Stay one step ahead

Distributing Policy Management to the Backhaul

Written by Dr. Shahar Gorodeisky
CTO & VP Marketing Celtro

February 2010

Introduction

The objective of the present paper is to demonstrate the need to distribute policy management to the mobile backhaul segment. To elaborate this point, we describe the challenges facing mobile operators in light of rapidly surging volumes of data vis-à-vis declining revenues per bit transferred. These challenges are exacerbated (1) by the limits of the relief that centralized policy management at the core network can achieve, and (2) by the special characteristics of traffic and traffic handling at the backhaul network. Both these considerations mandate the distribution of policy management if meaningful optimization is to be achieved for the overall mobile network.

Between a *Rock* and a Hard Place: Mobile Operators and the Broadband Explosion

In 2009 the rock tune *Don't stop believin'* topped 2 million digital downloads, driving hundreds of thousands of dollars in profit to Apple, Yahoo, and other OTTs. But the operators who carried the gigabytes of data around the globe saw precious little of this profit.

New broadband applications have resulted in an explosive increase in the volume of mobile traffic. The leading *content* category pushing traffic volumes is multimedia, especially video (as provided by YouTube and the BBC iPlayer); the dominant *device* that pulls in all the traffic is the smart phone (especially the iPhone); the *market force* that makes the smooth delivery of the traffic mandatory is user demand for high-quality experience, with best-effort service delivery no longer being acceptable. As a result, mobile carriers are caught literally between a rock and a hard place: the market demands that they supply ever higher bandwidths and better user experience (which come at a high investment), but the price they can charge for the increased traffic keeps falling.

Unable to share in any meaningful way in the fruits of the exploding new broadband applications and traffic, mobile carriers must attempt to maximize their revenue stream by keeping network capacity and traffic in perfect balance and utilizing existing infrastructure resources as close to their maximum potential as possible.

Enter Policy Management

The tool that mobile operators hope will help them balance their books is policy management, one of the important building blocks in the operator's business model. Policy management is a collection of tools and functions centrally located at the core of the mobile network that manage the allocation of existing network resources (bandwidth) to maximize operator income. Policy rule functions (PCRF) and policy enforcement functions (PCEF) draw on a database of information about users, billing, services, etc. to shape traffic and manage network bandwidth. Being centralized, these functions lack real-time information on bandwidth availability at each link and switch across the network, and therefore they are limited to enforcing quasi-static rules of priority and network resource allocation based on such variables as time of day and user requests. At present, policy management is unaware of mobile backhaul status and available resources, and has only limited capability of backhaul resource allocation and congestion management. As a result, its model of operation is to limit users to given capacities by time of day.

The Limits of Centralized Policy Management

At present, policy management tools enforce static bandwidth management rules based on predefined settings and prevailing conditions in the core network. For example, policy management tools can prevent customers of a given service evel from exceeding a certain amount of traffic between, say, 9 and12 in the morning. This is a blunt tool for several reasons. First, it unnecessarily punishes users at peak hours even if the network capacity happens to be available to carry their traffic. Second, it cannot prevent congestion because the backhaul segments of the network are invisible to the policy management tools, and the predetermined restrictions captured in the policy management rules may not be enough to prevent congestion in particular circumstances. Thus, static bandwidth management rules unnecessarily limit network capacity, reducing and degrading user experience, and at the same time cannot prevent network congestion events.

This method is flawed in principle, because ideally, and given the required bandwidth is available, users should be provided unlimited capacity: it is good for customers and good for operators. Users should be limited only if their aggregate demand overuses network resources and threatens to create congestion. Automatic and blind decisions (based on quasi-static rules) to restrict users without real-time knowledge of what is happening on the network are not effective, and are damaging in more than one way. First, users cannot carry on business as usual because of the lack of resources. Second, restricting any given user, or even all users, at specific peak hours contributes minimally to avoiding congestion and its devastating consequences on user experience and network performance. This is because RAN and especially backhaul dimensioning practice calls for overbooking the existing capacity by 10 to 1 or more, which makes congestion scenarios in the backhaul sporadic and very difficult to predict.

The Perils of Congestion

Congestion is the bane of mobile service delivery. Under regular switching, with no smart policing of user flow and applications, when traffic exceeds throughput, packets are being dropped randomly, across the board, so that all users lose packets. Because the traffic of most users, and of **all** mobile Internet users, runs over the TCP transport layer, a congestion event with packet drop closes down of the TCP window, resulting in a sharp decline in each user's throughput as well as in overall network capacity. Therefore, a congestion event produces serious degradation of user experience and network performance.

Theoretically, the ideal solution is to allow users unlimited access, and restrict capacity only in problem areas and only for as long as the problem persists, in other words, to implement smart congestion control. To do so, it is imperative that the policy management function will have access to the real-time status of the network and that smart congestion control be implemented in highly congested areas of the network, with policing and shaping capabilities per user and per application. Because most bottlenecks occur in the backhaul portion of the network, without distributing policy enforcement to the backhaul, policy management cannot be effective.

Aggregation and Overbooking on the Backhaul

With HSPA and LTE technologies, the radio access portion of the network, between the base station and the end user's mobile device, has been successfully keeping up with demand for capacity, and the bottleneck has shifted to the backhaul segment. Operators have found that on heavy sites, at any given moment, the relative load is by far the highest on the backhaul among all network components.
The backhaul is a complex network of aggregation points, with interconnected segments assembling traffic from a distributed network of pipes. At each aggregation point, the larger pipe is "overbooked" based on the statistical assumption that the smaller pipes feeding into it are not all being utilized to the maximum at the same time.
Because of the random nature of traffic, it is impossible to accurately predict congestion events in the backhaul. The only solution is to track backhaul traffic in real time and manage backhaul resources using a smart congestion control mechanism.

Solution: Backhaul-aware Policy Management

Advanced policy management is reaching into new areas. At one end, it is expanding into the IT realm, where it manages not only bandwidth but also users and services, monitoring customer choices and managing personalized services based on dynamic variables. But this is not enough. To deliver the required customer experience at a cost that mobile operators can afford, policy management must be aware of the real-time status of the network and must change from the blunt tool it is today to a surgeon's scalpel that can intervene with pinpoint precision in all areas that require corrective action, including the backhaul network.

Advanced distributed policy management replaces the quasi-static bandwidth management rules with dynamic ones that take into account service usage and real-time bandwidth requirements on the backhaul. It uses smart policy enforcement points that determine the real-time availability of backhaul network resources and perform intelligent congestion management to improve network performance and user experience for scenarios of network congestion. The smart backhaul policy enforcement points communicate with the central policy management function, delivering backhaul resource availability and receiving user flow applications and priorities. The central policy management system uses the real-time backhaul status to achieve service admission and BW management based on actual available network/backhaul resources. When congestion does occur at the backhaul, the smart policy enforcement points provide intelligent congestion management by policing and shaping traffic based on policy management level information per user flow priority. This type of congestion management can restrict heavy, low-priority users (e.g., P2P traffic or high-volume downloads) during congestion periods, allowing the traffic of others to flow unaffected through a congested link or switch. Transferring most user flows unaffected under congestion, and shaping only the flows of a small number of low-priority, heavy users

also increases network capacity and overall performance by reducing network traffic oscillations that occur when many user flows are degraded indiscriminately.

Under an advanced distributed policy management regime, smart policy enforcement points within the backhaul network are connected to the centralized policy management and extend policy management to the backhaul. The smart backhaul policers alert the policy management function about the real-time availability of backhaul resources. In case of coming congestion, they perform smart congestion management by policing and shaping a small number of heavy users precisely where they can cause real damage and exactly for the period needed, giving all other customers access to unlimited bandwidth. This reduces congestion, ensures fair usage, and limits the number of degraded customers. Above all, it keeps the network close to 100% utilization.

Within current technology, this is the best hope for smart distributed policy management that applies fair usage policing and shaping only for congestion avoidance, otherwise allowing users to utilize unlimited BW.