Disruptive Analysis

Don't Assume

Towards a new generation of holistic DPI & policy management technology

Introducing a series of Disruptive Analysis *thought-leadership* papers

Commissioned by Continuous Computing (now Radisys)
November 2010

Author: Dean Bubley
Contact: information@disruptive-analysis.com

Background

This document is intended as an introduction to a new series of three separate papers on policy management and the role of Deep Packet Inspection (DPI) in mobile broadband networks.

The basic premise is that first-generation "silo" approaches to controlling data traffic have been highly fragmented, and aimed at solving urgent "fire-fighting" problems – the next generation of solutions will need to be more "holistic" and joined-up. Shifts in network technology, smartphones, regulator environment and applications and use cases are together prompting a re-think.

As the industry looks more strategically at how to manage the data flowing across 3G/4G networks, it is apparent that the different elements will need to work together much more closely, in order to create compelling user experience, and enable new business models. The radio network, backhaul, core, devices, applications and billing/charging systems will need to be brought together to optimise the overall mobile data ecosystem. Trying to force onto users policies that seem unreasonable, inflexible and arbitrary will lead to dissatisfaction and churn.

In particular, future policy management platforms will need to be:

- Bearer- and radio-aware
- Device-aware
- Capable of managing complex and contextual offload

The paper has been written by the independent industry analyst & consulting firm Disruptive Analysis, and sponsored by Continuous Computing (now Radisys), as part of an initiative to promote thought-leadership, differentiation and innovative networking concepts for the mobile broadband and network policy-management marketplace. The opinions expressed are Disruptive Analysis' own, and are not specific endorsements of any vendor's or operator's products or strategy.

Introduction

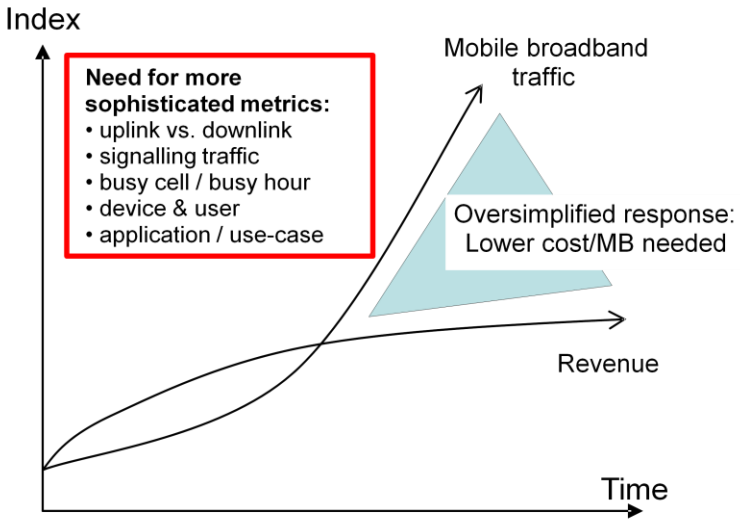
Over the past two years, everyone in the mobile industry has seen the famous "scissors" diagram, with mobile data traffic growing exponentially, despite revenues only creeping upwards slowly. Most regulars to industry conferences have seen variations of the same picture probably dozens, or even hundreds of times. Disruptive Analysis is as guilty as anyone else here, having used that same image over several years, with the simplistic conclusion that either "costs per Megabyte" need to be reduced, or that total volumes need to be managed down.

But while that chart has been useful for raising awareness of the trends and risks, it is becoming clear that it represents a quite naïve and unsophisticated analysis of the actual situation, which is much more complex. Many operator networks have been impacted more by signalling load than actual "tonnage" of data. Applications such as P2P BitTorrent can be more damaging because of relatively high *uplink* traffic rather than the sheer amounts downloaded.

And just because 80% of *traffic* comes from 20% of users, that doesn't necessarily mean that they cause 80% of the actual problems or 80% of the real-world congestion. The person downloading 2GB of video in quiet cell at 3am is probably not one of the 20,000 people with smartphones, each downloading 1MB of email at the airport at 8am the next morning.

The risk to operators is that they focus on the easy targets, the so-called "data hogs" rather than actually identifying and fixing the real pressure points. Yes, there is some correlation between outright heavy usage and costs, but cause-and-effect is far from perfect. Applying tiers and caps makes prudent sense compared with truly "unlimited" data plans – but heavy-handed approaches to overage charges, or harsh network discrimination against particular traffic types (e.g., video) risk a range of negatives – from consumer dissatisfaction and churn, through to unwelcome scrutiny from regulators. Creating fear amongst customers is never a winning strategy – especially when it seems arbitrary or vindictive, rather than fixing specific problems.

Figure: The data "scissors" chart: too simplistic for 2011 onwards



Source: Disruptive Analysis

Shifting applications and usage cases

Underlying the need for better traffic management is the blistering pace and extent of the adoption of mobile broadband. Initially catalysed by cheap 3G dongles and the first iPhones in 2007-8, the past two years have seen a further explosion of both data-consuming devices (smartphones, growth of Android, iPad, etc.) and demand for high-bandwidth applications such as video.

These shifts have changed the patterns of usage of data networks on an incredibly fast time-scale. Whereas voice traffic tends to grow steadily and predictably over time, data demand tends to be "spiky" and much less easy to plan for. It is also used in different places to voice – and these locations can also shift with application fashions. Given constraints on CapEx, spectrum and cell-sites, this means that networks can easily get overwhelmed by sudden new trends.

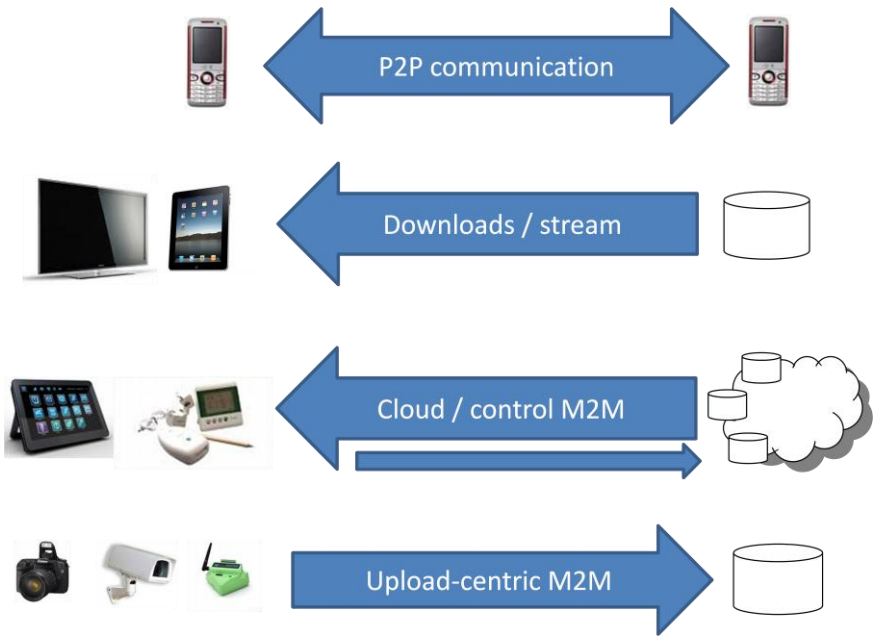
This is not just about new classes of devices, either. It reflects the move to particular computing and communications models, which might incorporate devices or handset software as elements. At a lower level, problems might be caused by new software being adopted rapidly – perhaps the "viral adoption" of a new and easily-downloaded application by millions of users, or even just a seemingly minor update to a handset OS which changes the profile of signalling traffic.

Realistically, operators cannot predict many of these changes, nor plan their networks to accommodate all possibilities. Neither can they expand their capacity or coverage fast enough to match these trends, some of which (e.g., OS updates) can happen literally overnight. Some form of more active monitoring and/or management of their effects is therefore necessary.

Looking to the future, the potential shift to "cloud services" aimed at mobile users will have a further profound impact on networks, as they inherently depend on good connectivity to better-distribute processing between end-point and server. Such services may be either operator-controlled, or open-Internet based. Either way, more granular control of the network will be desirable.

Other shifts are also anticipated, notably the vision of tens of billions of machines and other new (mostly data-consuming) devices added to the network. These will have yet more effects on resources, further mandating a shift in the old style of deploying, managing and upgrading mobile networks.

Figure: Mobile data flows will evolve further with new devices and M2M



Source: Disruptive Analysis

Ultimately, the core of problem is that the mobile industry is suddenly being exposed to the full power of Moore's Law. Devices are becoming much more powerful – some smartphones have GHz processors already – while the Internet and cloud / Web 2.0 domains being enabled by virtualisation and blade servers.

So, even if the network is "smart" rather than merely "dumb", it will still need to be improved in capability, if it is to cope when sitting between genius-grade devices, and almost god-like clouds.

Radio and network technology evolutions

Along with usage models and devices / applications, the other obvious driver of policy management sophistication and DPI is the evolution of the networks themselves, especially the radio domain. In the past, mobile data networks have been comparatively simple, such as the broad use of HSPA (High Speed Packet Access), on normal cell grids, at 2.1GHz in much of the world.

We are now moving to a much more layered and heterogeneous environment:

- Introduction of Long Term Evolution (LTE), but in a very fragmented way, with many frequency bands, FDD and TDD (Frequency and Time Division Duplexing) variants, and often only in certain locations.
- Rollout of femtocells and other small-cell variants
- Use of WiFi as a complementary indoor access technology, and for offload
- Refarming of 2G spectrum for 3G and LTE
- Certain operators using WiMAX
- Ability of devices to use multiple radios simultaneously, or connect to multiple cell sites
- Focus on power management and energy efficiency, both in the infrastructure and for devices
- Solutions for coping with massive increases in signalling traffic
- Overlays optimised for M2M (Machine to Machine) applications
- Richer wholesale and network-sharing models
- White-space and band-sharing models of spectrum management

Taking all these together – as well as inevitable future enhancements as well – it is clear that this diversity will have a huge impact on the optimum way to structure and organise data traffic and services, as well as how it is supported, billed and monitored. Ultimately, there will need to be some measure of "awareness" of all these variables built into the system.

This will happen gradually and progressively, but it highlights just how much intelligence will be needed in the mobile broadband value chain to maximise revenues and customer experience, while controlling costs and performance.

Regulation & its impact on network policy

A further driver for future mobile networks to have flexible policy infrastructure relates to the constantly-shifting legal and regulatory situation. Most countries around the world are still wrangling about Net Neutrality, lawful interception, privacy, Internet universal service obligations and other contentious topics, which have a direct impact on how telcos run their networks. While it is probably impossible to have a network architecture that could accommodate all future options, there still needs to be consideration given to likely shifts. Given the typical political / electoral cycle, and the onward march of both technology and consumer opinion, it seems doubtful that we will ever reach a completely steady-state situation.

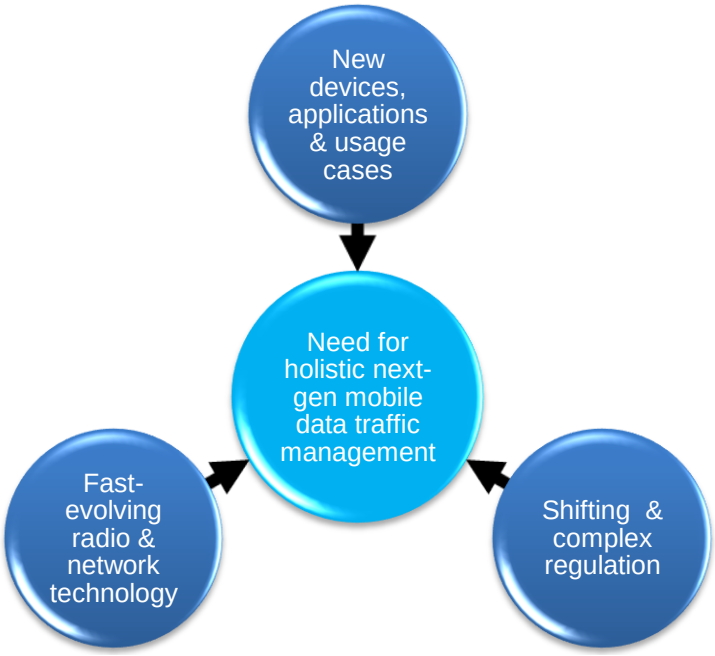
The debates tend to have distinct regional characteristics, though. A number of markets are moving towards permitting active traffic management, but only that which is the minimum required to operate the networks smoothly. This might put the onus on the operator to monitor and prove how and when congestion occurred, and that their actions to manage it were indeed proportionate and reasonable. Other nations may allow more invasive management of particular flows – but mandate clear transparency to end users, and accurate reporting. Some other countries have very strict and authoritarian national laws on permissible applications or usage.

It is also important to not that regulatory shifts are not solely about *restricting* operators in how they treat data. Other developments are likely to yield new opportunities and underlying challenges. For example, many countries are now moving to allow carriers to pool and share parts of their network – backhaul, radio access and so forth. This will require sophisticated policy management, especially where two or more notional competitors share part of a network, yet each wishes to manage its own partition differently.

Conclusion: holistic traffic management

Up until now, it has been possible for operators and their vendors to use simplistic approaches to traffic management – caps, tiers, throttling and quite shallow inspection and analysis of data flows. Often, the debate around traffic management gets mired in emotive discussions around Net Neutrality and the ethics of blocking or favouring certain websites or services.

Figure: Three trends driving the need for holistic traffic management



Source: Disruptive Analysis

This introduction – and the three more detailed papers in the series – show that the challenges and opportunities are much deeper. The interplay between the "big picture" trends is going to mean much greater need for intelligent DPI and advanced traffic management techniques, irrespective of the superficial controversies.

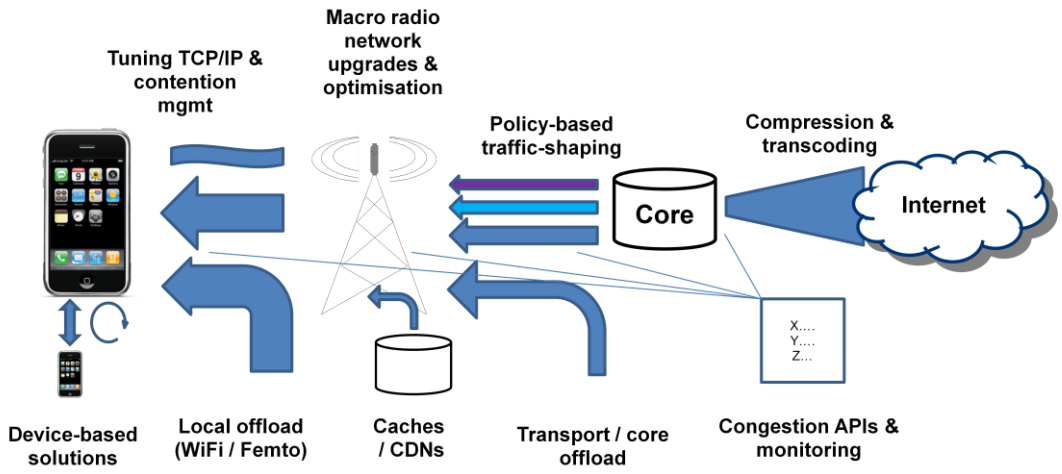
In general, the older approaches to controlling data traffic have been relatively simple from a technology standpoint – for example using "hard" approaches to tiering and thresholds, or perhaps some historic information about busy cells and times of day to apply throttling to particular users. Various compression and optimisation servers have been used at the main interface between the core network and the Internet.

Over the past two years, virtually every vendor in the industry has tried to contribute to solving the "traffic tsunami" problem, but generally just focusing on their own small silo of the network and extended ecosystem. The radio specialists have looked at optimisation, the device companies on WiFi and connection managers, the DPI vendors on particular "localised" examples of policy control.

What is going to be needed are much more intelligent, fine-grained solutions to mobile data traffic management and policies. These will need to search out the *real* root causes of any problems, and apply *proportionate* fixes, in a way that generates acceptance and even genuine loyalty from users and application/content providers. Given the rapid pace of development of radio technology, continued evolution of handset capabilities and the appearance of new usage scenarios and applications, this requirement will likely continue in perpetuity.

These more intelligent approaches to traffic management will need to be what could be termed "holistic", embracing *multiple* touch-points in the network and down to the device. This series of white papers examines a range of these scenarios, tying in the status of the handset, the actual congestion of the radio network, and the varying options for offload.

Figure: Holistic traffic management will link diverse silos & elements



Source: Disruptive Analysis

Ultimately, there is another challenge that parallels the technical issues outlined here: how to represent all this added complexity to the customer. How much can operators use tariff structures and "open" policies to affect user behaviour? And how much can mainstream users really understand, if confronted with complex choices about devices, offload, applications and times / locations? Often, many of the areas discussed in these papers will need to be "behind the scenes". Understanding the psychology of mobile data use is at least as difficult as getting the technology right.

About Disruptive Analysis

Disruptive Analysis is a technology-focused advisory firm focused on the mobile and wireless industry. Founded by experienced analyst Dean Bubley, it provides critical commentary and consulting support to telecoms / IT vendors, operators, regulators, users, investors and intermediaries. Disruptive Analysis focuses on communications and information technology industry trends, particularly in areas with complex value chains, rapid technical / market evolution, or multi-sided business relationships.

Currently, the company is focusing on mobile broadband, network policy-management, operator business models and services, voice and personal communications applications including VoIP and IMS, smartphones, Internet/operator/vendor ecosystems and the role of governments and regulation in next-generation networks.

The company produces research reports and white papers, conducts consulting projects on technology strategy and business models, and provides speakers and moderators for workshops and conferences.

For more detail on Disruptive Analysis publications, workshops and consulting / advisory services, please contact information@disruptive-analysis.com

Disruptive Analysis' motto is *"Don't Assume".*

Website: www.disruptive-analysis.com
Blog: disruptivewireless.blogspot.com
Twitter: @disruptivedean

Intellectual Property Rights / Disclaimer

Every reasonable effort has been made to verify research undertaken during the work on this document. Findings, conclusions and recommendations are based on information gathered in good faith from both primary and secondary sources, whose accuracy it is not always possible to guarantee. Disruptive Analysis Ltd. disclaims all warranties as to the accuracy, completeness or adequacy of such information. As such no liability whatever can be accepted for actions taken based on any information that may subsequently prove to be incorrect. The opinions expressed here are subject to change without notice.