Internet Traffic: Visualization, Discovery, and Very Large Displays

William S. Cleveland

Statistics Research
Bell Labs
http://stat.bell-labs.com/wsc

Abstract

For a decade, the ruling common wisdom for Internet traffic held that it was everywhere bursty: over periods lasting tens of milliseconds to hundreds, the traffic was either much below its average rate or much above. In other words, the traffic was not smooth, not staying at all times close to its average. It was bursty on the cable running down a street, carrying the merged traffic of a small number of cable modem users in one section of a town. It was bursty on the core fiber of an Internet service provider, carrying the merged traffic of thousands of users from all over the country.

The Internet was designed to accommodate the bursty traffic. The routers and switches that forward traffic from one place to the next were designed for burstiness, and Internet service providers allocated traffic loads on the devices based on an assumption of burstiness.

Recently, it was discovered that the old common wisdom is not true. Visualization played a fundamental role in the discovery. The old wisdom held up for links with a small numbers of users. But as the number of users increases, the burstiness dissipates, and the traffic becomes smooth. Design of the high-load part of the Internet needs to be rethought.

The old wisdom had persisted for high-load links because the databases of traffic measurements from them are immense, and the traffic measurements had not been studied in their fullest detail, which is necessary to see the smoothing. Visualization tools allowed the detail to be seen, and allowed the verification of a mathematical theory that predicts the smoothing.

To see the detail, individual visual displays were created that take up an amount of virtual screen real estate measured in hundreds of pages. It is a simple idea: if you have a lot of data, and you want to see it in detail, you need a lot of space. What is needed now is a rich set of ideas and methods for navigating such very large displays.
Bill Cleveland is currently a Distinguished Member of Technical Staff in the Research Division at Bell Labs, the R&D arm of Lucent Technologies. Before that, he was on the faculty in the Department of Statistics at the University of North Carolina. He received a B.A. in mathematics from Princeton University and a Ph.D. in statistics from Yale University.

Bill works as a statistician. He has been involved in projects requiring the statistical analysis of data from many fields, and has developed new statistical methods, including many visualization methods, that are widely used in engineering, science, medicine and business. He has published over 120 papers and four books on this work.

Bill’s book, The Elements of Graphing Data, is about principles of graph construction, the visual communication of data, and visualization methods. The principles and methods are supported by a rigorous, scientific discussion of graphical perception, the visual decoding of information from data displays. His book, Visualizing Data, is about visualization methods. It presents a strategy for data analysis that stresses the use of visualization to thoroughly study the structure of data and to check the validity of mathematical and statistical models fitted to data.

For this work, Bill has been elected a Fellow of a number of professional societies and has received several prizes for papers. In 1996 he was selected Statistician of the Year by the Chicago Chapter of the American Statistical Association.