March 1, 2010

Scientists Strive to Map the Shape-Shifting Net

By JOHN MARKOFF

SAN FRANCISCO — In a dimly lit chamber festooned with wires and hidden in one of California's largest data centers, Tim Pozar is changing the shape of the Internet.

He is using what Internet engineers refer to as a "meet-me room." The room itself is enclosed in a building full of computers and routers. What Mr. Pozar does there is to informally wire together the networks of different businesses that want to freely share their Internet traffic.

The practice is known as peering, and it goes back to the earliest days of the Internet, when organizations would directly connect their networks instead of paying yet another company to route data traffic. Originally, the companies that owned the backbone of the Internet shared traffic. In recent years, however, the practice has increased to the point where some researchers who study the way global networks are put together believe that peering is changing the fundamental shape of the Internet, with serious consequences for its stability and security. Others see the vast increase in traffic staying within a structure that has remained essentially the same.

What is clear is that today a significant portion of Internet traffic does not flow through the backbone networks of giant Internet companies like AT&T and Level 3. Instead, it has begun to cascade in torrents of data on the edges of the network, as if a river in flood were carving new channels.

Some of this traffic coursing through new channels passes through public peering points like Mr. Pozar's. And some flows through so-called dark networks, private channels created to move information more cheaply and efficiently within a business or any kind of organization. For instance, Google has privately built such a network so that video and search data need not pass through so many points to get to customers.

By its very nature, Internet networking technology is intended to support anarchic growth. Unlike earlier communication networks, the Internet is not controlled from the top down. This stems from an innovation at the heart of the Internet — packet switching. From the start, the information moving around the Internet was broken up into so-called packets that could be sent on different paths to one destination where the original message — whether it was e-mail, an image or sound file or instructions to another computer — would be put back together in its original form. This packet-switching technology was conceived in the 1960s in England and the United States. It made delivery of a message through a network possible even if one or many of the nodes of the network failed. Indeed, this resistance to failure or attack was at the very core of the Internet, part of the essential nature of an organic, interconnected communications web with no single control point.

During the 1970s, a method emerged to create a network of networks. The connections depended on a communication protocol, or set of rules, known as TCP/IP, a series of letters familiar to anyone who has tried to set up their own wireless network at home. The global network of networks, the Internet, transformed the world, and continues to grow without central planning, extending itself into every area of life, from Facebook to cyberwar.

Everyone agrees that the shape of the network is changing rapidly, driven by a variety of factors, including content delivery networks that have pushed both data and applications to the edge of the network; the growing popularity of smartphones leading to the emergence of the wireless Internet; and the explosion of streaming video as the Internet's predominant data type.

"When we started releasing data publicly, we measured it in petabytes of traffic," said Doug Webster, a Cisco Systems market executive who is responsible for an annual report by the firm that charts changes in the Internet. "Then a couple of years ago we had to start measuring them in zettabytes, and now we're measuring them in what we call yottabytes." One petabyte is equivalent to one million gigabytes. A zettabyte is a million petabytes. And a yottabyte is a thousand zettabytes. The company estimates that video will account for 90 percent of all Internet traffic by 2013.

The staggering growth of video is figuring prominently in political and business debates like the one over the principle of network neutrality — that all data types, sites and platforms attached to the network should be treated equally. But networks increasingly treat data types differently. Priority is often given to video or voice traffic.

A study presented last year by Arbor Networks suggesting that traffic flows were moving away from the core of the network touched off a spirited controversy. The study was based on an analysis of two years of Internet traffic data collected by 110 large and geographically diverse cable operators, international transit backbones, regional networks and content providers.

Arbor's Internet Observatory Report concluded that today the majority of Internet traffic by volume flows directly between large content providers like Google and consumer networks like Comcast. It also described what it referred to as the rise of so-called hyper giants — monstrous portals that have become the focal point for much of the network's traffic: "Out of the 40,000 routed end sites in the Internet, 30 large companies — 'hyper giants' like Limelight, Facebook, Google, Microsoft and YouTube — now generate and consume a disproportionate 30 percent of all Internet traffic," the researchers noted.

The changes are not happening just because of the growth of the hyper giants.

At the San Francisco data center 365 Main, Mr. Pozar's SFMIX peering location, or fabric, as it is called, now connects just 13 networks and content providers. But elsewhere in the world, huge peering fabrics are beginning to emerge. As a result, the "edge" of the Internet is thickening, and that may be adding resilience to the network.

In Europe in particular, such connection points now route a significant part of the total traffic. AMS-IX is based in Amsterdam, where it is also run as a nonprofit neutral organization composed of 344 members exchanging 775 gigabits of traffic per second.

"The rise of these highly connected data centers around the world is changing our model of the Internet," said Jon M. Kleinberg, a computer scientist and network theorist at Cornell University. However, he added that the rise of giant distributed data centers built by Google, Amazon, Microsoft, IBM and others as part of the development of cloud computing services is increasing the part of the network that constitutes a so-called dark Internet, making it harder for researchers to build a complete model.

All of these changes have sparked a debate about the big picture. What does the Internet look like now? And is it stronger or weaker in terms of its resistance to failure because of random problems or actual attack.

Researchers have come up with a dizzying array of models to explain the consequences of the changing shape of the Internet. Some describe the interconnections of the underlying physical wires. Others analyze patterns of data flow. And still others look at abstract connections like Web page links that Google and other search engine companies analyze as part of the search process. Such models are of great interest to social scientists, who can watch how people connect with each other, and entrepreneurs, who can find new ways to profit from the Internet. They are also of increasing interest to government and law enforcement organizations trying to secure the Net and use it as a surveillance tool.

One of the first and most successful attempts to understand the overall shape of the Internet occurred a decade ago, when Albert-László Barabási and colleagues at the University of Notre

Dame mapped part of the Internet and discovered what they called a scale-free network: connections were not random; instead, a small number of nodes had far more links than most.

They asserted that, in essence, the rich get richer. The more connected a node in a network is, the more likely it is to get new connections.

The consequences of such a model are that although the Internet is resistant to random failure because of its many connections and control points, it could be vulnerable to cyberwarfare or terrorism, because important points — where the connections are richest — could be successfully targeted.

Dr. Barabási said the evolution of the Internet has only strengthened his original scale-free model. "The Internet as we know it is pretty much vanishing, in the sense that much of the traffic is being routed through lots of new layers and applications, much of it wireless," said Dr. Barabási, a physicist who is now the director of Northeastern University's Center for Network Science. "Much of the traffic is shifting to providers who have large amounts of traffic, and that is exactly the characteristic of a scale-free distribution."

In other words, the more the Internet changes, the more it stays the same, in terms of its overall shape, strengths and vulnerabilities.

Other researchers say changes in the Internet have been more fundamental. In 2005, and again last year, Walter Willinger, a mathematician at AT&T Labs, David Alderson, an operations research scientist at the Naval Post Graduate School in Monterey, Calif., and John C. Doyle, an electrical engineer at California Institute of Technology, criticized the scale-free model as an overly narrow interpretation of the nature of modern computer networks.

They argued that the mathematical description of a network as a graph of lines and nodes vastly oversimplifies the reality of the Internet. The real-world Internet, they said, is not a simple scale-free model. Instead, they offered an alternate description that they described as an H.O.T. network, or Highly optimized/Organized tolerance/Trade-offs. The Internet is an example of what they called "organized complexity." Their model is meant to represent the trade-offs made by engineers who design networks by connecting computer routers. In such systems, both economic and technological trade-offs play an important role. The result is a "robust yet fragile" network that they said was far more resilient than the network described by Dr. Barabási and colleagues.

For example, they noted that Google has in recent years built its own global cloud of computers that is highly redundant and distributed around the world. This degree of separation means that Google is insulated to some extent from problems of the broader

Internet. Dr. Alderson and Dr. Doyle said that another consequence of this private cloud was that even if Google were to fail, it would have little impact on the overall Internet. So, as the data flood has carved many new channels, the Internet has become stronger and more resistant to random failure and attack.

The scale-free theorists, Dr. Alderson said, are just not describing the real Internet. “What they’re measuring is not the physical network, its some virtual abstraction that’s on top of it,” he said. “What does the virtual connectivity tell you about the underlying physical vulnerability? My argument would be that it doesn’t tell you anything.”