from *SIAM News*, Volume 33, Number 2

Oh, What a Tangled Web We've Woven. . . .

By Barry A. Cipra

What is the most rigorous law of our being? Growth. No smallest atom of our moral, mental, or physical structure can stand still a year. It grows—it must grow; nothing can prevent it.

—Mark Twain

Everybody talks about the Internet, Mark Twain might say if he were on line today, but nobody does anything about it.

Is traffic on the information superhighway an act of man, or is it a force of nature? For that matter, does it make a difference?

Speakers at a "hot topics" workshop, Scaling Phenomena in Communication Networks, held October 22–24 at the Institute for Mathematics and Its Applications at the University of Minnesota, aren't sure about the shallow, metaphysical side of things. But graph after graph of "fat tails" and power-law fits make a persuasive argument that the Internet has more in common with earthquakes and forest fires than with, say, an assembly line or a symphony orchestra.

Moreover, many of the statistical concepts commonly used by scientists and engineers, not to mention policymakers, to des-cribe large systems lose much of their meaning when it comes to the Internet. Even the well-worn notion of "average" fails to explain much of anything.

"Nothing is 'typical,'" says Vern Paxson, a computer scientist at the AT&T Center for Internet Research at ICSI (ACIRI—ICSI being the Berkeley-based International Computer Science Institute). In October 1992, he says, the size of the median FTP data transfer at Lawrence Berkeley National Laboratory was 4500 bytes; five months later, the median was a mere 2100 bytes. These days, of course, 1992 counts as ancient history. But five years later, in March 1998, the median was up to 10,900, while in December 1998, it was down again, to 5600.

In other words, Paxson concludes, there is no typical time period for Internet traffic. Means mean nothing when variances are infinite. That insight is "the painful lesson of many studies," he adds.

Paxson was a co-organizer of the IMA workshop, along with Ashok Erramilli of Telcordia/Netmetrix Inc., Iraj Saniee of Lucent Technologies, and Walter Willinger of AT&T Labs Research.

The Fractalnet

Instead of statistical approaches, researchers say, the mathematics of self-similarity and fractals may prove to be the best tools for understanding the Internet.

Paxson points to a 1991 paper by Henry Fowler and Will Leland, both then at Bellcore, as containing the first (metaphorical) hint of self-similarity on the Internet. Fowler and Leland's self-described "intuitive interpretation" was that "traffic 'spikes' (which cause actual losses) ride on longer-term 'ripples,' that in turn ride on still longer term 'swells.'" The metaphor was nailed down (so to speak) three years later, in a ground-breaking paper by Leland, Willinger, and Daniel Wilson, all then at Bellcore, and Murad Taqqu of Boston University.

The researchers had analyzed detailed logs of Bellcore's Ethernet traffic from four separate periods (ranging from 21 to 48 hours) over a four-year span. The statistical patterns were found to be essentially the same for all four periods, despite tremendous changes in the network. Moreover, the measured traffic satisfied the criteria for self-similarity: The fluctuations in the number of packets or bytes transmitted per unit time were statistically the same, whether the unit of time was minutes or milliseconds (see Figure 1).

Self-similarity has a more precise mathematical definition, of course. Actually, it has several definitions, depending on what's being called self-similar. The snappiest of the definitions is for a continuous-time stochastic process $X(t)$: It's self-similar, with "Hurst parameter" H, if $X(at) = a^H X(t)$ for all $a > 0$. (The equal sign, however, hides some probabilistic technicalities.)

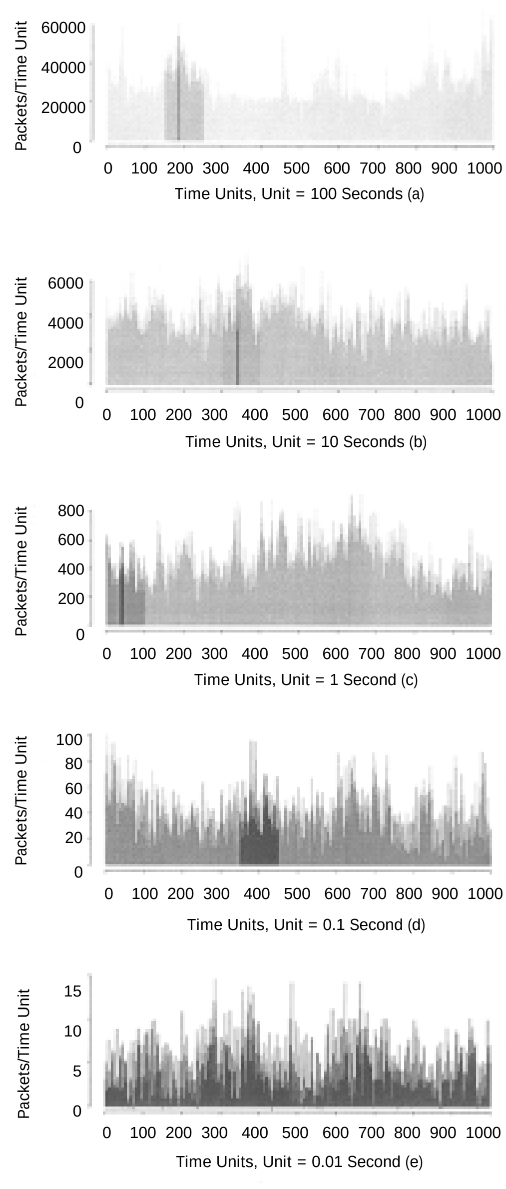

Figure 1. *Pictorial "proof" of self-similarity: Ethernet traffic (from August 1989) on five different time scales.*

In the more practical case of a time series $X(k)$, with $k = 1, 2, 3, \ldots$, the equality is between $X(k)$ and $X_m(k) = (X(mk - k + 1) + \ldots + X(mk))/m$. In particular, the autocorrelation functions for X and X_m should be identical for all m (or at least asymptotically equal for large m).

One tell-tale sign of self-similarity is the presence of "fat tails." Also known as heavy tails, they appear when larger-than-average events—precipitous, one-day drops on Wall Street, sudden surges in Internet traffic, humongous Web pages, or forest fires the size of Rhode Island—occur more often than "normal" statistics allow. The name refers to the "tail" of the probability distribution, which for the classic bell curve tapers quickly to insignificance. Fat tails make hash of any policy based purely on expected value.

Self-similarity implies overweight tails. The tails show up strikingly as straight lines on log–log plots of cumulative probability functions vs. size (or time). Mark Crovella, a computer scientist at Boston University, showed examples from a 1995 BU study of transmission times and file sizes for World Wide Web traffic. The researchers had collected data from 4700 sessions by 591 users during a two-month period at a BU computer science lab. The users—mostly students—clicked on 575,775 URLs, representing 46,380 unique files (so the average file was requested approximately 12 times). Because of caching, there were only 130,140 file transfers. The obesity of the tails is clear from the log–log probability plots (see Figure 2).

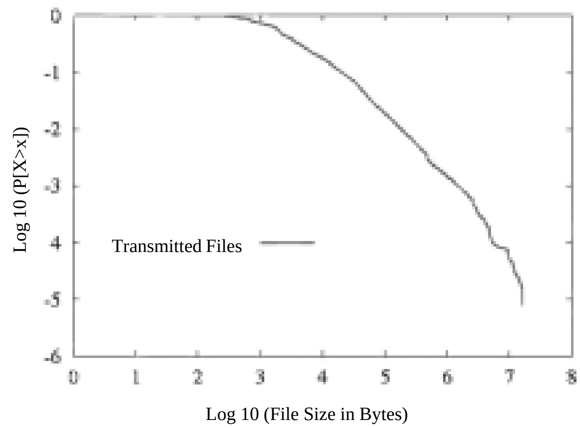

Figure 2. *Fat tails on the Web. A 1995 BU study found a power law distribution in file size at Web sites visited by undergraduates.*

So What?

Where does the Inter-net's self-similarity come from, and why does it matter, anyway?

One possible source of self-similarity is the irregular, on/off nature of Internet communications. Anyone who's ever used a browser knows that the flow of data is more like the flow of water from a faucet than of air through a ceiling fan. As Benoit Mandelbrot, the father of fractals, found as early as the 1960s, a large number of independent on/off sources, when taken together, produce self-similarity, provided the on and off cycles of each source are themselves fat-tailed—precisely what Crovella and colleagues found to be the case in the BU study.

The protocols used for Internet traffic, such as TCP (Transmission Control Protocol), may cause departures from self-similarity, at least on small time scales. (For an overview of TCP and several alternative approaches to the control of Internet congestion, see Ramesh Johari's article on page 1 in this issue.) The purpose of the protocols is to establish an orderly transfer of information. They do this in part by adjusting the rate at which packets are sent to correspond to the rate at which they arrive. Anna Gilbert of AT&T Labs Research described a wavelet-based scaling analysis of simulated Internet traffic in which multifractal behavior was found to result from TCP. Multifractals go beyond simple self-similarity; they model the volatility that characterizes systems like stock markets.

The Internet simulations, Gilbert explained, offer a controlled environment for investigating the effects of parameters like user demand and file sizes, or of network parameters like delays, bandwidth, and topologies. The study, conducted by Gilbert, Willinger, and Anja Feldman of AT&T Labs Research and Polly Huang of the University of Southern California, used two data sets of measured Internet Protocol (IP, the one standard for the entire Internet) traffic as benchmarks and "reality checks" for the simulations. Their key finding is that user-related variability results in self-similar scaling over large time scales, whereas the presence of TCP and other flow control algorithms underlies a transition to more complex, multifractal scaling over small time scales.

The latter result, however, is mainly empirical. "One of the most intriguing open issues that remains is how precisely TCP-like congestion control algorithms give rise to multifractal scaling," they write in the conclusion of a paper presented at SIGCOMM'99 (available at Gilbert's Web site, http://www.research.att.com/~agilbert). "A mathematically rigorous and intuitively appealing construction and explanation that makes sense in the networking context still eludes us."

As for the question of why self-similarity (or multifractality) matters, Gilbert and her colleagues mentioned the key word: congestion. What everyone is seeking is a way to keep traffic flowing smoothly on a rapidly growing and technologically evolving system with no central control, but rather thousands of independent Internet service providers (ISPs). In an "idiosyncratic view" of congestion control, Sally Floyd, a network expert at ACIRI, described various ways the Internet can get all balled up. In "classical congestion collapse," for example, Internet traffic slows to a crawl when a substantial fraction of the traffic consists of unnecessary retransmissions of packets. A slew of such slowdowns occurred in the mid-1980s, but the problem was solved by the introduction of better TCP algorithms, which adjust the rate at which they send packets out according to the rate of acknowledgment of their receipt.

More pressing today, Floyd thinks, is collapse from undelivered packets—which, of course, give rise to *necessary* retransmissions. Data may not make it all the way from sender to receiver for lots of reasons, one of which is that an intermediary may simply not bother to forward things. (Twain once described a western trip during which the stage coach driver stopped and dropped a heavy mailbag in the middle of nowhere.) These necessary retransmissions can further clog an already overwhelmed intermediary.

Three further potential sources of congestion collapse are packet "fragmentation," which means that packets arrive in unusable form, unwanted packets (which is bad enough when you wander into a poorly organized, zillion-byte Web site, and will become even worse when the e-tailers latch onto "push" technology), and control traffic itself, which requires increasing numbers of bytes for packet headers, routing updates, and other attempts to keep tabs on network status.

Floyd argues that end-to-end congestion control, in which transmission and retransmission rates are determined mainly by the sender and receiver, is the best way to avoid congestion collapse. (Other approaches include centralized scheduling and pricing mechanisms. There's also the rosy view that infinite bandwidth will eliminate congestion altogether—balanced by the apocalyptic view that collapse is imminent and unavoidable.) The end-to-end approach, Floyd says, can be promoted by router me-chanisms designed to restrict the flow of transmissions that are trying to hog bandwidth during periods of congestion.

"So far, the success of the Internet has rested on the IP architecture's robustness, flexibility, and ability to scale, and not on its efficiency, optimization, or fine-grained control," Floyd points out. "The rather decentralized and fast-changing evolution of the Internet architecture has worked reasonably well to date. There is no guarantee that it will continue to do so."

Barry A. Cipra is a mathematician and writer based in Northfield, Minnesota.