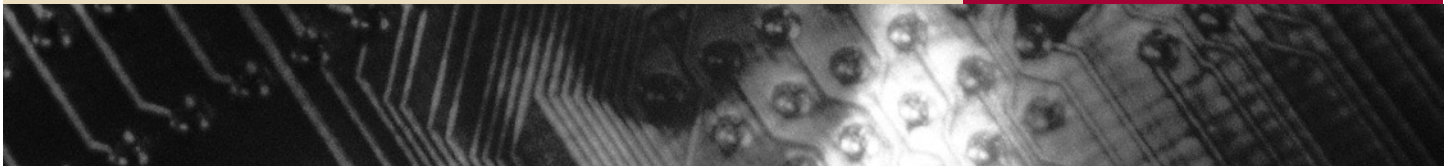


Deep Packet Inspection Optimizes Mobile Applications

Mobile data provides a notable growth opportunity for cellular operators, but adoption will be slow unless networks can provide good service to all subscribers. Deep packet inspection enables operators to understand the applications their subscribers are using and to build appropriate networks.

The communication industry widely recognizes mobile broadband as the growth engine for wireless carriers. As cell-phone penetration reaches saturation and voice revenue plateaus, mobile-data revenue is one segment that continues to grow. Mobile data is available through 3G (third-generation) networks, which are increasingly exhibiting the strain of the increased traffic. Most operators have announced plans to move to LTE (long-term-evolution) networks, which promise greater-than-100-Mbps data rates, a fivefold increase over 3G HSPA+ (high-speed-packet-access) networks. A quick look at customer-traffic patterns makes it clear that even network upgrades from 3G to LTE won't deliver sufficient data rates to guarantee good service to all users. However, additional technologies such as DPI (deep packet inspection) can ensure that prioritized and managed traffic optimizes the user experience.

*By Mike Coward,
VP of Strategy
& Innovation*



Network operators can use one of several techniques for adding capacity and meeting the growing demand for mobile broadband. First, they can increase spectral efficiency. LTE promises a fourfold improvement - in bits per seconds per hertz - over its WCDMA (wideband-code-division-multiple-access) predecessor. However, data rates are growing too fast for improvements in spectral efficiency to make more than a dent in the problem. It will take many years from the original release of LTE for full network rollouts and adoption to occur. Cisco expects mobile-data traffic to increase 66-fold between 2008 and 2013 (Reference 1).

Another technique is to use more spectrum; wider channels mean more bandwidth. However, spectrum is an expensive resource, and most operators have little available spectrum to spare. Alternatively, operators can use cell-splitting techniques; adding smaller cell sites reduces subscriber density in a given cell site. By shrinking the cells, the number of subscribers per cell decreases, and the average bandwidth per subscriber accordingly increases. This approach is both the easiest and the most expensive option because radio access is the priciest portion of the operator network, with costs directly proportional to the number of cell sites.

Another alternative is to more efficiently allocate user bandwidth by employing new DPI technologies. DPI manages data networks and optimizes data traffic. It reaches beyond the IP (Internet Protocol) headers and examines the packet contents. Although you can use this technique to look for keywords or other content that most consumers feel violates privacy norms, the most common use of DPI is simply to determine the application of the packet: Is it an e-mail, Web, video, or P2P (peer-to-peer) packet? The shortcomings of other approaches increase the attractiveness of more efficiently using

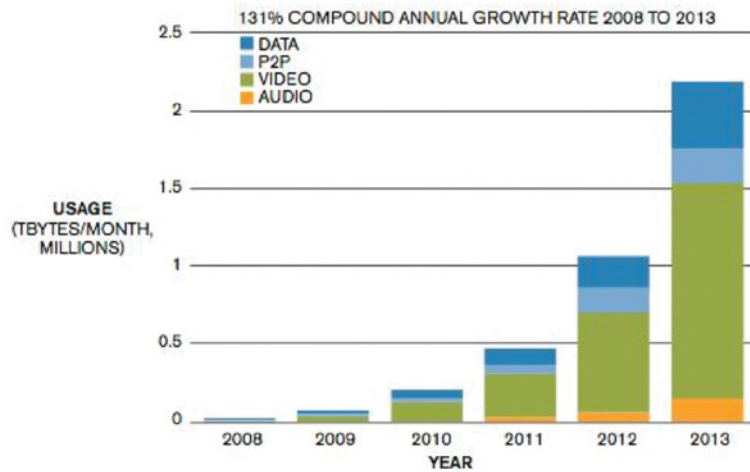


Figure 1. Cisco's forecast indicates that video services will see usage growth in the coming years

the overall bandwidth and ensure that applications and subscribers fairly share bandwidth. DPI enables the implementation and enforcement of these policies.

The Real Threat

Media coverage of mobile broadband tends to suggest that smartphones, such as the iPhone and BlackBerry, are driving the bulk of the congestion on the network. However, most of the data growth actually comes from laptop computers with data cards and USB (Universal Serial Bus) dongles. Operators that want to encourage mobile-data growth have been promoting the idea of mobile broadband, and some are even subsidizing networks with built-in cellular modems. Although a smartphone generates as much data traffic as 30 regular phones, laptops with data cards generate 15 times more traffic than smartphones, or as much traffic as 450 regular phones (Reference 2). Laptops are more conducive to Internet browsing and have larger, higher-resolution screens that demand higher-quality content. Laptop users are also more likely to run P2P applications that can consume huge swaths of bandwidth.

Shared Data Channel

The developers of both 3G and LTE networks employed the common concept of a single data channel that all subscribers in a given cell share. They chose this architecture because they assumed that users would employ mobile data primarily for "bursty" activities, such as Web surfing and e-mail. In these activities, a shared data channel means that users get high bandwidth for downloads but also that users can employ the channel while other users read e-mail or scroll through Web pages. The shared data channel performs poorly when it encounters large sustained transfers. These transfers fill up the shared channel and lead to dropped packets and long latencies for every subscriber in the cell. With the growth in mobile data and broad usage of wireless-data cards, video and P2P activities, which fit the profile of large sustained transfers, are becoming more common. Video and P2P traffic currently accounts for 60% of all data and should grow to 74% by 2013, according to Cisco (Figure 1).

In the early days of the Internet, IP headers clearly marked applications, but a combination of numbering limitations and corporate firewalls has encouraged application developers to mask the type of traffic. Most traffic today looks like Web traffic if you examine only the headers. Once operators determine the application and the subscriber, they can apply a range of policies to ease network contention, implement new data-service plans, or block traffic that violates the operator's terms of service. DPI allows operators to offer tiered data-service plans that they base on a range of criteria. Most mobile operators now offer one data plan and cap data usage at approximately 5 Gbytes/month. One simplistic alternative approach is to offer data plans with size caps of 5, 10, 25, and 100 Gbytes/month, for example. Operators can also offer plans that cap the mobile-broadband speed at levels such as 128 kbps, 256 kbps, and 1 Mbps.

By using DPI, operators can build plans showing an understanding of how customers use their service. You might be able to optimize some sample plans for Web surfing and e-mail sessions of approximately 64 kbps but with a tight bandwidth cap on any P2P traffic. Others might offer a service that allows YouTube-style video streaming at approximately 250 kbps but with limits on high-definition video streaming of 4 Mbps or higher. Another service might attract gamers, offering low latency for gaming packets. Corporations might choose a premium package, which offers traffic priority in any cell site for e-mail, CRM (customer-relationship management), and other corporate applications. A service for P2P users would offer unlimited bandwidth during off-peak hours but tightly cap the P2P bandwidth during peak-usage times.

Technical Challenges

DPI technology is simple in concept but complex in practice. At a conceptual level, looking at a packet to determine the application and subscriber and then taking action on that identification sounds easy. The complexity in DPI arises from network line rates and rapidly evolving applications. The packet rates in carrier networks are staggering. A single 10-GbE (gigabit-Ethernet) channel can support 30 million packets/sec with minimum-sized packets. Even with a more realistic traffic profile and packet sizes of 200 bytes, that 10-GbE channel has 10 million packets/sec.

At that speed, the system has only 100 nsec to receive and inspect the packet, determine its application, modify it if necessary, and send it. Assuming a modern, 3-GHz, single-core processor, this time frame equates to only 300 instructions' worth of execution, which is usually not enough to even receive the packet. This reality has driven the adoption of multicore, multithreaded processors for packet inspection. With 32 cores, or threads, attacking the problem, even at a more modest 1-GHz core-clock speed, that same system now must process only 300,000 packets per core/sec and a more reasonable 3200 clock cycles/packet, enough for inspection, classification, and modification.

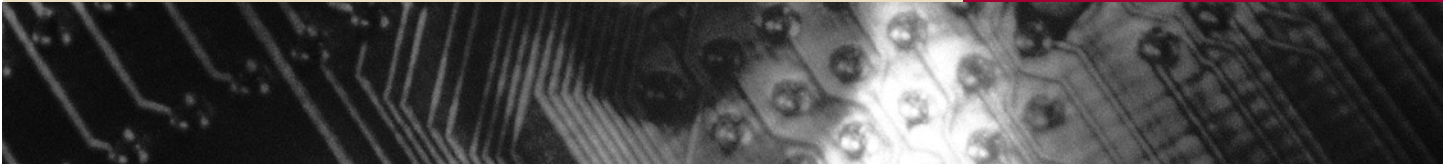
Even with this performance, carriers are demanding systems that can today process 40 Gbps of traffic, increasing to 100 or 200 Gbps by 2011. This trend has pushed the industry toward bladed approaches because a single appliance might handle a few gigabits per second or even 10 Gbps, but not 20, 40, or 100 Gbps. A blade-based DPI system can scale to 80 Gbps of traffic handling, with each DPI blade handling 10 Gbps of traffic.

The second challenge in DPI is reliably identifying applications based solely on the traffic flow. This feature is paramount because the DPI system might be using the application type to set the priority of the packets, decide which packets to drop if congestion arises, bill a customer based on different applications, and even block certain applications. Given the potential impact of misidentification, it is critical to accurately classify as many of the applications as possible and to minimize the number of false application matches. Yet this requirement must take place in an environment in which developers create applications every day and in which different applications are popular in different parts of the world.

To make matters worse, certain applications, particularly some P2P variants, actively try to disguise their identities to thwart corporate firewalls and traffic-shaping systems. Solving this problem requires a database of application signatures as well as a team that can actively update these signatures upon detection of new applications. The DPI-system vendor can address the requirements; alternatively, third parties that specialize in traffic identities are often willing to license application signatures.

Adaptive Traffic Shaping

One example of an advanced DPI application is adaptive traffic shaping. In any mobile network, peer-site cells load others at any time. Cell-site loading depends on location, the number of users, and the types of applications, and it varies over time. As an extreme example, a cell site serving a sports stadium will see huge surges in traffic during a game but is quiet the rest of the time. Similarly, a cell site covering a business park is most active during the day, whereas a cell site covering a suburb gets busier during the evening.



Traffic shaping in wireless networks is inherently more complex than in wire-line networks because of the dynamic of variable loading and capacity over time. Adaptive traffic shaping enhances wire-line traffic shapers by providing them with visibility into this additional dimension, thereby allowing operators to enact policies that employ knowledge of cell-site loading. Operators might block P2P traffic and downgrade video traffic during peak stadium hours, for example, to ensure that enough capacity remains for users to download e-mail or browse the Web. Conversely, bit-heavy services can run at full speed when the network is not busy.

References

1. "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update," January 2009.
2. Roberts, Mike, Future Mobile Broadband, Third Edition, Informa Telecoms and Media, April 2009.



radisy's®

Corporate Headquarters

5435 NE Dawson Creek Drive
Hillsboro, OR 97124 USA
503-615-1100 | Fax 503-615-1121
Toll-Free 800-950-0044
www.radisy's.com | info@radisy's.com