Beyond data caps

An analysis of the uneven growth in data traffic

By Monica Paolini

Senza Fili Consulting

0 Introduction

The growth of mobile data traffic over the last few years and projected for the near future is breathtaking (Figure 1), following a curve similar to that of the fixed internet in the 1990s. The latest Cisco Visual Networking Index (VNI) estimates that mobile data traffic grew by a factor of 2.6 during 2010, and the index forecasts a 26-fold increase over the next five years. At the end of that period, growth may decelerate as traffic generated by individual subscribers stabilizes, yet it should remain strong because the total number of subscribers will keep increasing.

Most of the discussion has focused on the overall increase in data volume and on the network wide tools available to meet the new level of demand. As a result, most of the action has been limited to introducing traffic caps or diverting traffic to Wi-Fi. Wi-Fi off-load has been a great boon to mobile operators. However, in the long term and as it is managed today, Wi-Fi off-load alone will not be able to shoulder the expected increase in data traffic.

Caps are more problematic. Generous caps can address abusive usage, but do not significantly affect

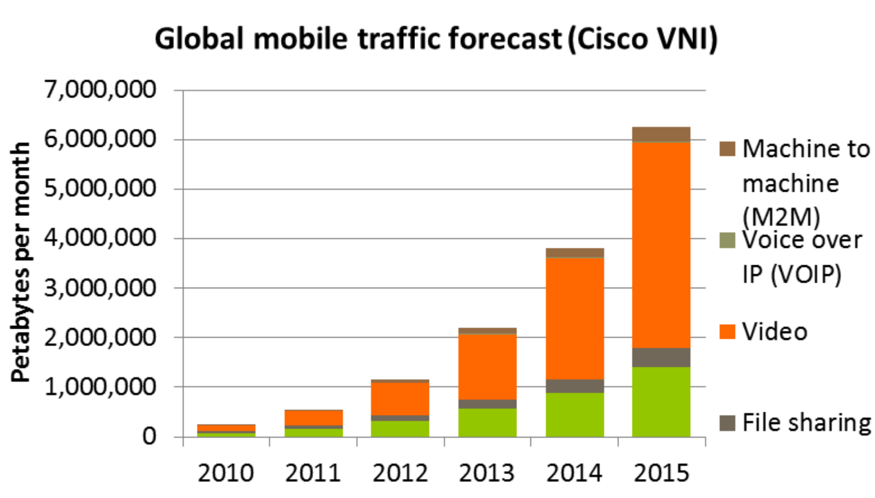

Figure 1. Global mobile data traffic forecast (Cisco VNI). Source: Cisco

legitimate use. Restrictive caps may generate resentment among subscribers, especially those used to unlimited internet plans, and risk dampening demand and use of mobile services at a critical phase in market growth. Perhaps even more importantly, restrictive traffic caps are ineffective at coping with data growth, because they target any type of data usage, regardless of time, location, or application, whereas operators need to selectively control the traffic going through the high-pressure points in their networks, where base stations are at or near capacity. Traffic elsewhere costs the mobile operators nothing. Even worse, it is quite possible that traffic caps have a stronger impact on off-peak usage, as subscribers are likely to engage in less urgent or valuable activities at that time—the type of activities that would be eliminated first if subscribers need to curb their data usage.

The inadequacy of traffic caps stems from the fact that growth in data traffic is not uniform across the network, or across subscribers. Instead, narrowly targeted solutions are required in different environments to address the specific characteristics of data traffic.

To determine which solutions are needed—and when and where—to cope with increasing traffic loads, we need to take a closer look at the distribution of traffic growth and where pressure points are within the overall network. We do this in the following pages by examining seven factors, each with specific traffic distribution dynamics, and the impact they will have on network traffic requirements. The analysis of the best suited solutions to address the challenges that these factors present is the topic of a subsequent report by Senza Fili Consulting.

1 Geography

Growth in third-generation (3G) wireless networks has been driven largely by developed markets. Only when established there did it move to developing markets—and the process is still ongoing. The initial explosion in the adoption of mobile data, too, has started in developed markets—but the dynamics are markedly different. Emerging markets are catching up more quickly, and they show a huge potential for data services growth in the short term.

Growth rates in emerging markets are stronger than in developed markets. The Vodafone group has recorded its strongest revenue growth rate in the last year in Turkey (99%; for comparison, in both Germany and the UK, the rate was 27%, and Spain's was 26%). The Cisco VNI predicts that the compound annual growth rate (CAGR) of data traffic over the next five years will be 101% for the Middle East and Africa, and 91% for Latin America, compared to 32% in Japan and 41% in North America (Figure 2).

One reason is that smartphone penetration is still low, limited mostly to users who are not price sensitive (top-end smartphones in some markets cost US$1,000); in addition, these users generate higher traffic volumes on average. For instance, among WiMAX operators, Yota's mobile subscribers in Russia generate on average more traffic than subscribers in the US (Clearwire) and Japan (UQ), even though they use similar devices and subscription plans without traffic caps.

Growth in emerging markets is set to expand rapidly to wider consumer segments as soon as the network infrastructure is deployed, and more affordable devices and data plans become available. This growth is partly driven by the lower availability and penetration of the fixed internet. For many subscribers in emerging countries, the mobile internet is or will be the first and only internet connection they have access to from a device they own. Furthermore, emerging countries often have a stronger demand for content, because it is less readily available through alternative channels like broadcast or cable TV than it is in developed markets.

Growth in emerging markets is going to have a salutary effect on mobile operators in developed markets. To be profitable, operators in emerging markets have to be ruthless in keeping costs low, and innovative in offering services that attract a broad subscriber base that is extremely price

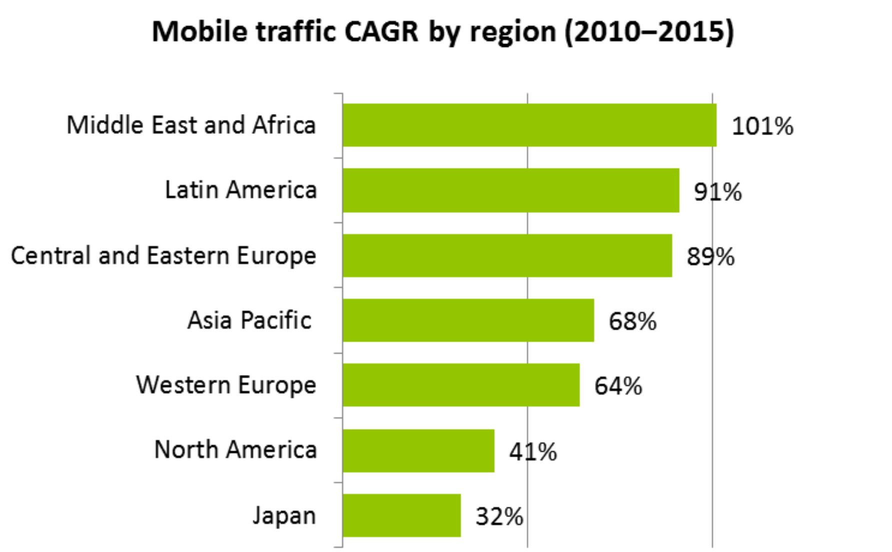

Figure 2. Mobile traffic CAGR by region (2010—2015). Source: Cisco

sensitive. This will pressure vendors to develop infrastructure solutions that will lower per-bit costs, which will benefit mobile operators in developed markets as well.

It is unlikely that data plans will have unlimited usage, in emerging as well as in developed markets. Comfortable data caps that are not punitive are needed as a protection against abuse—and they are indeed common among wireline service providers as well, even though their network resources are not as constrained as those for wireless networks. However, to keep plans affordable and to segment the market efficiently, mobile operators need to do more than offer plans with different traffic allowances. We expect to see more emphasis on application- and service-based packages, different service classes, and more prepaid options than currently available in developed markets.

2 Subscribers

Similarly to what happens in the fixed internet, a few subscribers account for a disproportionately high percentage of mobile traffic. In Europe, 6% of Vodafone subscribers account for 54% of overall traffic (Figure 3). The thriftiest 75% of subscribers use less than 200 MB per month and account for only 17% of traffic. A similar distribution is found across most mobile operators. In the US, for instance, 5% of users account for 68% of traffic, according to Sandvine.

In some markets, tiered services have been introduced to increase average revenue per user (ARPU) among heavier users, and to create lower-price plans that entice new, more price-sensitive subscribers. Most of these plans differ only on traffic allowances, even though for many subscribers the availability of a reliable broadband connection is more important that traffic allowance per se. For instance, light users may less frequently use bandwidth-heavy applications but be more sensitive to service availability and speed—for example, if they are business users with the need to receive or

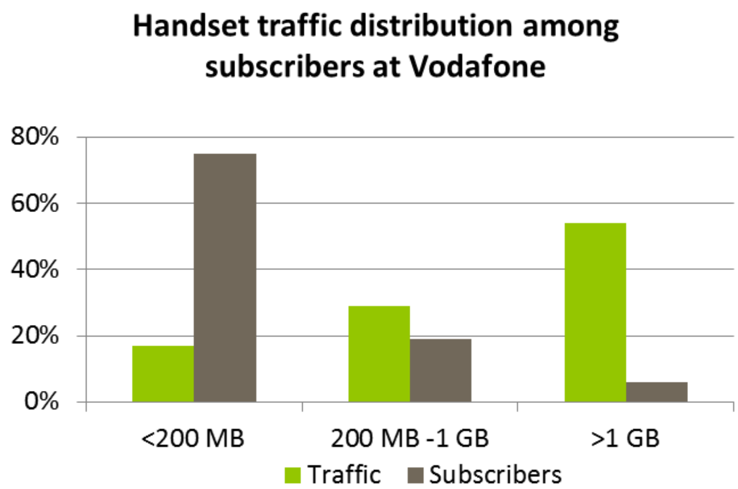

Figure 3. Handset traffic distribution among subscribers at Vodafone. Source: Vodafone

send information quickly. These users are likely to be willing to pay more for prioritized access, but not necessarily for large traffic allowances.

Subscribers who use their mobile internet connections primarily for entertainment, however, may prefer to have a higher traffic cap, but may be willing to accept a limited service availability at peak time—i.e., slower connections, lower video resolution, or even no video availability—if this means a lower-price plan.

It is still common to hear that the mobile internet is driven by business users—in most cases the point is made by people in the telecom sectors, who happen to be business users themselves. While it is true that business users are the main revenue contributors today on a global basis, the usage pattern is rapidly shifting toward consumer usage, and growth prospects in that segment are the more promising ones in developing markets.

The distinction between the consumer user and the business user is becoming less predictive of usage—and as a result less useful. Subscribers increasingly use the same device for both work and entertainment (Figure 4), which results in a more even distribution of activity across the day (more on this later).

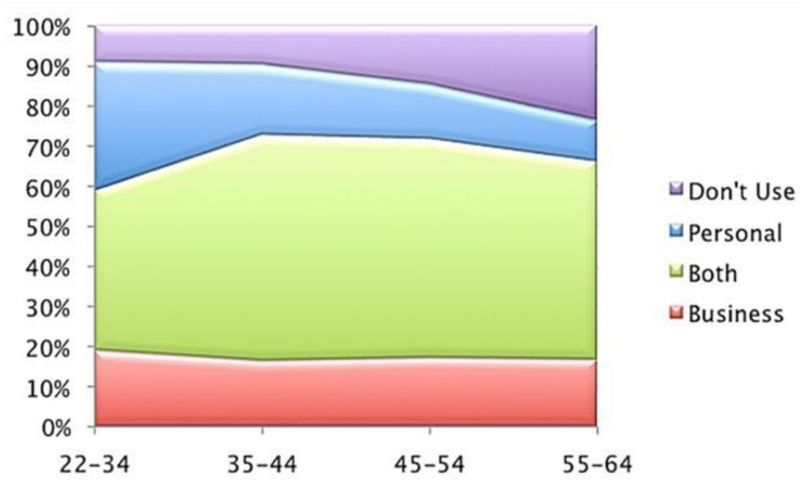

Figure 4. Smartphone adoption and use by age. Source: iPass

3 Devices

Devices also account for huge differences in data load usage across subscribers, with laptop subscribers generating by far the most traffic volume (Figure 5). At Vodafone, laptops account for 85% of the overall data traffic, according to Andy McLeod, the company's Director Group Networks.

Mobile operators are rightly ecstatic about the growth of smartphone penetration—and more specifically about smartphone adoption among feature phone and laptop users. Smartphones generate substantially lower traffic loads than laptops[1] and higher ARPUs than feature phones. That

1. One worrisome feature of smartphones is that they can overload the signaling channels. The intensive background processes required

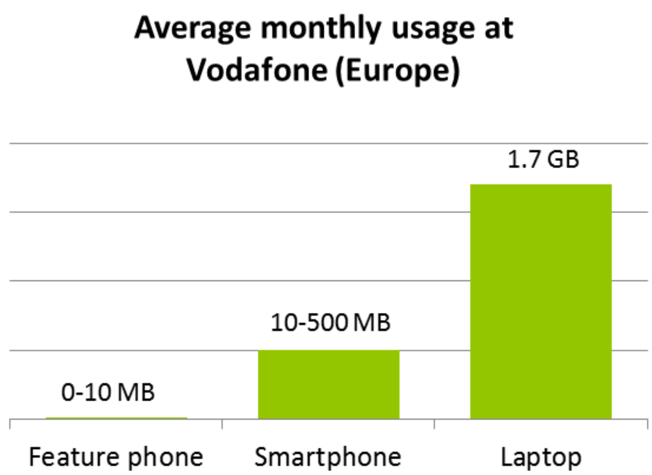

Figure 5. Average monthly usage at Vodafone (Europe). Source: Vodafone

means that laptop users are the least profitable subscribers, and in many cases they are not profitable at all—yet they play a major role in causing congestion.

Figure 6 shows how data profitability is intrinsically tied to widespread smartphone adoption: as subscribers adopt smartphones, the revenues per GB will on average grow. One reason is that the decreasing percentage of mobile broadband laptop subscribers will account for a decreasing percentage of traffic. Furthermore, as already mentioned, early smartphone adopters mostly belong to the high-tier segment that is not as profitable, because of the high usage levels[2]. As smartphone penetration increases and affordable smartphones are introduced over the next few years, the mid-tier smartphone segment will become dominant, increasing the overall smartphone market profitability.

by smartphone applications (e.g., email, software updates, some social networking and location-based applications), combined with the need to keep battery consumption to a minimum, lead to the frequent establishment and tear-down of connections. The resulting signaling traffic can overload cellular networks if operators do not take the necessary steps to allocate the signaling resources as the penetration of smartphones increases.

2. Even within the smartphone category, there has been some traffic differentiation, with iPhone users generating more traffic than other smartphone users in the past. However, to a large extent the higher traffic load from iPhone users reflects differences in demographics, market segment, and price sensitivity. This is demonstrated by the fact that Android smartphones that target a similar audience drive similar usage profiles.

As a further advantage, the increased penetration of smartphones may result in a deceleration in the growth of overall traffic. Until recently, subscribers needed a laptop dongle to have mobile broadband access. Since the introduction of the iPhone and other smartphones, this has changed dramatically, increasing the range of devices that provide a good mobile Internet experience. Smartphones and new devices like tablets will, to some extent, become alternatives to laptops. They are easier to carry around and more convenient to use, and require expensive monthly plans. In developed markets, smartphones account for half or more of device sales, and their share among subscribers is growing quickly. In June 2010, smartphones accounted for 21% of mobile devices in the US and laptops for 4%, according to the Cellular Telecommunications and Internet Association (CTIA). At Vodafone, the current penetration of smartphones in Europe is 17%, and the operator expects it to reach 35% by 2012–2013.

Device	Usage	Profitability
Feature phone	Low	Similar profitability
High-tier smartphone	Medium/high	
Mid-tier smartphone	Low/medium	Higher profitability
Laptop	Very high	Lowest profitability

Figure 6. Data profitability by device and usage. Source: Vodafone (adapted)

4 Applications

Despite the initial worries about peer to peer (P2P) and voice over Internet Protocol (VoIP) traffic dominating the mobile internet and cannibalizing revenues, it has become clear that it is video (e.g., YouTube, Netflix, or Hulu) and other real-time streaming content (e.g., Pandora or Spotify) that have become the main traffic sources. This is true across geographies, devices, operators, and technologies (Figure 7 with most data from cellular operators; Figure 8 for an example of a mobile WiMAX network). Video traffic is expected to grow to 66% of overall traffic by 2015, according to Cisco VNI. The portion of P2P traffic is getting smaller, with P2P traffic share decreasing by a fourth over the last 18 months. This may be because P2P traffic is more prevalent on laptops and the share of traffic from other devices is growing, combined with the increased awareness among subscribers about

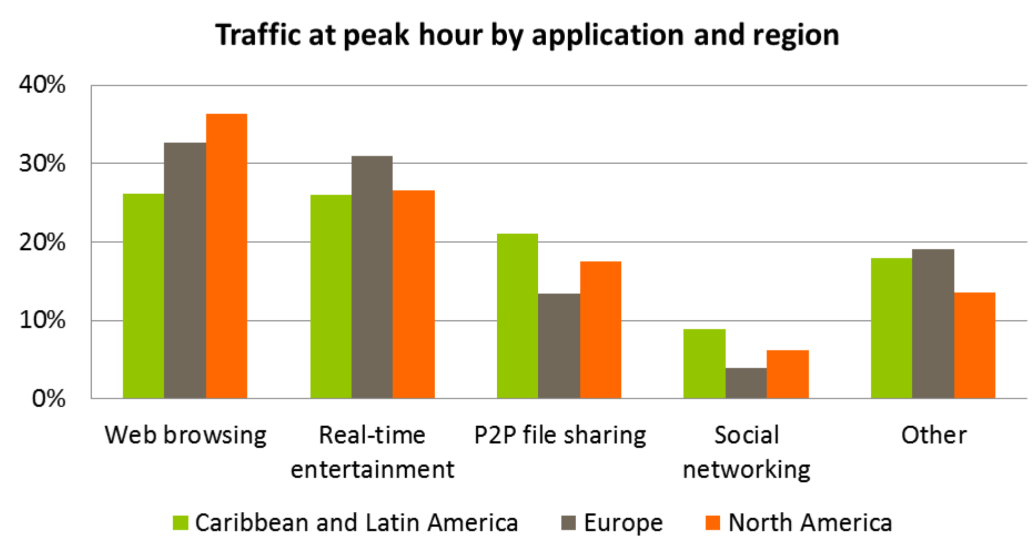

Figure 7. Traffic at peak hour by application and region. Source: Sandvine

traffic caps.[3]

Real-time applications account for a large percentage of traffic but a relatively small proportion of time spent online, because they are much more bandwidth intensive than applications like email or social networking. In addition, the real-time applications (video or otherwise) burden network resources more than applications like email, messaging, or web browsing that are more tolerant of latency, jitter, and bandwidth variation.

To provide a satisfactory user experience in locations that are at or near capacity, operators increasingly have to use a wide array of traffic management tools, such as deep packet inspection (DPI), policy, quality of service (QoS), and traffic optimization and compression, as well as subscriber management tools like traffic caps, throttling, and increasingly tiered services. Of course, increased radio access network (RAN) capacity is often necessary, too, but the potential for massive increases in video downloads, especially if high-definition videos are streamed, creates a big incentive for operators to go well beyond the RAN to manage and optimize video and other real-time content, on the basis of policy, service plan, and, perhaps more importantly, device.

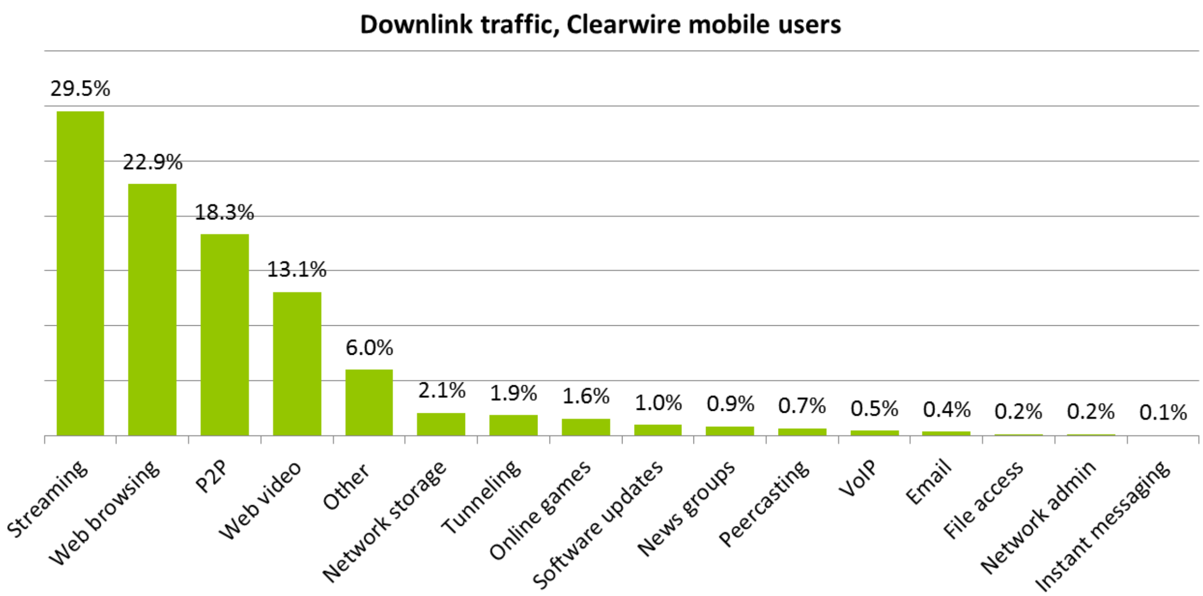

Figure 8. Downlink traffic, Clearwire mobile users. Source: Clearwire

3. As a survey from Tekelec shows, subscribers do not assign high priority to P2P applications, and they may be more willing to voluntarily reduce P2P than other types of traffic.

5 Time of day

Data usage is not uniform throughout the day. Base station traffic load peaks around 9 to 10 pm (Figure 9), and the time of the peak is surprisingly constant across countries. The uneven traffic distribution through the day means that only part of the base station capacity is used, but to accommodate demand, mobile operators have to size their networks to peak traffic.

The ratio of peak to average traffic is a crucial factor when a mobile operator computes the cost per bit to build and operate the network. If the distribution of traffic were flat, the cost per bit would be substantially lower. Conversely, any traffic that is added to nonpeak times (and that of course does not exceed the peak threshold) uses resources that otherwise would remain idle, and therefore it is effectively free to the operator.

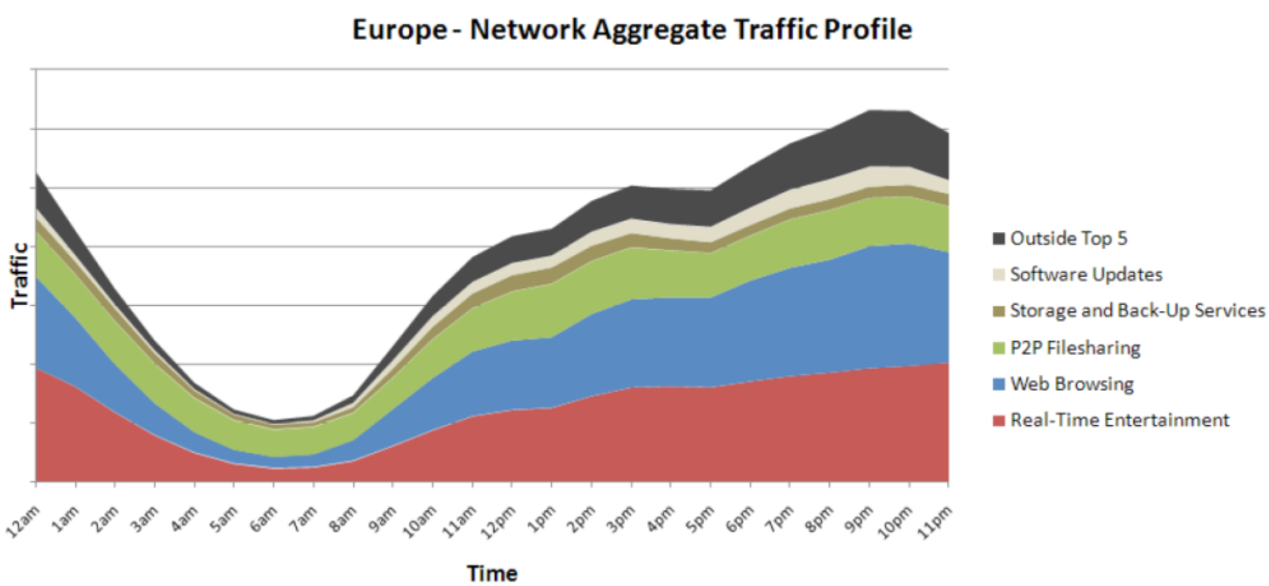

Figure 9. Europe—Network aggregate traffic profile. Source: Sandvine

6 Location

Mobile operators strive to provide coverage across the highest percentage of their territory that is financially justifiable. Yet population is usually concentrated within small portions of a territory, and people often use their data services more intensely within an even more restricted set of locations where these services can be accessed most conveniently—such as airports, stadiums, shopping malls, and coffee shops.

When deploying a new network, mobile operators first aim at establishing coverage, and then add base stations to increase capacity where needed. For many mobile operators ubiquitous Long Term Evolution (LTE) coverage is not a requirement, because they rely on previously built 2G and 3G networks to provide wide-area coverage, and use new technologies like LTE for high-traffic areas. Inevitably, however, mobile operators cannot maximize the utilization of all their base stations, because space availability and coverage requirements typically restrict the selection of cell sites, and because subscribers' movements can be only partially predicted.

The combination of time-of-day and location accounts for what is often perceived as a low network utilization of cellular capacity. At Vodafone's Western Europe operators, for instance, only 7% of sites run at 90% or higher capacity during peak hour. Across the network, the average busy-hour utilization was 38% during fiscal year (FY) 2010–2011, up from 35% in FY 2008–2009. Given that today's networks are mostly planned for coverage, these utilization levels indicate an efficient use of network resources for a mature network. These rates are likely to be lower for new entrants with a smaller subscriber base, for operators in less densely populated countries such as the US or Canada, and for operators that have a lower market share (but still have a footprint comparable to larger operators in their markets).

Traffic concentration in a small part of their footprint is good news for mobile operators. Most of their networks (in terms of number of base stations, or coverage area) are still minimally affected by the data traffic increase. As a result, they need to concentrate on capacity enhancement in just a limited part of their network to meet the expected traffic loads in the short term. This is likely to change in the longer term as traffic continues to rise, but in the meantime operators have the opportunity to experiment with different approaches, and to spread out over time the capex funding needed for infrastructure upgrades.

7 Indoor/outdoor

Cellular networks excel at outdoor coverage. Indoor coverage is more difficult to achieve. In densely built areas or where indoor coverage is essential (e.g., in airports, malls, or subway stations), mobile operators often deploy indoor equipment, such as picocells or distributed antenna systems (DASs).

Even more so than voice traffic, data traffic mostly originates from indoor locations. This is unlikely to change substantially, because we do spend much of our time indoors and because, for many applications and devices, indoor usage is simply more convenient—e.g., there is less glare, it is easier to find a place to sit down, and safety is less of a concern. According to a study from Cisco, subscribers use mobile data services 40% of the time from home, 25% from work, and 35% from public locations—with at least 80% of the traffic coming from indoor locations. Jaime Lluch Ladron of Telefonica expects that 95% of data traffic will come from indoor locations in a few years' time.

This is bad news for mobile operators. Not only is it more difficult to provide reliably good indoor coverage without installing equipment in the buildings to be covered, but where there is coverage, mobile devices in indoor locations connected to an outdoor macro base station typically use more network resources than devices used outdoors, other things being equal, because indoor devices use a modulation that is less spectrally efficient. Indoor usage reduces the base station throughput and, consequently, increases the capacity requirements for cellular networks.

The evening peak mentioned earlier suggests substantial residential usage. This is a rather unexpected outcome, as it was initially predicted that wireless internet connections would be mostly used by subscribers in a mobile or nomadic environment—i.e., outside the home or the office, where most subscribers lack a wireline broadband connection, which is often cheaper and faster. Why would subscribers use their smartphones to look at a YouTube movie at home, when they have a faster connection from their desktop or laptop? In most cases, subscribers find their mobile devices more convenient, more accessible, and faster—the cellular connection may have less bandwidth than the wireline connection, but it is usually faster to check email or watch a short video on a smartphone than on a laptop that sits in the office and has to be rebooted.

Increased residential usage helps mobile operators, because subscribers' homes tend to be in residential areas that are not as crowded as downtown areas. This makes it possible to off-load traffic to the home Wi-Fi network or to femtocells that use the wireline broadband connection instead of the cellular network to backhaul traffic, thus relieving the need to add base stations in the neighborhood.

In the context of increased residential data usage, data caps are largely inefficient, as they may have the effect of lowering residential usage (because there users have the option to switch to their wireline connection) and of increasing the pressure on high-density public location (where subscribers would still use their connection without having to worry about data caps). But for mobile operators, it is cheaper to increase capacity in the homes of their heavy users with Wi-Fi off-load or femtocells than in public locations that are already affected by traffic congestion.

Concluding remarks: why data caps are not enough

The explosive growth in data traffic is not the result of a ubiquitous and evenly spread evolution in usage patterns. Its uneven distribution across many interacting dimensions—geography, devices, applications, location, subscribers, time of day, and indoor/outdoor environment—adds remarkable complexity to the process of identifying the most effective solutions to increase data throughput in cellular networks.

There is no single solution that can alone address this issue. Data caps, while needed to preserve an even distribution of network resources among subscribers, are not well suited to address the sources of data traffic increase.

The goal for operators is not to decrease overall traffic levels. This would be a risky strategy as it may alienate the very subscribers mobile operators want to entice to use mobile access. Instead, mobile operators need to increase the efficiency of their network so that they can accommodate increasing levels of traffic while keeping the per bit costs at a minimum. Traffic caps at best cut traffic across the board, thus delivering no improvement in the network efficiency.

A key strategy for increasing efficiency (and thus ability to support higher traffic loads) is to increase network utilization by trying to flatten the time-of-day curve (e.g., encouraging off-peak use for non–time-sensitive applications like P2P or providing financial incentives for off-peak access), increase use at locations where traffic can be off-loaded (e.g., at home) , and create more locations where Wi-Fi, picocells, or femtocells off-load is available.

Changes in the RAN are not sufficient to provide the increase in efficiency needed. An equal effort is required to manage traffic more actively, using tools such as DPI, policy, and video optimization—taking into account real-time network conditions, subscriber preferences, operator policy, devices, and applications.

The complexity may be daunting. It requires profound changes in the way mobile services are provisioned and managed. Gone are the days of the price per minute or buckets of minutes. But the increased understanding of the sources of traffic gives operators the ability to attack the problem on multiple fronts, giving them more flexibility and room to innovate.

About Senza Fili

Senza Fili provides advisory support on wireless data technologies and services. At Senza Fili we have in-depth expertise in financial modeling, market forecasts and research, white paper preparation, business plan support, RFP preparation and management, due diligence, and training. Our client base is international and spans the entire value chain: clients include wireline, fixed wireless and mobile operators, enterprises and other vertical players, vendors, system integrators, investors, regulators, and industry associations.

We provide a bridge between technologies and services, helping our clients assess established and emerging technologies, leverage these technologies to support new or existing services, and build solid, profitable business models. Independent advice, a strong quantitative orientation, and an international perspective are the hallmarks of our work. For additional information, visit www.senzafiliconsulting.com or contact us at info@senzafiliconsulting.com or +1 425 657 4991.

About the author

Monica Paolini, PhD, is the founder and president of Senza Fili. She is an expert in wireless technologies and has helped clients worldwide to understand technology and customer requirements, evaluate business plan opportunities, market their services and products, and estimate the market size and revenue opportunity of new and established wireless technologies. She has frequently been invited to give presentations at conferences and has written several reports and articles on wireless broadband technologies. She has a PhD in cognitive science from the University of California, San Diego (US), an MBA from the University of Oxford (UK), and a BA/MA in philosophy from the University of Bologna (Italy). She can be reached at monica.paolini@senzafiliconsulting.com.