the development of large peripheral storage devices, the computer became a repository of large amounts of data that could be modified, manipulated, and queried. This is reflected in the current dictionary definition of the computer, as “a programmable electronic device that can store, retrieve, and process data” [1]. Recently, a new category of emerging telecommunications applications offers other media in addition to voice (including data, video, animation, graphics, etc.), incorporates the computer as an interface to these media, and renders these dictionary definitions of telecommunications and computers archaic. For the larger populace, and in terms of societal impact, this is an important class of applications.

The impact of the computer for computations or data manipulation is largely as a productivity-enhancement tool that greatly extends the reach of an individual. However, we would argue that this is primarily a quantitative difference; that is, the desktop computer used in this way has not really changed *what* we do, but rather *how* we do it. We argue below that leveraging the computer as a telecommunications tool allows people to do entirely new things. Similarly, incorporating the power of the computer into telecommunications expands the medium to incorporate graphics, animation, video, text, and data. Since people naturally want to communicate by means other than just voice, these developments make telecommunications more useful for people interacting at a distance.

Communications is at the heart of what constitutes a society and a civilization, and any deep-seated change in our ability to communicate affects not only the individual, but also the essence of what constitutes our society. In earlier ages, the printing press and the telephone each had a dramatic impact on society, as did the automobile and the airplane for similar reasons. The emergence of the computer as a telecommunications tool will have an equally great impact.

2.0 THE TECHNOLOGY

It is useful to briefly discuss the technological background, including some of the most distinguishing capabilities of a networked computer infrastructure.

2.1 A common representation

The physical world has many ways of representing information, including for example a molecular structure (as in DNA), the swinging of a pendulum (in a clock), or as a pressure wave in a gas (as in human speech). The modern computer and telecommunications world have agreed on a single elemental representation of all information: the *bit*. A bit is an abstract atom of information that can assume one of two possible states: zero or one. A collection of bits can represent an arbitrarily large number of states. When he developed the “information theory” in the 1940’s, Claude Shannon developed the notion that *any* information can be represented as a collection of bits, including text, data, music, speech, pictures, and video. What is remarkable is that these entities do not have a common representation in the *physical* world. Text is composed of icons on a page, music and speech as pressure waves in a gas, and pictures and video as light patterns. Yet, technologists have arrived at a common representation.

Following Shannon’s lead, all information in computing systems and modern communications systems adopt a “collection of bits” as the common representation of all information whether it be the numbers in a calculation, the speech in a telephone conversation, or a television picture. A communication and storage infrastructure that carries or stores collections of bits becomes readily applicable to any form of information. This is not to say the collection of bits is a *lingua franca* for communications in the technological world, any more than the human’s use of sound as a representation for their thoughts allows them to communicate without a common language. One must have a knowledge of the syntax and semantic content of the collection of bits to make use of them. An audio and a video signal look superficially similar when represented by bits, but their interpretation is quite distinct.

2.2 A distributed platform

The modern computer incorporates many inventions, but arguably the most significant among them is the idea of cleanly separating physical realization from functionality. This is accomplished through *programmability*: the application (intended use) is defined in software, which is a sequence of instructions represented, naturally, as a collection of bits. The computer executes this series of instructions. The functionality is thus embodied in the software at the time it is used, and not in the hardware at the time of manufacture. This separation of function from physical realization is not unique to the computer world; for example, a child’s profession is not defined at the time the child is born. However, the computer has great flexibility in that its functionality can be changed *quickly* and *easily* by software substitution.

Computers were once expensive, and thus few in number and centralized. The advent of the desktop computer in the early 1980’s was significant because it freed the user from the bureaucracy of the centralized computer center, allowing him or her to take direct control. From the perspective of the industry, the desktop computer became a *platform* for applications. By “platform”, we mean the basic physical realization, plus bundled software that performs routine and universal tasks (the “operating system”), that provides a foundation for new software-defined applications. Once there is a large installed base of these platforms (because many users have purchased them), entrepreneurial companies, large or small, can develop new applications and sell them into an economically significant installed base.The barriers to entry for applications development are dramatically reduced, and the computer center bureaucracy is eliminated. Thousands of new companies were started to developed desktop computer applications, and the rate of innovation was expanded dramatically.

The key technical advance in the convergence of computing and telecommunications has been the *computer network*, such as the famous Internet. A computer network enables any computer to transport an arbitrary collection of bits to another computer, exploiting the consistent abstract representation of information used in telecommunications. The implication is that applications are no longer restricted to a single computer, but rather a single application can share two or more computers connected by an appropriate network, exploiting the communications among them. There are many ways to leverage this capability, but for our purposes the most important is that the computer becomes a useful tool for telecommunications among people, rather than merely a tool for processing data.

The telecommunications industry during its century-long existence has relied on specialized terminals such as the “telephone” or “video conferencing set” or “radio” or “television set”. In each case, these are specialized boxes dedicated to a single telecommunications function. The computer industry went through a similar stage of evolution, with dedicated limited-functionality products such as “calculators” and “word processors”. The significance of the networked computer is that it becomes a *programmable* terminal for telecommunications applications, including many of the most familiar applications like telephony and television. It brings to telecommunications many of the same benefits of programmability long enjoyed by computing. As a result, the barriers to entry for new purveyors of telecommunications applications are reduced, and the user will have greater choice. Any applications developer can market directly to users without the intervention of the telecommunications service provider. The networked computer thus offers the telecommunications application similar advantages and capabilities as the desktop computer offered the computing application.

If an application is defined in software, which is represented in the same way as any other information -- a collection of bits -- that application can be transported on demand from one computer to another. A networked computing infrastructure thus represents an important new infrastructure for the distribution of (software defined) products. The barriers to entry for new applications are reduced further, since the functions of the manufacturer, wholesaler, and retailer are eliminated. Reflective of this, there are a remarkable number of public-domain applications for which the developer does not even request remuneration.

Communications networks suffer from a problem economists call “network externality”. That is, the network is useful to an individual only to the extent that there are other entities on the network offering some reason to communicate. Network externality presents a tremendous obstacle to commercial exploitation of a network. Who, for example, will be the first person to buy a new video conferencing product, given that there is no one else with whom to conference? Distribution of applications over the network (in the form of software definitions) will neatly bypass this problem, since a new application can be dynamically loaded into the shared platform consisting of network and programmable terminals, and thus lead to a vastly greater rate of innovation in telecommunications applications.

The collection of mutually networked computers clearly represents an important infrastructure for society, no less significant than our telephone system or our advanced transportation systems.

3.0 NEW CAPABILITIES FOR HUMAN ENDEAVOR

Imagine, for a moment, a human society composed of individuals who cannot communicate. Now suppose that communications between and among individuals is suddenly invented. The individuals can now share information, coordinate their actions, and even participate collectively in enterprises. Communication among previously isolated computers has no less dramatic implications to the technological world. However, our interest here is not on technology, but the impact on society. We thus focus here on fundamental new capabilities provided by the use of the computer as a telecommunications tool.

3.1 Integrated information processing and telecommunications

The development of writing systems allowed people to communicate accurately from one point in time to another, as well as one geographic location to another, by placing thoughts on the semi-permanent paper medium. This allowed people to accurately communicate their knowledge, ideas, and activities to future generations, or to their contemporaries geographically far removed. The printing press led to *mass* communications, or the ability to communicate to many people. Writing and printing incur a substantial delay from creation to receipt of information due to the delay in physically transporting the paper medium. Writing allows two-way communication (the letter), albeit with a considerable delay in receiving a response, due again to physical roundtrip transportation.

Telephony offers complementary strengths, such as instantaneous two-way communications. However, it has limitations, such as the absence of archival capability and the inability to handle other media like graphics, images, and video. More recently some of these capabilities have appeared in limited form in voiceband data modems and facsimile machines.

In contrast, incorporating the computer as a telecommunications tool allows complete flexibility in the medium of communication, including text, images, video, audio, or mixtures thereof. Under the banner of *multimedia,* computers can store, display, process and transport audio and video signals over the network. We can expect computers to routinely have video cameras as well as microphones. When such computers are networked, they can serve as the basis of traditional telecommunications applications like telephony and video conferencing.

More interesting possibilities arise from the integration of distributed information processing and telecommunications into a single networked computing platform. Where two people are collaborating on some task via a video conference, they can also arbitrarily share information, or have shared access to information databases. For example, two or more people collaborating on a document will be able to look at it and jointly edit it while they discuss it. Alternatively, one person can voice- and video-annotate a document, and send it to another person for his or her editing. Or perhaps the document *itself* should be partially represented by voice and video rather than completely by text. Speech recognition (the conversion of speech to text) and voice response (conversion from text to speech) will move from laboratory curiosities to widespread application in human-to-computer interfaces.

Less obvious but no less important, since a computer program is nothing but information represented by a collection of bits, documents can also incorporate programs. That is, a document can have executable as well as passive content. For example, a document may query its viewer as to his or her interests, and tailor its visible representation of content accordingly.

The history of computing, as well as paper-based information processing before it, is one of the users accommodating to the medium. Writing allowed people to communicate across distance or through time, but constrained them to use text and precluded other media such as speech and video. Computers used to similarly constrain users to textual representations. Computers or typewriters could generate paper documents, but the only way to transport them was by the relatively slow post office. In contrast, the networked computer infrastructure should adapt itself to the preferred modality of the user. If a “video document” is preferred to a “text document”, the local computer will allow the user to author it, and the network will allow the user to send it virtually instantaneously to others. If the user prefers to input a text document via speech rather than a keyboard, fine. Overall, the goal should be to make the technology of processing documents and information more accommodating to people.

3.2 Spatial and temporal consistency of information

The networked computer provides an important *consistency* of information in both *time* and *space*. Because the transport of information is quick, even on a global scale, the information available to anyone, anywhere, at any point in time is identical. When information changes, its latest rendition is available to anyone, anywhere, simultaneously. These forms of consistency have a substantial impact on many areas of human endeavor, such as knowledge generation and commerce. As one example, scientists once communicated their latest results by letters -- with perhaps weeks to months of delay -- but can now communicate by telephone or electronic mail with insignificant delay. Since the process of science is largely one of building incrementally on the results of other scientists, the delay in communication alters the rate at which new knowledge is uncovered. As another example, an auction marketplace depends critically on all participants knowing the actions of other participants as a guide for their own action. Networked computing allows such a marketplace to have a global reach.

Strictly speaking, there are some limitations imposed by the speed of light. Certainly the delay due to the speed of light would be a major issue in communicating with extraterrestrial civilizations, but even back here on earth the round-trip delay can be as large as a third of a second. This turns out to be a minor annoyance for interactive applications (like telephony or video conferencing), but could be a significant impediment in a global automated auction marketplace, for example. An exciting workaround to this problem is *intelligent agents*, which are executable programs transported in lieu of messages or queries. To participate in an auction, for example, I might send such an agent to negotiate on my behalf (by being executed in the remote computer), avoiding speed-of-light delays associated with multiple roundtrip transactions.

3.3 Publishing

The old style of publishing based on printing and distributing copies of a work requires a significant investment. This discourages the publication of works with a limited audience, or unconventional works that don't attract the interest of a mainstream publisher. Similar obstacles have faced the artist, musician, or performer in making their works available to a broad audience. In the world of networked computing, publishing requires no incremental investment for the owner of a networked desktop computer. The World-Wide Web today allows virtually any such user to author a work in any electronic medium (text, images, audio, video) and make it immediately available to all networked computers worldwide. In principle, this virtually eliminates publication and distribution as an obstacle to new artists, performers authors, and information providers. This is another example of the reach of technology in eliminating the "middleman".

3.4 Global communities

The dictionary definition of community is "a group of people with a common characteristic or interest living together within a larger society" [1]. Observing what is happening within the Internet, the dictionary should remove the words "living together", if that implies a geographic locality. Increasingly, global communities are forming around special interests, whether they be hobbies, professions, specialties, or politics. While telephone network directories are organized by name and location, in the Internet the directories are largely organized by interest or function. This makes is easy to locate people who share a common interest, regardless of where they reside. Further, it allows any member of the community to make their ideas or wishes available in a form that is available to the entire community, and receive responses back from interested members of the community, similar to the "want ads" of the geographically based community newspaper.

Traditional telecommunications in the form of telephony and television have had a significant impact on political structures, for example making it more difficult to sustain totalitarianism. However, their reach has been limited by restricted access to broadcast communication mediums like radio and television. (Indeed, a primary technique for maintaining totalitarian political structures is precisely such restrictions.) Networked computing cleanly removes this restriction. It has the potential to significantly shift the political organizational paradigm from one based on geographic locality to one based on shared interest or agenda. Clearly this has already happened in global commercial enterprises: successful large companies are today presumed to be global in reach, and their country of origin is less and less identifiable.

4.0 SOME CHALLENGES

Every technology raises problems as well as solutions. Our recourse is to either ban the technology (like weapons) or develop other mitigating technologies (like automobile airbags). It is expected then that the convergence of telecommunications with computing creates its own set of difficulties and undesirable side effects.

4.1 The ownership of information

The market system of commerce is based on the concept of ownership of property. The essence of ownership is the owner's ability to control the use of property. Information (where we include such things as software and audio and video performances) is an increasingly important commodity in world commerce, and yet the privileges of information ownership are difficult to exercise because of the inherent ease with which information can be copied. When ownership becomes ineffective, the market incentives to generate and enhance information are negated.

This problem arose first with the photocopier and the tape recorder, but the accumulation of noise and distortion in the process of copying limits the potential for mischief. When information is represented by a collection of bits, it can be copied indefinitely without degradation. All we have to do is faithfully reproduce the bits. Worse, networked computers allow us to distribute these copies widely at lightening speed. Attempts to address this problem thus far have been *ad hoc* and not very effective [2].

Ownership implies use only with the permission of the owner. There are encryption techniques to insure that this permission is required, and computer networking provides the means for users to request and owners to grant this permission. However, this becomes significantly invasive to the user, and requires a significant infrastructure for the distribution of secret "keys". In a market system, exploiting information also requires payment for its use. "Digital cash" will be available for a payment system, although this also requires an infrastructure.

On the positive side, networking allows "metered" or "usage-based" access to information. The most effective model for the pricing of goods has been based on the utility to the consumer, but this is difficult to implement in older information technologies. Books and software programs are typically fixed priced because of the impossibility of monitoring their use. This discourages the consumer who wants to "try out" a book or program, and gives the consumer who uses the information resource extensively a free ride. With networking, information access can be priced based on actual usage, by coupling that usage with communication to the owner.

Even though we are in the early stages of developing a networked computing infrastructure, some radically new business models are beginning to emerge. In a "give away the razor blades in order to sell the razor" paradigm, it is increasingly common to distribute a World-Wide Web browser or document viewer application over the network for free, in order to sell and derive revenue from a compatible server application. It is becoming clear that networked computing will ultimately have a profound impact on the ownership, distribution, and commerce of information, as well as the computer and telecommunications industries themselves.

4.2 Authentication and trust

The ease with which information can be published over the network raises many issues of trust and authentication. One role of a publisher is to validate the authenticity of the author and work. In contrast, if anyone can publish their information to a large audience, what assurance is there that the information is correct, or the source of the information is who they claim to be? Forgery or misrepresentation is an age-old problem, but becomes a serious threat in a networked environment. There are fortunately partial technological solutions to this problem, enabling for example one to determine that a document was generated by a certain party, at a certain time, and has not been subsequently modified.

4.3 Too much information

Improved telecommunications results in a rapidly increasing rate of new knowledge and information generation and accumulation [2]. Unimpeded publication will result in more information providers, and networked access makes it easier for the user to access the global information store. From an individual's perspective, an information overload will likely be the result.

There are partial technological responses to this problem, but they have so far fallen short. We can imagine the technological analog of the publisher, who selects useful or valid information at the source, or the librarian, who finds useful information on behalf of a user. Today the networked computer can systematically search for desired information, although the results are rarely precise. The challenge here -- and it is a severe one -- is that the system must *understand* the information to be fully effective [4]. It is likely that the role of the (human) publisher and librarian -- increasingly assisted by automation tools -- will be preserved for the foreseeable future, and in fact will become more critical functions. The particulars of their role will be changed, of course. For example, the publisher will maintained networked indexes of useful and authenticated information sources for user. The librarian will cruise the network rather than the card catalog or the stacks for relevant information sources. Libraries will shift from paper-based to electronics-based, and from geographic-based to discipline-based.

4.4 The tyranny of time

The computer industry activity in platform and applications can only be described as chaotic. This presents a serious challenge in the use of electronic storage as a medium for the archiving of information over long periods of time. Reliably storing collections of bits and guaranteeing their retreivability decades from now is surely technically feasible. However, these bits are useless without knowledge of how to interpret them, and this knowledge is likely to be lost over time. Storing programs embedding this associated language along with the information seems like a solution, but it is unlikely that such programs would run on the computer systems and operating systems decades from now. Electronic storage thus appears to be inadequate to the task of archiving information over long periods of time. Before relying on electronic storage for our civilization's historical archives, this problem must be solved.

4.5 Too much communications

The increasing globalization of commerce, largely enabled by modern telecommunications and transportation, dramatically increases the number of people with whom we may have occasion to communicate. Emerging technologies like personal communications are a double-edged sword. While they make communications easier, they also impede personal effectiveness by generating constant interruptions, and in their extreme have the potential to decrease the quality of life.

To understand this phenomenon better, it is useful to examine our own makeup. At the risk of over-simplifying an elegant picture, we can divide behaviors into several categories:

- *Conscious* behavior (response to external environment), which includes all the things that we have to explicitly think about in our day-to-day life.
- *Subconscious* learned (not innate) behavior, such as forming words or reaching for an object, which is initiated consciously but which does not require detailed conscious control.
- *Autonomic* (not learned) basic functioning of the organism, like control of the heart rate and breathing.

Our physiological systems delegate as much routine behavior as possible to the subconscious, while dealing with unexpected situations at the conscious level. Communications is intrusive because almost all communication and interaction with other people, verbal or written, desired or not, is performed at the conscious level. It has the cost of interrupting some other conscious activity.

In this context, it is helpful to divide human communications into two classes:

- *Deferred*, meaning a user is interacting with another user or an information source in a manner in which the delay between when one user originates a communication and another user accesses it is not critical, or even predetermined. Examples include electronic mail and voice mail, where the originator does not have a predetermined expectation of when they will receive a reply.

- *Immediate*, meaning that the users are interacting in a manner that makes the delay in the communications critical. Examples include a telephone call and a video conference, where round trip delays of even a second can be disruptive of the conversation.

Deferred communication is less intrusive, since it gives users to the freedom to chose *when* to communicate. This partially explains the popularity of electronic mail or voice mail. However, deferred communication is less effective for many purposes requiring many back-and-forth interactions, such as a complex negotiation, and thus does not completely displace immediate communication.

Effective individuals prioritize and selectively perform tasks. A similar approach applies to communications. The problem is arising in that an excess of intrusive demands for immediate communications causes people to rely too much on deferred communications. For example, if two people rely on voice mail systems to answer their telephone calls, and only personally originate calls, they will never manage to participate in an actual phone conversation, no matter how hard they try! On the other hand, answering the phone offers no way to prioritize. Everyone hiring a personal assistant to screen communications is not viable for any but the wealthiest individuals.

The effectiveness of immediate communication is undercut to some extent by time zones and the sleep and work habits of people. If the person I want to communicate with halfway around the world is sleeping while I am working, and *vice versa*, immediate communication becomes difficult to accomplish.

With the globalization of commerce, these problems become more serious, and better technological remedies are needed. Technological solutions to the prioritizing of communications is actually considerably more difficult than the screening of information, because the priorities and intent of *all* participants matters. The intelligent agents mentioned earlier may be able to manage the details of prioritizing communications by negotiating with similar agents representing parties desirous of communicating with us. Similarly, we can imagine such agents scheduling immediate communications sessions. The goal should be to make such negotiations unobtrusive, and limit people to the more important task of the communication itself.

5.0 A POSTSCRIPT

Our computers and networks can keep track of vast amounts of data, do computations at inconceivable rates, and communicate information over great distances at inconceivable speeds, and yet are rather unremarkable when it comes to skills like creativity or problem solving or negotiation. The history of technological innovation and industrialization has been to relegate an increasing set of tasks to machines, and simultaneously upgrade the skills and tasks performed by people. People and technology work most effectively as a team, not in isolation, because they complement one another.

To date we have largely viewed telecommunications and computing systems as *tools*. However, as these systems increasingly take on sophisticated tasks previously performed by people, they should be viewed as an integral part of the organization, just like the people they sometimes replace. This is a somewhat different perspective from viewing them merely as tools. Organizational management will be most effective when it views the technological systems and the people as an integral whole. The goal should be to make them work most effectively together, leveraging the unique but complementary strengths of each. Of course, the humanists will resist this view, and will need to be reminded that all the subsystems, technological and human, were created by people for the ultimate benefit of people.

For the future, the managers of our enterprises must have a more sophisticated knowledge of the capabilities of computing and communications technology, with a deeply imbued understanding of its capabilities and strengths, and equally importantly its limitations. Only then can these managers creatively leverage the full potential of both their people and the technology.

6.0 REFERENCES

1. *Merriam-Webster's Collegiate Dictionary*, Tenth Edition, Merriam-Webster, Incorporated, 1995.
2. Robert Lucky, *Silicon Dreams*, New York: St. Martin Press, 1989.
3. Nicholas P. Negroponte, “Products and Services for Computer Networks”, *Scientific American*, Sept. 1991.
4. Lawrence G. Tesler, “Networked Computing in the 1990s”, *Scientific American*, Sept. 1991.
5. Mark Weiser, “The Computer for the 21st Century”, *Scientific American*, Sept. 1991.
6. D.G. Messerschmitt, ``The convergence of communications and computing: What are the implications today?'', submitted to the *Proceedings of the IEEE*, 1996.
7. D.G.Messerschmitt, ``Telecommunications in the 21st Century'', invited paper in the special section on ``Dreams of Future Communications'' of the *Institute of Electronics, Information, and Communication Engineers English Transactions (Japan)*, January 1993.

About the author

David G. Messerschmitt is a Professor in the Department of Electrical Engineering and Computer Sciences at the University of California at Berkeley, and from 1993 to 1996 was Department Chair. Prior to 1977 he was at AT&T Bell Laboratories in Holmdel, N.J. Current research interests include issues overlapping signal processing (especially video and graphics coding) and transport in broadband networks with wireless access, network services and protocols for multimedia, wireless multimedia computing, and the economics of networks. He has served as a consultant to a number of companies, and is a co-founder and member of the Board of Directors of TCSI Corporation (NASDAC). He received a B.S. degree from the University of Colorado, and an M.S. and Ph.D. from the University of Michigan, is a Fellow of the IEEE, a Member of the National Academy of Engineering, and a Member of the Computer Sciences and Telecommunications Board of the National Research Council.