WHITE PAPER

www.ipsphereforum.org

Author: Keith Dickerson

Version: 1.0

27 November 2006

Response to ERG Consultation on IP Interconnection

ERG(06)42

IPsphere FORUM

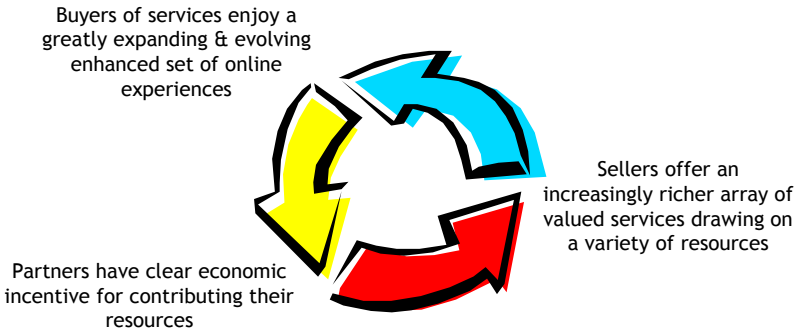
THE BUSINESS OF IP

Introduction – the IPsphere Framework

The IPsphere Forum (IPSF) is specifying a set of mechanisms that will enable added-value, differentiated, IP-based services to be provided by service providers over an NGN infrastructure. Therefore, the regulatory model to be applied for IP Interconnection is central to the concerns of IPSF members.

The IPsphere Forum aims to enhance the commercial framework for IP Services to provide a win-win scenario for all stakeholders (Figure 1). Within this 'Virtuous Circle', buyers of network-facilitated services will be able to enjoy an expanding and evolving set of experiences, sellers will be able to offer a richer array of valued services, and cooperating service providers will have a clear economic incentive for contributing their network, IT and associated resources.

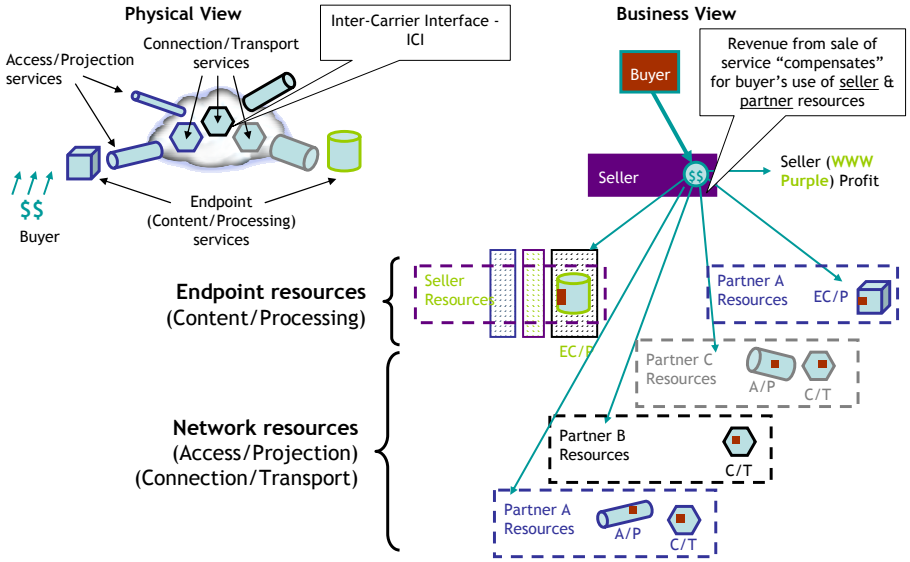
Figure 1: IP Services Virtuous Circle



A crucial next step for the online industry is, therefore, to develop a flexible commercial framework that rewards investment and innovation, and opens the way for many other models of commercial partnering alongside the existing Internet commercial model. Rather than regulating these commercial exchanges, this framework will allow participating service providers to offer whatever resources they choose, at whatever price they consider compelling. As in any market, the IP services market itself will then determine which offerings are worth buying and at what price, enabling the best commercial model for any IP service experience to evolve freely.

The IPsphere Framework specifies mechanisms by which next generation IP services and their market and revenue potential can be realized. Together these mechanisms create a framework that allows resource improvements that add value to service delivery to be offered - and the investments associated with them rewarded. Such a framework is already under development and has been implemented for trials.

Figure 2: The IPsphere Framework



Further information on the IPsphere Framework Architecture is given in [1].

Response to text of ERG(06)42

This Executive Summary should describe the desired outcomes of any regulation of IP interconnection in terms of public policy, competition policy, encouragement of investment, etc.

Separation of functional levels

A much richer set of Interconnect services will be provided in the NGN compared to the PSTN/ISDN. Interconnect services will be provided at both the service and control layer as well as at the transport layer as in legacy networks. A 3-layer model is proposed similar to that shown in Figure 1 on Page 10 of ERG(06)42:

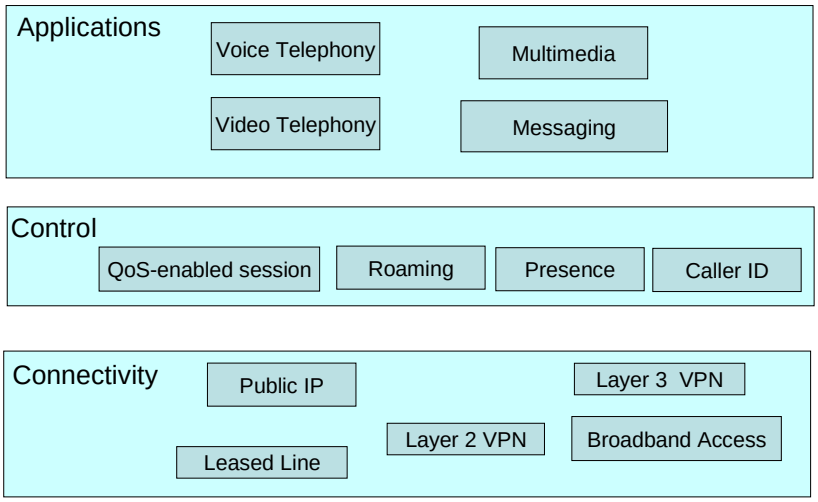
- Application (or Service) layer (e.g. Voice telephony, Messaging, etc).
- Intelligence layer (covering session control, authentication, directory, presence etc).
- Transport Connectivity layer (covering roughly OSI layers 1-4).

A simple categorization of services into a layered structure is shown in Figure 3. The purpose of layering services is to allow the definition of a set of common services at one layer that can support multiple services at a higher layer. As a simple example, a best-efforts IP-based service can support e-mail, web browsing and file transfers but not voice or video telephony services.

A generic "Session with QoS" service is proposed as a common capability to be used by session-based applications (e.g. Voice telephony, Video telephony etc.) as shown in Figure 3.

The IPSF proposes that a service structuring stratum (layer) is created above the control and intelligence layer of an NGN to allow the effective provision of end-to-end services across federated networks. The will enable customer specified levels of QoS, Security and Privacy with guaranteed bandwidth to be provided end-to-end across federated networks.

Figure 3: Service Categorisation



Quality of Service

Certain types of applications require defined QoS parameters in order to function properly, for example:

- Streaming multimedia (e.g. Video on Demand (VoD), Internet Television (IPTV)) may require guaranteed throughput with low jitter and low packet loss.
- IP telephony or Voice over IP (VoIP) will require strict limits on jitter and delay (both voice and video quality seriously declines when the underlying bearer performance degrades; p.e., when connecting to the 'delay challenged' mobile network).
- Video Teleconferencing (VTC) requires low jitter and low round trip delay.

- Dedicated link emulation requires both guaranteed throughput and imposes limits on maximum delay and jitter.

In general, these types of services are called inelastic, meaning they require a certain level of bandwidth with quality requirements to function properly. By contrast, elastic applications can take advantage of however much or little bandwidth is available.

The breakdown of service categories that will be proposed in NGN UK is:

a) Conversational (very low round trip delay)
b) Streaming (low jitter)
c) Priority data
d) Best effort

Besides technical requirements for differentiated services there might be commercial reasons to offer differentiated IP-based connectivity. For example, today the handling of Internet traffic does not differentiate between free content and paid content. However, from a commercial perspective it makes sense to treat traffic from paid content differently to traffic of free content.

Introduction of more quality classes would greatly improve the capabilities of IP-based services compared to the Internet.

The ability to provide differentiated services should be in the interest of all involved parties in the value chain as described below:

End-user's perspective
The end-user of paid content expects a better experience (e.g. connectivity and performance) of the service he is paying for. However, to access free content the end-user can not expect more than to be provided with a service on a "best-efforts" basis.

Content provider's perspective
Internet access with differentiated CoS offerings can be attractive for content providers. A differentiated offering is also in the interest of a content provider that offers both free and paid services. The prioritisation of packets of paid content compared with packets with free content provides the Content Provider with additional value propositions for marketing its premium services.

Carrier's perspective
Internet Access with differentiated QoS would enable a carrier to differentiate pricing for IP traffic according to the service type. End-to-end QoS functionality would lead to:

- more paid service offerings,
- new types of applications to end-users.

With a QoS based compensation model, carriers could get a portion of the revenue either directly from the end-user or the content provider when the business model and value chain roles allow carrier application “awareness”, once the QoS based bearer supports the requested application QoS parameters. It needs to be discussed how this can be translated into agreements among interconnecting networks.

From a carrier’s perspective, the implementation and support of QoS require network designs which allow marking and prioritising of IP packets according to the type of service. Many Tier 1 and Tier 2 carriers already have such functionality in place. However, the challenge will be the implementation of interconnects among carriers that support QoS.

Structural implications for the IC regime

The nature of IP means that the number of physical points of interconnection can decrease in an NGN environment. A smaller number of higher capacity PoIs should be used rather than a large number of lower capacity PoIs. This will optimize the use of network resources and reduce prices to customers.

Charging Principles

The current retail charging models for PSTN Voice are likely to remain appropriate in an NGN world, at least in the short term. The alternatives have significant downside impacts on end-users, who understand current "Caller Pays", and on network operators, whose business models are reliant on a stable model that does not result in imbalances in cost recovery between unequal operators. "Bill and Keep" or peering arrangements for Voice would be too radical a change for most, if not all, operators in Europe, and would result in unpredictable changes in business models and cash flows.

While the "currency" of traffic will become packets (and associated quality attributes), the use of minutes is likely to be retained at least in relation to voice calls originated on handsets connected to the fixed network. Customers need metrics that they can relate to and that result in charges that relate to apparent cost causality.

Just as with current networks, geographic differentiation in costs will result in more pressure for geographic wholesale pricing, but the extent to which this results in geographic pricing for end users will depend on the nature of the services provided and the degree to which national "postage stamp" pricing structures are preferable from a marketing and customer service perspective.

We believe, that IP interconnections should migrate into relationships where parties are being compensated for usage of their network resources.

To avoid “hot potato” routing, the best business model for IP traffic exchange is considered to be distance related, where compensation is based on the distance IP packets travel on a carrier’s network. This model is also service independent. It can be used for both existing Internet traffic in one class, as well as for IP-based traffic in different quality classes.

Distance compensation model

We believe that the party interconnecting with another network should pay for the traffic sent to it. In that case, building of the network to the end destination is avoided by handing off traffic to someone else. The amount paid should be determined by the ‘distance’ that traffic is carried by the other network. The exact meaning of ‘distance’ has to be defined and should probably be some set relation (e.g. between cities, countries or continents), rather than number of kilometres the traffic has to travel. It also needs to be taken into account that, in a wider IP 'carriage' scenario, with generalized mobility (and IMS facilities), the end users location (and associated call carriage 'distance') may not be known. In this case, presence on the visited network and mobility services may be a premium charged by the terminating network or subject to agreement between originating and terminating networks. A charging model based on distance would allow rational decisions to be made about building versus buying in order to grow the network and customer base. Networks would be paid according to the destination the traffic is transmitted to/from. The rate should be agreed based on network costs or topology.

This could be compared to the business model for PSTN Voice. Voice operators pay more if they have only a small network. As they build out their physical network, to interconnect with other large voice operators in more locations, they move expense to their own network. Although large voice operators still own most of the destinations, the new carrier gets his own customers and wants to carry calls from them on his own network for the greatest distance possible avoiding higher external costs. If the volume on certain routes is too low to justify network investment, calls can be handled to the large operator farther away for higher costs. A similar approach should be used in the IP interconnection world.

Today, in the Internet world traffic is handed over to the other network at the nearest interconnect (“hot-potato” routing) because there are typically only two types of settlements: free peering or transit. Today, there is no difference whether someone carries traffic across a whole continent or the traffic is terminated in the same city.

If the business model is changed all carriers would be paying all other carriers. Of course, like in the voice world, we would want relationships to balance so the net transfer of money is zero.

Billing should be based on the distance the traffic is transported on a carrier’s network. The carrier should be compensated for the cost of transfer and delivery of traffic with respect to distance

carried, quality level and capacity required. A system allowing distance and relation billing will be needed for such purpose.

There may be a lot of overhead required to measure the 'distance' in an IP-based network but in today's voice switches a high proportion of the functionality is actually related to the billing of the voice call. Settlement free peering was originally introduced because the value associated with the exchange of traffic was deemed to be far less then the overhead of creating a system to settle. This is no longer the case for many services.

The IPsphere mechanisms, as described in the Introduction to this response, will make it easier to realise a distance compensation model because the end points will be considered when the service is set up.

Class of Service based compensation model

A Class of Service model (as described above under Quality of Service) could be used as an extension either to the existing peering/transit or to the new distance related business model. The key issue here is that for high QoS services e.g. conversational, sufficient bandwidth needs to be provided for the call so that other traffic cannot impact on its quality. E.g. in a bandwidth managed network a session setup will reserve bandwidth whether it is used or not, therefore a session duration / session bandwidth model reflects the use of the resources employed.

Conclusions

It should be acknowledged that it is opportunity that enables competition. Regulation can help ease the way but regulation can also serve to discourage investment, requiring relief if investment is deemed to be important and in the public interest. Demand density is also a crucial localized factor that is the primary determinant of opportunity. This means that an EU-wide regulation approach needs to embrace/allow for differences in the demographics, demand densities etc of different countries and regions and must ensure the NRAs have flexibility to deal appropriately with these factors. NRAs need to pay attention to demographics (demand density etc) of the country/region and to have the flexibility to regulate accordingly.

The IPsphere Framework [1] will help to provide a standard way for operators to expose/harmonize technical and commercial interconnect arrangements and so will help to ensure that all types of Interconnect are possible. All the regulator needs to set are the limits/bounds for commercial parameters in wholesale & interconnect Element templates. The whole is abstracted/virtualized and therefore more easily implemented.

Finally, an EU wide regulation approach needs to ensure that investment is encouraged to best ensure end-to-end connectivity as well as allowing users to access services provided by another operator.

Questions for the Consultation

The position of the IPsphere Forum on each of the specific questions for consultation listed at the end of the executive summary in ERG(06)42 is given below:

- **How should the transition from the PSTN number of interconnection points to the probably reduced number of interconnection points in NGNs look like? Which are the implications for the price structure and price level of interconnection rates?**

  In the NGN, the price structure will reflect the number of Points of Service Interconnect (PoSI). A PoSI represents the location at which services would be handed over from one operator to another if transport costs and bearer modularity were irrelevant. In practice, given the move towards higher capacity bearers for interconnect, one physical Point of Handover may support a number of PoSI. Calls *not* delivered to the correct PoSI would incur an additional charge to cross the core network. A smaller number of PoSIs will result in lower costs and hence lower charges to customers. However many PoSIs there are, there will also be a significant cost difference in delivering calls to customers in heavily populated as opposed to those in lightly populated areas. This will need to be reflected in any pricing structure.

- **What is the equivalent to "local" interconnection in NGNs?**

  An example of the "local" equivalent of NGN interconnect could be either interconnect at the MSAN (Multi-Service Access Node) level in the BT 21C network or at the MDF (Main Distribution Frame). Interconnect at the MSAN is not recommended because it would be uneconomic to provide physical interconnect at over 10,000 points in the network. Interconnect at the MDF is equivalent to LLU.

  In order to realise the economic benefits of the NGN to both incumbent and other operators, the number of physical points of interconnect should be reduced as much as possible. Therefore, the most appropriate point of (physical and service) interconnect in the BT 21C network example is at the Metro node.

- **Reflecting the transition towards NGNs what are the implications for existing SMP products and bottleneck facilities? Does this technological change remove existing SMP positions or bottlenecks or could new ones emerge in NGNs?**

Access bottlenecks in the NGN environment have the potential to be much reduced compared to legacy networks because of the many different access technologies available, e.g. DSL, Ethernet, 2G, 3G, WiFi, WiMAX, etc.

Although it could be argued than new bottlenecks could emerge due to the centralised provision of intelligence functions that all services are dependent on, e.g. IMS, number translation and DNS, this can be alleviated by appropriate dimensioning of these functions as well as for the opportunities to invest and participate in the new services markets

Eventual regulatory remedies should take into account a necessary revision on the real bottlenecks remaining in an NGN environment. The opportunity and willingness to invest by service providers shall not be conditioned by similar obstacles as could be the case in the former PSTN environment, which dictated the existing regulatory policy. Therefore, such a policy should be fully revisited, specially when referring to the core NGN investment and deployment and, consequently, related services. Existing bottleneck to the participation of services providers in a much more open value chain should allow to remove currently identified bottlenecks.

- **How do you evaluate the advantages and disadvantages of different charging principles discussed in the paper?**

  The advantages and disadvantages of different charging principles should be evaluated based on the range and quality of services that are made available to customers as a result of the application of these charging principles. In particular:

  - in the short to medium term the retail principle of Calling Party Pays (CPP) and wholesale principle of Calling Party's Network Pays (CPNP) should continue
  - minutes are the unit of currency for calls and will continue to be used at the wholesale level
  - retailers will want to continue to differentiate by medium (fixed, mobile, internet) and usage (volume, distance and time of day)
  - wholesalers will want to differentiate geographically in relation to costs
  - bill and keep for all call types in the wholesale market would drive retailers to charge their customers in relation to the value and costs of making and receiving calls which would remove the equalization of costs between calling and receiving parties that results from CPP. This is unlikely to be acceptable to retail customers who would pay more e.g. mobile customers, fixed line customers in lightly populated areas

  A charging model based on Bill & Keep will not encourage infrastructure investment because there will be no opportunity for a network provider to obtain additional revenue by providing an increase in network capacity. Any charging model must enable revenues to be earned in proportion to the investment made in the infrastructure.

Section 3.3 Implications of NGN architecture for IP-Interconnection

Section 3.3 (Page 19) seems to indicate that BGP is the only way to "manage QoS on an end-to-end basis". In particular, the text indicates that, "across two independent networks, QoS can only be achieved on a bilateral basis, using SLAs and specifying the BGP." This statement is highly arguable for various reasons.

1. BGP is a path-vector protocol that has been initially designed for exchanging reachability information between domains, where a domain is defined as a set of routers operated by a globally unique administrative entity. From a QoS management perspective, there is nothing that prevents the investigation of other protocols, such as RSVP and its traffic engineering extensions, to participate to the enforcement of a global QoS policy across domains. The current specification effort conducted by the IETF pce (Path Computation Element) working group is another example of a protocol - namely the PCEP protocol - that could be used for dynamic inter-AS path computation and selection purposes, as part of the enforcement of a pan-provider QoS policy.
2. If the aforementioned notion of "two independent networks" means two distinct BGP domains, it is untrue to state that QoS can only be achieved by means of BGP on a bilateral basis. Indeed, the manipulation of BGP attributes like the MULTI_EXIT_DISCRIMINATOR, the AS_PATH or the LOCAL_PREF attributes allow for a finer enforcement of a BGP-based QoS policy where service providers that operate BGP domains can influence the way traffic will be forwarded from one AS (domain) to another.
3. There are indeed research efforts that investigate the use of BGP for carrying QoS-related information between domains, by means of specific BGP capabilities that would allow for the announcement of several routes towards a given set of destination prefixes associated to the use of a specific, optional and presumably transitive BGP attribute that would propagate route-specific QoS information. Details about such work can be found at http://www.ist-tequila.org/publications/wtc2002-idte.pdf and http://bgp.potaroo.net/ietf/all-ids/draft-jacquenet-bgp-qos-00.txt-37966.txt.

Section 3.5 Implications for interconnection products and network/service provision

With regard to the issue of cost determination it should be taken into account that in an NGN environment the level of investment carried out by those who enact the NGN, and particularly a Next Generation Access Network (NGAN), should reasonably lead to a revisited regulation (applicable to the new access network).

Undue regulation of access infrastructure, and particularly cost orientation, could induce a reduction or negation to investment. It should not be assumed, per-se, that new access networks represent a bottleneck. This networks are different from existing networks and cost conditions as well as conditions for entry are also different. Therefore, the cost of interconnection products should be dealt carefully since it simply provide a bottom line and the conditions for selling such products in a increasingly competitive environment should cover the risk as well as the opportunity cost that the altnet enjoy when postponing its build or buy decisions.

References

[1] Creating a Commercially Sustainable Framework for IP Services - Realizing Next Generation Revenues, IPSF White Paper, May 2006.