TECHNOLOGY WHITE PAPER

Which takes Precedence: your NGN or your Current Business Model?

Historically, many incumbent telecommunications operators have enjoyed the ability to price their service offerings via market–based pricing strategies instead of the more limiting cost–based valuation methods. Traditionally this price discrimination ability has been primarily supported through a combination of offering both differentiated support services (for instance business hours only versus 24/7 support), and additionally offering clearly differentiated service performance, via the use of multiple overlay network infrastructures. Whilst the former service support differentiation remains unchanged in a Next–Generation Network (NGN) world, the latter service performance differentiation becomes vitally important as carriers collapse their multiple, overlay infrastructures into a single, high–performance NGN.

This article addresses some of the primary implications and considerations service providers face, as they must decide whether to preserve their ability to continue offering market–priced differentiated service performance or to adjust their business practice models based on reduced NGN capability offerings. In particular, this article examines one key implementation feature, aggregate queuing, which is required within an NGN, in order to preserve this ability to price–discriminate without risk of market arbitration.

M. COX

Which takes Precedence: your NGN or your Current Business Model?

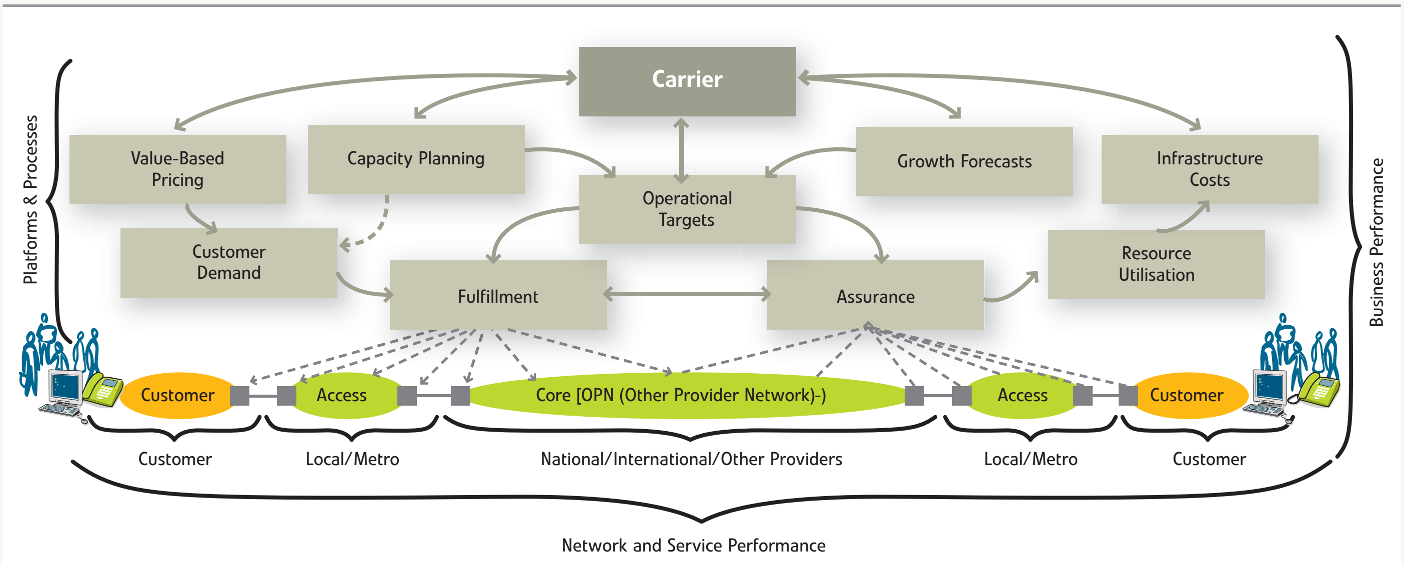
Does how you deploy your new NGN have the ability to directly and fundamentally impact the way your business may operate in the future?

Historically, many incumbent telecommunications operators have enjoyed the ability to price their service offerings using flexible, market–based pricing strategies instead of the more constrained, cost–based valuation methods. Traditionally, this price discrimination ability has been supported through a combination of offering both differentiated support services (for instance, business-hours-only support versus 24/7 support), and offering clearly differentiated service performance and capabilities, via the use of multiple overlay network infrastructures. Whilst the former practice is likely to remain unchanged in an NGN world, the latter becomes vitally important as carriers aim to collapse their legacy, overlay infrastructures into a single, high–performance NGN.

As carriers worldwide deploy these new, advanced NGN infrastructures, often insufficient consideration is given to the impact that various architectural and implementation choices may have on long–term business operational models. There is strong motivation to let these NGNs not only consolidate, but ultimately replace, all of the traditional overlay networks operated by most incumbent carriers today. However, this potentially introduces a tremendous risk to the business, since incumbents have traditionally leveraged each infrastructure platform in order to support their market–based price discrimination strategies.

By refocusing some attention on these long–term implications, it is possible to strategically determine whether the incumbent carriers wish to preserve their present operational models, with the possible consequence of increasing their NGN's operational complexity; or whether they are willing to sac–

Figure 1: Typical business interactions between the carrier, network, suppliers and customers



rifice some of their ability to price services within their market based on their value, and thus correspondingly simplify the deployment and management of their new NGN infrastructure. This article illustrates just some of the choices that are available and highlights the impact on both the business and the network that the corresponding choices may have. This is especially important, considering that these impacts and interactions may extend to and influence the entire market ecosystem that a carrier may be operating in, as illustrated in *Figure 1* (Note: In the figure, the core network may also be an Other Provider's Network (OPN)).

Business economics – an introduction

Fundamentally, a modern business is about providing goods and/or services that are perceived to have utility and value to customers who are willing to 'pay' for them (*Figure 2*). This difference between the 'price' of a particular product or service and all the associated 'cost' in creating or providing it naturally leads to a consideration of the options available in determining pricing, and ultimately adopting an appropriate pricing strategy. These pricing strategies can be roughly categorized into two groups: *cost–based* strategies, where unit prices are determined on the basis of costs, margin and revenue targets; and *market–based* strategies that define prices based on other concerns.

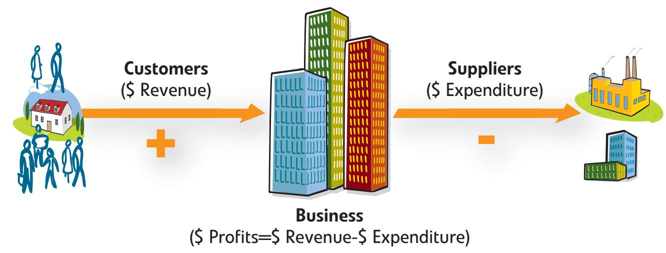
Cost–based pricing strategies

Cost–based pricing is *relatively* straightforward, and is characterized by calculating unit prices as some function of the unit's modeled cost. The resulting per–unit prices may accommodate appropriate break–point pricing for bulk purchases, which market aggregators can leverage in order to create secondary markets where they can add value with further price and service differentiation. Although cost–based pricing strategies are simpler to manage at the network level, they effectively reduce many price differentiation opportunities and hence are better suited for adoption by competitive carriers targeting specific market segments[1], or common carriers simply aiming to wholesale raw capacity, rather than for full–service incumbent carriers.

In fact, if an incumbent carrier is only able to price access to its infrastructure based on costs, then premium aggregators will invariably establish themselves, leveraging market–based pricing capabilities by offering differentiated services through corresponding performance differentiation capabilities. This provides the opportunity for service aggregators to gain some competitive advantage within the market population that is covered by the incumbent's infrastructure footprint by value–adding price and service discrimination. However, this has the overall side–effect of increasing the total capitalization required to support any given population footprint, thus increasing the total costs incurred and hence ultimately increasing the pricing of services to the end consumers within that market. This results in a sub–optimal market capitalization, where ultimately the consumer is forced to pay extra for the market's inefficiencies.

In fact, if an incumbent carrier is only able to price access to its infrastructure based on costs, then premium aggregators will invariably establish themselves, leveraging market–based pricing capabilities by offering differentiated services through corresponding performance differentiation capabilities. This provides the opportunity for service aggregators to gain some competitive advantage within the market population that is covered by the incumbent's infrastructure footprint by value–adding price and service discrimination. However, this has the overall side–effect of increasing the total capitalization required to support any given population footprint, thus increasing the total costs incurred and hence ultimately increasing the pricing of services to the end consumers within that market. This results in a sub–optimal market capitalization, where ultimately the consumer is forced to pay extra for the market's inefficiencies.

Figure 2: A simplified abstract model of a business



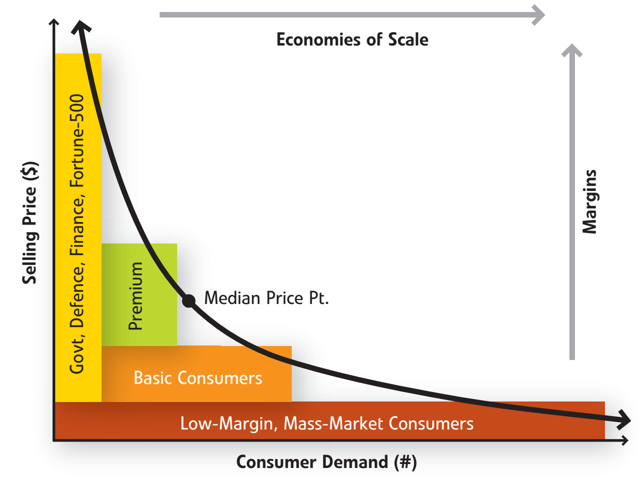
Market–based pricing strategies

Conversely, market–based pricing strategies, to all intents and purposes, can practically ignore the costs associated with providing goods and services and focus instead on pricing according to market expectations and/or goals. This pricing strategy is common amongst most incumbent operators today, as it provides greater flexibility for determining prices, despite being correspondingly more complex to manage. In the absence of hard cost–based constraints being applied, market–based pricing is often free to leverage cross–subsidization in a creative and flexible manner. Some more common examples of market–based pricing strategies include:

- Value–Based Pricing: Where prices can be subsidized across market segments by establishing prices on the basis of derived or perceived value (Residential versus Business, for instance).
- Cross–Bundled Pricing: Where prices can be subsidized across service components by pricing fixed-configuration product bundles (Triple Play services bundling Voice, Video and Data, for instance).
- Floor, Premium, Penetration or Parity–Based Pricing: Where prices are established to target a specific market segment and gain market share (targeting top 100 enterprise customers with premium pricing, or entering a consumer market segment with a consistently lower price, for instance).
- Price–Leadership Pricing: Where prices are set or discounted based on established reference pricing for a given market segment (traditional PSTN pricing and discounts, for instance).

1. Where they often actively attempt to reduce value differentiation that may exist, by decreasing prices and/or increasing performance standards.

Figure 3: Consumer demand curve (# customers) against willingness to pay (selling price)



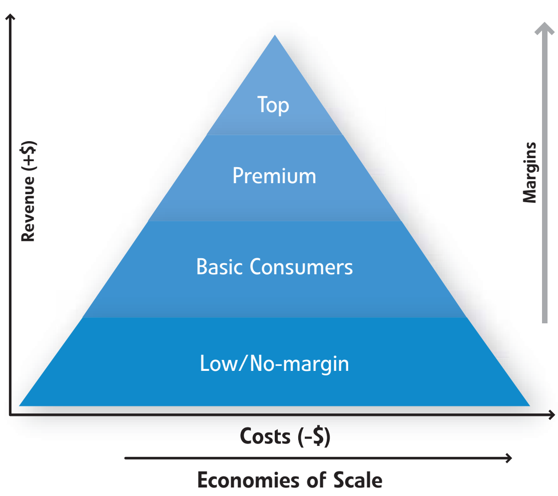
In comparison to fixed–price strategies (such as cost–based pricing), market–based price discrimination not only provides optimized revenue opportunities for the provider, but it can also ensure that consumers gain fair access to service resources at optimized price points[2]. This ability to sell similar services to different market segments at differentiated prices (*Figure 3*) allows incumbent operators to maximize the utilization of their fixed assets.

Apart from an inherent lack of demand that would result from excessive and unrealistic pricing, the next single biggest risk[3] that needs to be managed in a market–based pricing environment is that of arbitrage. Arbitrage occurs whenever the opportunity exists to purchase basic goods or services at some fixed price $x, and be able to resell them as competitive premium goods or services at a higher price $y. This situation will naturally occur if the basic goods/services perform consistently as well as the supposedly premium goods/services. This means that for a carrier to enjoy clearly differentiated pricing benefits (to support maximized revenues), the performance of the underlying services must also be clearly differentiated in order to reflect those pricing differences - and it is here that the linkage to the underlying network becomes so critical, as historically this differentiation was primarily achieved through the use of separate and distinct overlay networks.

What's different about incumbent carriers?

Incumbent carriers find themselves in a unique position, which often contrasts quite starkly with that of competitive carriers. Unlike a competitive carrier, whose primary strategy is often to target, usually aggressively, specific market segments that represent small subsets of the general population, an incumbent carrier is normally required to achieve near 100% population coverage or footprint. Hence they need to accommodate the requirements of all the market segments that exist within their deployment footprint. Whilst this situation has its advantages, namely the incumbent carrier enjoys the full benefits to be gained from large economies of scale and service reach, it also introduces a number of disadvantages that must be carefully managed. As a result of their infrastructure needing to cover a large and fixed portion of any given population, the incumbent carrier therefore inherits a large, fixed cost base associated with servicing and maintaining that infrastructure, regardless of the market penetration levels achieved.

In addition, often due to legislative or regulatory constraints, the incumbent carrier is usually required to provide basic service access at market–floor prices to any and all customers seeking such services within the infrastructure footprint. This is despite the fact that various customer segments within the population have different value assessments, hence different price/performance expectations of the carrier's products and services. Although incumbent carriers benefit from the larger deployment scales to achieve better price breaks on their infrastructure, it is still necessary for them to carefully manage their products and services, so as not to exclude potentially valuable market segments from consideration (see value pyramid diagram). Nor should they allow significant over–performance within their basic service offerings, since this can potentially cannibalize and/or marginalize their premium business opportunities.



Network engineering – an introduction

At the most fundamental level, an NGN can be modeled as nothing more than a simple[4] packet transfer function that accepts packets for processing via ingress interfaces and ultimately forwards those packets to a set of egress interfaces. Packets that form part of a common point–to–point stream, and are also considered equivalent from a forwarding transfer perspective, are termed a packetstream (a stream of packets that share a common forwarding policy/action). An NGN transfer function treats all packets of a packetstream in an equivalent manner. This concept is illustrated in *Figure 4*, which illustrates a simple First–In, First–Out (FIFO) packet transfer function exhibiting a single example of "tail discard". It should also be noted that it is possible to model a series or network of NGN nodes in a similar abstract manner (for instance, an entire core could be modeled as a

2. This is only true if the markets are not subject to pricing collusion and/or pure monopoly pricing.
3. Other noteworthy risks include externally subsidized competition, general market failures (which can occur if resource utilization is not managed in manner consistent with the resource's scarcity), and the aforementioned over–pricing and/or significant under–performance (thus stifling demand).
4. Once again, this is a gross oversimplification; however for the purposes being illustrated here it is sufficient to view an NGN in this manner.

Figure 4: An NGN packetstream interacting with an NGN FIFO transfer function

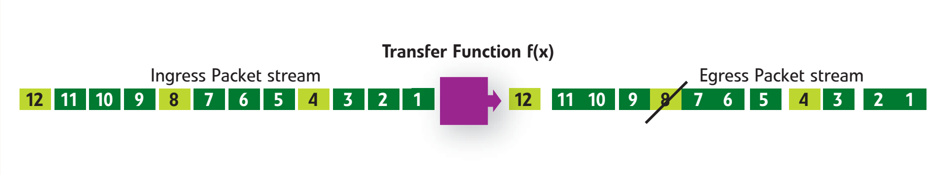
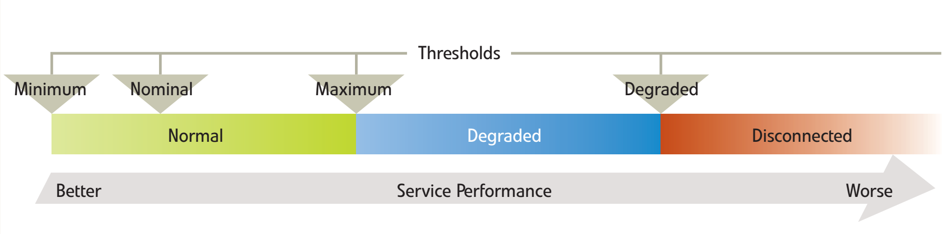


Figure 5: NGN Service Connectivity Concept



single, logical node with several ingress/egress interfaces to geographically diverse edge and aggregation nodes). Amongst the most important characteristics and consequences of modeling an NGN as a packet forwarding function is that for any given packetstream, the packet transfer function may induce one or more of the following performance impacts:

- **Packet Discard:** The failure of one or more packets from a packetstream to successfully egress the transfer function is deemed packet loss or discard.
- **Packet Delay:** The minimum duration of time it takes for one or more packets of a packetstream to be processed by the transfer function and to egress successfully.
- **Packet Delay Variation:** The significant[5] difference between the maximum and minimum packet delays experienced by all packets of a packetstream as they are processed successfully by the transfer function.
- **Stream Connectivity:** If a packetstream is impacted in a severe enough manner (excessive loss and/or delay), it may be deemed that connectivity across the transfer function has been lost completely (*Figure 5*). This situation is roughly analogous to the service availability measures in use today.

These parameters form the basis for measuring the intrinsic NGN packet performance, hence it is important to understand how and why each of these performance characteristics may be affected in a network. As a general rule, performance variations only occur in the event of some network level resource contention and/or failure. Furthermore, the degree (mild or heavy), duration (instantaneous or sustained) and the nature of the contention, intra–packetstream (within) or inter–packetstream (between), are also important considerations. Finally, the manner in which degradation affects particular packetstreams is highly dependent on specific internal implementations of the packetstream transfer function – in particular there exist three packet forwarding model abstractions, each of which exhibit different performance impact characteristics under resource contention. These models, referred to as simply Model–A Packetstream Transparent; Model–B Packetstream Priority; and Model–C Packetstream Transport, differ solely in terms of the internal implementation of the packetstream transfer function.

What does a generic packet transfer function look like?

RFC2475 "An Architecture for Differentiated Services" provides a good abstraction model for specifying the handling of packet–based traffic streams via a group of mechanisms collectively referred to as a "traffic conditioner" (*Figure 6*). Traffic conditioning involves several distinct phases that identify ingress traffic flows, classify them, apply policies to meter and mark each packet appropriately as being either "in–profile" or "out–of profile", queue the packets that are acceptable for forwarding and discard those that are not, and finally service packets from each of the various queues to the egress. In this diagram, traffic flows are associated with a logical traffic class based on classification and metering. Each of these traffic classes are then serviced for egress through a simple hierarchical scheduler supporting one or more queues.

The maximum latency variation induced by a particular queue and scheduler hierarchy is a combination of the maximum size of the queue and the minimum servicing rate of the scheduler. The average latency variation is somewhat more complex to determine as it is dependant on interactions between the application protocols, which change their behaviors over time, and the specific AQM implementation and configuration.

$$Latency_{MAX}(s) = \frac{Buffer\_Size\,(bits)}{Servicing\_Rate\,(bps)}$$

A small, yet highly significant consideration is the difference in the intra–packet delay variation performance induced by a Queuing System as opposed to that induced by a Scheduler System. In particular, schedulers only induce a limited form of virtual packetstream serialization delay in a manner roughly analogous to the bitstream serialization effect, whilst queuing systems can induce a shared and common delay to all packets of a packetstream up to the maximum depth of the queue. With these differences in mind, the behavioral characteristics of each of the various models can be highlighted.

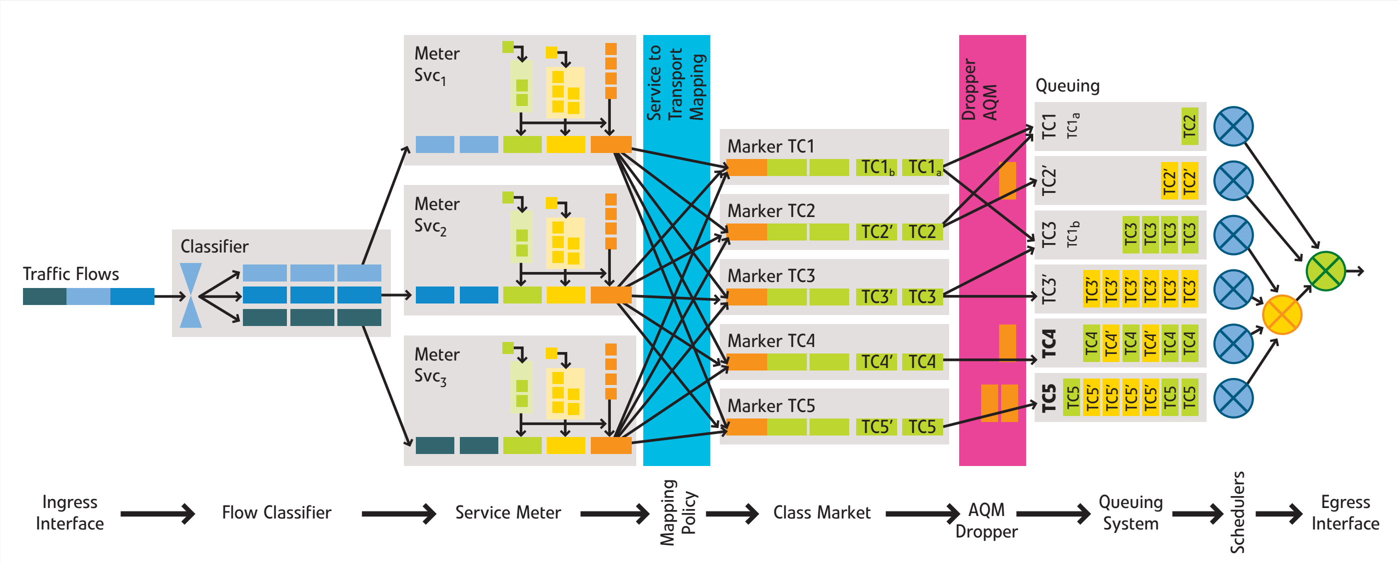
Note: that in all of these reference models, the location of the queues, policers and schedulers is deliberately obscured as it is the association of each that has more significance to the architecture. For instance, the fact that a per–destination queue is associated with an ingress interface is more important than whether or not the queue is implemented as an ingress and/or egress queue.

Model A: Packetstream Transparent Mode

In this model (**Figure 7**), the distinguishing feature is that all packets at the point of ingress are associated with a single, aggregate packetstream and are thus con-

5. Significant is used in a statistical manner here and is usually calculated using a percentile measure (for instance the 99^{th} percentile).

Figure 6: Stages of traffic conditioning within a congestion management system (transfer function)



sidered equivalent. This model closely approximates the traditional 'bit–pipe' circuit models associated with transport layer networks (see FIFO example in Figure 4). Capacity is usually dimensioned within the transfer function to support the pre–agreed packetstream profiles at the ingress and egress interfaces. In an NGN it is possible to permit an additional level of differentiation, based on 'in–profile' and 'out–of–profile' packet markings/classifications, with the latter being subject to some packet discard performance differentiation. Users of these types of service may choose to manage their packetstreams such that 'out–of–profile' traffic is not presented at the ingress interface (a strategy referred to as 'shaping'), and hence would subsequently enjoy the functionally equivalent performance of a traditional Constant Bit–Rate (CBR) based pipe.

Model B: Packetstream Priority Class Mode

In this model, the distinguishing feature is that an NGN's ability to classify individual packets into several priority classes is leveraged in an attempt to add value by adding network capability. Although the ingress traffic rate is subject to an all traffic classifier, each priority class packetstream may be individually classified at ingress and per–priority class policies can be established. An example of how individual packets of each priority class packetstream may be affected is illustrated in *Figure 8* – clearly, in these cases, higher priority packets may be serviced ahead of lower–priority packets, and likewise, under heavy congestion, higher classes can induce increased loss within the lower classes due to queue exhaustion. Although individually queued, the aggregate priority classes logically share a common scheduler hierarchy, allowing significant flexibility in how users of the service may choose to manage and 'shape' their individual and aggregate packetstreams. Herein lies the risk associated with this particular model as users (including service aggregators and wholesalers) can directly manage their ingress packetstreams to ensure that their aggregate rate is always kept 'in–profile', thus reducing this model to the functional equivalent of Model–A (CBR Pipe). This ability to arbitrage reduces the added value of the priority handling, especially when dealing with service aggregators, wholesalers and/or large customers.

Model C: Packetstream Transport Class Mode

In this model (*Figure 9*), the distinguishing feature is that all packets, of all ingresses, that are considered as 'belonging' to the same Transport Class (TC), share a common aggregate queue as part of the transfer function. This distinction becomes critically important when considering the risk of arbitrage. Single customers (either aggregators or individuals) are quite capable of managing their own packetstreams to

Figure 7: Model–A Packetstream Transparent Mode

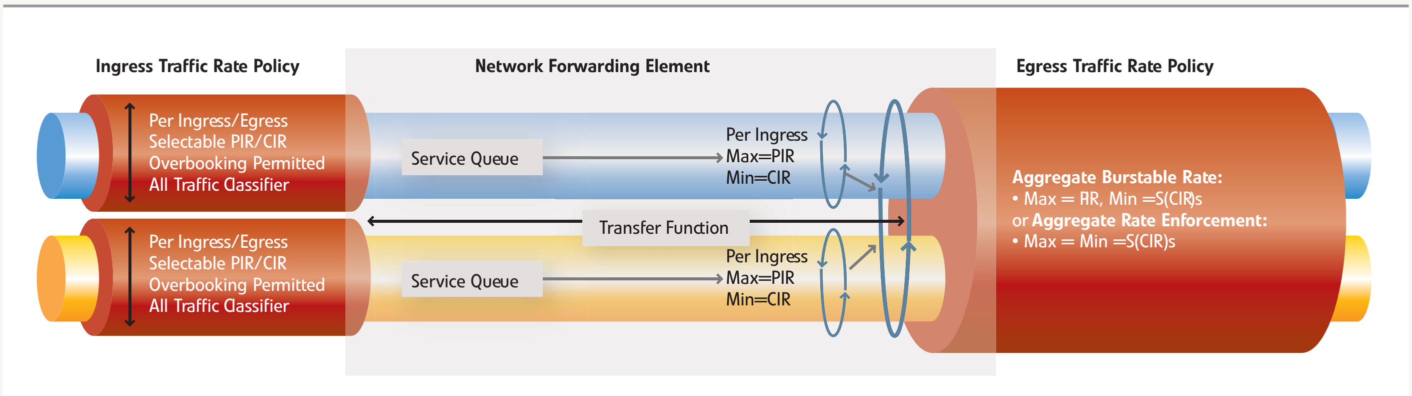


Figure 8: Model–B Packetstream Priority Class Mode

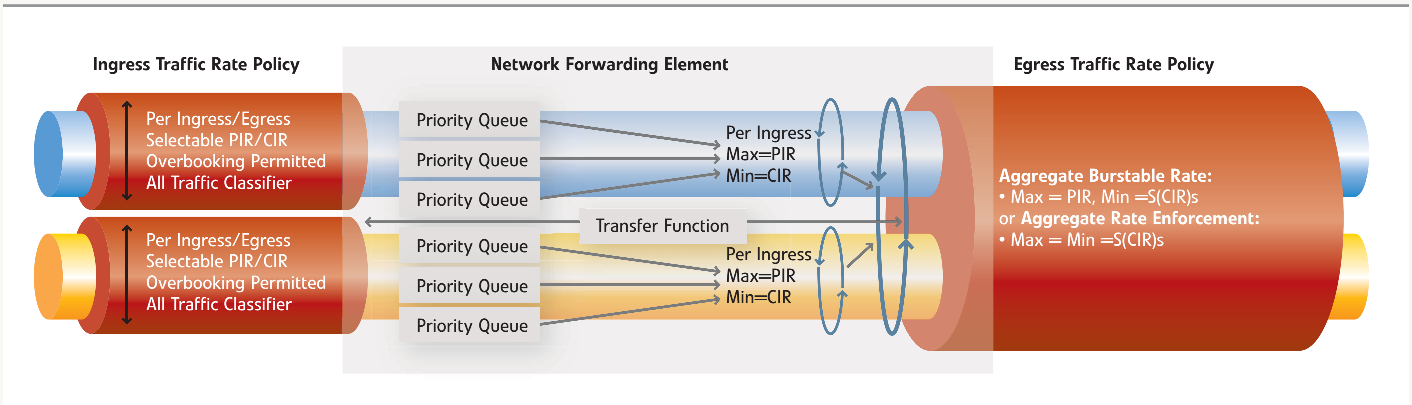
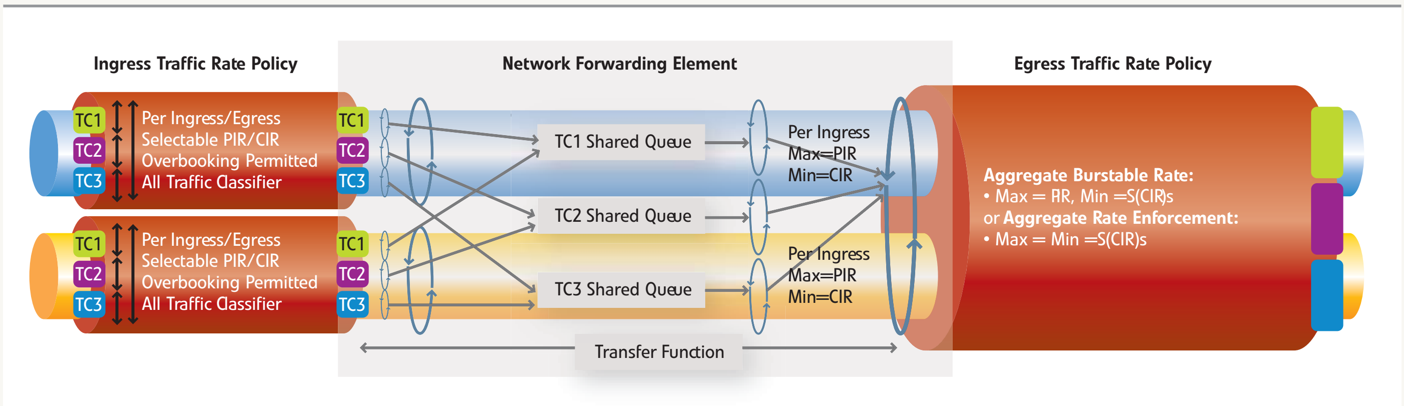


Figure 9: Model–C Packetstream Transport Class Mode



ensure they fall within the default policies established by the service provider, thereby alleviating any resource contention effects within their own packetstreams. However, if multiple customer packetstreams are treated as a single aggregate packetstream within a common Transport Class, and they share a common queue within the queuing system, then the likelihood of all customers successfully colluding to manage their aggregate packetstreams to fall within the default policy established by the service provider is improbable at best. Hence Model C is the only model that natively prevents service arbitrage, thus ensuring a level of absolute service performance differentiation, and ultimately protecting any market–based price differentiation strategies the provider may wish to adopt. In a similar manner to Model B, individual packets of each Priority Transport Class packetstream can be affected is illustrated in Figure 8 above—however with the distinction that no single user of a Transport Class is able to reliably arbitrage performance to achieve guaranteed CBR–like service within the lower priority classes. Since each transport class can be clearly differentiated in terms of "base" (or native) performance characteristics, the service provider is now able to independently price and sell capacity within each logical transport class. Likewise, customers are free to individually choose, based on performance/value assessments, which transport class they are prepared to purchase and pay for to meet their respective service requirements for each communication application they utilize.

Conclusion

It is clear that in this new world, many incumbent carriers need to identify whether they wish to be all things to all people (and hence adopt a market–based differential pricing strategy), or whether they are more comfortable adopting a carrier–of–carriers position, almost exclusively selling premium services to themselves and others, who are then responsible for differentiating performances and thus able to offer market–differentiated pricing for end–user services. The key is that the outcome of this decision will *automatically* imply a specific design model for the architecture of their NGN, and this is the critical linkage between the business and the engineering disciplines within a carrier.

Abbreviations

NGN	Next–Generation Network
OPN	Other Provider Network
FIFO	First–In First–Out
TC	Transport Class
AQM	Active Queue Management
DS	Differentiated Services
SIR	Sustained Information Rate

Alcatel·Lucent