

To Flat or To Cap?

No that's not a question about Australian coffee tastes and the critically important difference between a flat white and a cappuccino. This is a question about the differences in ISP retail models for broadband Internet access and the choice between a retail model of a "unlimited" flat fee that has no volume component, and a "capped" model where the service fee provides for a certain data volume and when that volume is reached either the user is exposed to an incremental fee, or the service is throttled back to a narrowband service for the remainder of the billing period. It seems that this is once more a critical question in the ISP world, and maybe this time the topic is best approached through television.

I don't think it's a surprise to anyone, but it's the Christmas season again and doubtless a large number of television sets will be sold as part of the annual retail festivities. But these days the devices for sale in the shops are not just televisions: today's television is perhaps better described as a media computer with a very large display. Sure, the device can tune in to radio transmissions and display them, as one would of course expect from a conventional television, but the device also is equipped with either a WiFi or an Ethernet jack, or both. This alone sounds like a relatively innocuous addition to the television, but it's providing to be a highly disruptive change in the traditional Internet market space. Behind that network interface lurks a highly capable computing environment, with embedded applications and services that turn the television into a highly capable communications device. And embedded into the device is a set of interfaces into a world of streaming video content, all provided over the Internet. All this is starting to be all very disruptive to a number of broadband access providers business plans.

As an illustration of just how disruptive this can be, it's interesting to review some events that occurred South Korea at in February 2012. At that time Korea Telecom (KT) made public its quite surprising move to block Samsung's "Smart TVs" from downloading streaming content over KT's consumer broadband network. In essence, KT's blocking move transformed the device back into a "dumb" TV, and needless to say neither Samsung, nor the hapless consumers who had purchased these devices to use with a KT broadband connection, were overly impressed.

South Korea is a country that proudly proclaims its effective saturation of its domestic population with high speed broadband access services, and rightly so, as this is a notable achievement. Megabit speeds are common and these days experimental deployments of a gigabit access service are underway in parts of the country. So it's not without some small element of surprise to hear a KT representative claim at a recent OECD meeting at the OECD that deploying a device that actually makes use of this broadband infrastructure is in some fashion "unfair," even "damaging," and indeed so "damaging" to the network that KT felt it necessary to pull the plug on these devices.

Evidently, according to Korea Telecom, Samsung opted to take a "very negative response" to KT's actions. Samsung obtained a court injunction to lift KT's block on their TVs and an associated order

for KT and Samsung to enter into arbitration. At the same time Samsung filed a lawsuit against KT. In due course the temperature of the dispute abated and all the parties backed down. KT discontinued its block, and Samsung dropped its lawsuit. However, there was evidently some residual bad feeling here as Samsung expressed their desire for the national regulator to convey a "strict warning" to KT over its actions.

What KT was after is quite simple: they were insisting that Samsung, and other local "Smart TV" vendors in the Korean market, must pay a levy to KT to have their devices deliver their content over KT's broadband access network. Predictably, Samsung officials said in response that they had no intention of paying KT for network access for their devices.

Samsung remains publicly confident that the Korean regulatory position will continue to support its position, but it raises a larger spectre across the generally buoyant Internet consumer content industry. The threat here is that if the incumbent carrier is able to carry out its threat and block these devices from the network unless the manufacturer comes to a prior agreement with the carrier to pay some form of levy, then it would set an unfortunate precedent that would have repercussions across the entire Internet. This contretemps potentially extends way beyond Samsung and KT, and draws in LG and Panasonic and also potentially draws in Microsoft and its Xbox, the Sony Playstation and the Apple TV, to mention some of the more prominent vendors of the current generation of streaming content devices.

Video is by no means novel and video over the Internet is also by no means novel. Why has this become an issue in 2012? Why didn't it surface years ago with, say, the emergence of YouTube in 2005?

A combination of various factors is certainly placing some new pressures on local Internet access infrastructure, and the shift from broadcast television to streaming video is central to the picture. Are these carriers' claims of "overuse" of the network justified? Just how much data does a streaming video TV pull through the access network in order to generate a picture?

Today's television sets are typically 1080 lines with a 16:9 widescreen aspect ratio, so that the screen is 2.1 megapixels in size, with a display rate of some 24 frames per second. Without compression, using a three color system with 16 bits per color, this display is equivalent to a data rate of 2.4Gbps in a raw (uncompressed) format. A typical video codec can reduce this data rate considerably, and a high quality HDTV video stream can generate a sustained data stream of between 10Mbps to 20Mbps through a high quality codec, although it is more common to see compressed HD video content using streaming rates of some 3Mbps - 4Mbps. Even this rate is of course far higher than that used with the small format video streams used for video display on computer screens, which are typically around 10 times smaller at 300Kbps - 500Kbps.

As well as the data volume, there is also the factor of the transport protocol used to pass the streaming video data to the consumer. Streaming protocols are not exactly the most social of protocols on the wire. They are typically based on the Real-time Transport Protocol (RTP), and typically use a unicast UDP streaming transport model. Unlike DCCP (DCCP is not feasible due to the high density of deployment of basic edge firewalls which would effectively filter out DCCP as an unrecognized transport protocol) or even TCP, these unicast streams do not conventionally perform any form of congestion-based data adaptation.

That's not all in terms of the factors that make the current round of streaming video content uncomfortable for the network. Consumers tend to behave in similar ways, such that there are pronounced peak periods in the day. Like the physical transportation infrastructure in a city, it's not the average load that matters. What matters for most users is the match of the peak load to the infrastructure capacity. In the case of these "smart" TVs the peak load considerations in the network occur the 6pm to 9pm evening time slot.

A further factor here is that consumers tend to treat a "smart" TV as a TV and not as a computer, and when they leave the TV, they tend to leave it running, rather than shut down the video stream when they are not viewing it. This is particularly prevalent where there is no marginal cost associated with leaving the streaming device on, such as is the case in a flat rate tariff environment.

And finally there is the factor of the access provider's network provisioning model. Access networks are not engineered on a zero contention model. When an access provider connects 100 consumers each with a 100Mbps broadband service, then it is not the case that the feeder network would be provisioned with 10Gbps of back end capacity dedicated to these 100 consumers. While published details of the precise nature of the engineering in access networks are scant, contention ratios of 100:1 are not uncommon in this area, where one unit of back end feeder capacity is provisioned for every 100 units of access capacity delivered to the consumer. While gigabit networks are now in the area of commodity systems, higher speeds in the back end of these access network, such as 100Gbps, just do not exist, and even 40Gbps systems attract an unwelcome price premium simply because they are some years ahead of the technology curve. And the older the broadband deployment, such as is the case in Korea, the more likely that the back end networks tend to lower speeds and higher contention ratios. So while the outward statistics of the broadband network may look impressive, with provisioned speeds of up to 100Mbps, the contention ratio in the provisioning model may be very high, so that if every consumer attempting to pull down 100Mbps of content at the same time the network would simply be unable to cope.

If the back end of these broadband access networks are so heavily over-committed, then why was this not such a public problem for many years? Much of the answer lies in the evolution of usage of the network and the difference in behaviour between TCP and UDP protocols.

For many years the Internet was a predominately TCP network. The main data volume in the network was various forms of file transfers that were parts of web pages. parts of a peer-to-peer file sharing network, shared data sets, or just about anything else that involved the movement of data from one machine to another. None of these applications were "real time" applications, and in general the network transactions that passed these data elements around were based on the TCP protocol. Recent years have seen a shift in data volumes on the access networks such that video streaming has supplanted all other applications as the major application by data volumes on the access network in many parts of the world. And video streaming is a UDP application.

So if we look at protocol behaviour for a second, TCP is a rate adaptive protocol, and over the course of long held sessions multiple TCP users tend to equilibrate their use of the common network and each TCP session receives an approximately equal share of the constrained network resource. One TCP stream cannot "shut down" any other TCP stream. Under conditions of network congestion each TCP application will reduce its data transfer rate to a level that alleviates the congestion pressure. This is generally not directly visible to the consumer, in so far as the vagaries of second-by-second file transfer rates are not normally prominently displayed as part of the user's interface to the network. In essence, TCP performs its rate control function quietly and with direct visibility to the end user. On the other hand, UDP has no such adaptive flow controls, and in a video streaming context each stream will attempt to push a relatively constant data rate through the congested network bottleneck irrespective of the congestion level in the networks through which the streams are being pushed. Network contention implies lost UDP data, and in the case of video streaming this lost data means compromised picture quality in the streaming video application. In other words this saturation condition in the back end of the access network becomes highly visible to all video streaming users.

The Carriage Perspective

KT, like many carriers, has its own IPTV service, but this service is evidently not madly popular with consumers. The IPTV service is conventionally modeled on a broadcast TV model, where a single

stream is fed to all consumers simultaneously via multicast. This is distinct from the content streaming model, which is more like a DVD library model where each consumer can program their own content in their own time. The content streaming video models have proved to be extremely popular with consumers, but now there are no carriage efficiencies to be had. Instead of multicasting a fixed number of IPTV channels through the network each consumer is receiving their own unique content stream. Consequently, video streaming traffic levels are on the rise in the carriage network, and this has some potentially interesting implications on the contention levels in the back end of broadband access networks.

This shift in the consumers traffic patterns with high definition streaming video content and smart TVs is presenting new challenges for the carriage provider of broadband access services. It's no longer a case of a conventional "heavy tail" of distribution, where 10% of customers are responsible for 80% of the traffic, such as was the case when file sharing was the predominate traffic component and the so-called "super seeders" were the high profile users of the network. In a streaming content environment the peak profile of usage is such that a much higher proportion of consumers are consuming large traffic volumes at peak times, and the access network is failing under congestion load during these peak usage periods. In other words as well as observing a small number of users contributing to the average traffic volumes for the entire network, we now also see a broad set of users equally contribute to the peak traffic load, and now the claim is being made that this peak traffic load is overwhelming the network's capacity and compromising service quality for all the network's users in these peak periods.

In many consumer markets we are used to the good being sold using an incremental tariff. Purchasing two apples will normally cost twice the amount of a single apple. More generally, if a consumer consumes a greater quantity of the good, then the consumer is charged a higher tariff in proportion to the quantity consumed. The higher tariff provides incentive for the producer to produce more of the good, and the market equilibrates the unit price of the good between the consumer's perception of value for a given quantity of the good and the producer's estimation of an efficient production price. But where the good is sold on a flat rate basis, such as unlimited flat rate broadband services (such as those retailed by KT), then these conventional market incentives do not work. The consumer is incited to consume more, as there is no marginal cost associated with consumption, but the producer is motivated to produce less, as there is no marginal revenue associated with higher demand. Where demand rises in a flat rate tariffed market then, according to one industry presentation I heard last week, the producer reports that: "we have a decoupling of revenues from traffic."

The obvious response to this escalation in traffic volumes would be to construct higher capacity back end subsystems in the access network. But in a flat rate tariff environment the business problem is that any such investment in the network would be funded by their existing revenue margins, as the flat rate tariff implies that for as long as any increment in network capacity is consumed by the existing customer base then the costs of that increment in infrastructure capacity are being funded by the business, not by the customer base.

An obvious response would be to introduce volume-based tariffs, or "data caps" as they are often called. In some markets, such as Australia, this retail model of data caps is so widely used that unlimited flat rate offerings are viewed with some suspicion as being of compromised quality by the consumer.

In other markets, including Korea, the flat fee model is so ubiquitous across the broadband retail market that the claim is made that any attempt to introduce data caps would be tantamount to commercial suicide, or so the operators in these markets claim. They see the introduction of any form of volume-based retail tariff to be simply not an option for them. So deeply held is this opinion that in these "flat rate" markets a number of carriers are trying to engage the content industry in what the carriers would call "cost sharing" models, but the prospects of any mutually satisfactory outcome from such engagements are dim at best.

It seems like these flat rate access service providers have managed to wedge themselves between a rock and a hard place. On the one side we see the content providers exploiting this flat rate tariff structure

with a streaming video content model that imposes high volumes of data upon the back ends of their networks at peak times, and they claim that this additional traffic does not generate any revenues from the customer base, so any efforts to add further capacity to their networks is going to be funded by the carriage provider out of the existing revenue margins. On the other side they firmly believe that efforts to introduce data caps after many years of operating on a flat fee structure would push their customer base to use competitor's products and drive them into business failure for this particular activity, as well as attracting a strongly unsympathetic consumer reaction, which would be a public relations disaster.

The Content Perspective

What about the picture from the other side? A good a perspective as any comes from Netflix, an entity that has wholeheartedly embraced streaming video content delivery. Today, according to a report from Netflix to the same recent meeting at the OECD, Netflix has more than 30 million customers, predominately in the Americas, but also in the United Kingdom, Ireland and, most recently in the Nordic countries.

The Netflix offering is a flat fee system that allows the customer to stream videos without incremental cost per session. They have taken an earlier DVD library model, where the entire library is available to the customer and reproduced this in an online environment. This model has been so successful commercially that Netflix is now following in the footsteps of HBO in producing its own content, releasing the entire series at once, allowing the customers to select how they wanted to view the series.

Not surprisingly Netflix's business model is based on a retail broadband offering that is essentially an unlimited flat fee offering. In this way the customer is not exposed to any incremental marginal cost in choosing to watch streamed video content, as compared to broadcast, cable or DVD material. And equally unsurprisingly Netflix is supportive of a position that is opposed to the introduction of volume caps in retail broadband tariffs in those markets where Netflix is active, and on a consistent theme, Netflix strives to ensure that in an ISP's offering streamed Netflix content is not tarified within the volume caps.

Of course Netflix is not the only such "over the top" service provider in this area, and there are now a number of such streaming providers offering services in these markets. The increasing prominence of this form of service in the market place the more weight is placed behind the pressures for flat fee based broadband services, or at the very least exemptions for video streaming services from volume caps.

Netflix, like many content provides, appears to be strongly resistant to any suggestion that they subsidies or fund the delivery of content to the user. They argue that they have already funded their Content Distribution Network (CDN), and at their cost have brought content close to the user through the deployment of CDN Points of Presence at major exchange points and the execution of peering arrangements with those service providers willing to enter into such arrangements. I have heard the content folk argue that to enter into financial relationship with a service provider with whom they do not necessarily have any direct network interconnection relationship, and over traffic flows that are initiated and maintained by the service provider's users rather than by the content provider, would conventionally be considered as extortion in other contexts.

To Flat or Cap?

The recent calls for the introduction of "sender pays" into the network's commercial landscape, championed in recent times from European Telecommunications Network Operators (ETNO) show that KT's perceived plight is not just an isolated case. It appears that many of the broadband access carriage providers, perhaps notably those that invested heavily in "triple play" and other forms of

bundled IPTV offerings, are finding that the own business models are foundering. The flat fee is not covering their costs of investment in network infrastructure, and the expanding data volumes arising from the shift towards content streaming models bypass the service providers own multicast IPTV models just as much as they bypass traditional broadcast TV. The expectation that users would augment the basic flat fee offering by spending money on the service provider's own premium content offerings have proved ill-conceived. The associated business structure that assumed that this premium content income stream would cross-subsidize a loss-leading flat fee entry tariff is proving to be an expensive mistake and yesterday's highly fashionable triple play is proving to be today's toxic liability. Consumers simply purchase the low-priced flat rate tariff and then purchase their content from third parties who evidently provide a combination of a massive array of popular content and extremely low flat rate fees.

It seems that when you have a business failure of this scale there are a number of options available, but some appear to be more sensible than others. One is to try and get the content providers to take the place of the service provider's own premium content offering, and force these entities to cross-subsidize the providers' basic flat rate broadband access tariff by paying the service provider to allow the service provider to pass the content to the consumer. And if the content industry is unwilling to pay then perhaps its time to invoke a regulatory impost, as seen by ETNO's recent attempts to introduce this measure into the negotiation process leading to the redrafting of the International telecommunications Regulations (ITR). But perhaps its folly to wander about making the claim that "you must pay for my poor choice of business models!" to anyone within hearing distance. Perhaps the problem here is that a poor choice of business models requires a change of the business model. If consumers are the source of revenue for a broadband access network then conventionally the tariff's levied on consumers need to cover the cost of the business. And where the consumer makes a greater use of the network by generating higher volumes of data on the network then there is a compelling case to expose this marginal cost to the customer. In the case way as my other utilities, such as water, electricity and gas are metered services, then perhaps an economically efficient model for the utility role of the provision of packets is by a metered service.

In any undistinguished commodity market, where there are incremental costs associated in the production of the good being traded in the market, then the long term prospects for a provider who addresses the market with a flat fee schedule are not good. The flat fee model provides little incentive to moderate the consumption of the good, and overconsumption causes failure of the provider's ability to sustain the production of the good. This is a situation that has a lot in common with the "tragedy of the commons".



The Internet has often been compared to the Commons, where a communal resource was owned by noone, yet it was commonly used to the benefit of all. It is not the concept of the commons itself that has become entrenched in our vocabulary, but the aspect of the "tragedy of the commons", where the unmanaged common resource was abused to the point of destruction. Each individual user stood to gain more through increasing their use of the common resource, and, as there was no governance of each individual's use of the resource, there was no penalty imposed for overuse. No single person

or entity was responsible for the proper maintenance of the commons and the cumulative problem of degradation of the resource to the point of collapse was not a problem that any individual user was equipped to tackle.

In old English law the "commons" were areas of land that were held in common by the general population, "the commoners," as opposed to specific tracts that were held by the nobility. The grounds may have been pasture lands, woodlands, or open space used by the general population. The word "commons" is derived from Latin "communis" and means the quality of sharing by all or many.

Fourteenth-century Britain was organized as a loosely aligned collection of villages, each with a common pasture for villagers to graze horses, cattle, and sheep. Each household attempted to gain wealth by putting as many animals on the commons as it could afford. As the village grew in size, more and more animals were placed on the commons, and the overgrazing ruined the pasture. No stock could be supported on the commons thereafter. As a consequence, village after village collapsed.

The analysis of this in a social context was explored in depth in the 1960's. These papers can be found at <http://dieoff.com/page95.htm>

The failure here is a failure of the flat fee access model. However, the underlying failure might possibly be attributed to a failure to a deep appreciation that the Internet is far more versatile than the telephone network it replaced, and the dynamics of change are a constant factor in the behaviour of the Internet. To base a network's engineering, and its business model on a single model of network use, and to assume that this will not change rapidly over time is perhaps the real folly here. To assume that carriage and content are so inextricably interwoven that when a consumer purchases a carriage product from a provider that the same consumer will be bound to also purchase premium content services from that same provider is part of that same folly. And to bind the two together in an intricate web of structural cross-subsidization simply adds to the problem, rather the offering any form of sustainable solution.

A commercially viable carriage provider needs flexibility to respond to changing usage patterns. When carriage providers use inflexible business models that trigger situations where revenues are disconnected from traffic volumes, what we are seeing is a knee jerk reaction to blame the generator of the increased traffic volumes, and try and make the content providers repair the revenue gap. However, such an approach does not have overly bright prospects for lasting success. It may be a more challenging, but a more sustainable, approach to expose to the consumer those points where incremental costs are incurred by the carriage provider by using a tariff structure that includes various forms of volume-based parameters., such as are used in a capped tariff structure.

However you look at it, a broadband access carriage industry response to smart TVs and the increasing proliferation of "over the top" video content streamers of "Your innovation has broken my business plan! You owe me money!" is not going to go anywhere productive. If the carriage provider's business plan is not working then perhaps its time to look at what went wrong and how it might be corrected, rather than blame someone else.

Disclaimer

The views expressed are the author's and not those of APNIC, unless APNIC is specifically identified as the author of the communication. APNIC will not be legally responsible in contract, tort or otherwise for any statement made in this publication.

About the Author

Geoff Huston B.Sc., M.Sc., has been closely involved with the development of the Internet for many years, particularly within Australia, where he was responsible for the initial build of the Internet within the Australian academic and research sector. He is author of a number of Internet-related books, and has been active in the Internet Engineering Task Force for many years.

www.potaroo.net