University of Dayton Law Review

Volume 22
Number 3 *Symposium: Copyright Owners' Rights and Users' Privileges on the Internet*

Article 5

3-1-1997

The Anatomy of the Internet Meets the Body of the Law

Andy Johnson-Laird

THE ANATOMY OF THE INTERNET MEETS THE BODY OF THE LAW

Andy Johnson-Laird

TABLE OF CONTENTS

PAGE

I. INTRODUCTION 467

II. THE ANATOMY OF THE INTERNET? 468
- *A. What Is the Internet?* 468
- *B. Transferring Data Between Computers* 469
  - 1. Digitized Data 470
    - a. Ordinary Computer Data 470
    - b. Digitized Analog Data 470
    - c. Why Is Digital Data Special? 471
  - 2. Data Transmission 472
    - a. Two Computers and a Wire 472
    - b. Data Packets 473
    - c. Identifying Each Computer 474
    - d. The Number of the Beast 475
  - 3. Data Packet Routing 475
  - 4. "I Am Not a Number! I Am a Free Man!" 477
- *C. How Big Is the Internet?* 479
- *D. What Does the Internet Do?* 480
  - 1. The World Wide Web 480
    - a. Global Publishing 482
    - b. Web Search Engines 483
    - c. Improving Web Performance by Caching Servers ... 484
  - 2. Electronic Mail 486
  - 3. Electronic Bulletin Boards—USENET 487
  - 4. Exchanging Computer Files—File Transfer Protocol (ftp) .. 490
  - 5. Internet Relay Chat 492
- *E. Who Owns the Internet?* 493

III. THE CHALLENGES OF LEGISLATING THE INTERNET 495
- *A. Jurisdiction* 496
- *B. Identification* 497
- *C. Venue* 498
- *D. Detection* 499
  - 1. Proscribed Information 499
  - 2. Where Might Detection Occur? 499
    - a. At the Sending Computer 499
    - b. At the Receiving Site 500

3. How Will We Recognize It When We See It? 501
E. Other Problems for Legislation 503
1. The Hydra Problem 504
2. The Greased Pig Problem 504
3. The Distributed Storage Problem 504
4. The Security Through Obscurity Problem 506
F. Assessment: Taxation and the Net 508
G. What Might Be Done About It? 508
IV. CONCLUSION 509

THE ANATOMY OF THE INTERNET MEETS THE BODY OF THE LAW[1]

Andy Johnson-Laird

I. INTRODUCTION

The Internet has been called many things: the information superhighway, the national information infrastructure, and other not-quite-right metaphors. Technically speaking, as this article will attempt to reveal, the Internet is "merely" a global copying machine. One is hard pressed to describe anything on the Internet that is the "original" of a work. Everything on the net is a copy of something—even if it is just a local copy of a World Wide Web page on a local Web server.

Furthermore, much of the technology that makes the Internet possible is single-mindedly devoted to make more copies, and send them further, faster sometimes, as with the so-called "caching server," by making temporary local copies in anticipation that the same information will be required in the near future, and thereby avoiding the need to obtain another copy from the original source. The general ethos of those on the Internet is that if you find something on the Internet that you do not have, then you *make* a copy and save it on your computer. Or, if you have something that the Net does not have, then you *put* a seminal copy on the Net so that others may copy it.

Not too much attention is paid, if any, to the issues of who owns the rights to information on the Internet. Compounding the problem is the fact that there is no convenient means for an individual to find out who the rights-holder of a given work might be. Absent this means, it is easier to copy first, and duck the later questions. This article provides the reader with the intellectual wherewithal to create a working mental model of what the Internet is, how it works, and how it is used and abused. Part II describes the "anatomy" of the Internet by detailing what the Internet is, how data is transferred between computers, and explaining the vast size of cyberspace. Part II further explains the numerous capabilities of the Internet, including the World Wide Web and electronic bulletin boards. Part III lays out specific examples of the challenges that must be confronted by intellectual property law as the citizens of the world become "netizens" of cyberspace. This section examines the nature of the Internet in relation to the nature of laws and suggests what might be done to aid the effectiveness of legislation. Part IV concludes that trying to legislate the Internet is a great task that is potentially impossible technically.

---

1. Parts of this paper have previously been published in The CyberSpace Lawyer, September 1996, Volume 1, Number 6. Parts of it have also been presented at the University of Dayton School of Law, Seventh Annual Advanced Computer Law Seminar, "Computer And Cyberspace Law," 1996.

II. THE ANATOMY OF THE INTERNET?

Astute readers will note the question mark that follows the heading above, "THE ANATOMY OF THE INTERNET?" That question mark is necessary because the very title implies that the Internet *has* an anatomy. In fact, it does not. It is a distributed, amorphous, changeling that shimmers at the fringe of our perception. There is a central mass of jelly that purports to be a "backbone" (indeed it is even called that), but when one looks closely, it is not as solid, and certainly not as visible, as one would wish.

A. What Is the Internet?

Imagine two computers connected to each other by a data cable with the necessary software to permit computer data files to be exchanged between them. This is the smallest practical network one can have. Next, imagine looking down on an office building from 300 feet above it. Now, extend that mental model to include fifty computers in the building with the necessary software to allow all fifty to exchange data files (and deal with the problems of more than one computer wanting to "talk" at the same time). This is a so-called Local Area Network (LAN). Extend the model further: increase your altitude to 5,000 feet. Imagine fifty different office buildings within a few city blocks, each with fifty computers on fifty LANs and the wiring necessary to connect the buildings. This is a Wide Area Network (WAN).

Finally, climb up to 15,000 feet and look down on a city with 5,000 office buildings each with fifty computers. Climb higher yet—up to 40,000 feet—and look down on one or more states, with dozens of cities connected by high speed data lines, thousands of office buildings connected by lower speed (but fast) data lines, and (by now) thousands of computers. Make the final leap upwards into space and look back at the entire planet rotating underneath. There are hundreds of countries, thousands of cities, and millions of office buildings and hundreds of millions of computers. All of the computers can, if their human masters wish, send and receive data from all of the other computers. An individual computer can broadcast digitized audio (including real-time telephone calls), still images, full motion video, documents, and computer software to one specific computer on the other side of the planet, or to millions of computers around the planet.

Welcome to the Internet.

The 'Net[2]—as it is known to those whose bodies live in the real world, but whose brains live in the cyberspace—has several mind-boggling character-

2. It is correct with or without the apostrophe.

istics that we have never seen before since we crawled out of the oceans and developed a serious oxygen habit:

(1) It is the largest structure that "we the people"[3] have ever created—largest both in the sense of its physical size and in terms of its impact upon us;

(2) It slices through the physical boundaries of county, state and international borders as though they do not exist. In fact, they do not exist on the Net;

(3) One person can speak to another without any knowledge of where the other person is on the planet, or what their skin color, social status, credit rating, religious beliefs, or political persuasion might be. Identity is optional. One can be anonymous, or can adopt a pseudonym, or can be an impostor—or all of the above from one minute to the next based on whimsical choices.

Access to the Net is often flat-fee based, so it can be perceived as essentially "free" and unlimited once the entrance fee has been paid. This fee arrangement destroys much of what we are used to in pricing models in which we have the notion of "use more, send more, costs more." The Net has created a frictionless, essentially "free" global distribution scheme. For example, high school children, in their bedrooms, can write computer programs and sell them into a global marketplace operating seven days a week, 365 days a year. Artists, authors, consultants, musicians, journalists, plagiarists, and pornographers can reach their clients, consumers, readers, voyeurs, and victims with similar ease. With the notion of military-grade encryption, they can all do it without anyone else knowing, thus creating global publicity, with total privacy. We know not what they do and with whom they do it.

Information is no longer "sent" in the conventional sense that one sends a book from point A to point B, with the book being at A, in transit, or at B. The Net replaces this physical model of moving a particular cluster of atoms from one place to another with the electronic model of prostitution: "You got it. You sell it. You still got it." The original data file is *copied*. The copy is transmitted. The recipient gets the copy. Both original and copy exist. Add to the list of Internet metaphors the notion of the global copying machine. In fact, on the Net one must struggle to define what an "original" work is. That struggle can only get worse.

B. Transferring Data Between Computers

While it is difficult, if not impossible, to embrace the "anatomy" of the Net as a whole, one can conceptually sketch out the extremities. These sketches, coupled with a stimulated imagination, hopefully will allow the creation of a mental model sufficiently accurate to grasp the import of what the

3. Note that it really is "we the people" not "we, the governments of the people," a fact that will continue to grow in significance as time passes.

vox populi, equipped with the personal computer acting as *box populi*, is actually saying.

The Internet "may fairly be regarded as a never-ending worldwide conversation." As with those "Netaphors" that preceded it, to a techie the description is almost, but not quite, accurate, but it does give one a clue as to where the journey to the center of the Net should start: What *do* computers say to each other?

1. Digitized Data

a. "Ordinary" Computer Data

Computer data, as most of us know, consists of just zeros and ones. Here, for example, is a snippet of a data file:

```
0000000000000010000000010000101100000000000000000100000000000000
0000000000000000001000000000000000000000000000000000111011111100
0000000000000000000011010000100000000000000000000100000001000000
0000000000000000000000000000000000000000000000000000000000000000
0010010000010111010001111110111100000000000001000010001000000010
1110010110000001010010011111001100011000000001000010001111001100
0000000000000010000000000000000001001000010101000100100001010011
0010111100000010010011011111100000000000000000000110000100100110
0100111010111001000000000000000001000001001100001001110101111001
0000000000000000001000001010000011011101111110000000000000001100
0010111100111100000000000000000000000000000000001001110101111001
```

What do these zeros and ones mean? They could mean anything. By analogy, what does a red light mean? It all depends on where one finds it: Is it on a stop light or on the instrument panel of an automobile or of a space shuttle? We can ascribe meaning to groups of these zeros and ones suitable to the moment. For one such group, we can call them letters and numbers and they become a bank statement, another group can represent the specific amounts of red, green and blue in the dots that make up a color photograph, or one frame of a film. It is perhaps no accident that life on this planet can be coded into the four nucleotides that make up DNA, and now, with the advent of computers, all of the creations of that life can be represented by just zeros and ones, albeit a vast quantity of them.

b. Digitized Analog Data

Analog information is "real world" information that can be represented on a continuum. A good example of this analog is an electrical clock that represents the passing of time by a continuous motion of its sweep second

hand. In contrast, a digital clock moves in discrete steps, "sampling" time and converting its continuous value into discrete digital values. This conversion is inaccurate because a digital clock is only accurate for an instant. It is only twelve hours fifty-six minutes and two seconds for a single quantum of time, but the digital clock shows this for a full second. This is like using a decimal number with only two digits to the right of the decimal to represent the number B—the true value cannot be represented accurately.

Of course, anticipating such "rounding" or "quantization" errors permits hardware and software engineers to design machines that can digitize analog information with sufficient accuracy that no one can perceive errors.

c. Why Is Digital Data Special?

Once real-world data has been digitized, it takes on characteristics very different from its analog counterpart:

a) A 100% accurate copy of the information can be made as many times as required without any degradation whatsoever;
b) Digital data can be transmitted in the blink of an eye from one computer to another; in fact, as stated above, "transmit" is a misnomer; what actually happens is that a copy of the original data is sent to the receiving computer, which then makes another copy of the data and places it into computer storage. Any time a computer does anything with digital data, it spawns a *copy* of the original data (to the delight of the computer scientist and dismay of the copyright lawyer);
c) Digital data on computer disks can be "erased" by overwriting it with different digital data. Only on magnetic tapes can the data be erased by exposure to strong magnetic fields—but with magnetic tapes, one cannot selectively erase information to remove, say, a specific privileged document. Instead, the entire tape must be erased;
d) At the other end of the spectrum, digitized data can be processed by computer programs in such a way that alterations are undetectable. Numeric and textual data can be altered without trace. Audio and visual data can be processed in subtle ways. Audio signals can be erased and altered in ways beyond the wildest dreams of the late President Nixon and Rosemary Woods. The camera must now be assumed to lie;
e) Digital data can be encrypted with military-grade encryption, putting its contents out of reach of all but those who happen to have a supercomputer or hundreds of high-speed workstations in order to decode the message;
f) Digital data, encrypted or "clear," can by digitally signed such that the signer's identity can be authenticated, and the alteration of even a single binary digit ("bit") anywhere in the document can be

detected in an instant. Such authentication and anti-tampering means are simply not available in the analog world of Xerox copies or type documents;

g) Digital computers can be guided to perform specific tasks by giving them instructions in digital form. A simple number-to-instruction correspondence is defined by the manufacturer of the central processor unit (CPU) in the computer. Programs are created by transforming real-world processes into a series of steps, much like instructing a child how to use a hand-held calculator to compute, say, the net present value of a cash flow. From the computer science point of view, there is no difference whatsoever between the digital data used to represent a computer program and the digital data used to represent the information upon which the computer program will operate; both are just zeros and ones. In fact, the only definition that appears to separate a computer program from the data upon which it operates is that the computer should never be called upon to execute data as instructions; to do so leads to results that are, to use the euphemistic understatement of the computer scientist, "unpredictable."

All of these special characteristics conspire to make the life of an intellectual property attorney particularly interesting—straining almost all of the basic concepts learned from the analog world.

2. Data Transmission

a. Two Computers and a Wire

Digital data can be sent (copied) from one computer to another by the simple expediency of converting each zero or one to a predetermined voltage that is transmitted from the sending machine to the receiving machine in the electrical equivalent to sending a message from one destroyer to another via a signalling lamp. As with comedy, timing is everything, and the computers must synchronize with each other so that the eight zeros or ones that make up a character of text or byte of data can be delimited appropriately. The two computers also need to coordinate their efforts so that the zeros and ones in the data file being sent are appropriately reconstituted back in a bit-for-bit accurate replica of the data file on the receiving machine.

By judicious use of special "control characters," the sending and receiving machines can establish a "dialogue" of electronic conversation. This can best be thought of as a real conversation taking place between two people, with the exception that the utterances shown below can be compressed to a single letter (for example, "Y" means "yes," "N" means "no"):

Sending machine: Are you ready?

Receiving machine: Yes.
Sending machine: "Here is 8,192 binary digits of data..."
Sending machine: "I added them all together as a check, and I got a value of 111,001,110,000,000,100. What did you make the total to be?"
Receiving machine: "I got the same value, so send the next block of information."
Sending machine: "Here is the next 8,192 binary digits..."
Sending machine: "This time the checksum is 101,010,110,111,111,001. Did you get that?"
Receiving machine: "No, I got a different number. Send it again."
Sending machine: "Here is a retransmission of the previous 8,192 binary digits..."
Sending machine: "The checksum is 101,010,110,111,111,001. Did you get that?"
Receiving machine: "I got the same value, so send the next block of information."
...and so on.

This "formal" exchange of information is called a "communications protocol" and, as shown, makes sure that the receiving machine has received a correct copy of the information being sent. If necessary, the receiving machine requests retransmissions when electrical clicks and pops on the telephone line corrupt the data.

b. Data Packets

The problem with sending large files from one computer to another is that all of one's binary eggs are in one basket. If the particular communications circuit between the two machines is momentarily interrupted, and assuming that remedial steps are not taken, the file received may be badly damaged and many retransmissions may be required—or in extreme cases, the entire file may be lost in transit.

In the late '60s, when there was grave concern about this country's ability to withstand a nuclear attack,[4] a new communications protocol technique was devised. Data files, instead of being transmitted as large blocks of characters, would be chopped up into small "data packets," each of which could be transmitted as independent chunks of the original file. The advantage of "packetizing" a file was that the amount of data at risk at any

4. The concern was also due to the giant electromagnetic pulses that accompany a nuclear explosion and disrupt or destroy communications.

moment in time was much smaller, and any retransmissions would thus take less time. Furthermore, each data packet, much like a real postal packet, could be sent from the sending computer to the receiving computer by different routes and these routes could be changed at a moment's notice. Further, redundant electrical circuits could be installed across the country, and packets could be redirected around individual failures, whether the failures were caused by a Russian 100-kiloton air-burst over the Lincoln Memorial or a heartlander's back-hoe slicing through a buried telephone cable.

c. Identifying Each Computer

The preceding discussion has conveniently ducked one critical issue: If there are millions of computers connected to the Internet, how can a data packet be sent from a computer at point A to a specific computer at point B? By what means can the electronic magic of data transmission be told how to send to one machine amongst millions?

The answer is that each computer on the Internet is given a number—in this case, each computer is given a unique numeric "Internet Protocol address." This IP address has some of the characteristics of a postal address and some of the characteristics of a telephone number. However, as a foretaste of things to come, the IP address also differs in fundamental ways from both postal addresses and telephone numbers. Perhaps the most important difference is that if one knows a computer's IP address, a data packet can be sent to it, but the sender has no idea where on the planet the actual receiving computer might be. Nor does the sender have any idea of who is responsible for the receiving computer and its operation. Similar to the telephone, each computer that has an IP address (i.e. telephone number) also has just over 65,000 different "port numbers" associated with the IP address. The port number relates to the IP address in the same way that a telephone extension number relates to the main telephone number used to get to the building that contains the telephone.

Individual computer programs "talk" and "listen" on predetermined port numbers. Therefore, to be certain that a given data packet is received by that program, the sending computer must also specify the recipient computer's IP address and the appropriate port number. A moment's reflection reveals that this is identical to person-to-person contact on the telephone system. If one dials the wrong number or one asks for the wrong extension, the conversation will not begin—one either gets no answer or the wrong person answers.[5]

5. Would be "crackers," wishing to break into a computer system often resort to "attack dialing" IP addresses and port numbers, trying each number in sequence just to see what software might be listening on a given port number. On finding a port number on which a program is indeed listening, a cracker can then try to persuade that program to let them into the computer system. "Cracker" is the technically correct name applied to individuals who illicitly break into computer systems. However, the popular press started calling such persons "hackers" and the name stuck. To veterans of the computer industry, "hacker" is an honorific

d. The Number of the Beast

An IP address has four groups of three digit numbers, but, unlike telephone numbers, leading zeroes are omitted. A typical IP address is 199.201.1.147. The separating periods are required when IP addresses are used with Internet software. Each group of three digits can have a value from zero to 255 (the range of numbers representable by an "octet" or eight zeroes and ones).

Unlike a telephone number, IP addresses are available at no charge. One merely applies to a central registry[6] for an address or a block of addresses to be allocated. Large organizations with thousands of subnetworks and host computers will apply for a so-called Class A block of addresses with the first octet in the range one to 126. For example, IBM has 9.0.0.0 to 9.255.255.255 assigned to it (amongst dozens of other IP addresses). Smaller organizations use a Class B network, with the first octet in the range 128 to 191, such as 128.6.0.0 to 128.6.255.255. For yet smaller organizations, a Class C network is used, with the first octet in the range 192 to 223. For example, 199.2.111.0 to 199.2.111.255. Finally, an individual user can apply for a specific IP address: 196.206.12.5.

The real problem is that IP addresses were handed out willy-nilly, merely for the asking. Compounding the problem, large organizations asked for blocks of numbers, essentially asking for entire "area codes" to be allocated to them, even though they knew they would not need all of the numbers in those areas immediately. The explosive expansion of the Internet has raised the scare that there are not enough Internet Protocol addresses to go around. The "Goldilocks Effect" has driven organizations to avoid Class C networks because they would be too small to handle growth, and Class A are too large, but Class B was deemed to be just right. Fortunately, a new IP addressing scheme is in the works. With a tip-of-the-hat to Star Trek, this new scheme, Internet Protocol Next Generation (IPng), is being phased in gradually to defer the problem for the foreseeable future (or, as was thought when the first IP addressing scheme was devised, for long enough to make it someone else's problem).

3. Data Packet Routing

An electronic data packet has a well-defined format with individual components that correspond, in part, to a real-world postal packet. The most important components are:

---

term applied to a particularly gifted programmer capable of solving difficult technical problems in particularly creative ways.

6. The e-mail address for the central registry is hostmaster@internic.net.

a) The sender's IP address and port number;
b) The recipient's IP address and port number; and
c) The data contents.

Additional data fields, such as a sequence number, allow the data packet to travel, not in convoy with the rest of the packets that make up a particular message, but in complete isolation. The sequence number is used by the recipient to rearrange the incoming packets (which may arrive out of sequence) back into their original order.[7]

Data packets cannot simply be converted into electrical signals, and sent off into the wide blue electrical yonder; they must be passed through a hierarchical series of "routers." A router corresponds to a postal sorting station: the router receives multiple incoming data feeds from individual computers in a LAN, and, based on the destination IP addresses, it forwards copies of the data packets to local or wide area routers who are "closer" (in the electrical sense, not necessarily in the geographical sense) to the destination. Packets destined for machines on the same LAN remain "in-house." Packets destined for "foreign" LANs are sent to the local ISP, either over a dedicated digital telephone line or by converting the digital data into analog tones transmitted over voice-grade telephone lines.

At the ISP, another router examines each data packet and references in-built "routing tables" describing the router's connections to the part of cyberspace for which it is responsible. The router then forwards a copy of the data packet. For example, a data packet from the West Coast, destined for the East Coast, may pass through several routers before it arrives at its destination. From this author's computer in Portland, Oregon, a data packet destined for the White House in Washington DC travels first to the in-house router, down a telephone line to the ISP's router, then to MCI's high speed line to Los Angeles, over to the East Coast, and then to Washington DC, then to the White House. The entire process of transmission and routing, though repeated at each of ten different computers across the country, takes less than a quarter of a second for the entire *round trip*. Europe and Asia take just a fraction of a second longer.

One of the biggest problems with routing is trying to keep every router's routing tables accurately reflecting reality. Each table must reflect to which

---

7. Data packets, whose design was conceived in more innocent times than the present, have no provisions for authentication of the sender's true identity. A computer can "easily" (given a geek-in-residence) send out a data packet from, say, Scappoose, Oregon, that appears to come from the White House (either the American one or the one in Moscow). This particular technological gem is known as "packet spoofing" and has been used in recent attacks on several Internet Service Providers (ISP). The malfeasants simply bombard the target sites with an unremitting stream of 150 or more such data packets per second, each with a different fake sender's IP address. The unfortunate target site is brought to its computational knees, experiencing the electronic equivalent of death by a thousand cuts. There is no defense against this kind of "packet bombing." Further, there are no means for backtracking through the Net to find out from where these data packets are coming or to determine that the data packets really are fake.

computer (by IP address) a given packet is to be sent based on the destination IP address, and, of course, this information is subject to change at a moment's notice depending on hardware or cabling failures.

4. "I Am Not a Number! I Am a Free Man!"[8]

While IP addresses work admirably for computers, they are not particularly memorable to humans. Therefore an optional naming system is used to relate site names to their IP addresses. This so-called Domain Name System (DNS, where "domain" simply means a group of computers) divides Internet users into several groups whose names may give some clue as to the type of organization they represent. Names ending in ".gov" are government organizations; for example, "president@whitehouse.gov." Educational institutions have names ending in ".edu," such as "princeton.edu." Commercial businesses have names ending in ".com," such as "apple.com." Non-profit organizations have names ending in ".org," such as the Electronic Freedom Foundation, "eff.org."

Cutting across these organization classifications are geographical domains which imply the geographical location of their registrants, not the type of organization. For example, "jli.portland.or.us" indicates the name of the organization, "jli," (which is an arbitrary choice made by the registrant); "portland," the city; "or", the state Oregon; and "us," the country United States.

Site name registration is on a first-come, first-served basis. The registering organization, the Internet Information Center (InterNIC), has absolutely no power to reject a requested name other than on the basis that it is already in use, although in recent months there has been an explosion of disputes involving domain names. Unlike trademarks, domain names are globally accessible and not divided into different classes. For example, "mcdonalds.com" and "mcdonalds.portland.or.us" are different domain names and can simultaneously exist with "mcdonalds.com.sg" (Singapore) as far as DNS is concerned. Needless to say, this distresses owners of strong trademarks who see those marks being eroded by anyone with the $100.00 it takes to register a domain name.

Competitors in a given industry segment can attempt a virtual domain name "land grab" by registering numerous variants of domain names that contain the names of competitive companies or products. Many companies have already registered numerous domain names for various dysfunctions of the human body and major diseases. It remains to be seen whether an electronic mail address such as andy@diarrhea.com (Proctor and Gamble), andy@anus.com (American Nihilist Underground Society) or andy@

8. The character played by Patrick McGoohan was known as Number 6 in the TV series *The Prisoner*.

anal.com (All National American Legions) would really be an asset that would stand the test of time. Other organizations have registered domain names based on *every* human organ by proper and street name (including those that this author cannot put in this paper in the interest of good taste). The final contribution to chaos is that if one is denied a given domain name, such as ibm.com (which, as one might expect was snapped up fairly early on), one can still name a specific computer "ibm" and use "ibm.jli.com" with no technological ill-effect.

Once one has registered a domain name, that information is broadcast around the world to various Domain Name Servers. These are computers arranged in a hierarchical structure that act rather like the telephone systems directory enquiry service, converting site names back into IP addresses. The hierarchical arrangement cuts down on the amount of data traffic on the Net. There would otherwise be a huge amount of traffic bottle-necking on centralized name servers, as the name/IP address relationship can be changed at any time. Therefore, to be prudent, the software must assume that the name/IP address has changed, and check it each time.

For example, when this author's computer, t4800 (expressed more correctly as the name/IP address t4800.jli.com) needs to send a data packet to a computer called clarity.princeton.edu (at Princeton University), it must first contact its designated Domain Name Server computer. This begs the question how t4800 knows which computer to contact—the answer is that the IP address of this computer was "hard-wired" into the software on the t4800 when the Internet software was installed.

What follows next is a journey up the hierarchy of the Domain Name Server system. Each computer en route is asked: "Do you know the IP address for clarity.princeton.edu?" If the DNS computer has this information in its cache, or has a specific entry for this in its so-called Name Tables, it sends back the IP address in question. If not, it forwards the request up the hierarchy—again, the knowledge of which machine to contact is built into the tables on each DNS server computer. Sooner or later, a DNS server will be found that has "authoritative" information for clarity.princeton.edu, and the IP address is transmitted back to jli.com and then to t4800.jli.com. As this IP is handed back down the hierarchy, each machine stores it in its cache so that additional enquiries can be handled more efficiently. The Internet software running in t4800, can now create a data packet identified as being from 199.2.111.12, and destined for 128.112.144.1 (the request was answered by the DNS server at princeton.edu, 128.112.128.1). It is worth bearing in mind that DNS is optional—t4800.jli.com could, if requested to, simply slam out the data packet to a given IP address, and could "fake" the sender's IP address to be something else (thus masking the fact that it was being sent from t4800.jli.com). Malfeasance is easy on the Net.

C. How Big Is the Internet?

The short answer is that no one really knows how large the Internet is. Because of its anarchic beginnings, and with no central controlling authority even to this day, there is no mandatory registration of individual users or computers. This lack of registration and authentication of humans and computers is the seed of a major problem in the Internet community. It is also a problem that might have been solved a decade ago, but has little or no hope of being solved now.

Estimating the size of the Internet is fraught with error. The number of *allocated* IP addresses cannot be used to estimate the number of active computers connected to the Internet. Many more addresses have been allocated than have even been registered with hostmaster@internic.net. In fact, so many addresses have been allocated that there is now a serious concern that there are not enough addresses for future years. Some attempts have been made to conduct a census of the Internet using electronic means to send out interrogation signals to each IP address in turn and then looking to see if a computer responds. This so-called "pinging"[9] reveals that less than thirty percent of allocated IP addresses will actually respond. Even this is a bogus number because it fails to consider how many computers or gateways to other networks might not be up and running at the time they were pinged.

Even if the number of active sites were known with some accuracy, it does not lead to any estimate of the number of users. There is no way of knowing how many users share access to a given computer or groups of computers. Furthermore, there is no easy method of determining the number of *different* networks that are themselves connected to the Internet. For example, networks such as CompuServe, America Online, Counsel Connect, and ABANet are all connected. But each of them requires only a single IP address, for they have their own users' addresses within their networks.

All of these factors conspire to ensure that there really is no way of knowing how many different networks, computers, and people are connected, directly or indirectly, to the Internet. The best that can be said is that the Internet is big. Very big. Very, very big. And growing fast. Very fast.

The table below shows some current research performed by Network Wizards.[10] It shows the number of "host" computers, *i.e.* computers that are connected to the Net:[11]

Date	Hosts
Jul 96	12,881,000
Jan 96	9,472,000

9. "Pinging" refers to using the UNIX command *ping*.
10. (visited April 16, 1997) <http://www.nw.com>.
11. (last visited July 12, 1997) <http://www.mit.edu/people/mkgray/net/internet-growth-raw-data.html>.

Jul 95	6,642,000
Jan 95	4,852,000
Jul 94	3,212,000
Jan 94	2,217,000
Jul 93	1,776,000
Jan 93	1,313,000

This kind of growth rate is truly exponential and has created an investor feeding frenzy in recent months. The mere presence of the three letters "N," "E," and "T" in a company name foretells of an initial public offering with a concomitant inability to quote a P/E ratio for lack of any E.

The numbers above are merely the number of host computers connected directly to the Internet and do not include the number of computers connected via local area networks to these hosts. Many organizations have dozens of other computers connected to their host computers. If this rate is correct and continues for the next nine years, it will exceed the present world human population.[12] Many a survey organization is presently trying to find the actual numbers of Net users, but there does not appear to be any particularly reliable means for determining the numbers for the global Net community.

D. What Does the Internet Do?

1. The World Wide Web

While there any many different communications subsystems on the Internet, all dealing with the transmission of digital information from one computer to another (or one to many), of all the systems the most phenomenal growth in recent years has been the World Wide Web (WWW). Conceptually, the Web is based upon the idea of an electronic book in which text, images, and sounds can be stored. Citations to other books, cross-references, and footnotes can be linked to pages within the same book, or to pages of other electronic books anywhere on the planet. When the reader uses the mouse pointer to click upon such a link, he or she is immediately transported to the destination regardless of where it is. Links contained within these destinations link to yet more and more electronic books around the globe. Within five minutes, the reader (better known as a Web-surfer) can sign on in Iowa, and end up reading information from the south of France or Japan, merely by following thematic citations on a particular topic.

Behind the screens, the Web operates on a simple text-based language called Hypertext Markup Language (HTML). For example, here is the

12. Bob Metcalfe, *Counting Users Isn't Easy on the Incredible (Shrinking?!) Internet*, INFOWORLD, Aug. 22, 1994, at 46.

welcoming page for the Copyright Office as it would be seen using a "Web browser" program such as Netscape:

The text that creates this screen image (or more correctly put, causes the Web browser program such as Netscape, running in the user's computer, to create this screen), is shown below.

The text contained in "<" and ">" brackets are the HTML directives that determine, in a manner similar to the embedded codes used by WordPerfect, how the text is to be formatted and the places in the text into which a graphic should be included.

```
<html>
<HEAD>
<TITLE>Welcome from Register of Copyrights Marybeth Peters </TITLE>
<base href="http://lcweb.loc.gov/copyright/mb.html">
</HEAD>
<BODY bgcolor="#ffffff">
<i>U.S. Copyright Office, Library of Congress</i>
<hr size=1 color=xxx>
<p>
<CENTER>
<table width="575">
<tr>
<td>
<blockquote>
<CENTER><img src="http://lcweb.loc.gov/copyright/mb100.gif">
<a href="ftp://ftp.loc.gov/pub/copyright/cpypub/mbpbio.html">Bio</a>
</CENTER>
<CENTER><h2>Welcome from Register of Copyrights<br>
Marybeth Peters</h2></CENTER>
```

```
<p>
Welcome to the Copyright Office.
We in the Copyright Office are proud to be part of a long
tradition of promoting the progress of the arts and...
```

The real power of HTML is illustrated by just one statement above:

```
<img src="http://lcweb.loc.gov/copyright/mb100.gif">
```

This statement contains a link to another computer system and directs the Web browser program on the user's computer as follows:

> This is a citation to an image source (img src). Use the HyperText Transfer Protocol (http), to go to the computer site named "lcweb" in the domain loc.gov (the Library of Congress, a government site), and in the subdirectory "copyright," retrieve the graphic image file (gif), called "mb100" ("mb" could well mean Marybeth, and the 100 may refer to a 100 by 100 dot image).

More to the point, as this Web page is being displayed on the computer screen, its constituent components are being retrieved from different computers. There is no limit imposed on how many such remote sites can be used, nor on the physical locations of those sites. A document in Monterey, California can contain links to a document in London, Paris, Rome, and Moscow.

a. Global Publishing

Using a $59.00 program from Microsoft (or any one of several publicly available programs published on the Internet), *anyone*, be they a ten-year-old child or an ninety-eight-year-old grandmother, *anywhere* in the world, can create a "home" page and publish *anything*. Such a home page (which is in effect, a globally accessible electronic book) can be reached by anyone on the Internet from anywhere on the planet, using the facilities of the World Wide Web.

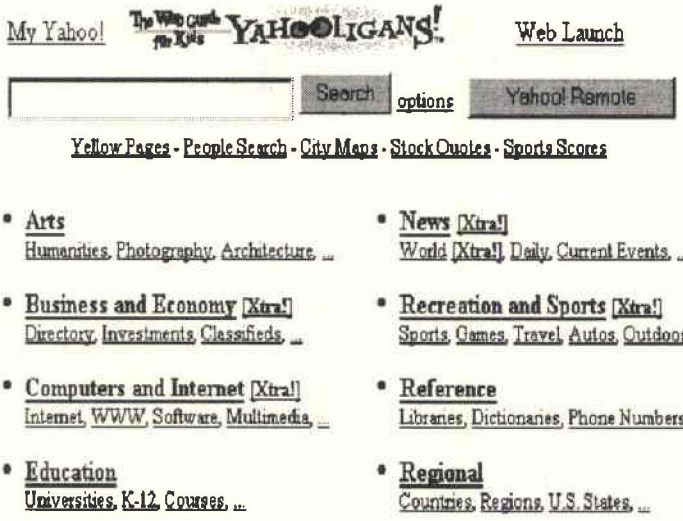
As might be imagined, if anyone can publish anything, they do. Millions of people do. This does not necessarily mean they do it well. Merely giving a computer to someone does not make them an author, a journalist, or an accountant. All it means is that they have the *capability* to do it, regardless of the actual *ability* to do it. Nor does it mean that these new-age publishers pay sufficient attention to the intellectual property rights of others either directly (when they place material on their home pages to which

they have no rights), or when they link (behind the screens) to another computer site where there is misappropriated material.[13]

b. Web Search Engines

Given that anyone can create a Web page and that there are some thirty million computer sites with in excess of 275 million Web pages around the world, how can one find an informational needle in this global electronic haystack? The answer was to create giant electronic indexes of all the publicly accessible Web pages and place these, where else, but on the Web! Two basic techniques are used: topical searches and content searches.

The first of the topical search engines was Yahoo!,[14] started by two Stanford University students and now a publicly traded company. The Yahoo! home page contains a giant hierarchical index to Web sites organized by topic:



The most notable content search mechanism was originally created by Digital Equipment Corporation, and is called Alta Vista.[15] Alta Vista's search engine travels around the Internet, searching for Web pages, and, when it finds them, indexes every word of every page. It then creates a keyword searchable index along the same lines as Lexis/Nexis. Here, for example, is a fragment of the

13. There are also no guarantees that a remote site will remain static or will continue to contain the information that attracted a Web page author to link to it in the first place. It is not even immediately apparent to a Web page author that there may be a link (or thousands of links) to their site.

14. (last visited March 21, 1997) < http://www.yahoo.com/>.

15. (last visited March 21, 1997) <http://www.altavista.digital.com>. The most desirable site name, http://www.altavista.com, was registered by someone else.

first results screen showing the results of searching for "Computer Law Association:"[16]

c. Improving Web Performance by Caching Servers

This huge popularity has, not surprisingly, made the Web suffer from its own success. Alta Vista, for example, is visited some twenty million times each weekday. This success brings with it speed brownouts during peak times.[17] Because the Web is a public place and anyone can broadcast and receive information, everyone does, with the net result (no pun intended) being that the technology cannot move enough data far enough and fast enough.

A valid technical solution is to establish "caches"[18] of the popular Web pages at certain strategic points around the Net, and to dispense copies of frequently accessed pages (such as the Yahoo! home page) from local caches rather than necessitating contact with the "real" home page. This concept of caching is an ancient computing concept, and an even older human concept.[19] Modern computers have several such electronic caches to stave off the need

16. Of particular note, apart from the fact that there are about 400 Web pages around the world that contain "computer law association," is the fact that the search took less than four seconds.

17. In a ghostly echo of the past, we are witnessing the electronic version of the tragedy of the commons.

18. Pronounced "cashes," it is derived from the French, *cacher*: To hide or conceal.

19. The caveman's larder was a form of cache. A supply of previously killed sabre-tooth tiger in a corner of the cave could stave off the need to hunt (and be hunted) until the cache was empty.

to access a slow storage device like a hard disk; the disk itself has internal memory that stores the contents of a block of data read from the disk's surface. Should the computer program request the same block of data again, the disk will respond instantaneously with the information stored in the cache, rather than wait for the disk's platter to rotate to the correct position from which to read the data.

On the World Wide Web, a user's computer keeps a *copy* of the last few Web pages that have been loaded into it. A subsequent request from one of these same Web page results in the local copy being displayed, without the need to contact the original Web server. This has the benefits of improving the perceived response time, reducing the amount of data transferred over the Internet, and the computational workload on the Web server. A variant of this is a strategically placed computer on the Internet that deliberately acts as a substitute server for large numbers of Web pages. For example, one of these so-called "proxy" servers exists in England, and caches Web pages received from the U.S.A. A user in England who requests a Web page from the U.S.A. will, more likely than not, actually receive a copy of the Web page from the English proxy server rather than from the original Web server in the U.S.A. The amount of transatlantic Web traffic is significantly reduced.

No special authorization is required to connect a Web caching computer to the Internet, resulting in an indeterminate, but probably very large, number of proxy servers. A byproduct of a caching is that, at any given instant of time, numerous copies of a given Web page can exist on computers scattered around the globe. This begs the question whether all such caching computers are licensed to make and preserve the copies of all of these Web pages, and, if so, under what legal principle is such copying allowed?

Caching can also create other interesting side-effects. For example, given that a particular Web page is discovered to have badly erroneous information on it, how can all the "downstream" caching servers be provoked into abandoning their local copies of this Web page and obtaining a "fresh" copy from the original Web server? While this may not seem to be a particularly troublesome situation, this author has seen a situation where a particular company's "home" page was altered by computer vandals to include a pornographic image. Users on the East Coast of the U.S.A. saw this pornographic image, but users on the West Coast did not, nor did the company whose home page was altered (the damage apparently having been repaired). This could create some interesting questions of liability for the original company. What you see on the World Wide Web is like star light—it represents what *was* on the web server, not what *is* on it now—with the difference that one cannot tell what the current information is. Schemes are presently being discussed whereby content owners can indicate to caching servers whether or not to cache specific Web pages, and, if cached, for how long cached pages would be valid. However, all such schemes are predicated

on "well-behaved" caching servers that detect and respect these kinds of conditions attached to Web pages.[20]

2. Electronic Mail

Electronic mail (e-mail) is fast becoming the FAX machine of the 90's. Notwithstanding the burdens of literacy, both keyboard and linguistic, that e-mail places on the sender, e-mail does cut through the telephone tag and voice mail systems that rival the great maze of the Los Angeles freeway system. A text message can be created and sent to recipients who are continents and time zones away. Encryption software ensures that only the recipient can decode the message and, having done so, can also authenticate that the author of the message is who she or he represents themselves to be.

Initially, the corporate view was that e-mail is an incredibly liberating tool. E-mail can be prepared at the convenience of the sender's schedule. Once transmitted, it can be received and responded to at the convenience of the recipient's schedule. Perhaps less than obviously, neither the sender nor the recipient need be at the physical location where the other thinks they are. Sender and recipient can access their e-mail account from any place in the world with telephone lines capable of sustaining electronic transmissions; neither need be aware of (nor care about) the other's physical location in the same way that the reader need not be aware of this author's altitude and location when this sentence was written.[21]

In the last few years, though, the initial corporate enthusiasm has been tempered with the realization that people type e-mail far more slowly than they speak and therefore spend a considerable amount of time responding to e-mail. Furthermore, because it is easy to send copies of e-mail, people do. Dozens, hundreds and thousands of redundant "cc" messages flash around the Net (the Internet and the Intranets). Subordinates copy supervisors and managers with e-mail intended more to be tokens demonstrating how hard the subordinate is working, not because the supervisor or manager really needs to be involved. Vacationing managers return from a week out of the office to find 700 messages in their in-boxes, all requiring a few moments even to delete. The messages must, at least, be scanned to ensure that they do not contain important information concerning a liability, wrongful termination, or a harassment suit in the making. Some companies turn off their e-mail systems for two hours a day just to provide their employees with a respite during which they can concentrate on getting "conventional" work done. Electronic mail has not come without a price tag.

---

20. See Professor Post's subsequent paper, "*Bargaining in the Shadow of the Code: File Caching, Copyright, and Contracts Evolving in Cyberspace*," revisiting the topic of caching servers and considering the legal issues surrounding them.

21. This sentence was written at an altitude of 33,000 feet aboard a Boeing 757 traveling between Denver, Colorado and Portland, Oregon.

Despite the negative effects of globe-spanning Internet e-mail, synergies occur that were not possible previously. Groups of people in disparate parts of the world can collaborate as closely as if they were in the same town. Recently some 600 widely separated computer users combined to form a single massive computer system to crack an otherwise intractable problem. Each took 1/600th of the numerical calculation to be performed and ran it on their own computer, e-mailing the results obtained to a central site. Together they performed a calculation that would have otherwise taken hundreds, if not thousands, of *years* to perform. Programmers working in India can develop software for clients in California (and, unfortunately for American programmers, they can do so more cheaply and more cost-effectively than programmers residing in California). Residents of Sri Lanka can trade on the stock exchanges of London, New York, Tokyo, Chicago, and San Francisco. Using distribution lists, copies of the same message, document, or data file can, with equal ease, be broadcast to dozens, or hundreds, of people scattered around the world. Such a message can contain newsflashes, revised price lists, software upgrades, reports of disasters, earthquakes, or human rights violations.

The other side of the coin is that e-mail is a rich source of forensic information in a law suit. Something about the medium makes people cast aside their caution and say things in a message that they would never, ever say if they were face-to-face with the recipient. On the Internet as a whole, it creates the feeling of "speech without consequence, action without liability." In some cases, this is an illusion (if the recipient's identity can be determined), but in most cases it is very real—the sender can remain completely anonymous.

3. Electronic Bulletin Boards—USENET

The Internet, via a system called USENET, takes electronic mail one stage further—it breaks through the implicit barrier created by the need to know a recipient's electronic mail address by providing the electronic equivalent of broadcasting. USENET consists of some special software along with the concept of several thousand specialized "newsgroups." A newsgroup is essentially the equivalent of a radio station broadcasting messages on one particular frequency and being heard by all those who happen to select that frequency. The added twist, however, is that all of the recipients on USENET can, if they choose, broadcast responses right back to all other recipients.

A given recipient of a USENET newsgroup is therefore confronted with a series of messages: the original message broadcast plus numerous responses. By convention, a response includes that part of the original message text that the respondent is referring, making it easy to select any message and get the general thread of the conversation. Special USENET software actually

threads together successive messages and responses by topic to form a coherent chain of communication.

As before, the sheer scale of the Internet creates a magic not previously possible. People around the world with a specific interest, from the conventional to beyond the bizarre, can exchange ideas and viewpoints, or can pose questions and receive answers from those who have already solved the problem. There are presently some 15,000 newsgroups in existence around the world, with new ones being created daily in response to the ebb and flow of the electronic democracy/anarchy on the Internet. Any individual can create a newsgroup, although whether or not anyone else will post messages to it, is uncertain. Old groups or unwanted new groups, like desert puddles, evaporate and dry up quickly for lack of messages.

The contents of these newsgroups fall into several broad classifications: computers, alternate (a euphemism for "anything else"), scientific, miscellaneous, and so on. A random sampling of the newsgroups that exist reveals a complete cross section of human kind. Of particular note in the listing of the alternate ("alt.") groups shown below, are those that are described as "[c]opyright violations...."[22] These are primarily newsgroups that contain pornographic images. The name of each newsgroup is given, followed by its description written by an unknown author. As befits the Internet, many of the descriptions (and indeed the newsgroups) are very much tongue-in-cheek:

alt.3d: Discussions of 3 dimensional imaging.
alt.activism: Activities for activists.
alt.agriculture.misc: General discussions of agriculture, farming, etc.
alt.alien.visitors: Space creatures ate my modem.
alt.bacchus: A newsgroup for the non-profit 'BACCHUS' organization.
alt.backrubs: Lower...to the right...aaaah!
alt.binaries.multimedia: Sound, text and graphics data rolled in one.
alt.binaries.pictures.erotica: Gigabytes of copyright violations.
alt.binaries.pictures.erotica.blondes: Blonde copyright violations.
alt.binaries.pictures.erotica.d: Discussion about erotic pictures.
alt.binaries.pictures.erotica.female: Copyright violations mostly females.
alt.binaries.pictures.erotica.male: Copyright violations mostly males.
alt.binaries.pictures.fine-art.digitized: Art from conventional media.
alt.binaries.pictures.fine-art.graphics: Art created on computers.
alt.binaries.pictures.tasteless: Binary postings of especially tasteless.
alt.binaries.sounds.misc: Digitized audio adventures.
alt.bitterness: No matter what it's for, you know how it'll turn out.
alt.bonsai: For discussion of Bonsai gardening.
alt.cable-tv.re-regulate: Re-regulation of the cable television industry.

22. This list was found by the author on a computer system at CERN, the large nuclear research facility just outside Geneva, Switzerland.

alt.cad: Computer Aided Design.
alt.california: The state and the state of mind.
alt.callahans: Callahan's bar for puns and fellowship.
alt.conspiracy.jfk: Discussion of the JFK assassination.
alt.current-events.bosnia: Discussion of the situation in Bosnia.
alt.current-events.somalia: Discussion of the situation in Somalia.
alt.fan.monty-python: Electronic fan club for those wacky Brits.
alt.fan.noam-chomsky: Noam Chomsky's writings and opinions.
alt.fan.ronald-reagan: Jellybeans and all.
alt.fan.rush-limbaugh: Fans of the conservative activist radio announcer.
alt.games.mk: Struggling in Mortal Kombat!
alt.internet.access.wanted: Information about connecting to the Internet.
alt.internet.services: Information about services available on the Internet.
alt.kill.the.whales: This newsgroup is evidence for the coming apocalypse.
alt.meditation.transcendental: Contemplation of states beyond the teeth.
alt.missing-kids: Locating missing children.
alt.pave.the.earth: One world, one people, one slab of asphalt.
alt.personals: Geek seeks Dweeb. Object: low-level interfacing.
alt.planning.urban: Urban and regional planning concepts.
alt.politics.clinton: Politics of Bill Clinton.
alt.politics.org.cia: Politics of the U.S. Central Intelligence Agency.
alt.politics.usa.constitution: U.S. Constitutional politics.
alt.politics.usa.republican: Discussions of the U.S.A. Republican Party.
alt.sex.bestiality: Why do you think it is called "petting"?
alt.sex.pictures: Gigabytes of copyright violations.
alt.sex.pictures.female: Copyright violations featuring mostly females.
alt.sex.pictures.male: Copyright violations featuring mostly males.
alt.society.foia: Discussion regarding the Freedom of Information Act.
alt.sport.bowling: Discussion of the art and the sport of bowling.
alt.sport.bungee: Like alt.suicide with rubber bands.
alt.support.cancer: Support mechanism for those dealing with cancer.
alt.support.diet: Seeking enlightenment through weight loss.
alt.sys.sun: Technical discussion of Sun Microsystems products.
alt.toon-pics: More copyright violations.
alt.tv.barney.die.die.die: Discussions about the purple dinosaurs demise.
alt.tv.mash: M*A*S*H, the Korean Conflict's 4077th army hospital.
alt.tv.northern-exp: Discussion of the TV series Northern Exposure.

Offsetting the bizarre nature of the "alt." groups are the "sci." (scientific) and "comp." (computer) newsgroups, where the messages are, by and large, between professionals, and the discussion is very technical and focused on the specialty of the newsgroups. As a parenthetical note, it is also interesting to see the relatively prominent place in the USENET traffic taken by talk.politics.guns, alt.atheism, and alt.fan.rush-limbaugh.

USENET depends upon a secondary channel of communication, co-existing in large part with electronic mail. Designated news servers act as both concentration points for outgoing messages and dissemination points for incoming messages. Hundreds (or maybe thousands) of these news servers exchange a relentless stream of messages, attempting to ensure that all recipients receive all outgoing messages. In the time it took to read this sentence, millions of characters of USENET newsgroup information will have been transmitted around the world.

4. Exchanging Computer Files—File Transfer Protocol (ftp)

In addition to acting as switching yards for electronic mail and USENET newsgroup messages, many sites on the Internet act as archives for information, including software, reference documentation, images, sounds, music, and video fragments. Many of these sites—and no one really knows how many exist—are academic institutions such as universities, or government institutions like NASA. Each site usually has megabytes[23] of information available at no charge; cumulatively, there is an enormous amount of information available, and it is growing daily as new material is placed on the archive sites.

This information is accessed using a program called "ftp," which stands for "file transfer protocol." Protocol means simply a convenient convention used between two computers to make sure that users have controlled access to files, and can get copies of files from the archive site onto their local computers, or can place new files onto the archive sites. Many sites offer uncontrolled access, at least to get selected files *from* the site, using a system called "anonymous ftp." By logging into the archive site with a user identifier of "anonymous," anyone can access the available information.

Perhaps the hardest aspects of ftp to comprehend are the ease with which one can roam around the world, visiting one site after another in just a few *seconds*, and the vastness of the amount of information that is available. The act of ftping[24] files from remote sites requires little more than two or three commands.

The following is an example of accessing a site in Krakow, Poland, known by the strange name of *cyf-kr.edu.pl*, to get a particular program, *mars.exe*. Characters typed by this author are underlined with explanation comments added in Helvetica font:

> ftp cyf-kr.edu.pl Request the ftp program to contact the particular site in Poland.

23. A megabyte is 1,000,000 characters of information. This number of characters is equivalent to approximately 33,000 typewritten pages.

24. Pronounced "eff tee peeing"!

After a pause of about 2 seconds, the site in Poland responds and identifies itself:

```
Connected to lajkonik.cyf-kr.edu.pl.
220- > Welcome user at jli.portland.or.us!
220- >
220- >
220- >                This is ftp server at
220- >
220- >       A c a d e m i c   C o m p u t e r   C e n t e r
220- >
220- >
220- >
220- >
220- >
220- >
220- >
220- >
220- >
220- >                    Poland
220- >
220- >
220- > You can log in typing 'anonymous' or 'ftp'.
220- >
220- > lajkonik FTP server (Version wu-2.1c(5) Fri Feb 4 15:06:35 MET
1994) ready.
```

The Polish site requests a log in, using either the user id. of 'anonymous' or 'ftp' (which is easier to type). The user types anonymous:

```
Name (ftp.cyf-kr.edu.pl:andy): anonymous
```

The user then, by convention and as a courtesy (it is not mandatory) enters their full e-mail address as a password—merely to record their 'visit.'

```
331-Guest login ok, send your complete
   e-mail address as password: andy@jli.portland.or.us
230-You are 6th user on your class (max 50).
230-Local time is Sun Sep 4 03:35:05 1994
```

230-Please report problems to yskarock@cyf-kr.edu.pl
230-
230-Guest login ok, access restrictions apply.

ftp> 200 PORT command successful.

Retrieving a file requires the 'get' command.

ftp> get mars.exe
226 Transfer complete.
95700 bytes received in 100.5 seconds (2.62 Kbytes/s)

Once the file has been retrieved, the user logs out of the site in Poland.

ftp> quit
221-Goodbye.

That was from Krakow, Poland; the entire operation took less than three minutes. Perhaps there are more interesting files in New Zealand, Australia, Japan, Korea, the Netherlands, or Switzerland? Jules Verne would have to modify the title to "Around The World In 80 *Seconds!*"

5. Internet Relay Chat

It was not long before netizens grew tired of the delays that even electronic mail introduces. From that realization was born the idea of a real-time method of communication over the Internet—Internet Relay Chat or IRC. IRC is, in effect, a means for numerous people to participate in real-time conversations, typing in messages that are broadcast to all others who are on the same channel with appropriate software mediation to ensure that each participant's typewritten line appears in proper order on all other participants' screens. IRC supports an unlimited number of channels as each channel is given a name, not a number. The effect of IRC can be riveting: During the Gulf War several people in the Gulf were broadcasting real-time reports as events occurred. It was textual CNN. But unlike CNN, "viewers" could interactively pose questions to those in the Gulf.

Of course, traffic jams do occur and it can be particularly daunting to be confronted with 400 questions in rapid succession; however IRC is another inchoate system and such logistical details will be overcome with suitable software. IRC represents just one more step along the road towards a giant, world-wide community, capable of providing an immediate global view and reaction to events. One can only ponder the effects if Rush Limbaugh, Ross Perot, and President Clinton discover that the electronic town hall is already

old hat.[25] That concept has, of course, already long since been considered and talked to death on IRC and USENET.[26]

E. Who Owns the Internet?

We in the western world are used to having the works of man owned by someone, some corporate body, or at the most, "we the people." Confronted, therefore, with the gigantic, if dimly perceived, Internet, it is particularly difficult to believe that no one "owns" it *in toto*. Each of the individual computers that are connected, directly or indirectly, to the Internet, are indeed owned, as are the data communication lines and voice-grade lines that connect them all together.

At the top of the ownership data chain (presuming it to be analogous to a food chain) are the common carriers such as AT&T, Sprint, and MCI. They own the so-called T1, T2, and T3 ultra high-speed data links that form the backbone of the Internet and connect the major parts of the network together in the U.S. using Metropolitan Area Exchanges in San Jose (dubbed MAE-West) and Washington DC (MAE-East). Recent data for these MAE's show they transmit a near-unremitting data stream of 250 million characters *per second* in both directions every hour of every day. The data volume only drops between 3:00 and 4:00 a.m.

Commercial enterprises such as IBM, Digital Equipment Corporation, and major universities throughout the U.S., Europe, Australia, and other significant countries, usually democracies, own the main-frame computers that form the nodes along the backbone. In almost all cases, the universities have been funded, partially or totally, from government agencies—in the U.S., from the National Science Foundation and in England, by government grants. The last few years, however, have seen a gold-rush to cyberspace and there has been a huge infusion of money into anything that provides infrastructure for cyberspace: namely content providers and hardware vendors specializing in modems and routers. The arrival of AT&T in the Internet Service Provider business also threatens to alter the present pricing charged by the smaller ISPs, many of whom will simply get eaten alive in the process. Arguably, if, in some doomsday scenario, all funding for the Internet were to cease overnight, a network of volunteers would spring up to relay the traffic via local voice grade lines. This particular info-genie is out of the bottle and there is no putting him back.

Cynics have observed that if there had been some giant guiding hand that had attempted to create the Internet, the Net simply would not have

25. Barbara Kantrowitz & Debora Rosenberg, *Ready Teddy? You're Online*, NEWSWEEK, Sept. 12, 1994, at 60.

26. Like the WWW, the IRC can be addictive. The newsgroup alt.irc.recovery provides adequate testament to that.

happened—indeed, *could* not have happened. Too many dedicated students and netizens donate their time to create, maintain, and cherish the information available on the Internet and they would not have been permitted to do this in a governmentally or corporately run system.

Furthermore, the Net provides a freedom rarely experienced in other parts of society: practical anonymity, along with its brothers, pseudonymity and impersonation. There is yet no means for determining whether I am me and you are you. This exposes one very unpleasant truth about humankind: If we are sure that no one knows who we are, that no one is watching us, and the risk of getting caught is vanishingly small, many of us have a distinct tendency to behave like boorish, belligerent bigots and scoff-laws. And that is precisely what is happening on the Net. In fact, the Net is nothing more than a mirror, and can bear no more responsibility than our bathroom mirrors when reflecting our early morning ugliness. The Net simply provides the space and the means for us to misbehave, and like the proper English Public Schoolboys in *The Lord Of The Flies*, absent parental authorities, many netizens behave like complete savages, saying things that they would never say, and doing things they would never do, if their identities were known.[27] This concept is certainly not new; Alexis de Tocqueville, writing in 1840, said "The best laws cannot make a constitution work in spite of morals; morals can turn the worst laws to advantage. That is a commonplace truth, but one to which my studies are always bringing me back. It is the central point in my conception. I see it at the end of all my reflections."[28] Stare hard at the Net and we see our own reflections.

The fact that legislating the Net will be hard or presently impossible should not and must not deter lawmakers. Laws must be made even if, *pro tem*, they might be unenforceable. This paper does no more than highlight the present technical problems facing those who would legislate the Net.

That said, it must be emphasized that the Net is not a place; it is a space. It is not a country; it is a planet. It is not owned by someone or something; we all own it. It is not a thing; it is made of people. In fact, the Net is run (if that does not imply too much organization) by dedicated computer technicians, many of whom do what they do, not for profit, but for the doing of it. And as has been observed by other commentators, the Net is self-healing—it views attempts to restrict it as damage and responds with technically ingenious "repairs" to heal its wounds. Any laws enacted that fail to take this into account will be vitiated—the only variables being the amount of time before this occurs and the amount of scorn heaped on the proponents.

---

27. This is not strictly true. Sadly, there are just a few bad people who use the Net to proliferate what they say and do in real life. They can just shout it louder and with wider effect on the Internet.

28. ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 769 (Harper & Row trans., 1st ed. 1966).

III. THE CHALLENGES OF LEGISLATING THE INTERNET

It would be a disservice to the reader, having described in overview the putative anatomy of the Internet, not to point out the major legal problems that the future holds for those practicing computer law. Since the Net is a sufficiently novel phenomenon, it seems appropriate to start from an epistemological point of view. Before discussing how legislation can be applied effectively, one must ponder precisely what one wishes to legislate.[29]

There appear to be two primary areas that can be legislated: the transmission of information and the storage of information. All other legislatable acts are subsumed within these. For example, the Computer Decency Act, almost passed into law in the U.S., legislates the transmission of information, and the Child Pornography laws legislate the mere possession of it. The law of patents, copyright, and trade secrecy applies, at least to electrons on the Net, in the form of transmission of information.

Technology makes both "transmission" and "storage" ambiguous terms. Transmission is presumed to include the act of making information available for someone else to retrieve, even though to a technician there is a world of difference between sending something and reaching out to retrieve it. Consider also the case where a particular digitized image is transmitted from someone in England to someone in Oregon using a CompuServe account.

There are many entities responsible for transmission and storage: the person originating the transmission, the numerous entities in between the originator and the recipient, and the final recipient. The intermediary entities operate a reception, storage, and transmission facility, although the actual dwell time of the information in transit usually is, well, transitory, and the entire information may not actually exist on an intermediary machine at any one moment. However, online services such as CompuServe act more as a post office box, holding the information pending the arrival of the individual and therefore, there is little technical doubt that acts of storage and transmission occur.

Caching Web servers further complicates matters as they store non-transitory copies of information that *was* in transit.[30] The original sender performs acts of storage of a copy of the information in the first place (presuming, of course, that the data is not simply generated on demand), and an act of making a copy that is transmitted. The recipient only stores the information. Fundamentally, therefore, for law to be applied to the Net, it must address the issues of transmission and storage, and these are specifically

29. This author is not an attorney. Therefore, these views are respectfully offered as a forensic software analyst's perception of the law.

30. An additional issue is whether the human operators of these computers were implicitly licensed to make these copies.

the two areas in which the law will find that the Net is a very slippery and hard to grasp fish.

Presently, it appears that law to a very large degree is determined by latitude and longitude. What is illicit one inch south of the 49th parallel (which separates the U.S. from Canada) may or may not be illicit one inch to the north, and *vice versa*. In England, there is no concept of "sweat of the brow" nor "structure, sequence and organization." In Europe, Asia, or anywhere else outside the U.S., the words "First Amendment" do not merit initial capital letters and are greeted with the genuine question: To what? With these thoughts in mind, ponder what the situation might be, if by magic, all of the governments of the world decreed that as of midnight last night, all county, province, canton, region, state, and international borders were abolished and the use of latitude and longitude was a thing of the past.

Further ponder: The magic worked. It happened. The Net has made a world with no borders; latitude and longitude are indiscernible, yet there are still geographically-based laws. That is where we stand today. The Net provides five specific challenges to the would-be legislator and law-enforcer, namely the questions of:

a) Jurisdiction: Whose law applies?
b) Identification: Who are the participants?
c) Venue: Where can the defendants be found in terms of filing litigation?
d) Detection: How can the illicit transmission and storage of information be detected?
e) Assessment: Of what value is the information?[31]

In all of these, the Net excels at taunting us with half-truths, and inept metaphors that worked for the other side of life, the Outernet of the real world.

A. Jurisdiction

To understand the question of jurisdiction, consider this situation: A programmer in France accesses a computer system in Kuwait, via intermediary computers in the U.S., and makes a copy of a computer program, the rights to which are owned by an American software company. The programmer places that software on a computer in Guatemala. Has copyright infringement occurred? If so, in which jurisdiction? Neither Kuwait nor Guatemala appear to be signatories to any of the major copyright conventions such as Berne or the Universal Copyright Convention. The

31. This pertains specifically to taxation of transmission and storage.

passage of the apparently illicit copy of the software through the U.S. was fleeting—perhaps to the point that at any given moment in time no more than 512 characters of information were resident in the U.S. The apparently illicit copy, once made, comes to rest in Guatemala.[32] Even if the hypothetical is brought conceptually and geographically closer to home and the source and destination of the software are presumed to be Mexico and the U.S. respectively—the illicit copy ends up on a computer in the U.S.—the question still stands: Which country's laws would be applied to the French perpetrator?

It is tempting to answer that U.S. law would apply, because that is where the final illicit copy ends up. This leads to an interesting converse: If the situation were reversed and the illicit copy were to end up in Mexico, would Mexican copyright law then apply? And if that is true, then make one final twist: the illicit copy ends up in France; does French law now apply? Does that mean that France's vigorous application of Moral Rights, "Droits des Auteurs," will also apply? If that is the case, it may well lead to some interesting times in the U.S., as well as bruised national egos, when we discover that what we say and do on the Net in the U.S. may be subject to French law merely because copies end up in France.

To those unfamiliar with the Net, these hypotheticals may seem far-fetched. The problem of talking and writing about the Net is that it fails to communicate the immediacy with which one can bounce around the world. It takes approximately 1/3rd of a second (300 milliseconds) to reach to France or Japan from the U.S. Transferring files of considerable size might take several minutes, but small images and programs take a scant few seconds. From *anywhere*. Many is the time that experienced netizens make copies of information with barely a glance to see from which part of the globe information arrives. With the World Wide Web, it is commonplace to find a home page in Japan pointing to pages in England, Italy, and New Zealand, for example.

History has not yet had the opportunity to record the problems of computers connected to the Net from the High Seas or from Outer Space (where there does appear to be global law), but time will remedy this ommission.

B. Identification

At the root of much behavior and misbehavior on the Net is the fact that it is difficult, if not impossible, to establish who someone is and where they are. While some domain names, such as *jli.portland.or.us*, appear to give a

32. The only good news in this scenario is that there are not too many computers in Kuwait or Guatemala on the Internet. However, that is the only aspect of this hypothetical that can be argued as being unrealistic.

hint as to where the site is located, all that is really foretold is that the name was registered in the Portland, Oregon, U.S. part of the *name space*.[33] It says nothing about the location of the computer that responds to that name, nor the identification nor venue of the perpetrator.

First, the computer in question does not actually respond to the name *per se*, but responds only to an Internet Protocol (IP) address. For example, this author's main computer's IP address is 199.2.111.1, and that must be derived by converting the name *jli.portland.or.us* from a domain name using the global Domain Name Service. While the registrant submits information about the domain name, it only need be sufficiently accurate to pass peremptory validation. It would be tempting to infer that, given a particular IP address, one can unambiguously find the computer to which it has been assigned, but this is a fallacy. There is no law that prevents an individual living in an isolated cabin in Montana,[34] having been assigned an IP address, from assigning that IP address to a telephone line that snakes off to New Guinea where it is connected to a computer.

In summary, one cannot identify the precise location of a given computer on the Net. One cannot readily identify in which country a computer is located. Furthermore, one cannot identify a given individual on the Net. One person can have many different *personae*, none of them corresponding to his or her own. Men can be women, women can be men, straights can be gays, gays can be straights, adults can be children, children can be adults—all from one minute to the next. To quote the now famous New Yorker cartoon showing one dog seated at a personal computer keyboard, and speaking to another dog looking up at him from the floor: "On the Internet, no one knows you're a dog." Furthermore, no one knows whether you are a dog, a chien, or a hund or where your kennel is.

C. Venue

Finding the geographical location of the computer system(s) involved does not imply that the perpetrators will be in the same place(s). If they access their computers remotely via dial-up telephone lines, the determination of their precise venue for legal purposes will be quite difficult—they could be anywhere on the planet—and nothing about the Net or computer technology will make this problem any easier to solve.

---

33. "Name space" is a term of art that views all possible names as though they were corralled together into a single bounded space.

34. Where after the alleged Unibomer was apprehended and the Freemen Militia standoff with the FBI occurred, it is asserted that at least the cows are sane.

D. Detection

For any law to be effective on the Net, it must denominate, to some degree of specificity, the nature of the information whose transmission or storage is proscribed. A proximate question is consideration of where on the Net the transmission and storage can be detected.

1. Proscribed Information

In overview, it appears as though the U.S. authorities, in connection with other governments around the world, wish to control the transmission and storage of pornography, materials pertaining to terrorism, extortion, assassination, bomb making, gambling, and other illicit materials. Other more authoritarian states may choose to detect anti-Government or seditious materials. Taxation authorities need to be able to detect the sales and/or purchases of information to be able to levy taxes on such transactions. Customs agencies need to be able to detect the transit of information as it crosses international borders.

The fundamental question that flows from these needs is whether or not, other than by inference, the transmission and storage of this information can be detected. Inference can be used in various ways to detect these materials indirectly: a file name of anarchists.cookbook.txt implies that the file might describe how to make bombs, a file of clinton.jpg, implies a picture of the president. But what does teen106.jpg imply? Or aa1049.jpg? There is only one way to find out and that is to download that file and view it on the screen.

2. Where Might Detection Occur?

a. At the Sending Computer

Clearly, if a particular computer site is noted for the storage of illicit information, then, as a bright spot in this whole Net mess, law enforcement authorities can inspect the site remotely—as any other member of the Net community can.[35] Generally, if inappropriate information is being stored on

35. This occurred when the FBI investigated CompuServe after a complaint from the American Family Association in Tupelo, Mississippi. The complaint alleged that CompuServe had failed to provide adequate safeguards to keep children from viewing sexually explicit digitized images and movies in the area called MacGlamour. (last visited July 12, 1997) <http://www.interactivesports.com/backissu/may281996/fmlegal.htm>.

a computer or the operator of the computer is "transmitting" inappropriate information, and the authorities have appropriate know-how, software tools, time, and the good fortune for the computer to be within their jurisdiction, appropriate measures can be taken to shut the computer down.

b. At the Receiving Site

Given a computer site that is distributing, or making available for distribution, inappropriate information, it would appear relatively simple for law enforcement authorities to monitor which computers transfer information from that site and track backwards to the individual who is "transmitting" the information, and possibly storing it.[36] In fact, this simplicity is illusory and may stem from the overworked and inappropriate model of the Net being a superhighway, along with the notion that one can monitor traffic from on-ramps, off-ramps, or even from overpasses or a helicopter in the sky.

The reality is that one cannot identify the sender or the recipient easily, if at all. Even if one could, there is no guarantee that the geographic location of either is at the location of the domain name registrant and within any given jurisdiction—indeed either the sender or recipient or both could be accessing their respective computers from a continent away. Furthermore, the traffic ostensibly flowing from the sending computer to the receiving computer may neither be starting nor ending its journey at those computers—one or both computers may be running *proxy servers.* Proxy servers merely act like middlemen and receive the data from one computer and pass it on to some other computer. A specialized form of proxy server is the anonymous remailer, which deliberately strips off any identification of the sending computer. These machines, although relatively few in number now, will blossom in popularity around the world if the Net perceives that data monitoring is being performed routinely by law enforcement agencies.

The superhighway metaphor is further confusing because electronic superhighways, the high-speed data lines that form the backbone of the Net, do not travel as logically as geography would make us think. For example, transmitting a data packet from Portland, Oregon, to Beaverton, Oregon (a distance of about 12 miles geographically) may involve an electronic journey to California and back (about 1,200 miles). This *routing* of data makes it particularly difficult to predict where in the Net to monitor the passing traffic—the only two places where data from a given sending computer to receiving computer will pass by is *at* those computers and the nearest *router*

36. There is a school of thought that it is pointless to store something on one's own computer if it is freely available on demand via the Net. The newest Internet Computer, which sells for less than $600 and has little or no permanent storage, is living testament to this notion.

(the computer that routes the data—usually located at the nearest point of presence (POP) run by the Internet Service Provider).

Individual messages are chopped up into smaller data packets, the better to share the available information capacity of the data transmission lines. Thus, there is no guarantee that all of the data packets for a given message will travel by the same route from the sending computer to the receiving computer. Therefore, the notion of monitoring passing traffic from an information superhighway overpass is not certain to work—not all the packets of a particular message may come past that point. Indeed, *none* of them may come by.

3. How Will We Recognize It When We See It?

There is a fundamental problem with digitized information. Nothing is what it seems. Text is not really text. Images are not images. Sounds are not sounds. Everything is merely zeros and ones that must be interpreted by computer software before it abandons its cyberspace form and assumes one that can be recognized by humans. It is therefore extremely difficult to predict in what form proscribed information might appear; in fact, there are some encrypted forms that will defy even the most powerful computers' attempts to decrypt them back into something perceptible. Put simply: One cannot recognize proscribed information because, at best, you need to run computer programs to convert it into a form that can be *perceived* by a human. At worst, not even a computer and a human can recognize it without the correct decryption key.

This need for actual human recognition of digitized images or sounds merits an example because we are too used to thinking about recognizing proscribed text by searching for key words in the text—and this metaphor serves to obscure the technical problem of recognizing digitized images and sounds. By way of illustration (no pun intended), consider the example of a digitized image stored in a very common format: JPEG (a compressed image format adopted by the Joint Photographers Engineering Group, whence it gets its name). Computer technicians normally *dump* out the contents of files in a notation called hexadecimal, in which the numbers 0 through 9 and the letters A through F are used. This is a convenient method of viewing the zeros and ones that the computer understands insofar as a single textual *digit* represents a cluster of four binary digits. What follows is a hexadecimal dump of the first part of an actual image captured by the author from the Net and shown as it might be *seen* were it to be transmitted or stored on the Net. A would-be legislator must answer the following questions:

a) Is this a pornographic image?[37]

37. Yes it is.

b) Is this a pornographic image depicting children?[38]

c) Is this an image that offends "community standards?"[39]

Here are the first ten lines of the hexadecimal dump of this image (in reality the total image is several thousand such lines):

```
ff d8 ff e0 00 10 4a 46  49 46 00 01 00 01 00 60
00 60 00 00 ff fe 00 17  55 2d 4c 65 61 64 20 53
79 73 74 65 6d 73 2c 20  49 6e 63 2e 00 ff db 00
84 00 03 02 02 03 02 02  03 03 03 03 04 04 03 04
05 09 06 05 05 05 05 0b  08 08 06 09 0d 0c 0e 0e
0d 0c 0d 0d 0f 11 16 12  0f 10 14 10 0d 0d 13 1a
13 14 16 17 18 18 18 0e  12 1b 1d 1a 18 1c 16 18
18 17 01 04 04 04 05 05  05 0b 06 06 0b 17 0f 0d
0f 17 17 17 17 17 17 17  17 17 17 17 17 17 17 17
17 17 17 17 17 17 17 17  17 17 17 17 17 17 17 17
```

Clearly, it is impossible for a human to comprehend this *image*. A special program must be used to convert the JPEG format file into colored dots on a screen for it to be seen for what it is.

Another common format for image storage is a compressed form called a Zip file. Here is the first few lines of this same image in Zip format. To view the image, it must first be uncompressed, and then viewed using a JPEG viewer.

```
4b 03 04 14 00 00 00 08  00 3a a3 3d 1f d9 52 41
1e ee 09 ae 00 00 36 b6  00 00 0b 00 00 00 54 45
45 4e 31 30 36 2e 4a 50  47 94 bb 67 58 5b 69 96
2d 5c 7d 6f cf 74 cf ed  ae e0 44 06 44 06 8c 0d
26 67 04 08 49 64 91 24  a1 9c b3 84 24 04 08 05
24 a1 48 ce 39 e7 9c 4d  30 60 c0 39 67 57 b9 5c
a9 ab aa ab 73 cf 4c cf  cc bd cf 3c 77 be fb dc
fa b6 ba 7e 7d 3f 3f bc  7d 9e c3 41 b6 d0 3a 6b
af bd f6 7e df f3 e3 e7  3f 7e fb c1 27 79 e8 5c
```

Alternatively, if the contents of the file have been encrypted using publicly available, free, military-grade encryption software called PGP ("Pretty Good Privacy"), the information will look like this:

```
-----BEGIN PGP MESSAGE-----
      Version: 2.6.2
```

38. No it is not.

39. Consult your local community. This may prove difficult since an important consideration is determining which community is relevant to the inquiry.

hGwD1ImKm4VPMd0BAwCZl/6NPWzmd+GjhIOW6Wvyz4gWAwnxrxt
8WBccaTUuReV/
9abhHEqs04GNQQxsWUWXLXWWoTRtSAWN3xVDI0MObQg4jImcQ
9UPKrrNypCt+Csz
YJR/Kta0H+TceprJ5zmmAACuwuRP3h9BG0cF0lh1jZEp4eNdxz7f5+WN
NfnmBIF1
lHzhKG8duB/uceVDxA5aeRkHx7DSSUhje08ZnAeF8NijhzRF0OXK7jyK
QmdDwphv
mpUEwjwaYfxYXdT3AoLt2mAINE2E7IcoQEFzRJE1A+Q9u59OdESQc
8olTJ5Pezn+
tXI6/ty/08KBrW9ncdbmAS9+1HBwdnsKrDuCay3sfQC+NRC+1ZX2+Hc
XbAA0XB4+
RO5wfv7RypGL+ijBq7OJwDTlm9EAmH0Cix2GK99DoEOoDiZVKXVL
qpFfv8exn0Xa

These three easily prepared examples illustrate some of the very troubling issues that confront the legislators and the law enforcement community:

a) The same information can appear very differently depending on its storage format.
b) None of the representations give any clue as to the underlying image.
c) It is not possible to determine whether data represents proscribed information such as pornographic images or offensive sounds without using software to display the images or play the sounds.[40]
d) Unless a law states that mere possession is an offense, there is no means for detecting whether or not a particular file of information on a computer has ever been viewed, or played. It simply exists as information.

E. Other Problems for Legislation

The Internet moves a staggering amount of data. Estimates suggest that from fourteen to thirty terabytes (30,000,000,000,000 characters) move across the fast data transmission lines of the Net each day. Given that this estimate is not completely accurate, the point is that there is simply too much data moving too fast, between too many different places, for any immediate hope of effective monitoring.

---

40. Some existing software, quite ingeniously, purports to do this by determining the amount of flesh tone in images. However, family snapshots taken on the beach in the summertime will be difficult to differentiate from pornographic images.

1. The Hydra Problem

The best way to ensure that a particular piece of information survives for decades to come is to place that information out onto the Net. Depending on the means used, this information will, for all practical purposes, become immortal because of the number of different copies that will be stored around the world. On the Net, information spreads out like an electronic ripple around the world—as is shown clearly by the global availability of PGP (Pretty Good Privacy) military grade encryption software on hundreds of different computers even though just one copy was originally placed on the Net. Anyone doubting this need only obtain a copy of the latest version of PGP from a site such as ftp.kiae.su, which happens to be in Moscow—in spite of the fact that encryption software is illegal in Russia and the act of exporting it from the U.S. is illegal under the ITAR regulations (which classify encryption as a munition).

Legislators and law enforcement authorities must therefore confront the possibility that for any one site with proscribed information on it, there may well be many other such sites around the globe with digitally perfect copies of that same information, and all such sites are equally accessible—and all it will take is one such site outside the particular jurisdiction to provide a completely viable alternative source for that information.

2. The Greased Pig Problem

If one jurisdiction adopts draconian or repressive laws for particular kinds of information, the Net will repair this *damage* to itself by driving that information into friendlier climes further away. If, for example, the U.S., as if by magic (and that may be what it takes), really clamps down on pornography, the sites offering these images and digitized videos will simply move outside the U.S. Those seeking this kind of information will then have to contact the site pornsite.nl (in the Netherlands) instead of pornsite.com and suffer a delay of a few seconds for the data to arrive at their computers. If the Netherlands, using other magic, then clamps down, it will move to another country, perhaps Finland, or Thailand.

3. The Distributed Storage Problem

Stretching the imagination further by the application of yet more magic, suppose that well-funded authorities acquire sufficient computer skills and processing power to be able to monitor the Net at random points, and can detect inappropriate information when they see it flying by—netizens will then play another trump card: distributed storage.

Here are the first few bytes of the earlier example image. In that prior example, the bytes were all contained in the same file on the same computer.

ff d8 ff e0 00 10 4a 46 49 46 00 01 00 01 00 60

In distributed storage, the contents of the information are systematically and repetitively scattered around the globe. Specifically:

ff stored on a computer in London.
d8 stored on a computer in Paris.
ff stored on a computer in Oslo.
e0 stored on a computer in Bonn.
00 stored on a site in Madrid.
10 stored on the same computer in London.
4a stored on the same computer in Paris.
46 stored on the same computer in Olso.
49 stored on the same computer in Bonn.
46 stored on the same computer in Madrid.
. . . .

The significance of distributed storage is that the entire information does not exist at *any* of the computers upon which it is stored. Any one of these computers only contains 1/5th or 1/32nd of the entire information, depending on the number of computers, and therefore it is impossible to search the information for any recognizable characteristics. If the information were an image, as it is in this case, the image cannot be reconstructed for viewing. The would-be human viewer of this image must reach over the Net to each of the computers, retrieve a copy of the information, and, only when all the various parts of the information are present on the local computer, combine them into a viewable image. The Net being what it is, this may take *less* time than if the entire information were on just a single computer in London—information could be transmitted in parallel from the different computers. If the entire distribution of this information is under computer control, the number of computer sites used for distributed storage could be increased to, say, sixty-four, and backup copies of the information could be stored redundantly on another sixty-four sites to guard against the possibility that one or more of the computers are inaccessible when the information is required.

The primary question raised by distributed storage is whether or not a given fragment of the information can be treated under law as though it were the whole information. For example, is that part of the image stored on the computer in London still truly a pornographic image? Would the transmission of that file from London to the recipient's site constitute an offense under the CDA (or other law with the same intent)? Clearly it could be argued that the

intent of the would-be viewer was illicit, however it is less than clear that the available computer-based evidence would support the act—this is especially true if the viewer's computer merely assembled the image in its memory and displayed it from there without ever preserving a copy on the computer's hard disk. The final *coup de grace* is added if one postulates that the original information is encrypted and *then* distributed around the world. The only period of time when it is reconstituted into its original form would be those fleeting seconds while it was on the screen of the viewer's computer.

4. The Security Through Obscurity Problem

One of the most secure ways to conceal something is to hide it in full view and steganography permits just this. Users can hide information within digitized images and sounds. Stego (as it is known for short) does this by distributing the hidden information throughout an image.

Each dot in the image might be represented by eight binary digits, each of which can be either 0 or 1, with a value of black being 00000000 and white 11111111. Assume that the first dot, at the top left-hand corner of the image, is a shade of light gray and has the value 00001101. The next dot, is slightly darker and has the value 00010000.

Stego simply uses the right-most binary digit of each dot in which to store information. Therefore, if each dot in the first row has the values:

00001101 00010000 00100111 etc.,

and the user wants to hide a particular message in the picture, say a message that starts with three binary digits of 011..., stego puts these binary digits into each successive value's right-hand digit position. The resulting picture would be as follows (the underscoring marks those bits that are now the message and not the original digitized data):

00001100 00100001 00100111 etc.

In the first dot, the right-hand binary digit changes from 1 to 0—in effect the gray scale value is now slightly darker (1/256th darker in fact). In the second dot, the right-hand digit changes from a 0 to 1 making it 1/256th lighter than it should be, and in the third dot, the right-hand digit is already 1 so there is no change. Stego therefore hides the data in the picture and the errors that are introduced are imperceptible to the human eye—nothing looks wrong, especially since the original photograph is probably not available for comparison. Unless you know that a given picture has steganographically hidden information, there is no way of detecting it, yet the information is (literally) in full view. For example, compare the two images that follow:

The image on the right exhibits some artifacts on the broad expanses of similar tones (such as the dress of the lady dancing on the left of the image and the shoulder of the lady in the center foreground).

Hidden in the image on the right is this recently declassified satellite image of a Russian airfield with the bombers visible in the ramp area at the top of the image:

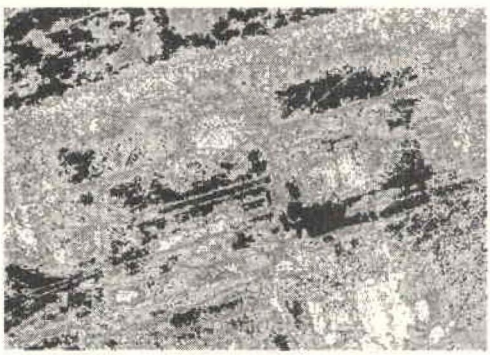


Figure 4: Long-Range Aviation Airfield[5]

The stego software, which is in the public domain and available at no charge on the Internet, can also extract the message from the picture on the recipient's computer.

While it might appear inefficient to store just one binary digit of the message in every eight binary digits of the image, a typical image might be 300,000 binary digits, and could therefore hide a 37,500 character message within it (about 12 typewritten pages of text). Stego security can further be enhanced by compressing or encrypting the message before it is hidden in the image and re-arranging the message data so "standard" headers cannot act as tell-tales that could be the object of a computerized search.

F. Assessment: Taxation and the Net

The problems of jurisdiction and detection described above will make life more difficult for taxation authorities. As more and more purchases of electronic information occur via the Net, taxation authorities fret over the loss of income and are therefore contemplating ways in which they can assess taxes on the information.

There is nothing to prevent a U.S.-based company from establishing electronic distribution centers outside the U.S. in places where sales tax does not exist. Would-be purchasers could then reach out to these distribution sites, either directly or indirectly through anonymizing servers in Europe or the Pacific Rim, and obtain their purchases. Local, state, and federal authorities will then have to determine how best to assess the taxes on the sales of everything from news, images, video, and software, this begging some questions of whether one assesses purely by market value, by number of binary digits (compressed or uncompressed?), or some other measure.

With the advent of digital cash, which can be made payer and payee anonymous, not even tracking the flow of money will lead the authorities back to the seller. It is not without reason that some proponents of digital cash say with glee that it will make taxation voluntary.

G. What Might Be Done About It?

It appears that technology may be outstripping the legislators, and, to a certain degree, this is appropriate. A saying in parts of the U.S. is, "Don't pour the concrete for the paths until you know where the folks are going to walk." This is particularly poignant for the Net. An unfortunate complexity is that by the time the legislators have discovered where the folks are going to walk, it is perhaps too late for them to pour the concrete—and even if they do, the folks will go around it if it turns into a toll-road.

Notwithstanding that the U.S. Constitution appears to offer citizens the right of anonymity, one common denominator for many of the problems caused by the Net is that of the anonymity of netizens. One cannot help but ask whether or not there should be a requirement to register one's identity in order to be on the Net—much like the FCC controls those who would transmit on other parts of the electromagnetic spectrum. Given that we could establish who was really who on the Net, perhaps we could then find out who and where the dogs are? The downside is that such a move would cause a public outcry on the Internet and the Outernet. Such freedoms are not given up easily in the U.S., nor those countries in Europe where citizens still alive recall the words "Papieren bitte. Macht schnell."[41]

41. "Papers please. Make it quick." Often said by Nazi Storm Troopers to the local populace in occupied countries during World War II.

An associated thought would be that of requiring every computer that will transmit information via the Net to be registered, and in a very limited sense that is merely the continuation of the process started by the InterNIC. However, it remains to be seen whether or not such a suggestion will sit comfortably with those in cyberspace, or whether it will be viewed as an Orwellian move to control netizens rather than merely hold them accountable for what they say and do.

One thing is for certain—any action that is taken to legislate the Net must be effective and enforceable in every country in the world where computers are connected to the Net. Not to do so merely creates electronic versions of tax havens, and the information-wealthy will merely move their information offshore to bask in computers in fairer climes. It remains to be seen whether or not all nations of the world can be persuaded to legislate the Net. Some of the more restrictive governments will want regulations that contradict the U.S. Constitution, and others may not yet be ready to subject themselves to such treaties. So much will depend on the still-strong national ethos of each country. As the apoplectic British colonel remarks in the movie *Those Magnificent Men And Their Flying Machines*, "That's the trouble with these International Events—they're full of foreigners." Treaty regulation of the Net demands we think more in terms of Earthlings than foreigners.

IV. CONCLUSION

Perhaps the most stunning thing about the Internet is that it has exploded across the planet in total silence like a videotape of a nuclear test in the Nevada desert with the sound turned off. The fireball went unnoticed by governments, politicians, bureaucrats, judges, and lawyers. By the time the majority of the populace read their newspapers and saw the TV specials on the Internet, it was too late; the shock-wave was upon them. The Internet already existed and was controlled by thousands of computer techies who had (and will continue to have, by all accounts) absolutely no intention of letting the "suits" run it. Early democracy was borne in the hands of the Greeks, and its electronic future now rests in the hands of the Geeks.

It is worth pondering that what we have today is the product of just a few years of explosive growth. What will the Net be like when the first real Captains Kirk, Picard and Janeway take their ships out of Earth orbit? Will it have been pried loose from the fingers of those whose hands presently rest on the "ASDF JKL;" keys? Imposing law and order on the Old West was a piece of cake in comparison. Cyberspace will have to be invaded by force, and the natives are known to be hostile. Thousands of them are scattered around the world, and they are armed to the teeth with more powerful weapons than all the legislators, judges, and law enforcers put together: computers and the Internet.