Income-based billing:
Dealing with limited capacities by valuing the customer's time

For many in the telecommunications industry, the harsh reality remains that a decade after data started its inexorable march into the network, billing is still about little more than getting some money in the door. But, as George Huitema and Doug Zone ask, for how long can the telecoms industry ignore the lessons from energy and mass transit (among other verticals)? These lessons suggest that the real win isn't just bringing in the revenue but influencing end-user behaviour so that you can maximise margins.

The authors are George Huitema, Senior Strategist, TNO, (pictured top) and Doug Zone, Chief Technology Officer, MetraTech Corp

Any commercial enterprise will say, "if you have unlimited capacity, you maximise the amount of that service that people consume." But if you have finite resources and don't commercialise them accordingly, people will use them right to the edge of their availability and the result – over-consumption – inevitably becomes a problem.

Privatisation and the advent of ubiquitous competition in voice services have led the telecoms industry to forget that, at its root, it shares the advantages and perils of any utility as well as corresponding business models. The crisis in capacity driven by the iPhone is not dissimilar to 'brown outs' caused by affordable air conditioners. Unfortunately, while most utilities remain regulated monopolies – making capacity investment an easy and profitable remedy – the telecommunications industry is typically made up of lightly regulated duopolies with considerable commercial freedom – capital expenditure is not guaranteed a return.

Dysfunctional models
Purely revenue- or cost-based billing and settlement models are dysfunctional in an environment where there is a limited capacity such as Cloud, wireless broadband (3G, 4G, etc.), or energy. Income-based billing – recognising both revenue and cost management – which is largely forgotten in telecoms, is needed. The telecoms industry, which to date has steadfastly failed to get to grips with the problem of limited capacity, should care about this greatly.

Billing strategies over the past 10 years have evolved with the advent of the network as a sales channel for bandwidth-consuming services. Though these strategies are income-based – with an objective to maximise margins

(sales price less the cost of goods sold) – revenue-based billing has sufficed as the number of units sold if anything reduced unit costs (i.e. via volume purchasing agreements).

Subscription prices climb
Network services on the other hand, which are key to the industry's future as a shop front for the virtual world, deteriorate as the network gets close to capacity. This is more than a quality issue – it directly speaks to the economics of providing the service. As the number of bytes (or cloud services or kilowatts) delivered per minute falls, relative price goes up. Though the actual price of subscribing to get access to the network remains constant, the cost goes up as you approach the edge and network speeds slow – even on 'all you can eat' plans with one monthly fee.

Why? Simply put, time costs money. Every minute a consumer waits on the network, he could be doing something else. Given that once a monthly subscription is paid a user perceives the next download as essentially free, the only economic price of using the network is the opportunity cost of time.

If demand is not controlled by explicit 'by the minute' pricing the mechanics of a service will implicitly manage demand, regardless of the service provider's intentions. Though cost perceived by the end user gets higher and higher (in time required rather than money), the service provider gets no additional return as the real price stays the same.

Threat to telecoms
The telecommunications world must grasp the potential threat this situation poses: when this happens – and increasingly, it is happening – eventually the end user will churn as he seeks a ▶

supplier who is less expensive. The original supplier's strategy to maximise revenues by maximising the number of subscribed users will at this point become self-defeating because, when capacity nears, churn will become the result of 'success'.

To make matters worse, churn will hit for those users that have the highest perceived opportunity cost of time – those that lose the most by waiting around – businesses and high income consumers. Maximising revenues thus becomes a pyrrhic victory. What can the telecoms industry do? One way of addressing the dilemma would be to set the subscription price high enough to moderate demand. But this only achieves one thing – it scares off subscribers who now cannot afford the service. It does not change the price, which remains essentially free once the subscription is paid.

Another way is to place 'quotas' on the service – making the service free up to a fair use limit and then essentially cutting users off with punishingly high tariffs thereafter. Economic theory shows that from a consumer's point of view this is sub-optimal. It equally impacts users that really need the extra capacity and those who don't but will use it if it is available – the former being the most valuable customers.

Of course, one approach might be to charge per unit: in the case of 4G 'by the byte', the Cloud 'by CPU utilised', energy by the kilowatt, etc. The argument would run, "if people consume too much, I'll raise the price."

That's a workable policy in theory but when tested in the real world (in California with energy services for instance) it proved to be a disaster because, while on average energy service provision was good, at peak times it was awful and extensive 'brown-outs' resulted. Why?

The average price is sufficient to have users save at average times, but was too low at peak times. This was largely because the meters used were only capable of collecting end of month information; they noted how much network resources were consumed on average, but not when. Managing demand doesn't work or can't be achieved on the average. So, obviously the demand side is not being managed optimally. This has direct implications for the supply side. As customers begin to churn, the only alternative is to invest in the network – to increase capacity. But the economics of demand will continue to drive up usage to the point where it is no longer 'free' – to the point where network contention and the opportunity cost of time take over. Only when

there is over-capacity on dimensions – at all times, across all geographies, for all classes of service – will the need for new capital expenditure cease. So, by not managing demand through effective pricing – the telecoms industry is not investing on its own terms. For monopolies and duopolies this is truly ironic.

The network's 'sweet spot'

To manage demand appropriately, the network's availability 'sweet spot' – where contention is minimised – and consumption need to match each other. How can this be achieved? Simply by pricing on the dimension that matters most to consumers – time. Charging consumers by units consumed and by when they are consumed. It is natural to ask, "isn't this approach going to make my customers miserable?" Experience and economic theory says no. At peak time, premium or real-time business customers will happily pay more to ensure that they get good service with low latencies – thereby saving on time – their most valuable asset. Average leisure or business batch users – those with low opportunity cost of time and who would download huge files without caring if it took 10 minutes or so – would be better off as they could meet their needs off-peak at a saving.

Income-based billing's objective is to manage demand so as to manage supply – to manage prices so as to manage costs. The premise is that billing is more than revenue maximisation – it is a key demand management tool – it is used to maximise income. With income-based billing, service providers bill for continual services on a continual basis, as they do in cloud and energy. It sounds simple, but with telecoms data services this doesn't happen. Why?

Because conventional billing processes are not configured to handle the volumes for continual services. Rather, they are designed around a traditional, all-you-can-eat or event-by-event business model, which was built on the premise that no resource is finite.

In the days when voice was predominant and little bandwidth was required, this status quo was not problematic but with voice and (expanding) data sharing the same bandwidth, it's becoming a real problem. The reality is that if you have finite capacity, at some point you have to do income-based billing as practiced in energy with smart meters and now with the cloud to achieve higher revenues and lower costs. And to maintain, let alone grow, your customer base. The imperative for dynamic billing in the telecoms industry continues to grow daily. $

"Isn't this approach going to make my customers miserable?" Experience and economic theory says no.