DISCOVERING

MATHEMATICS

Joint Policy Board for Mathematics

public and government affairs for

American Mathematical Society

http://www.ams.org/

Mathematical Association of America

http://www.maa.org/

Society for Industrial and Applied Mathematics

http://www.siam.org/

Mathematics and the Internet

by Paul Davis

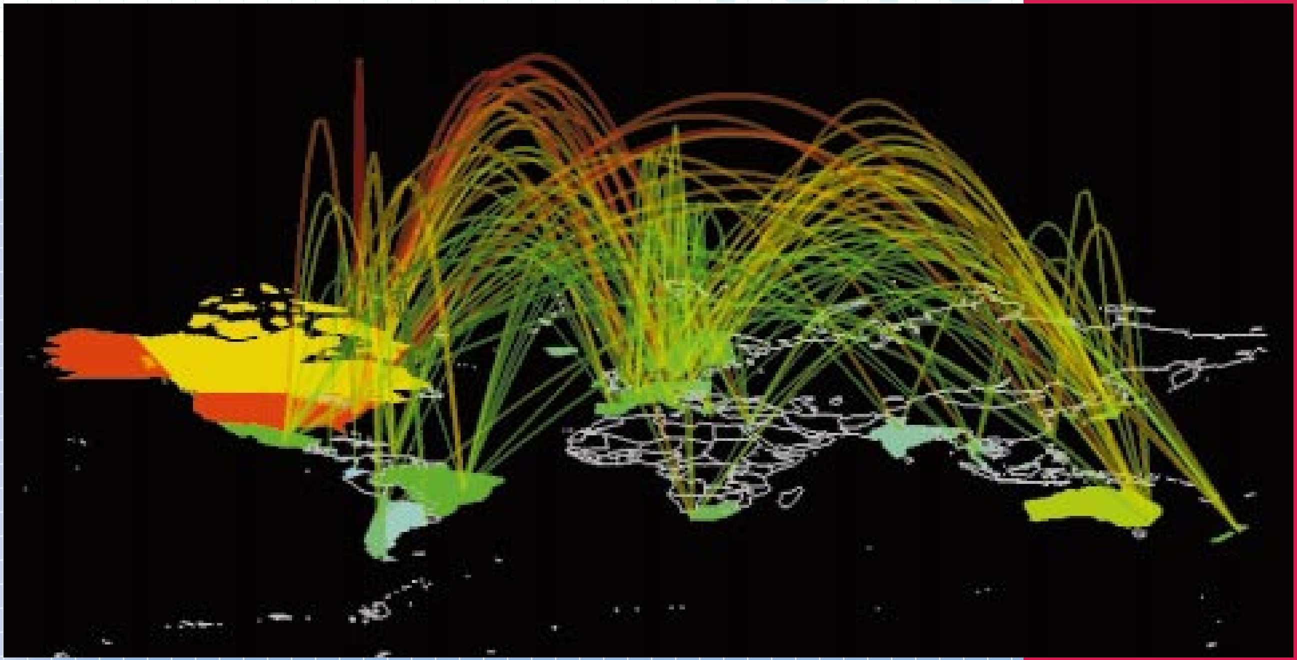
A longer version of this essay with World Wide Web links can be found at http://forum.swarthmore.edu/maw/97/articles/theme.essay.html.

The relationship between mathematics and the Internet is like that between language and the works of Shakespeare: his work could not have been conceived without language, and his poems and plays have contributed to the evolution of language.

Computers were born in the language of mathematics. Binary numbers let computers represent words, music, images and more so that machines can now communicate across the Internet with an alphabet of 0's and 1's. The impartial rules of mathematical logic govern computer operations, Internet addressing, and even Web search engines.

public affairs office
telephone: 301-942-9595
e-mail: piojpbm@deans.umd.edu

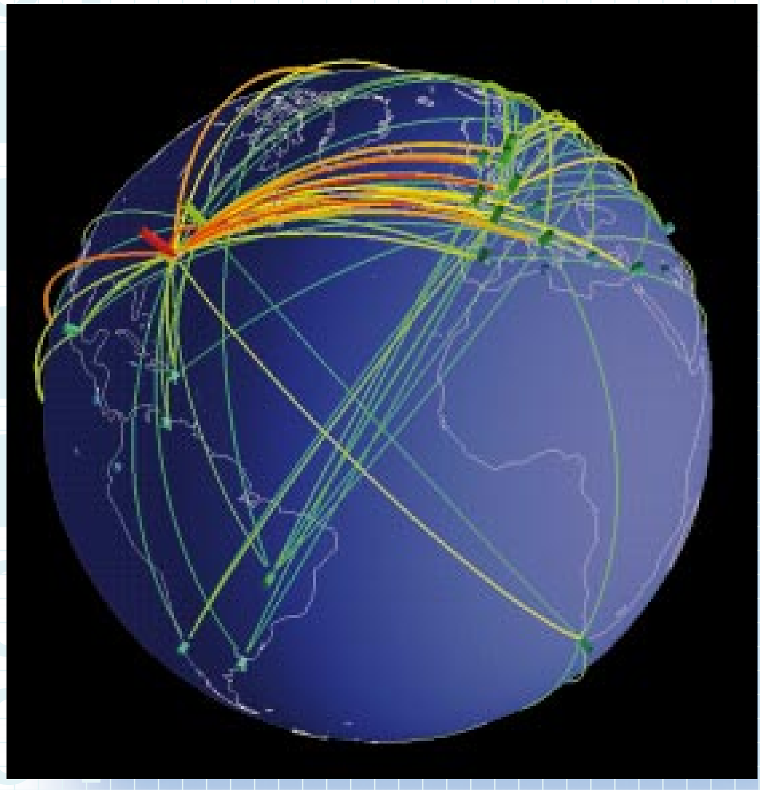
government affairs office
telephone: 202-234-9570
e-mail: jpbm@math.umd.edu



Worldwide Internet traffic over a two-hour period, with the color and thickness of the lines representing the traffic. Image provided by Bell Laboratories.

Within the Internet, mathematics is at the heart of security for messages and financial transactions. It is the basic tool of data compression, coding, and error correction for transmitting large files. It is the foundation of databases for managing email addresses and for searching the World Wide Web, and it is the agent for routing messages and managing networks.

The Internet is also helping advance mathematical research and education. Groups of educators and researchers communicate through email, newsgroups, and special World Wide Web sites. The Internet also supports distributed computing such as the recent cooperative effort which linked computers across dozens of countries to crack a code once thought secure for 20 millennia.


Global view of the data shown in the arc map on page 1. Image provided by Bell Laboratories.

Managing data on the Internet

As most people know, Internet messages — email, graphics, sound, the results of database searches — are transmitted as strings of 0's and 1's. Mathematics is central to two parts of this digital translation and transmission:

- accurately transmitting a text message, say, that has been translated into binary numbers requires codes for detecting and correcting errors (not to be confused with secret codes), and
- reducing the volume of data in an image, for example, that must be transmitted and then reconstructed as a reasonable likeness of the original, uses the tools of data compression.

When massive strings of 0's and 1's are forced over computer networks, some errors are inevitable, and even small losses of data can be catastrophic. Error detecting codes introduce mathematical tools to detect many of those losses, much like counting the number of pages in a long letter as a way of determining if anything was lost in the mail.

Data compression ideas are shared across a wide range of technologies, including the forthcoming digital television. (One second of high definition, uncompressed video would require more than *seven hours* to arrive over a conventional home modem!) The challenge of data compression is to reduce by many orders of magnitude the volume of data, and hence the transmission time, while preserving all the visually important parts of the image.

Good data compression schemes help World Wide Web graphics appear quickly and attractively on a computer screen. The same tools bring sound files that please the ear, even though selected parts have been removed or reconstructed. Some of the latest data compression ideas use wavelets, a kind of multiscale analysis tool.

Security on the Internet

Security on the Internet is as important as the security of a bank vault. Security concerns encompass privacy of messages, integrity of computers connected to the Internet, and trust in financial transactions, among many other issues. The rapidly growing Internet marketplace, for example, depends heavily on clever secret codes that combine centuries-old number theory with discoveries of the past two decades.

Moreover, efforts to break such codes use the Internet to distribute the computing burden over a wide array of machines. That distributed computing in turn depends in a crucial way on modern extensions of an old idea of Fermat for methodically searching for prime factors of large numbers.

Internet security can be seen in two complementary parts. One is the problem of sending a message that only the recipient can read, insuring both confidentiality of the message and its fidelity. The other is verifying the identity of the sender of a message. The first amounts to finding a code which is hard to crack while still permitting rapid transmission and decoding. The second is the problem of digital signatures: how can an Internet merchant be sure that the signature on an electronic check is genuine? The solutions to both problems rest squarely on the shoulders of number theory, a deceptively deep branch of mathematics.

Databases and searching

Powerful Web search engines like AltaVista and Yahoo! let Internet users find specialized nuggets of information hidden all over cyberspace. The heart of most of these search tools is an index of key words; each index entry lists the Web sites that contain that key word. (The entry for "mathematics" in one search index lists 332,966 sites!) Ideally, the search engine returns not just the intersection of all index entries for the given key words but also a priority score reflecting the potential relevance of each listed site to the searcher's needs.

In reality, search engines do not explicitly manipulate matrices with hundreds of thousands of rows and columns. Instead, they rely upon clever computational implementations of databases.

Many databases are built around the mathematical object known as a tree. These trees are like family trees that record relations among parents and children and their ancestors and descendants. An index, for example, might consist of twenty-six family trees, one for each letter of the alphabet. The first level of children would be all legal two-letter combinations, and so on; "aardvark" would, for example, be a distant descendant of "aa."

Beyond the parent-child connection, relational databases define additional relationships among their entries. The power of a relational database comes from its ability to manipulate those relations; e.g., performing an intersection operation that can find a common string of letters appearing in two different words. The rules for those manipulations are mathematically defined in a relational algebra or relational calculus specific to that database structure. Mathematics is the framework for describing database constructs, and mathematical tools are the basis for improving their efficiency and reliability.

1630—1750

1630-1750
P. Fermat and L. Euler lay number theoretic foundations which are now used in public key cryptography, the basis for secure messages on the Internet.

1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000

1944-45
J. von Neumann develops methods of translating mathematical procedures into machine-language instructions, and later, designs and constructs a computer.

1950-52
R. Hamming, D. Huffman et al. introduce the basic ideas of error detecting and correcting codes using tools from the algebra of polynomials over finite fields.

1956-59
R. Bellman, L. Ford, and E. Dijkstra develop the first shortest-path algorithms, essential for packet routing on the Internet.

1962
P. Baran devises a new kind of communications network that is net- or web-like, rather than point-to-point.*

1965
B. Mandelbrot publishes the first attempt to use a self-similar mathematical model in communications traffic.

1970
ARPANET hosts start using Network Control Protocol (NCP).*

1972
R. Tarjan and J. Hopcroft refine graph search algorithms for finding connected segments of networks and other applications.

1974
V. Cerf and R. Kahn develop a protocol for packet network intercommunication which specifies in detail the design of a Transmission Control Program (TCP).*

1976
W. Diffie and M. Hellman propose notion of public key cryptography based on modular arithmetic and discrete logarithms.

1978
R. Rivest, A. Shamir and L. Adelman, devise a method for obtaining digital signatures and public key cryptosystems based on modular arithmetic and properties of prime numbers.

1982
D. Anick, D. Mitra and M. Sondhi develop a mathematical model that can be used for a data-handling switch in a computer network.

DCA and ARPA establish the Transmission Control Protocol (TCP) and Internet Protocol (IP), as the protocol suite, commonly known as TCP/IP, for ARPANET.*

1986
V. Jacobson and M. Karels develop "slow start" protocols to prevent congestion collapse on ARPANET.

NSFNET created (backbone speed of 56Kbps).*

1991
NSFNET backbone upgraded to T3 (44.736Mbps); traffic passes 1 trillion bytes/month and 10 billion packets/month.*

1994
NSFNET traffic passes 10 trillion bytes/month.*

W. Leland, M. Taqqu, W. Willinger, and D. Wilson conclude that Internet traffic is fractal in nature and suggest new models.

1996
President Clinton proposes next-generation Internet.

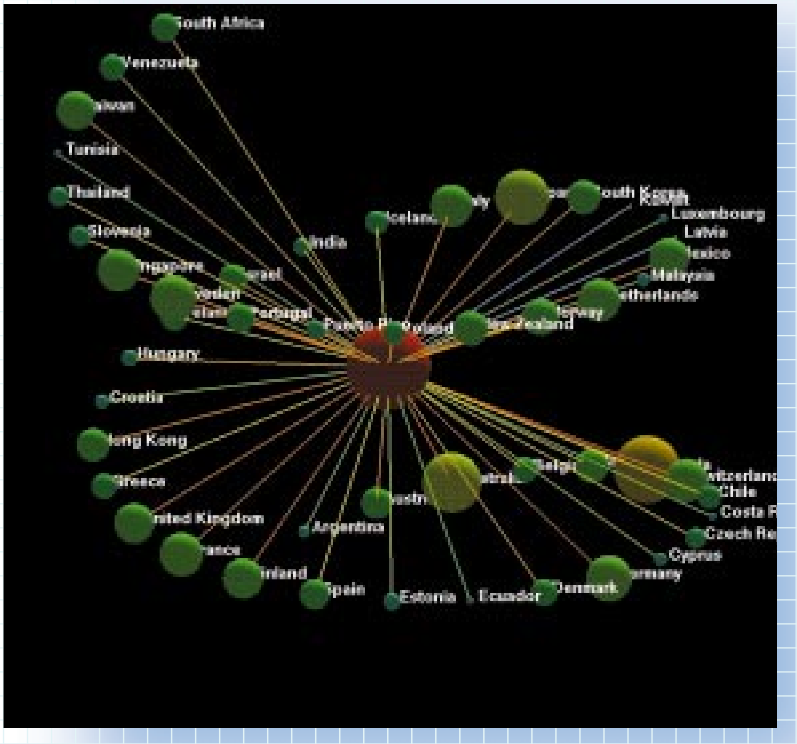
** from Hobbes' Internet Timeline v2.5 by Robert Zakon*

Routing and network configuration

A local area network of moderate size might have 10,000 pairs of nodes that communicate with one another. The messages they share are like trains running at the speed of light on the tracks of the network. Each car in the train carries part of one message, as if a long letter had been written on a series of postcards, one card per car. Typically, cards from many messages are mixed in one train.

The performance of the network depends on the length of the trains — the size of the message packets — and on the space between the trains. For example, a long message train that arrives at the wrong time can delay many other messages until it passes; short messages properly spaced can be slid in among one another.

The mathematical ideas of queuing theory can predict the behavior of message handling protocols based on information about the size and arrival patterns of these message packets. (The classic application of queuing theory is estimating the waiting time at a bank, given the arrival patterns of customers and the service time of the bank teller.)



Traffic to and from the USA with nodes positioned in a helix. Image provided by Bell Laboratories.

But investigations of alternate message handling protocols are based on mathematical models of the message traffic. Good models assure that a new protocol will perform as well in practice as queuing theory predicts; bad models can lead a protocol developer to make performance promises that can't be fulfilled.

Mathematics on the Web

Mathematicians take full advantage of the Internet and the World Wide Web. These tools let them share ideas, techniques, and resources across geographic and disciplinary boundaries to advance both teaching and research.

Central hubs for a wide range of mathematical activity, including considerations of the role of mathematics in society, are the Math Forum and the home pages of the three sponsoring societies for Mathematics Awareness Week: the American Mathematical Society, the Mathematical Association of America, and the Society for Industrial and Applied Mathematics.

Examples of more specialized sites are the Math Archive, which specializes in educational issues, and the Geometry Center, whose focus is computation and visualization of geometric structures. Number theorists interested in the search for so-called Mersenne primes pool their resources through the Great Internet Mersenne Prime Search. For many years, computational scientists have shared problems, solutions, and methods through Netlib, where the best public-domain numerical analysis software is available for downloading.

Mathematics and the Internet

Mathematics is the language of Internet operation, from the binary numbers that describe text and images to the complex data structures of search engines for the World Wide Web. Adroit combinations of old and new ideas from fields like number theory have enabled such key Internet technologies as data encryption for secure financial transactions. At the same time, the Internet has given birth to worldwide collaborations among mathematics teachers and researchers, collaborations that are advancing both education from kindergarten through university and our understanding of some of the most difficult problems of pure and applied mathematics.