For more information
call +1.877.427.5076
www.helloneustar.biz

10 Keys to Optimizing IP Traffic

By Jon Peterson

Fellow, Neustar, Inc.

neustar™

Executive Summary

Carriers today find themselves at a turning point. Internet traffic relentlessly increases in volume year after year, and today the undisputed driver of that increase is video. In addition to the gradual migration of television and home video delivery to the Internet, the proliferation of sites that host user-uploaded media and steady volumes of peer-to-peer video traffic consume a substantial portion of Internet media bandwidth. With no end in sight, some forecasters predict that the very infrastructure of the Internet will succumb to voracious user demand for media that is of the highest quality and delivered immediately. Some carriers have been forced to implement network management technologies to curtail downloads in areas where resources are strained, but ultimately, carriers stand the most to gain from increased demand and hope to foster controlled growth, not constrain it.

It is difficult to strike a balance between enabling the applications that drive increased demand and preserving the stability and responsiveness of the network. In order to keep the ravenous appetite of users for media bandwidth in check, one must:

1. Understand the Problem
2. Know the Network
3. Follow Usage Patterns
4. Solve for the General Case
5. Implement Application-Layer Traffic Optimization (ALTO)
6. Localize and Conserve
7. Reconcile Policies
8. Enhance the Value of Caches
9. Help Content Providers Succeed
10. Give Users What They Want

In order to meet the growing demands for bandwidth, a carrier must leverage its deep understanding of its own network — the way the network is used, how it is structured, where peering points lie, and where congestion habitually occurs. From the study of these properties of the network, a carrier can advise applications how best to use the available resources of the network. This includes finding ways to avoid congested areas, preferring short paths to media sources rather than long ones, and crossing into other carriers' networks in cases that incur the least cost to the originating networks. Collectively, the technologies that enable Internet applications to leverage these carrier policies are called Application-Layer Traffic Optimization, or ALTO.

To take advantage of the promise of ALTO, carriers must:

1. Understand the Problem

Tomorrow's media will come from many sources — from massive data centers hosting the dedicated media servers of content providers to the humblest home network connections. Tomorrow's media also comes in many shapes and sizes. It may be a single giant movie file put together by a studio, or a real-time interactive audio stream whose contents are made up as it goes along. Neustar, as a player in the Digital Entertainment Content Ecosystem (DECE), knows well that tomorrow's media will be diverse and will appear on the Internet anywhere and everywhere.

From the perspective of the underlying network, the most important quality of media is its elasticity. The delivery of elastic media can be stretched out over an arbitrary period of time, without impairing the user experience (except, perhaps, by making them wait longer for a download to complete). Inelastic media, however, breaks if it is not delivered at a regular and consistent speed. Streaming media is a good example of this type of media, as is a voice over IP (VoIP) call. If, in the middle of a download of a multi-gigabit movie, there is a 10-second interruption in transmission, the user loses nothing except 10 seconds of download time. Yet, a 10-second delay in a phone call would certainly result in both parties hanging up in frustration and retrying the call. Media with different degrees of elasticity can therefore be treated differently by the network, optimizing for the desired user experience.

2. Know the Network

In order to optimize the flow of IP traffic, one must understand the topology of the network. Generally speaking, the more detailed your understanding is of the network, the better you can judge how to optimize traffic. The most effective optimizations find a better path for a particular traffic flow to cross the network — or, if a particular media service is available from several sources on the network, find the optimal source for a particular client. For example, is one peer in New York City acquiring media from a source in Japan when they could be acquiring it from New Jersey? Remember that to applications on the Internet, all IP addresses look alike. If an application has a choice of four IP addresses that are potential sources of media, today there is no standard way for it to tell which one is in Japan and which is in New Jersey. Carriers know better, however; they know which ranges have been assigned to them, and they know where (and through which peer) they send traffic outside their jurisdiction. While no carrier has a detailed vision of the entire Internet, all certainly can differentiate the topologically "local" traffic from that which is not topologically local. Once you can ascertain the topological proximity of endpoints, you can improve on their situation.

3. Follow Usage Patterns

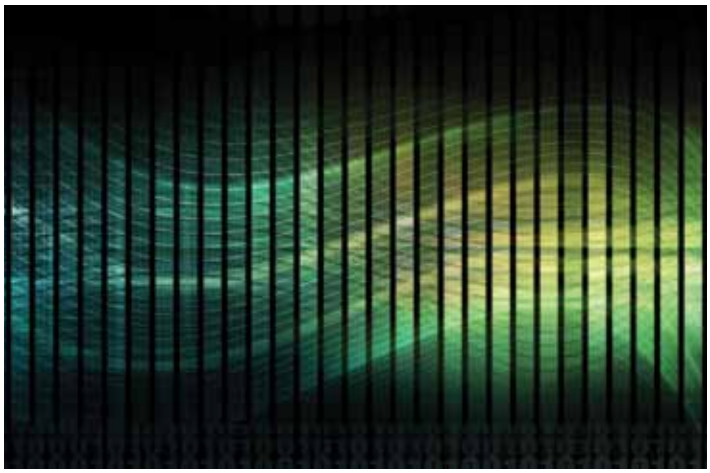
Beyond the topology, one must further monitor and process the state of the network. When problems arise in the delivery of media, do they happen at the edges of the network or in the core? Problems at the extreme edges, such as the last mile to a particular customer premise, cannot be solved by the sort of optimization described in the preceding section; there is no alternative route for the traffic to go in that case. If congestion arises at the aggregator for a number of premises, however, even then some optimization may be possible. The farther you get from the edges, the more alternative routes there are for traffic. If usage patterns are currently sending too much media traffic over an expensive peering point with another carrier, this is precisely the sort of problem that optimization can address.

4. Solve for the General Case

Rather than designing a point solution intended to fix a tactical problem with a particular source or type of media (for example, the BitTorrent application), it is essential that carriers address the general case of optimizing the delivery of media traffic. While VoIP, peer-to-peer file sharing, and massive content delivery networks vary significantly in the elasticity of the traffic they carry and the sorts of places from which traffic is likely to originate, the problem of finding an optimal path from one place on the Internet to another for the delivery of media remains the same. There is simply no way to predict what forms the bandwidth crunch of tomorrow will take, or what applications might cause it. What we can do is provide core infrastructural tools that address the underlying problem, no matter what form media distribution takes.

5. Implement ALTO

The solution to this problem is a general framework for Application-Layer Traffic Optimization (ALTO). Recently, the Internet Engineering Task Force (IETF) has undertaken an effort to standardize an ALTO protocol to support the network management needs of diverse applications on the Internet. The ALTO architecture enables clients to ask topological questions



of an ALTO server in order to discover the optimal way to access a resource on the Internet. For example, if a client wants to download a movie, and it knows of four possible sources on the Internet, it could ask the ALTO service which source would be best. The ALTO service relies on its understanding of topology and usage patterns (see 2 and 3 above) to identify the best path for that media to traverse, or for a set of potential sources to rank them in order from "most preferred" to "least preferred."

An ALTO client might be embedded in a web browser that downloads video, or in a VoIP client, or a peer-to peer file-sharing client — all of which might be operated by an end user or integrated into a media distribution service operated by a content provider. In all of these cases, the ALTO client asks the service the same question: to which network location should this application connect, given the position of the client on the network? Depending on the nature of the application, this question can be posed in a number of specific ways. A web browser that learns the IP addresses of four potential media sources in a content distribution network, for example, could deliver those addresses to an ALTO service along with its own address, and ask the ALTO service to return an ordered list from most to least preferred based on the ALTO service's evaluation.

In order to answer these sorts of questions, the ALTO service must have access to the necessary topological information and carrier policies, both of which must be provisioned at the service by the carrier. These two critical sources of data are described in the following two keys.

6. Localize and Conserve

Topological information allows an ALTO service to determine how close two endpoints are to one another on the Internet. Thus, one key to optimizing IP traffic based on topology is localization — finding media services that are topologically close to the endpoints whose traffic must be optimized. For elastic media traffic, choosing local sources conserves bandwidth on the backbone by reducing the number of bit miles that media travels from its source to its consumer. If a user in New York gets his or her media from Chicago rather than Los Angeles, they have halved the bit miles consumed by their traffic, and thus, between Chicago and Los Angeles, the backbone provider conserves that amount of capacity. These same principles of localization apply on a smaller scale to flows even within a metropolitan area. Figuring out how to find the closest preferred source for media results in savings for the carrier, and typically in faster downloads for customers as well.

Localization results in similar benefits for inelastic real-time services like VoIP. If a VoIP call from New York City to New Jersey needs to pass through a media relay (for NAT traversal purposes, for example), clearly a media relay in Philadelphia would be preferred over one in Singapore. This is another question that an ALTO client integrating into a VoIP application might ask the ALTO service, which shows the benefit of solving the general case.

7. Reconcile Policies

Topological information alone cannot truly capture the preferences of carriers, however. Policy plays an equal role in determining carriers' preferences as to where media streams travel. Carriers may, for example, be charged different rates by different peering partners, and the cost of traversing a particular peering point may change over time or with utilization. Consequently, carrier policies can also inform the ranking decisions made by ALTO services as they order the most preferred sources of media. In some cases, a carrier might prefer that users find a source for media within their own network, for example.

Of course, carriers closely guard policy information critical to their business, and ALTO accordingly exposes to applications only what carriers choose to show. Typically, this means showing only the results of applying the policy and topological considerations to a set of candidate addresses rather than exposing the policy itself. Carriers must also recognize that their peer might not share the same preferences for directing media, and thus different ALTO services operated in different networks might give different answers.

Neustar, as a longstanding trusted and neutral third party in the delivery of communications services, understands how critical policy can be to managing traffic, and recognizes the necessity of reconciling the potentially conflicting incentives of carriers who jointly provide the underlying infrastructure of the Internet. Once carriers begin to express their policies through ALTO, a natural need arises for a neutral third party to help normalize these policies into directions for the network that serve the best interests of users, carriers and content providers.

8. Enhance the Value of Caches

Today, many providers use caching as a means of duplicating a source of media that is topologically closer to the clients that want to download that media. For example, ISPs deploy web caching servers that automatically remember copies of popular web pages, and content providers contract with content distribution networks that house copies of popular media files close to centers of consumer demand. ALTO is entirely compatible with the use of caches, and in fact makes caches more useful. A cache is just another source for a media file that a consumer wants to download, and if the ALTO service learns which caches have the file, it can tell consumers whether or not a cache would be a preferable place to acquire their media.

9. Help Content Providers Succeed

While the difficulties for ISPs in managing peer-to-peer networks motivated much of the initial work behind ALTO, ALTO is a potential key to enabling content providers deliver media to end users under increasingly strained network conditions. One can look at ALTO as a protocol that disaggregates traditional content distribution networks down into components that can be operated independently by different entities (or as part of a vertical silo). ALTO could help a content provider to harness a peer-to-peer network to distribute its content just as well as it can find the best choice among caches operated by ISPs. Consequently, ALTO opens up new ways for content providers to offer their products over the Internet, and with more choices come lower prices and more competitive services. As digital downloads grow into a larger share of the overall business of distributing media (due to the DECE and related efforts), ALTO will play an increasingly critical role in managing the utilization level of carrier networks and ensuring that users enjoy the best download experience possible.

10. Give Users What They Want

Ultimately, users of the Internet have many choices when it comes to consuming media. They naturally gravitate towards services that offer them ease of use and rapid, low-latency downloads. If the results of IP traffic optimization makes their download experience superior, they will use it; if, however, it makes their experience worse, they won't. Balancing the needs of ISPs, content providers and users poses challenges, but within the ALTO framework, users and carriers effectively collaborate to discover the optimal way to reduce carrier load and improve user experience.

The challenges that today's networks face will undoubtedly change as time goes on. Two years ago, peer-to-peer traffic volumes seemed poised to overwhelm the Internet backbone; today, while peer-to-peer utilization remains strong, the highest growth areas have moved back to traditional client-server web applications for video download. This clearly argues for general solutions to problems of constrained bandwidth, rather than solutions restricted to a single application that happens to be problematic today. Optimizing existing bandwidth resources will never take the place of expanding overall network capacity or implementing measures to curtail abuse — but in fact, it complements both of these practices. ALTO provides a versatile framework that plugs seamlessly into new applications — whether tomorrow's "bandwidth hog" is client-server, peer-to-peer or something as yet unimagined, ALTO will provide a way for that application to make optimal use of the existing network resources.

For further information about Neustar's products and services, call **1.877.427.5076** or email **SolutionsTeam@neustar.biz.**

About Jon Peterson

Jon Peterson is a Fellow at Neustar, Inc. and a member of the Internet Architecture Board. He currently co-chairs the Application-Layer Traffic Optimization (ALTO) Working Group of the IETF. Previously, Jon founded the SIMPLE WG of the IETF and served as chair of the SIP WG.

Jon is the author or co-author of more than 20 RFCs, including RFC3261, the core Session Initiation Protocol specification. His work has also extended to numerous other standards and technical coordination organizations, including the ITU-T, the Liberty Alliance, and ICANN, where he served on the Security and Stability Advisory Committee.

Jon's primary interests are immediate personal communications, geolocation, security, and privacy.

About Neustar, Inc.

About Neustar Inc.: Neustar (NYSE: NSR) solves complex communications challenges by providing market-leading and innovative solutions and directory services that enable trusted communication across networks, applications and enterprises around the world. Visit Neustar online at **www.neustar.biz.**

Trademarks